

*IBM SPSS Direct Marketing 22*

**IBM**

**Hinweis**

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 27 gelesen werden.

**Produktinformation**

Diese Ausgabe bezieht sich auf Version 22, Release 0, Modifikation 0 von IBM SPSS Statistics und alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuausgabe geändert wird.

---

# Inhaltsverzeichnis

<b>Kapitel 1. Direct Marketing (Direktmarketing)</b> . . . . .	<b>1</b>	<b>Kapitel 5. Rücklaufquoten nach Postleitzahlen.</b> . . . . .	<b>17</b>
		Einstellungen. . . . .	18
<b>Kapitel 2. RFM-Analyse</b> . . . . .	<b>3</b>	Erstellen eines kategorialen Antwortfelds . . . . .	19
RFM-Scores aus Transaktionsdaten . . . . .	3	<b>Kapitel 6. Kaufneigung</b> . . . . .	<b>21</b>
RFM-Scores aus Kundendaten . . . . .	4	Einstellungen. . . . .	23
RFM-Klassierung . . . . .	5	Erstellen eines kategorialen Antwortfelds . . . . .	24
Speichern von RFM-Scores aus Transaktionsdaten . . . . .	6	<b>Kapitel 7. Kontrollpakettest</b> . . . . .	<b>25</b>
Speichern von RFM-Scores aus Kundendaten . . . . .	7	<b>Bemerkungen.</b> . . . . .	<b>27</b>
RFM-Ausgabe . . . . .	8	Marken. . . . .	29
<b>Kapitel 3. Clusteranalyse</b> . . . . .	<b>11</b>	<b>Index</b> . . . . .	<b>31</b>
Einstellungen. . . . .	12		
<b>Kapitel 4. Profile über potenzielle Kunden</b> . . . . .	<b>13</b>		
Einstellungen. . . . .	14		
Erstellen eines kategorialen Antwortfelds . . . . .	15		



---

## Kapitel 1. Direct Marketing (Direktmarketing)

Die Option "Direktmarketing" bietet eine Reihe von Tools zur Verbesserung der Ergebnisse von Direktmarketing-Kampagnen durch die Identifizierung von Demografie-, Einkaufs- und anderen Merkmalen, die unterschiedliche Kundengruppen definieren, sowie durch Konzentration auf bestimmte Gruppen zur Maximierung positiver Rücklaufquoten.

**RFM-Analyse.** Dieses Verfahren identifiziert bestehende Kunden, die sehr wahrscheinlich auf ein neues Angebot antworten.

**Clusteranalyse.** Hierbei handelt es sich um eine explorative Prozedur zum Ermitteln von natürlichen Gruppierungen (Clustern) innerhalb Ihrer Daten. Damit können beispielsweise verschiedene Kundengruppen auf der Basis unterschiedlicher demografischer und Kaufverhaltensmerkmale ausgemacht werden.

**Profile über potenzielle Kunden.** Bei dieser Technik werden Ergebnisse aus einer früheren Kampagne oder einer Testkampagne verwendet, um beschreibende Profile zu erstellen. Diese Profile können bei zukünftigen Kampagnen für das Targeting bestimmter Gruppen von Kontakten verwendet werden. Weitere Informationen finden Sie im Thema Kapitel 4, „Profile über potenzielle Kunden“, auf Seite 13.

**Rücklaufquoten nach Postleitzahlen.** Bei dieser Technik werden Ergebnisse aus einer früheren Kampagne verwendet, um Rücklaufquoten nach Postleitzahlen zu berechnen. Diese Raten können bei zukünftigen Kampagnen für das Targeting bestimmter Postleitzahlbereiche verwendet werden. Weitere Informationen finden Sie im Thema Kapitel 5, „Rücklaufquoten nach Postleitzahlen“, auf Seite 17.

**Kaufneigung.** In diesem Verfahren werden Ergebnisse einer Testsendung oder einer früheren Kampagne verwendet, um Propensity-Scores zu erstellen. Die Scores zeigen an, bei welchen Kontakten die Wahrscheinlichkeit einer Antwort am höchsten ist. Weitere Informationen finden Sie im Thema Kapitel 6, „Kaufneigung“, auf Seite 21.

**Kontrollpakettest.** Dieses Verfahren vergleicht Marketingkampagnen, um herauszufinden, ob es hinsichtlich der Effektivität signifikante Unterschiede zwischen verschiedenen Paketen oder Angeboten gibt. Weitere Informationen finden Sie im Thema Kapitel 7, „Kontrollpakettest“, auf Seite 25.



---

## Kapitel 2. RFM-Analyse

Die RFM-Analyse (Recency – Aktualität, Frequency – Häufigkeit, Monetary – Geldwert) ist eine Technik, die verwendet wird, um bestehende Kunden zu identifizieren, die am wahrscheinlichsten auf ein neues Angebot reagieren werden. Diese Technik wird häufig im Direktmarketing eingesetzt. RFM-Analyse basiert auf der folgenden einfachen Theorie:

- Der wichtigste Faktor bei der Identifizierung von Kunden, die wahrscheinlich auf ein neues Angebot reagieren, ist **Aktualität**. Kunden, die kürzlich gekauft haben, kaufen wahrscheinlicher wieder ein, als Kunden, die weiter zurück in der Vergangenheit gekauft haben.
- Der zweitwichtigste Faktor ist **Häufigkeit**. Kunden, die in der Vergangenheit häufiger gekauft haben, kaufen wahrscheinlicher wieder ein, als Kunden, die weniger gekauft haben.
- Der dritt wichtigste Faktor ist der ausgegebene Betrag, der als **Geldwert** bezeichnet wird. Kunden, die in der Vergangenheit (für alle Einkäufe insgesamt) mehr ausgegeben haben, reagieren wahrscheinlicher, als Kunden, die weniger ausgegeben haben.

Funktionsweise der RFM-Analyse

- Kunden wird basierend auf dem Datum des letzten Kaufs bzw. des Zeitintervalls seit dem letzten Kauf ein Aktualitätsscore zugewiesen. Dieser Score basiert auf einer einfachen Einstufung von Aktualitätswerten in eine kleine Zahl von Kategorien. Wenn Sie zum Beispiel fünf Kategorien verwenden, erhalten die Kunden mit den neuesten Kaufdaten eine Aktualitätseinstufung von 5 und die mit den am weitesten zurückliegenden Kaufdaten eine Aktualitätseinstufung von 1.
- Auf ähnliche Weise wird Kunden dann eine Häufigkeitseinstufung zugewiesen, wobei höhere Werte eine höhere Kaufhäufigkeit bedeutet. In einem Einstufungsschema mit fünf Kategorien erhalten Kunden, die am häufigsten einkaufen, eine Häufigkeitseinstufung von 5.
- Schließlich werden die Kunden nach Geldwert eingestuft, wobei die höchsten Geldwerte die höchste Einstufung erhalten. In dem Beispiel mit fünf Kategorien würden die Kunden, die den höchsten Betrag aufwenden, eine Geldwerteinstufung von 5 erhalten.

Das Ergebnis sind vier Scores für jeden Kunden: der Aktualitäts-, der Häufigkeits-, der Geldwert- und der kombinierte RFM-Score, bei dem einfach die drei einzelnen Scores in einem einzigen Wert verkettet werden. Die "besten" Kunden (die am wahrscheinlichsten auf ein Angebot reagieren) sind diejenigen Kunden mit den höchsten kombinierten RFM-Scores. In einer Einstufung mit fünf Kategorien gibt es zum Beispiel insgesamt 125 mögliche, kombinierte RFM-Scores, der höchste kombinierte RFM-Score ist 555.

Erläuterung der Daten

- Wenn Datenzeilen Transaktionen darstellen (jede Zeile repräsentiert eine einzelne Transaktion und es kann mehrere Transaktionen für jeden Kunden geben), verwenden Sie RFM aus Transaktionen. Weitere Informationen finden Sie im Thema „RFM-Scores aus Transaktionsdaten“.
- Wenn Datenzeilen Kunden mit Auswertungsinformationen für alle Transaktionen darstellen (mit Spalten, die Werte für den Gesamtkaufbetrag, die Gesamtzahl der Transaktionen und das letzte Transaktionsdatum enthalten), verwenden Sie RFM aus Kundendaten. Weitere Informationen finden Sie im Thema „RFM-Scores aus Kundendaten“ auf Seite 4.

---

### RFM-Scores aus Transaktionsdaten

Erläuterung der Daten

Das Dataset muss Variablen enthalten, die die folgenden Informationen enthalten:

- Eine Variable oder eine Kombination von Variablen, die jeden Fall (Kunden) identifizieren
- Eine Variable mit dem Datum jeder Transaktion

- Eine Variable mit dem Geldwert jeder Transaktion

Erstellen von RFM-Scores aus Transaktionsdaten

1. Wählen Sie die folgenden Befehle aus den Menüs aus:  
**Direktmarketing > Verfahren wählen**
2. Wählen Sie **Hilfe bei der Erkennung meiner besten Kontakte (RFM-Analyse)** und klicken Sie auf **Weiter**.
3. Wählen Sie **Transaktionsdaten** und klicken Sie auf **Weiter**.
4. Wählen Sie die Variable aus, die Transaktionsdaten enthält.
5. Wählen Sie die Variable, die den Geldwertbetrag für jede Transaktion enthält.
6. Wählen Sie die Methode für die Zusammenfassung der Transaktionsbeträge für jeden Kunden: Summe (Summe aller Transaktionen), Mittelwert, Median oder Maximum (höchster Transaktionsbetrag).
7. Wählen Sie die Variable oder die Kombination von Variablen, die jeden Kunden eindeutig identifiziert. Zum Beispiel könnten Fälle durch einen eindeutigen Schlüsselcode oder eine Kombination aus Nachname und Vorname identifiziert werden.

---

## RFM-Scores aus Kundendaten

Erläuterung der Daten

Das Dataset muss Variablen enthalten, die die folgenden Informationen enthalten:

- Das letzte Kaufdatum oder ein Zeitintervall seit dem letzten Kaufdatum. Dies wird zur Berechnung der Aktualitätsscores verwendet.
- Gesamtzahl von Käufen. Dies wird zur Berechnung der Häufigkeitsscores verwendet.
- Gesamtgeldwertbetrag für alle Käufe. Dies wird zur Berechnung der Geldwertsscores verwendet. In der Regel ist dies die Summe aller Käufe, könnte jedoch auch der Mittelwert (Durchschnitt), das Maximum (größter Betrag) oder eine andere Auswertungskennzahl sein.

Wenn Sie die RFM-Scores in ein neues Dataset schreiben möchten, muss das aktive Dataset auch eine Variable oder eine Kombination aus Variablen enthalten, die jeden Fall (Kunden) identifizieren.

Erstellen von RFM-Scores aus Kundendaten

1. Wählen Sie die folgenden Befehle aus den Menüs aus:  
**Direktmarketing > Verfahren wählen**
2. Wählen Sie **Hilfe bei der Erkennung meiner besten Kontakte (RFM-Analyse)** und klicken Sie auf **Weiter**.
3. Wählen Sie **Kundendaten** und klicken Sie auf **Weiter**.
4. Wählen Sie die Variable, die das letzte Transaktionsdatum oder eine Zahl enthält, die ein Zeitintervall seit der letzten Transaktion darstellt.
5. Wählen Sie die Variable, die die Gesamtzahl der Transaktionen für jeden Kunden enthält.
6. Wählen Sie die Variable, die den Gesamtgeldwertbetrag für jeden Kunden enthält.
7. Wenn Sie die RFM-Scores in ein neues Dataset schreiben möchten, wählen Sie die Variable oder die Kombination aus Variablen, die jeden Kunden eindeutig identifiziert. Zum Beispiel könnten Fälle durch einen eindeutigen Schlüsselcode oder eine Kombination aus Nachname und Vorname identifiziert werden.



---

## RFM-Klassierung

Der Prozess der Gruppierung einer großen Zahl von numerischen Werten in eine kleine Zahl von Kategorien wird manchmal als **Klassierung** (Binning) bezeichnet. Bei der RFM-Analyse sind die Klassen Einstufungskategorien. Sie können die Registerkarte "Klassierung" verwenden, um die zur Zuweisung von Aktualitäts-, Häufigkeits- und Geldwertwerten zu diesen Klassen verwendete Methode zu ändern.

### Klassierungsmethode

**Verschachtelt.** Bei der verschachtelten Klassierung wird den Aktualitätswerten eine einfache Einstufung zugewiesen. Innerhalb jeder Aktualitätseinstufung wird Kunden eine Häufigkeitseinstufung zugewiesen. Innerhalb jeder Häufigkeitseinstufung wird Kunden eine Geldwerteinstufung zugewiesen. Diese neigt dazu, eine gleichmäßigere Verteilung von kombinierten RFM-Scores bereitzustellen, hat jedoch den Nachteil, dass sich die Interpretation der Häufigkeits- und Geldwerteinstufungsscores schwieriger gestaltet. Zum Beispiel kann eine Häufigkeitseinstufung von 5 für einen Kunden mit einer Aktualitätseinstufung von 5 nicht das Gleiche bedeuten wie eine Häufigkeitseinstufung von 5 für einen Kunden mit einer Aktualitätseinstufung von 4, denn die Häufigkeitseinstufung hängt von der Aktualitätseinstufung ab.

**Unabhängig.** Aktualitäts-, Häufigkeits- und Geldwerte werden einfachen Einstufungen zugewiesen. Die drei Einstufungen werden unabhängig zugewiesen. Die Interpretation jeder der drei RFM-Komponenten ist daher eindeutig. Ein Häufigkeitsscore von 5 für einen Kunden bedeutet das Gleiche wie ein Häufigkeitsscore von 5 für einen anderen Kunden, unabhängig von ihren Aktualitätsscores. Bei kleineren Stichproben hat dies den Nachteil, dass es zu einer weniger gleichmäßigen Verteilung der kombinierten RFM-Scores kommt.

### Anzahl an Klassen

Die Anzahl der Kategorien (Klassen) für jede Komponente für die Erstellung der RFM-Scores. Die Gesamtzahl der möglichen kombinierten RFM-Scores ist das Produkt der drei Werte. Zum Beispiel würden 5 Aktualitätsklassen, 4 Häufigkeitsklassen und 3 Geldwertklassen insgesamt 60 mögliche kombinierte RFM-Scores zwischen 111 und 543 erzeugen.

- Standard für jede Komponente ist 5, sodass 125 mögliche, kombinierte RFM-Scores zwischen 111 und 555 erzeugt werden.
- Die maximale Zahl an zulässigen Klassen für jede Scorekomponente ist neun.

### Bindungen

Eine "Bindung" sind einfach zwei oder mehr gleiche Aktualitäts-, Häufigkeits- oder Geldwerte. Idealerweise wünscht man sich ungefähr die gleiche Zahl an Kunden in jeder Klasse, aber eine größere Zahl an Bindungswerten kann sich auf die Klassenverteilung auswirken. Es gibt zwei Alternativen für die Handhabung von Bindungen:

- **Bindungen der gleichen Klasse zuweisen.** Diese Methode weist unabhängig von der Auswirkung auf die Klassenverteilung gebundene Werte stets der gleichen Klasse zu. So ergibt sich eine konsistente Klassierungsmethode: Wenn zwei Kunden denselben Aktualitätswert besitzen, werden sie stets demselben Aktualitätsscore zugewiesen. In einem extremen Beispiel haben Sie vielleicht 1.000 Kunden, von denen 500 ihren letzten Einkauf am gleichen Tag tätigen. In einer 5-Klassen-Einstufung würden 50 % der Kunden daher anstelle des Idealwerts von 20 % einen Aktualitätsscore von 5 erhalten.

Beachten Sie, dass es bei der verschachtelten Klassierungsmethode "Konsistenz" bei Häufigkeits- und Geldwertcores etwas komplizierter ist, da Häufigkeitsscores innerhalb von Aktualitätsscoreklassen und Geldwertcores innerhalb von Häufigkeitsscoreklassen zugewiesen werden. So haben zwei Kunden mit dem gleichen Häufigkeitswert gegebenenfalls nicht den gleichen Häufigkeitsscore, wenn sie nicht, unabhängig von der Handhabung gebundener Werte, auch über den gleichen Aktualitätsscore verfügen.

- **Bindungen willkürlich zuweisen.** Hierüber wird eine gleichmäßige Klassenverteilung gewährleistet, indem Bindungen vor der Einstufung ein sehr kleiner Varianzfaktor zugewiesen wird, sodass es zum Zweck der Zuweisung von Werten an die eingestufteten Klassen keine gebundenen Werte gibt. Dieser Prozess hat keine Auswirkungen auf die Originalwerte. Er wird nur eingesetzt, um Bindungen eindeutig zu machen. Zwar erzeugt dies eine gleichmäßige Klassenverteilung (ungefähr die gleiche Anzahl an Kunden in jeder Klasse), es kann aber auch zu vollständig unterschiedlichen Scoreergebnissen für Kunden führen, die ähnliche oder identische Aktualitäts-, Häufigkeits- oder Geldwerte haben, speziell, wenn die Anzahl der Kunden relativ klein und/oder die Anzahl der Bindungen relativ hoch ist.

Tabelle 1. Bindungen der gleichen Klasse zuweisen im Vergleich mit Bindungen willkürlich zuweisen.

ID	Letzter Kauf (Aktualität)	Bindungen der gleichen Klasse zuweisen	Bindungen willkürlich zuweisen
E	29.10.2006	5	5
Z	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	Z
8	13.8.2006	Z	Z
9	13.8.2006	Z	E
10	20.6.2006	E	E

- In diesem Beispiel führt die Zuweisung von Bindungen der gleichen Klasse zu einer ungleichmäßigen Klassenverteilung: 5 (10 %), 4 (40 %), 3 (20 %), 2 (20 %), 1 (10 %).
- Das willkürlich Zuweisen von Bindungen führt zu 20 % in jeder Klasse. Um dieses Ergebnis aber zu erreichen, werden die vier Fälle mit dem Datumswert 28.10.2006 3 verschiedenen Klassen zugewiesen und die 2 Fälle mit einem Datumswert von 13.8.2006 werden ebenfalls unterschiedlichen Klassen zugewiesen.

Beachten Sie, dass die Art, mit der Bindungen unterschiedlichen Klassen zugewiesen werden, absolut zufällig erfolgt innerhalb der Einschränkung, dass das Endergebnis eine gleiche Anzahl von Fällen in jeder Klasse hat). Wenn Sie eine zweite Menge an Scores mit der gleichen Methode berechnet haben, könnte sich die Einstufung für einen bestimmten Fall mit einem gebundenen Wert ändern. Zum Beispiel könnten sich die Einstufungen von 5 und 3 für die Fälle 4 und 5 beim zweiten Mal vertauschen.

## Speichern von RFM-Scores aus Transaktionsdaten

RFM aus Transaktionsdaten erstellt stets ein neues aggregiertes Dataset mit einer Zeile je Kunde. Verwenden Sie die Registerkarte "Speichern", um anzugeben, welche Scores und anderen Variablen Sie speichern möchten und wo Sie sie speichern möchten.

### Variablen

Die Schlüsselvariablen, die jeden Kunden eindeutig identifizieren, werden automatisch im neuen Dataset gespeichert. Die folgenden zusätzlichen Variablen können im neuen Dataset gespeichert werden:

- **Datum der letzten Transaktion für jeden Kunden.**
- **Anzahl der Transaktionen.** Die Gesamtzahl an Transaktionszeilen je Kunde.
- **Betrag.** Der Gesamtbetrag für jeden Kunden, basierend auf der in der Registerkarte "Variablen" gewählten Auswertungsmethode.

- **Aktualitätsscore.** Der jedem Kunden zugewiesene Score, basierend auf dem letzten Transaktionsdatum. Höhere Scores geben aktuellere Transaktionsdaten an.
- **Häufigkeitsscore.** Der jedem Kunden zugewiesene Score, basierend auf der Gesamtzahl an Transaktionen. Höhere Scores stehen für mehr Transaktionen.
- **Geldwertsscore.** Der jedem Kunden zugewiesene Score, basierend auf der ausgewählten Geldwert-Auswertungskennzahl. Höhere Scores stehen für einen höheren Wert für die Geldwert-Auswertungskennzahl.
- **RFM-Score.** Die drei Einzelscores, zu einem einzigen Wert kombiniert:  $(\text{Aktualität} \times 100) + (\text{Häufigkeit} \times 10) + \text{Geldwert}$ .

Standardmäßig werden alle verfügbaren Variablen in das neue Dataset aufgenommen. Inaktivieren Sie die, die Sie nicht aufnehmen möchten. Optional können Sie Ihre eigenen Variablennamen angeben. Die Variablennamen müssen den Regeln zum Benennen von Variablen entsprechen.

Ort

RFM aus Transaktionsdaten erstellt stets ein neues aggregiertes Dataset mit einer Zeile je Kunde. Sie können ein neues Dataset in der aktuellen Sitzung erstellen oder die RFM-Scoredaten in einer externen Datendatei speichern. Die Namen von Datasets müssen den Regeln zum Benennen von Variablen entsprechen. (Diese Beschränkung gilt nicht für Namen von externen Datendateien.)

---

## Speichern von RFM-Scores aus Kundendaten

Für Kundendaten können Sie die RFM-Scorevariablen dem aktiven Dataset hinzufügen oder ein neues Dataset erstellen, das die ausgewählten Scorevariablen enthält. Verwenden Sie die Registerkarte "Speichern", um anzugeben, welche Scorevariablen Sie speichern möchten und wo Sie sie speichern möchten.

Name der gespeicherten Variablen

- **Automatisch eindeutige Namen generieren.** Wenn Sie Scorevariablen dem aktiven Dataset hinzufügen, stellt diese Option sicher, dass neue Variablennamen eindeutig sind. Dies ist besonders nützlich, wenn Sie dem aktiven Dataset mehrere unterschiedliche Sets an RFM-Scores (basierend auf unterschiedlichen Kriterien) hinzufügen möchten.
- **Benutzerdefinierte Namen.** Über diese Option können Sie den Scorevariablen Ihre eigenen Variablennamen zuweisen. Die Variablennamen müssen den Regeln zum Benennen von Variablen entsprechen.

Variablen

Wählen (aktivieren) Sie die Scorevariablen, die Sie speichern möchten:

- **Aktualitätsscore.** Der jedem Kunden zugewiesene Score, basierend auf dem Wert des Transaktionsdatums oder der Intervallvariablen, die in der Registerkarte "Variablen" ausgewählt ist. Höhere Scores werden neueren Daten bzw. niedrigeren Intervallwerten zugewiesen.
- **Häufigkeitsscore.** Der jedem Kunden zugewiesene Score, basierend auf der Variablen "Anzahl der Transaktionen", die in der Registerkarte "Variablen" ausgewählt ist. Höhere Scores werden höheren Werten zugewiesen.
- **Geldwertsscore.** Der jedem Kunden zugewiesene Score, basierend auf der Variablen "Betrag", die in der Registerkarte "Variablen" ausgewählt ist. Höhere Scores werden höheren Werten zugewiesen.
- **RFM-Score.** Die drei Einzelscores, zu einem einzigen Wert kombiniert:  $(\text{Aktualität} \times 100) + (\text{Häufigkeit} \times 10) + \text{Geldwert}$ .

Ort

Für Kundendaten gibt es drei Alternativen für den Speicherort von neuen RFM-Scores:

- **Aktives Dataset.** Ausgewählte RFM-Scorevariablen werden dem aktiven Dataset hinzugefügt.
- **Neues Dataset.** Ausgewählte RFM-Scorevariablen und die Schlüsselvariablen, die jeden Kunden (Fall) eindeutig identifizieren, werden in ein neues Dataset in der aktuellen Sitzung geschrieben. Die Namen von Datasets müssen den Regeln zum Benennen von Variablen entsprechen. Diese Option ist nur verfügbar, wenn Sie eine oder mehrere Kunden-ID-Variablen in der Registerkarte "Variablen" wählen.
- **Datei.** Ausgewählte RFM-Scores und die Schlüsselvariablen, die jeden Kunden (Fall) eindeutig identifizieren, werden in einer externen Datendatei gespeichert. Diese Option ist nur verfügbar, wenn Sie eine oder mehrere Kunden-ID-Variablen in der Registerkarte "Variablen" wählen.

---

## RFM-Ausgabe

Klassifizierte Daten

Diagramme und Tabellen für klassierte Daten basieren auf den berechneten Aktualitäts-, Häufigkeits- und Geldwertcores.

**Verteilung des mittleren Geldwerts nach Aktualität und Häufigkeit.** Die Verteilung des mittleren Geldwerts zeigt den durchschnittlichen Geldwert für Kategorien, die durch Aktualitäts- und Häufigkeitsscores definiert sind. Dunklere Bereiche zeigen einen höheren durchschnittlichen Geldwert an.

**Diagramm der Klassenhäufigkeiten.** Das Diagramm der Klassenhäufigkeiten zeigt die Klassierung für die ausgewählten Klassierungsmethoden an. Jeder Balken steht für die Anzahl der Fälle, die jedem kombinierten RFM-Score zugewiesen werden.

- Auch wenn Sie sich in der Regel eine relativ gleichmäßige Verteilung wünschen, bei der alle (oder die meisten) Balken ungefähr die gleiche Höhe haben, sollte eine gewisse Varianz erwartet werden, wenn die Standardklassifizierungsmethode verwendet wird, die gebundene Werte der gleichen Klasse zuweist.
- Extreme Schwankungen in der Klassenverteilung und/oder viele leere Klassen können anzeigen, dass Sie eine andere Klassierungsmethode (weniger Klassen und/oder zufällige Zuweisung von Bindungen) versuchen oder die Eignung der RFM-Analyse überdenken sollten.

**Tabelle der Klassenhäufigkeiten.** Die gleichen Informationen, die sich im Diagramm der Klassenhäufigkeiten finden, nur in Form einer Tabelle mit Klassenhäufigkeiten in jeder Zelle.

Nicht klassierte Daten

Diagramme und Tabellen für nicht klassierte Daten basieren auf den Originalvariablen, die für die Erstellung der Aktualitäts-, Häufigkeits- und Geldwertcores verwendet wurden.

**Histogramme.** Die Histogramme zeigen die relative Verteilung von Werten für die drei Variablen, die für die Berechnung der Aktualitäts-, Häufigkeits- und Geldwertcores verwendet wurden. Diese Histogramme zeigen oftmals etwas verzerrte Verteilungen anstelle einer normalen oder symmetrischen Verteilung an.

Die horizontale Achse jedes Histogramms ist stets von niedrigeren Werten links zu hohen Werten rechts geordnet. Bei der Aktualität hängt jedoch die Interpretation des Diagramms vom Typ der Aktualitätsmessung ab: Daten oder Zeitintervall. Für Daten stellen die Balken links Werte dar, die weiter in der Vergangenheit liegen (ein weniger aktuelles Datum hat einen geringeren Wert als ein aktuelleres Datum). Für Zeitintervalle stellen die Balken links aktuellere Werte dar (je kleiner das Zeitintervall, umso aktueller die Transaktion).

**Streudiagramme von Variablenpaaren.** Diese Streudiagramme zeigen die Beziehungen zwischen den drei Variablen, die für die Berechnung der Aktualitäts-, Häufigkeits- und Geldwertescores verwendet wurden.

Es ist üblich, eine wahrnehmbare lineare Gruppierung der Punkte auf der Häufigkeitsskala festzustellen, da die Häufigkeit oftmals einen relativ kleinen Bereich diskreter Werte darstellt. Wenn zum Beispiel die Gesamtzahl der Transaktionen 15 nicht überschreitet, gibt es nur 15 mögliche Häufigkeitswerte (außer Sie zählen teilweise Transaktionen), während es Hunderte von möglichen Aktualitäts- und Tausende von Geldwerten geben kann.

Die Interpretation der Aktualitätsachsen hängt jedoch vom Typ der Aktualitätsmessung ab: Datum oder Zeitintervall. Bei Daten stellen Punkte näher am Ursprung Daten dar, die weiter in der Vergangenheit liegen. Bei Zeitintervallen stellen Punkte näher am Ursprung aktuellere Werte dar.



## Kapitel 3. Clusteranalyse

Bei der Clusteranalyse handelt es sich um eine explorative Prozedur zum Ermitteln von natürlichen Gruppierungen (Clustern) innerhalb Ihrer Daten. Damit können beispielsweise verschiedene Kundengruppen auf der Basis unterschiedlicher demografischer und Kaufverhaltensmerkmale ausgemacht werden.

**Beispiel.** In Einzel- und Fachhandel werden Clustermethoden regelmäßig auf Daten angewendet, die Kaufgewohnheiten, Geschlecht, Alter und Einkommensniveau der Kundschaft beschreiben. Ziel der Analyse ist eine Ausrichtung der unternehmenseigenen Marketing- und Produktentwicklungsstrategien auf einzelne Konsumentengruppen, um Umsatzsteigerungen und Markentreue zu erreichen.

Erläuterungen der Daten für die Clusteranalyse

**Daten.** Mit dieser Prozedur können sowohl stetige als auch kategoriale Felder analysiert werden. Jeder Datensatz (Zeile) stellt einen Kunden dar, der gruppiert werden soll, während die Felder (Variablen) die Attribute darstellen, auf deren Grundlage die Gruppierung erfolgt.












**Datensatzreihenfolge.** Beachten Sie, dass die Ergebnisse von der Reihenfolge der Datensätze abhängen können. Um die Auswirkungen der Reihenfolge zu minimieren, sollten Sie versuchen, die Datensätze in zufälliger Reihenfolge zu mischen. Prüfen Sie daher die Stabilität einer bestimmten Lösung, indem Sie die Analyse mehrmals durchführen, wobei die Datensätze in einer unterschiedlichen, zufällig ausgewählten Reihenfolge sortiert sind.

**Messniveau.** Es ist wichtig, das korrekte Messniveau zuzuweisen, da sich dieses auf die Berechnung der Ergebnisse auswirkt.

- *Nominal.* Eine Variable kann als nominal behandelt werden, wenn ihre Werte Kategorien darstellen, die sich nicht in eine natürliche Reihenfolge bringen lassen, z. B. die Firmenabteilung, in der eine Person arbeitet. Beispiele für nominale Variablen sind Region, Postleitzahl oder Religionszugehörigkeit.
- *Ordinal.* Eine Variable kann als ordinal behandelt werden, wenn ihre Werte für Kategorien stehen, die eine natürliche Reihenfolge aufweisen (z. B. Grad der Zufriedenheit mit Kategorien von sehr unzufrieden bis sehr zufrieden). Ordinale Variablen treten beispielsweise bei Einstellungsmessungen (Zufriedenheit oder Vertrauen) und bei Präferenzurteilen auf.
- *Stetig.* Eine Variable kann als metrisch (stetig) behandelt werden, wenn ihre Werte geordnete Kategorien mit einer sinnvollen Metrik darstellen, sodass man sinnvolle Aussagen über die Abstände zwischen den Werten machen kann. Metrische Variablen sind beispielsweise Alter (in Jahren) oder Einkommen (in Geldeinheiten).

Ein Symbol neben jedem Feld zeigt das aktuelle Messniveau an.

Tabelle 2. Messniveausymbole

	Numerisch	Zeichenfolge	Datum	Zeit
Metrisch (stetig)		entfällt		
Ordinal				
Nominal				

Sie können das Messniveau in der Variablenansicht des Dateneditors ändern oder das Dialogfeld "Variableneigenschaften definieren" verwenden, um ein geeignetes Messniveau für jedes Feld anzugeben .

Felder mit unbekanntem Messniveau

Der Messniveau-Alert wird angezeigt, wenn das Messniveau für mindestens eine Variable (ein Feld) im Dataset unbekannt ist. Da sich das Messniveau auf die Berechnung der Ergebnisse für diese Prozedur auswirkt, müssen alle Variablen ein definiertes Messniveau aufweisen.

**Daten durchsuchen.** Liest die Daten im aktiven Dataset und weist allen Feldern, deren Messniveau zurzeit nicht bekannt ist, das Standardmessniveau zu. Bei großen Datasets kann dieser Vorgang einige Zeit in Anspruch nehmen.

**Manuell zuweisen.** Öffnet ein Dialogfeld, in dem alle Felder mit unbekanntem Messniveau aufgeführt werden. Mit diesem Dialogfeld können Sie diesen Feldern ein Messniveau zuweisen. Außerdem können Sie in der Variablenansicht des Dateneditors ein Messniveau zuweisen.

Da das Messniveau für diese Prozedur bedeutsam ist, können Sie erst dann auf das Dialogfeld zur Ausführung dieser Prozedur zugreifen, wenn für alle Felder ein Messniveau definiert wurde.

Durchführen eine Clusteranalyse

Wählen Sie die folgenden Befehle aus den Menüs aus:

**Direktmarketing > Verfahren wählen**

1. Wählen Sie **Meine Kontakte in Cluster segmentieren** aus.
2. Wählen Sie die kategorialen (nominalen, ordinalen) und stetigen (metrischen) Felder aus, die Sie zum Erstellen von Segmenten verwenden möchten.
3. Klicken Sie auf **Ausführen**, um die Prozedur auszuführen.

---

## Einstellungen

Auf der Registerkarte "Einstellungen" können Sie Diagramme und Tabellen, die die Segmente beschreiben, anzeigen oder unterdrücken, ein neues Feld im Dataset speichern, das das Segment (Cluster) für jeden Datensatz im Dataset identifiziert, und festlegen, wie viele Segmente die Clusterlösung enthalten soll.

**Diagramme und Tabellen anzeigen.** Zeigt Tabellen und Diagramme an, die die Segmente beschreiben.

**Segmentzugehörigkeit.** Speichert ein neues Feld bzw. eine neue Variable, das bzw. die das Segment identifiziert, zu dem jeder Datensatz gehört.

- Die Feldnamen müssen den Benennungsregeln von IBM® SPSS Statistics entsprechen.
- Der Feldname der Segmentzugehörigkeit kann kein Duplikat eines Feldnamens sein, der bereits im Dataset vorhanden ist. Falls Sie diese Prozedur also mehr als einmal mit demselben Dataset ausführen, müssen Sie jedes Mal einen anderen Namen angeben.
- **Anzahl der Segmente.** Legt fest, wie die Anzahl der Segmente ermittelt wird.
- **Automatisch ermitteln.** Die Prozedur ermittelt automatisch die "beste" Anzahl der Segmente bis zum angegebenen Höchstwert.

**Feste Anzahl angeben.** Die Prozedur erzeugt die angegebene Anzahl der Segmente.



---

## Kapitel 4. Profile über potenzielle Kunden

Bei dieser Technik werden Ergebnisse aus einer früheren Kampagne oder einer Testkampagne verwendet, um beschreibende Profile zu erstellen. Diese Profile können bei zukünftigen Kampagnen für das Targeting bestimmter Gruppen von Kontakten verwendet werden. Das Antwortfeld zeigt, wer auf die frühere Kampagne bzw. die Testkampagne reagiert hat. Die Liste "Profile" enthält die Merkmale, die Sie zur Erstellung des Profils verwenden möchten.

**Beispiel.** Anhand der Ergebnisse einer Testsendung möchte die Marketingabteilung eines Unternehmens auf Basis von demografischen Informationen Profile der Typen von Kunden erstellen, bei denen die Wahrscheinlichkeit einer Antwort auf ein Angebot am höchsten ist.

Ausgabe

Die Ausgabe enthält eine Tabelle, die eine Beschreibung jeder Profilgruppe enthält und in der Rücklaufquoten (Prozentsatz der positiven Antworten), kumulative Rücklaufquoten sowie ein Diagramm der kumulativen Rücklaufquoten angezeigt werden. Wenn Sie eine minimale Zielrücklaufquote einschließen, wird die Tabelle farbcodiert, sodass erkennbar ist, welche Profile der Mindestanforderung an die kumulative Rücklaufquote entsprechen. Das Diagramm enthält eine Bezugslinie, die den Wert der minimalen Rücklaufquote kenntlich macht.

Erläuterung der Daten für Profile über potenzielle Kunden

**Antwortfeld.** Das Antwortfeld muss nominal oder ordinal sein. Es kann ein numerisches Feld oder ein Zeichenfolgefeld sein. Falls dieses Feld einen Wert enthält, der die Anzahl von Käufen anzeigt, müssen Sie ein neues Feld erstellen, in dem ein einzelner Wert sämtliche positiven Antworten repräsentiert. Weitere Informationen finden Sie im Thema „Erstellen eines kategorialen Antwortfelds“ auf Seite 15.

**Wert für positive Antworten.** Der Wert für positive Antworten bezeichnet diejenigen Kunden, die positiv reagiert haben (zum Beispiel, indem sie einen Kauf getätigt haben). Es wird davon ausgegangen, dass alle anderen nicht fehlenden Antwortwerte eine negative Antwort anzeigen. Falls es definierte Wertbeschriftungen für das Antwortfeld gibt, werden diese Beschriftungen in der Dropdown-Liste angezeigt.












**Profile erstellen mit.** Diese Felder können nominal, ordinal oder stetig (metrisch) sein. Es können numerische Felder oder Zeichenfolgefelder sein.

**Messniveau.** Es ist wichtig, das korrekte Messniveau zuzuweisen, da sich dieses auf die Berechnung der Ergebnisse auswirkt.

- *Nominal.* Eine Variable kann als nominal behandelt werden, wenn ihre Werte Kategorien darstellen, die sich nicht in eine natürliche Reihenfolge bringen lassen, z. B. die Firmenabteilung, in der eine Person arbeitet. Beispiele für nominale Variablen sind Region, Postleitzahl oder Religionszugehörigkeit.
- *Ordinal.* Eine Variable kann als ordinal behandelt werden, wenn ihre Werte für Kategorien stehen, die eine natürliche Reihenfolge aufweisen (z. B. Grad der Zufriedenheit mit Kategorien von sehr unzufrieden bis sehr zufrieden). Ordinale Variablen treten beispielsweise bei Einstellungsmessungen (Zufriedenheit oder Vertrauen) und bei Präferenzbeurteilungen auf.
- *Stetig.* Eine Variable kann als metrisch (stetig) behandelt werden, wenn ihre Werte geordnete Kategorien mit einer sinnvollen Metrik darstellen, sodass man sinnvolle Aussagen über die Abstände zwischen den Werten machen kann. Metrische Variablen sind beispielsweise Alter (in Jahren) oder Einkommen (in Geldeinheiten).

Ein Symbol neben jedem Feld zeigt das aktuelle Messniveau an.

Tabelle 3. Messniveausymbole

	Numerisch	Zeichenfolge	Datum	Zeit
Metrisch (stetig)		entfällt		
Ordinal				
Nominal				

Sie können das Messniveau in der Variablenansicht des Dateneditors ändern oder das Dialogfeld "Variableneigenschaften definieren" verwenden, um ein geeignetes Messniveau für jedes Feld anzugeben .

#### Felder mit unbekanntem Messniveau

Der Messniveau-Alert wird angezeigt, wenn das Messniveau für mindestens eine Variable (ein Feld) im Dataset unbekannt ist. Da sich das Messniveau auf die Berechnung der Ergebnisse für diese Prozedur auswirkt, müssen alle Variablen ein definiertes Messniveau aufweisen.

**Daten durchsuchen.** Liest die Daten im aktiven Dataset und weist allen Feldern, deren Messniveau zurzeit nicht bekannt ist, das Standardmessniveau zu. Bei großen Datasets kann dieser Vorgang einige Zeit in Anspruch nehmen.

**Manuell zuweisen.** Öffnet ein Dialogfeld, in dem alle Felder mit unbekanntem Messniveau aufgeführt werden. Mit diesem Dialogfeld können Sie diesen Feldern ein Messniveau zuweisen. Außerdem können Sie in der Variablenansicht des Dateneditors ein Messniveau zuweisen.

Da das Messniveau für diese Prozedur bedeutsam ist, können Sie erst dann auf das Dialogfeld zur Ausführung dieser Prozedur zugreifen, wenn für alle Felder ein Messniveau definiert wurde.

#### Abrufen von Profilen über potenzielle Kunden

Wählen Sie die folgenden Befehle aus den Menüs aus:

#### Direktmarketing > Verfahren wählen

1. Wählen Sie **Profile für die Kontakte erstellen, die auf ein Angebot reagiert haben**.
2. Wählen Sie das Feld aus, das anzeigt, welche Kontakte auf das Angebot reagiert haben. Dieses Feld muss nominal oder ordinal sein.
3. Geben Sie den Wert an, der eine positive Antwort anzeigt. Für Werte mit definierten Wertbeschriftungen können Sie die Wertbeschriftung aus der Dropdown-Liste auswählen, woraufhin der entsprechende Wert angezeigt wird.
4. Wählen Sie die Felder aus, die Sie verwenden möchten, um die Profile zu erstellen.
5. Klicken Sie auf **Ausführen**, um die Prozedur auszuführen.

## Einstellungen

Auf der Registerkarte "Einstellungen" können Sie die Mindestgröße der Profilgruppe angeben und eine minimale Rücklaufquotenschwelle in die Ausgabe einschließen.

**Minimale Profilgruppengröße.** Jedes Profil repräsentiert die gemeinsamen Merkmale einer Gruppe von Kontakten im Dataset (z. B. Frauen unter 40 Jahren, die im Westen leben). Standardmäßig ist 100 der

kleinste Wert für die Größe der Profilgruppe. Kleinere Gruppengrößen können zu einer größeren Anzahl von Gruppen führen, größere Gruppengrößen liefern jedoch verlässlichere Ergebnisse. Dieser Wert muss eine positive Ganzzahl sein.

**Informationen über minimale Rücklaufquotenschwelle in Ergebnissen einschließen.** Die Ergebnisse enthalten eine Tabelle, in der Rücklaufquoten (Prozentsatz der positiven Antworten), kumulative Rücklaufquoten sowie ein Diagramm der kumulativen Rücklaufquoten angezeigt werden. Wenn Sie eine minimale Zielrücklaufquote eingeben, wird die Tabelle farbcodiert, sodass erkennbar ist, welche Profile der Mindestanforderung an die kumulative Rücklaufquote entsprechen. Das Diagramm enthält eine Bezugslinie, die den Wert der minimalen Rücklaufquote kenntlich macht. Der Wert muss größer als 0 und kleiner als 100 sein.

---

## Erstellen eines kategorialen Antwortfelds

Das Antwortfeld sollte kategorial sein, wobei ein Wert alle positiven Reaktionen darstellen sollte. Es wird davon ausgegangen, dass alle anderen nicht fehlenden Werte eine negative Antwort anzeigen. Falls das Antwortfeld einen stetigen (metrischen) Wert enthält, beispielsweise die Anzahl oder den Geldwert der Käufe, müssen Sie ein neues Feld erstellen, das allen von Null abweichenden Antwortwerten eine einzelne positive Antwort zuweist.

- Falls negative Antworten als "0" (nicht leer, was als fehlender Wert aufgefasst wird) aufgezeichnet werden, kann dies anhand der folgenden Formel berechnet werden:

$$\text{neuerName} = \text{alterName} > 0$$

Dabei ist *neuerName* der Name des neuen Felds und *alterName* der Name des alten Felds. Dies ist ein logischer Ausdruck, der allen nicht fehlenden Werten größer 0 einen Wert von 1 und allen nicht fehlenden Werten kleiner oder gleich 0 den Wert 0 zuweist.

- Falls für negative Antworten kein Wert aufgezeichnet wird, werden diese Werte als fehlend behandelt und die Formel ist etwas komplizierter:

$$\text{NewName} = \text{NOT}(\text{MISSING}(\text{OldName}))$$

Bei diesem logischen Ausdruck wird allen nicht fehlenden Antwortwerten ein Wert von 1 und allen fehlenden Antwortwerten ein Wert von 0 zugewiesen.

- Falls Sie zwischen negativen (0) Antwortwerten und fehlenden Werten nicht unterscheiden können, kann kein korrekter Antwortwert berechnet werden. Falls es nur relativ wenig tatsächlich fehlende Werte gibt, muss dies jedoch keine großen Auswirkungen auf die berechneten Rücklaufquoten haben. Falls es jedoch viele fehlende Werte gibt – z. B. wenn die Antwortinformationen nur für eine kleine Teststichprobe des gesamten Datensets berechnet werden –, wird dies dazu führen, dass die berechneten Rücklaufquoten bedeutungslos sind, da sie deutlich niedriger sein werden als die tatsächlichen Rücklaufquoten.

Erstellen eines kategorialen Antwortfelds

1. Wählen Sie die folgenden Befehle aus den Menüs aus:

**Transformieren > Variable berechnen**

2. Geben Sie für "Zielvariable" einen neuen Feldnamen (Variablennamen) ein.
3. Falls negative Reaktionen als 0 aufgezeichnet werden, geben Sie als numerischen Ausdruck  $OldName > 0$  ein, wobei *OldName* der ursprüngliche Feldname ist.
4. Falls negative Reaktionen als fehlend (leer) aufgezeichnet werden, geben Sie als numerischen Ausdruck `NOT(MISSING(OldName))` ein, wobei *OldName* der ursprüngliche Feldname ist.

---

## Kapitel 5. Rücklaufquoten nach Postleitzahlen

Bei dieser Technik werden Ergebnisse aus einer früheren Kampagne verwendet, um Rücklaufquoten nach Postleitzahlen zu berechnen. Diese Raten können bei zukünftigen Kampagnen für das Targeting bestimmter Postleitzahlbereiche verwendet werden. Das Antwortfeld zeigt an, wer auf die frühere Kampagne reagiert hat. Das Postleitzahlfeld kennzeichnet das Feld, das die Postleitzahlen enthält.

**Beispiel.** Anhand der Ergebnisse einer früheren Postsendungskampagne erzeugt die Marketingabteilung eines Unternehmens Rücklaufquoten nach Postleitzahlen. Auf Basis verschiedener Kriterien wie der minimalen akzeptablen Rücklaufquote und/oder der maximalen Anzahl von Kontakten, die in die Postsendungskampagne eingeschlossen werden sollen, können daraufhin bestimmte Postleitzahlbereiche für die Kampagne bestimmt werden.

Ausgabe

Zur Ausgabe dieser Prozedur gehört ein neues Dataset, das die Rücklaufquoten nach Postleitzahl sowie eine Tabelle und ein Diagramm enthält, die die Ergebnisse nach Dezilrang zusammenfassen (oberste 10 %, oberste 20 % usw.). Die Tabelle kann auf Basis einer vom Benutzer festgelegten minimalen kumulativen Rücklaufquote oder maximalen Anzahl von Kontakten farbcodiert werden.

Das neue Dataset enthält folgende Felder:

- **Postleitzahl.** Falls die Postleitzahlengruppen auf nur einem Teil des Gesamtwerts basieren, ist dies der Wert dieses Teils der Postleitzahl. Die Kopfzeilenbeschriftung für diese Spalte in der Excel-Datei ist der Name des Postleitzahlfelds im ursprünglichen Dataset.
- **Rücklaufquote.** Der Prozentsatz der positiven Antworten in jeder Postleitzahlengruppe.
- **Antworten.** Der Anzahl der positiven Antworten in jeder Postleitzahlengruppe.
- **Kontakte.** Die Gesamtanzahl von Kontakten in jedem Postleitzahlbereich, die einen nicht fehlenden Wert für das Antwortfeld enthalten.
- **Index.** Die "gewichtete" Antwort auf der Basis der Formel  $N \times P \times (1-P)$ , wobei  $N$  die Anzahl von Kontakten und  $P$  die als Anteil ausgedrückte Rücklaufquote ist.
- **Rang.** Dezilrang (oberste 10 %, oberste 20 % usw.) der kumulativen Postleitzahlrücklaufquoten in absteigender Reihenfolge.

Erläuterung der Daten für Rücklaufquoten nach Postleitzahlen

**Antwortfeld.** Das Antwortfeld kann ein Zeichenfolgefeld oder ein numerisches Feld sein. Falls dieses Feld einen Wert enthält, der die Anzahl von Käufen oder ihren Geldwert anzeigt, müssen Sie ein neues Feld erstellen, in dem ein einzelner Wert sämtliche positiven Antworten repräsentiert. Weitere Informationen finden Sie im Thema „Erstellen eines kategorialen Antwortfelds“ auf Seite 19.

**Wert für positive Antworten.** Der Wert für positive Antworten bezeichnet diejenigen Kunden, die positiv reagiert haben (zum Beispiel, indem sie einen Kauf getätigt haben). Es wird davon ausgegangen, dass alle anderen nicht fehlenden Antwortwerte eine negative Antwort anzeigen. Falls es definierte Wertbeschriftungen für das Antwortfeld gibt, werden diese Beschriftungen in der Dropdown-Liste angezeigt.

**Postleitzahlenfeld.** Das Postleitzahlfeld kann ein Zeichenfolgefeld oder ein numerisches Feld sein.

Abrufen von Rücklaufquoten nach Postleitzahlen

Wählen Sie die folgenden Befehle aus den Menüs aus:

## Direktmarketing > Verfahren wählen

1. Wählen Sie **Postleitzahlbereiche mit den meisten Antworten identifizieren**.
2. Wählen Sie das Feld aus, das anzeigt, welche Kontakte auf das Angebot reagiert haben.
3. Geben Sie den Wert an, der eine positive Antwort anzeigt. Für Werte mit definierten Wertbeschriftungen können Sie die Wertbeschriftung aus der Dropdown-Liste auswählen, woraufhin der entsprechende Wert angezeigt wird.
4. Wählen Sie das Feld, das die Postleitzahl enthält.
5. Klicken Sie auf **Ausführen**, um die Prozedur auszuführen.

Außerdem sind die folgenden Optionen verfügbar:

- Anstelle des vollständigen Werts können Sie Rücklaufquoten auf Basis der ersten  $n$  Zeichen oder Stellen der Postleitzahl generieren.
- Sie können die Ergebnisse automatisch als Excel-Datei speichern.
- Anzeigeeoptionen der Ausgabe anpassen

---

## Einstellungen

Postleitzahlen gruppieren nach

Dadurch wird festgelegt, wie Datensätze gruppiert werden, um Rücklaufquoten zu berechnen. Standardmäßig wird dazu die gesamte Postleitzahl verwendet und alle Datensätze mit derselben Postleitzahl werden zur Berechnung der Gruppenrücklaufquote gruppiert. Alternativ können Sie Datensätze auch anhand eines Teils der vollständigen Postleitzahl gruppieren, welcher aus den ersten  $n$  Stellen oder Zeichen besteht. Dies ist nützlich, wenn Sie beispielsweise nur die ersten fünf Zeichen einer Postleitzahl aus zehn Zeichen oder die ersten drei Stellen einer fünfstelligen Postleitzahl für die Gruppierung verwenden möchten. Das Ausgabedataset wird einen Datensatz für jede Postleitzahlengruppe enthalten. Falls Sie einen Wert eingeben, muss es sich dabei um eine positive ganze Zahl handeln.

Numerisches Format von Postleitzahlen

Wenn das Postleitzahlfeld numerisch ist und Sie die Postleitzahlen auf Basis der ersten  $n$  Stellen anstatt des Gesamtwerts gruppieren möchten, müssen Sie die Anzahl von Stellen des ursprünglichen Werts angeben. Die Anzahl von Stellen ist die *maximal* mögliche Anzahl von Stellen der Postleitzahl. Falls das Postleitzahlenfeld beispielsweise sowohl fünfstellig als auch neunstellig Postleitzahlen enthält, sollten Sie als Anzahl von Stellen 9 eingeben.

Hinweis: Abhängig vom Anzeigeformat werden manche fünfstelligen Postleitzahlen unter Umständen mit nur vier Stellen angezeigt, wobei aber eine führende Null impliziert ist.

Ausgabe

Neben dem neuen Dataset, das die Rücklaufquoten nach Postleitzahl enthält, können Sie auch eine Tabelle und ein Diagramm anzeigen, die die Ergebnisse nach Dezilrang zusammenfassen (oberste 10 %, oberste 20 % usw.). In der Tabelle werden Rücklaufquoten, kumulative Rücklaufquoten, die Anzahl von Datensätzen sowie die kumulative Anzahl von Datensätzen in jedem Dezil angezeigt. Im Diagramm werden kumulative Rücklaufquoten sowie die kumulative Anzahl von Datensätzen in jedem Dezil angezeigt.

**Akzeptable Mindestrücklaufquote.** Wenn Sie eine akzeptable Mindestrücklaufquote oder eine Break-even-Formel eingeben, wird die Tabelle farbcodiert, sodass erkennbar ist, welche Dezile der Mindestanforderung an die kumulative Rücklaufquote entsprechen. Das Diagramm enthält eine Bezugslinie, die den Wert der Mindestrücklaufquote kenntlich macht.

- **Zielrücklaufquote.** In Prozent ausgedrückte Rücklaufquote (Prozentsatz der positiven Antworten in jeder Postleitzahlengruppe). Der Wert muss größer als 0 und kleiner als 100 sein.

- **Gewinnrate aus Formel berechnen.** Berechnen Sie die minimale kumulative Rücklaufquote anhand dieser Formel:  $(\text{Kosten der Postsendung}/\text{Nettoertrag pro Antwort}) \times 100$ . Beide Werte müssen positive Zahlen sein. Das Ergebnis sollte ein Wert größer 0 und kleiner als 100 sein. Falls die Kosten einer Postsendung beispielsweise 0,75 Euro und der Nettoertrag pro Antwort 56 Euro betragen, beträgt die Mindestrücklaufquote:

**Maximale Anzahl von Kontakten.** Wenn Sie eine maximale Anzahl von Kontakten angeben, wird die Tabelle farbcodiert, sodass erkennbar ist, welche Dezile die kumulative maximale Anzahl von Kontakten (Datensätzen) nicht übersteigen. Das Diagramm enthält eine Bezugslinie, die diesen Wert kenntlich macht.

- **Prozentsatz der Kontakte.** Das in Prozent ausgedrückte Maximum. Dies ist nützlich, wenn Sie beispielsweise die Dezile mit den höchsten Rücklaufquoten ermitteln möchten, die nicht mehr als 50 % aller Kontakte enthalten. Der Wert muss größer als 0 und kleiner als 100 sein.
- **Anzahl der Kontakte.** Das als Anzahl der Kontakte angegebene Maximum. Dies ist nützlich, wenn Sie beispielsweise nicht mehr als 10.000 Sendungen verschicken möchten; in diesem Fall würden Sie den Wert auf 10.000 festlegen. Der Wert muss eine positive ganze Zahl sein (ohne Gruppierungssymbole).

Wenn Sie sowohl eine minimale akzeptable Rücklaufquote als auch eine maximale Anzahl von Kontakten angeben, erfolgt die Farbcodierung der Tabelle abhängig davon, welche Bedingung als erste erfüllt wird.

Nach Excel exportieren

Bei dieser Prozedur wird automatisch ein neues Dataset erstellt, das Rücklaufquoten nach Postleitzahlen enthält. Jeder Datensatz (Zeile) im Dataset steht dabei für eine Postleitzahl. Sie können dieselben Informationen automatisch als Excel-Datei speichern. Sie wird im Format "Excel 97-2003" gespeichert.

---

## Erstellen eines kategorialen Antwortfelds

Das Antwortfeld sollte kategorial sein, wobei ein Wert alle positiven Reaktionen darstellen sollte. Es wird davon ausgegangen, dass alle anderen nicht fehlenden Werte eine negative Antwort anzeigen. Falls das Antwortfeld einen stetigen (metrischen) Wert enthält, beispielsweise die Anzahl oder den Geldwert der Käufe, müssen Sie ein neues Feld erstellen, das allen von Null abweichenden Antwortwerten eine einzelne positive Antwort zuweist.

- Falls negative Antworten als "0" (nicht leer, was als fehlender Wert aufgefasst wird) aufgezeichnet werden, kann dies anhand der folgenden Formel berechnet werden:

`neuerName=alterName>0`

Dabei ist *neuerName* der Name des neuen Felds und *alterName* der Name des alten Felds. Dies ist ein logischer Ausdruck, der allen nicht fehlenden Werten größer 0 einen Wert von 1 und allen nicht fehlenden Werten kleiner oder gleich 0 den Wert 0 zuweist.

- Falls für negative Antworten kein Wert aufgezeichnet wird, werden diese Werte als fehlend behandelt und die Formel ist etwas komplizierter:

`NewName=NOT(MISSING(OldName))`

Bei diesem logischen Ausdruck wird allen nicht fehlenden Antwortwerten ein Wert von 1 und allen fehlenden Antwortwerten ein Wert von 0 zugewiesen.

- Falls Sie zwischen negativen (0) Antwortwerten und fehlenden Werten nicht unterscheiden können, kann kein korrekter Antwortwert berechnet werden. Falls es nur relativ wenig tatsächlich fehlende Werte gibt, muss dies jedoch keine großen Auswirkungen auf die berechneten Rücklaufquoten haben. Falls es jedoch viele fehlende Werte gibt – z. B. wenn die Antwortinformationen nur für eine kleine Teststichprobe des gesamten Datasets berechnet werden –, wird dies dazu führen, dass die berechneten Rücklaufquoten bedeutungslos sind, da sie deutlich niedriger sein werden als die tatsächlichen Rücklaufquoten.

Erstellen eines kategorialen Antwortfelds

1. Wählen Sie die folgenden Befehle aus den Menüs aus:

### **Transformieren > Variable berechnen**

2. Geben Sie für "Zielvariable" einen neuen Feldnamen (Variablennamen) ein.
3. Falls negative Reaktionen als 0 aufgezeichnet werden, geben Sie als numerischen Ausdruck  $OldName > 0$  ein, wobei *OldName* der ursprüngliche Feldname ist.
4. Falls negative Reaktionen als fehlend (leer) aufgezeichnet werden, geben Sie als numerischen Ausdruck `NOT(MISSING(OldName))` ein, wobei *OldName* der ursprüngliche Feldname ist.



---

## Kapitel 6. Kaufneigung

Für die Kaufneigung werden Ergebnisse einer Testsendung oder einer früheren Kampagne verwendet, um Scores zu erstellen. Die Scores zeigen an, bei welchen Kontakten die Wahrscheinlichkeit einer Antwort am höchsten ist. Das Antwortfeld zeigt, wer auf die Testsendung oder die frühere Kampagne reagiert hat. Die Neigungsfelder sind die Merkmale, die Sie verwenden, um die Wahrscheinlichkeit einer Antwort seitens Kontakten mit ähnlichen Eigenschaften vorherzusagen.

Diese Technik verwendet die binäre logistische Regression für den Aufbau eines Vorhersagemodells. Der Prozess des Aufbaus und der Anwendung eines Vorhersagemodells umfasst die folgenden beiden Schritte:

1. Erstellen des Modells und Speichern der Modelldatei. Sie erstellen das Modell mithilfe eines Datensets, für das das relevante Ergebnis (oft als **Ziel**) bezeichnet) bekannt ist. Wenn Sie beispielsweise ein Modell erstellen möchten, mit dem vorhergesagt wird, welche Personen vermutlich auf eine Direktmailing-Aktion reagieren, müssen Sie mit einem Dataset beginnen, das bereits Informationen über die Personen enthält, die reagierten und die nicht reagierten. Dabei kann es sich beispielsweise um die Ergebnisse einer Testsendung an eine kleine Gruppe von Kunden oder um Informationen zu Reaktionen auf eine ähnliche Kampagne in der Vergangenheit handeln.
2. Anwenden des Modells auf ein anderes Dataset (für das das relevante Ergebnis nicht bekannt ist), um die vorhergesagten Ergebnisse zu ermitteln.

**Beispiel.** Die Direktmarketing-Abteilung eines Unternehmens verwendet die Ergebnisse einer Testsendung, um den übrigen Kontakten in ihrer Datenbank Propensity-Scores zuzuweisen, wobei verschiedene demografische Merkmale eingesetzt werden, um Kontakte zu ermitteln, bei denen die Wahrscheinlichkeit einer Antwort und eines Kaufs am größten ist.

Ausgabe

Mit diesem Verfahren wird automatisch ein neues Feld im Dataset erstellt, das Propensity-Scores für die Testdaten und eine XML-Modelldatei enthält, die zur Bewertung anderer Datensets verwendet werden kann. In der optionalen Diagnoseausgabe sind ein Diagramm zur Modellgesamtqualität sowie eine Klassifikationstabelle enthalten, die vorhergesagte Antworten mit tatsächlichen Antworten vergleicht.

Erläuterung der Daten zur Kaufneigung

**Antwortfeld.** Das Antwortfeld kann ein Zeichenfolgefeld oder ein numerisches Feld sein. Falls dieses Feld einen Wert enthält, der die Anzahl von Käufen oder ihren Geldwert anzeigt, müssen Sie ein neues Feld erstellen, in dem ein einzelner Wert sämtliche positiven Antworten repräsentiert. Weitere Informationen finden Sie im Thema „Erstellen eines kategorialen Antwortfelds“ auf Seite 24.

**Wert für positive Antworten.** Der Wert für positive Antworten bezeichnet diejenigen Kunden, die positiv reagiert haben (zum Beispiel, indem sie einen Kauf getätigt haben). Es wird davon ausgegangen, dass alle anderen nicht fehlenden Antwortwerte eine negative Antwort anzeigen. Falls es definierte Wertbeschriftungen für das Antwortfeld gibt, werden diese Beschriftungen in der Dropdown-Liste angezeigt.












**Kaufneigung vorhersagen mit.** Die Felder, die verwendet werden, um die Neigung vorherzusagen, können Zeichenfolgefelder oder numerische Felder und außerdem nominal, ordinal oder stetig (metrisch) sein – es ist jedoch wichtig, allen Feldern für Prädiktoren das geeignete Messniveau zuzuweisen.

**Messniveau.** Es ist wichtig, das korrekte Messniveau zuzuweisen, da sich dieses auf die Berechnung der Ergebnisse auswirkt.

- *Nominal*. Eine Variable kann als nominal behandelt werden, wenn ihre Werte Kategorien darstellen, die sich nicht in eine natürliche Reihenfolge bringen lassen, z. B. die Firmenabteilung, in der eine Person arbeitet. Beispiele für nominale Variablen sind Region, Postleitzahl oder Religionszugehörigkeit.
- *Ordinal*. Eine Variable kann als ordinal behandelt werden, wenn ihre Werte für Kategorien stehen, die eine natürliche Reihenfolge aufweisen (z. B. Grad der Zufriedenheit mit Kategorien von sehr unzufrieden bis sehr zufrieden). Ordinale Variablen treten beispielsweise bei Einstellungsmessungen (Zufriedenheit oder Vertrauen) und bei Präferenzbeurteilungen auf.
- *Stetig*. Eine Variable kann als metrisch (stetig) behandelt werden, wenn ihre Werte geordnete Kategorien mit einer sinnvollen Metrik darstellen, sodass man sinnvolle Aussagen über die Abstände zwischen den Werten machen kann. Metrische Variablen sind beispielsweise Alter (in Jahren) oder Einkommen (in Geldeinheiten).

Ein Symbol neben jedem Feld zeigt das aktuelle Messniveau an.

Tabelle 4. Messniveausymbole

	Numerisch	Zeichenfolge	Datum	Zeit
Metrisch (stetig)		entfällt		
Ordinal				
Nominal				

Sie können das Messniveau in der Variablenansicht des Dateneditors ändern oder das Dialogfeld "Variableneigenschaften definieren" verwenden, um ein geeignetes Messniveau für jedes Feld anzugeben .

#### Felder mit unbekanntem Messniveau

Der Messniveau-Alert wird angezeigt, wenn das Messniveau für mindestens eine Variable (ein Feld) im Dataset unbekannt ist. Da sich das Messniveau auf die Berechnung der Ergebnisse für diese Prozedur auswirkt, müssen alle Variablen ein definiertes Messniveau aufweisen.

**Daten durchsuchen.** Liest die Daten im aktiven Dataset und weist allen Feldern, deren Messniveau zurzeit nicht bekannt ist, das Standardmessniveau zu. Bei großen Datasets kann dieser Vorgang einige Zeit in Anspruch nehmen.

**Manuell zuweisen.** Öffnet ein Dialogfeld, in dem alle Felder mit unbekanntem Messniveau aufgeführt werden. Mit diesem Dialogfeld können Sie diesen Feldern ein Messniveau zuweisen. Außerdem können Sie in der Variablenansicht des Dateneditors ein Messniveau zuweisen.

Da das Messniveau für diese Prozedur bedeutsam ist, können Sie erst dann auf das Dialogfeld zur Ausführung dieser Prozedur zugreifen, wenn für alle Felder ein Messniveau definiert wurde.

#### Abrufen von Kaufneigungsscores

Wählen Sie die folgenden Befehle aus den Menüs aus:

#### Direktmarketing > Verfahren wählen

1. Wählen Sie **Kontakte mit höchster Kaufneigung auswählen**.
2. Wählen Sie das Feld aus, das anzeigt, welche Kontakte auf das Angebot reagiert haben.

3. Geben Sie den Wert an, der eine positive Antwort anzeigt. Für Werte mit definierten Wertbeschriftungen können Sie die Wertbeschriftung aus der Dropdown-Liste auswählen, woraufhin der entsprechende Wert angezeigt wird.
4. Wählen Sie die Felder aus, die Sie verwenden möchten, um die Neigung vorherzusagen.  
So speichern Sie eine XML-Modelldatei zur Bewertung anderer Datendateien:
5. Aktivieren Sie die Option **Modellinformation in XML-Datei exportieren**.
6. Geben Sie einen Verzeichnispfad und einen Dateinamen ein oder klicken Sie auf **Durchsuchen**, um zu dem Speicherort zu navigieren, unter dem Sie die XML-Modelldatei speichern möchten.
7. Klicken Sie auf **Ausführen**, um die Prozedur auszuführen.  
So verwenden Sie eine Modelldatei zur Bewertung anderer Datasets:
8. Öffnen Sie das zu bewertende Dataset.
9. Verwenden Sie den Scoring-Assistenten, um das Modell auf das Dataset anzuwenden. Wählen Sie die folgenden Befehle aus den Menüs aus:  
**Dienstprogramme > Scoring-Assistent**

---

## Einstellungen

### Modellvalidierung

Bei der Modellvalidierung werden zu Diagnosezwecken Trainings- und Testgruppen erstellt. Falls Sie die Klassifikationstabelle im Abschnitt "Diagnoseausgabe" auswählen, wird die Tabelle zu Vergleichszwecken in (ausgewählte) Trainings- und (nicht ausgewählte) Testabschnitte unterteilt. Wählen Sie die Modellvalidierung nur aus, wenn Sie auch die Klassifikationstabelle auswählen. Die Scores erfolgen auf Basis des Modells, das aus der Trainingsstichprobe erstellt wurde, deren Anzahl enthaltener Datensätze stets niedriger als die Gesamtanzahl verfügbarer Datensätze ist. Ein Beispiel: Die Standardgröße für Trainingsstichproben ist 50 %, und ein Modell, das auf Basis der Hälfte aller verfügbaren Datensätze erstellt wird, kann nicht so zuverlässig sein, wie ein Modell auf Basis aller verfügbaren Datensätze.

- **Partitionsgröße der Trainingsstichprobe (%)**. Legen Sie den Prozentsatz der Datensätze fest, die der Trainingsstichprobe zugewiesen werden sollen. Die übrigen Datensätze mit nicht fehlenden Werten für das Antwortfeld werden der Teststichprobe zugewiesen. Der Wert muss größer als 0 und kleiner als 100 sein.
- **Startwert zur Replikation von Ergebnissen festlegen**. Da die Zuweisung von Datensätzen zu den Trainings- und Teststichproben auf Zufallsbasis geschieht, erhalten Sie unter Umständen bei jeder Durchführung der Prozedur unterschiedliche Ergebnisse, es sei denn, Sie geben jedes Mal denselben Startwert für Zufallszahlen an.

### Diagnoseausgabe

**Modellgesamtqualität**. Zeigt ein Balkendiagramm der Modellgesamtqualität an, die als ein Wert zwischen 0 und 1 ausgedrückt wird. Ein gutes Modell sollte einen Wert größer 0,5 aufweisen.

**Klassifikationstabelle**. Zeigt eine Matrix an, die die vorhergesagten positiven und negativen Antworten mit den tatsächlichen positiven und negativen Antworten vergleicht. Die Gesamtgenauigkeitsrate kann Aufschluss darüber geben, die gut das Modell funktioniert, aber möglicherweise interessieren Sie sich mehr für den Prozentsatz korrekt vorhergesagter positiver Antworten.

- **Minimale Wahrscheinlichkeit**. Weist der Kategorie für vorhergesagte positive Antworten in der Klassifikationstabelle Datensätze mit einem Scorewert zu, der höher als der angegebene Wert ist. Die Scores, die durch die Prozedur erstellt werden, stehen für die Wahrscheinlichkeit, dass der Kontakt positiv reagieren wird (zum Beispiel indem er einen Kauf tätigt). Allgemein sollten Sie einen Wert angeben, der in der Nähe Ihrer minimalen, als Anteil ausgedrückten Zielrücklaufquote liegt. Falls Sie zum Beispiel an einer Rücklaufquote von mindestens 5 % interessiert sind, geben Sie 0,05 an. Der Wert muss größer als 0 und kleiner als 1 sein.

## Name und Beschriftung des umcodierten Antwortfelds

Dieses Verfahren codiert das Antwortfeld automatisch in ein neues Feld um, in dem "1" positiven Antworten und "0" negativen Antworten entspricht. Die Analyse wird für das umcodierte Feld durchgeführt. Sie können den Standardnamen und die Standardbeschriftung durch eigene Angaben ersetzen. Die Namen müssen den Benennungsregeln von IBM SPSS Statistics entsprechen.

## Scores speichern

Im ursprünglichen Dataset wird automatisch ein neues Feld mit Propensity-Scores gespeichert. Die Scores stehen für die Wahrscheinlichkeit einer positiven Antwort, welche als Anteil ausgedrückt wird.

- Die Feldnamen müssen den Benennungsregeln von IBM SPSS Statistics entsprechen.
- Der Feldname kann kein Duplikat eines Feldnamens sein, der bereits im Dataset vorhanden ist. Falls Sie diese Prozedur also mehr als einmal mit demselben Dataset ausführen, müssen Sie jedes Mal einen anderen Namen angeben.

---

## Erstellen eines kategorialen Antwortfelds

Das Antwortfeld sollte kategorial sein, wobei ein Wert alle positiven Reaktionen darstellen sollte. Es wird davon ausgegangen, dass alle anderen nicht fehlenden Werte eine negative Antwort anzeigen. Falls das Antwortfeld einen stetigen (metrischen) Wert enthält, beispielsweise die Anzahl oder den Geldwert der Käufe, müssen Sie ein neues Feld erstellen, das allen von Null abweichenden Antwortwerten eine einzelne positive Antwort zuweist.

- Falls negative Antworten als "0" (nicht leer, was als fehlender Wert aufgefasst wird) aufgezeichnet werden, kann dies anhand der folgenden Formel berechnet werden:

$$\text{neuerName} = \text{alterName} > 0$$

Dabei ist *neuerName* der Name des neuen Felds und *alterName* der Name des alten Felds. Dies ist ein logischer Ausdruck, der allen nicht fehlenden Werten größer 0 einen Wert von 1 und allen nicht fehlenden Werten kleiner oder gleich 0 den Wert 0 zuweist.

- Falls für negative Antworten kein Wert aufgezeichnet wird, werden diese Werte als fehlend behandelt und die Formel ist etwas komplizierter:

$$\text{NewName} = \text{NOT}(\text{MISSING}(\text{OldName}))$$

Bei diesem logischen Ausdruck wird allen nicht fehlenden Antwortwerten ein Wert von 1 und allen fehlenden Antwortwerten ein Wert von 0 zugewiesen.

- Falls Sie zwischen negativen (0) Antwortwerten und fehlenden Werten nicht unterscheiden können, kann kein korrekter Antwortwert berechnet werden. Falls es nur relativ wenig tatsächlich fehlende Werte gibt, muss dies jedoch keine großen Auswirkungen auf die berechneten Rücklaufquoten haben. Falls es jedoch viele fehlende Werte gibt – z. B. wenn die Antwortinformationen nur für eine kleine Teststichprobe des gesamten Datensets berechnet werden –, wird dies dazu führen, dass die berechneten Rücklaufquoten bedeutungslos sind, da sie deutlich niedriger sein werden als die tatsächlichen Rücklaufquoten.

## Erstellen eines kategorialen Antwortfelds

1. Wählen Sie die folgenden Befehle aus den Menüs aus:

**Transformieren > Variable berechnen**

2. Geben Sie für "Zielvariable" einen neuen Feldnamen (Variablennamen) ein.
3. Falls negative Reaktionen als 0 aufgezeichnet werden, geben Sie als numerischen Ausdruck  $\text{OldName} > 0$  ein, wobei *OldName* der ursprüngliche Feldname ist.
4. Falls negative Reaktionen als fehlend (leer) aufgezeichnet werden, geben Sie als numerischen Ausdruck  $\text{NOT}(\text{MISSING}(\text{OldName}))$  ein, wobei *OldName* der ursprüngliche Feldname ist.

---

## Kapitel 7. Kontrollpakettest

Dieses Verfahren vergleicht Marketingkampagnen, um herauszufinden, ob es hinsichtlich der Effektivität signifikante Unterschiede zwischen verschiedenen Paketen oder Angeboten gibt. Die Kampagnenwirksamkeit wird anhand von Antworten gemessen. Das Kampagnenfeld identifiziert unterschiedliche Kampagnen, zum Beispiel Angebot A und Angebot B. Das Antwortfeld zeigt an, wenn ein Kontakt auf die Kampagne geantwortet hat. Wählen Sie "Kaufbetrag" aus, wenn die Antwort als Kaufbetrag aufgezeichnet wird, zum Beispiel "99,99". Wählen Sie "Antwort" aus, wenn die Antwort nur angibt, ob der Kontakt positiv reagiert hat oder nicht, zum Beispiel "Ja" oder "Nein".

**Beispiel.** Die Direktmarketing-Abteilung eines Unternehmens möchte herausfinden, ob eine neue Verpackungsgestaltung mehr positive Antworten erzeugt als die bestehende Verpackung. Daher verschicken sie Testsendungen, um zu ermitteln, ob die neue Verpackung eine deutlich höhere positive Rücklaufquote erzeugt. Die Testsendung besteht aus einer Kontrollgruppe, die die aktuelle Verpackung erhält, und einer Testgruppe, an die die neue Verpackungsgestaltung geschickt wird. Die Ergebnisse der zwei Gruppen werden dann miteinander verglichen, um zu sehen, ob ein deutlicher Unterschied besteht.

Ausgabe

Die Ausgabe enthält eine Tabelle, in der Häufigkeiten und Prozentwerte von positiven und negativen Antworten für jede anhand des Kampagnenfelds definierte Gruppe sowie eine Tabelle, in der festgehalten wird, welche Gruppen stark voneinander abweichen.

Erläuterungen und Annahmen der Daten des Kontrollpakettests

**Kampagnenfeld.** Das Kampagnenfeld sollte kategorial (nominal oder ordinal) sein.

**Wirksamkeitsantwortfeld.** Wenn Sie für das Wirksamkeitsfeld "Kaufbetrag" auswählen, muss das Feld numerisch sein und das Messniveau sollte stetig (metrisch) sein.

Falls Sie nicht zwischen negativen (für den Kaufbetrag ein Wert von 0) Antwortwerten und fehlenden Werten unterscheiden können, kann keine korrekte Rücklaufquote berechnet werden. Falls es nur relativ wenig tatsächlich fehlende Werte gibt, muss dies jedoch keine großen Auswirkungen auf die berechneten Rücklaufquoten haben. Falls es jedoch viele fehlende Werte gibt – z. B. wenn die Antwortinformationen nur für eine kleine Teststichprobe des gesamten Datensets berechnet werden –, wird dies dazu führen, dass die berechneten Rücklaufquoten bedeutungslos sind, da sie deutlich niedriger sein werden als die tatsächlichen Rücklaufquoten.

**Annahmen.** Diese Prozedur geht davon aus, dass jeder Kampagnengruppe zufällig Kontakte zugewiesen wurden. Anders ausgedrückt besteht keine spezielle Gruppenzuweisung hinsichtlich Demografie, Kaufverlauf oder anderen Merkmalen und bei allen Kontakten ist die Wahrscheinlichkeit, einer beliebigen Gruppe zugewiesen zu werden, gleich hoch.

Durchführen einen Kontrollpakettests

Wählen Sie die folgenden Befehle aus den Menüs aus:

**Direktmarketing > Verfahren wählen**

1. Wählen Sie **Wirksamkeit der Kampagnen vergleichen** aus.
2. Wählen Sie das Feld aus, das anzeigt, zu welcher Kampagnengruppe jeder Kontakt gehört (zum Beispiel Angebot A, Angebot B etc.). Dieses Feld muss nominal oder ordinal sein.
3. Wählen Sie das Feld aus, das die Antwortwirksamkeit anzeigt.

Wenn das Antwortfeld ein Kaufbetrag ist, muss das Feld numerisch sein.

Wählen Sie **Antwort** aus, wenn das Antwortfeld nur angibt, ob der Kontakt positiv reagiert hat oder nicht (zum Beispiel "Ja" oder "Nein"), und geben Sie den Wert ein, der eine positive Antwort darstellt. Für Werte mit definierten Wertbeschriftungen können Sie die Wertbeschriftung aus der Dropdown-Liste auswählen, woraufhin der entsprechende Wert angezeigt wird.

Es wird automatisch ein neues Feld erstellt, in dem 1 positiven Antworten und 0 negativen Antworten entspricht; die Analyse wird in dem neuen Feld durchgeführt. Sie können den Standardnamen und die Standardbeschriftung durch eigene Angaben ersetzen. Die Namen müssen den Benennungsregeln von IBM SPSS Statistics entsprechen.

4. Klicken Sie auf **Ausführen**, um die Prozedur auszuführen.

---

## Bemerkungen

Die vorliegenden Informationen wurden für Produkte und Services entwickelt, die auf dem deutschen Markt angeboten werden.

Möglicherweise bietet IBM die in dieser Dokumentation beschriebenen Produkte, Services oder Funktionen in anderen Ländern nicht an. Informationen über die gegenwärtig im jeweiligen Land verfügbaren Produkte und Services sind beim zuständigen IBM Ansprechpartner erhältlich. Hinweise auf IBM Lizenzprogramme oder andere IBM Produkte bedeuten nicht, dass nur Programme, Produkte oder Services von IBM verwendet werden können. Anstelle der IBM Produkte, Programme oder Services können auch andere, ihnen äquivalente Produkte, Programme oder Services verwendet werden, solange diese keine gewerblichen oder anderen Schutzrechte von IBM verletzen. Die Verantwortung für den Betrieb von Produkten, Programmen und Services anderer Anbieter liegt beim Kunden.

Für in diesem Handbuch beschriebene Erzeugnisse und Verfahren kann es IBM Patente oder Patentanmeldungen geben. Mit der Auslieferung dieses Handbuchs ist keine Lizenzierung dieser Patente verbunden. Lizenzanforderungen sind schriftlich an folgende Adresse zu richten (Anfragen an diese Adresse müssen auf Englisch formuliert werden):

IBM Director of Licensing  
IBM Europe, Middle East & Africa  
Tour Descartes  
2, avenue Gambetta  
92066 Paris La Defense  
France

Trotz sorgfältiger Bearbeitung können technische Ungenauigkeiten oder Druckfehler in dieser Veröffentlichung nicht ausgeschlossen werden. Die hier enthaltenen Informationen werden in regelmäßigen Zeitabständen aktualisiert und als Neuausgabe veröffentlicht. IBM kann ohne weitere Mitteilung jederzeit Verbesserungen und/oder Änderungen an den in dieser Veröffentlichung beschriebenen Produkten und/oder Programmen vornehmen.

Verweise in diesen Informationen auf Websites anderer Anbieter werden lediglich als Service für den Kunden bereitgestellt und stellen keinerlei Billigung des Inhalts dieser Websites dar. Das über diese Websites verfügbare Material ist nicht Bestandteil des Materials für dieses IBM Produkt. Die Verwendung dieser Websites geschieht auf eigene Verantwortung.

Werden an IBM Informationen eingesandt, können diese beliebig verwendet werden, ohne dass eine Verpflichtung gegenüber dem Einsender entsteht.

Lizenznehmer des Programms, die Informationen zu diesem Produkt wünschen mit der Zielsetzung: (i) den Austausch von Informationen zwischen unabhängig voneinander erstellten Programmen und anderen Programmen (einschließlich des vorliegenden Programms) sowie (ii) die gemeinsame Nutzung der ausgetauschten Informationen zu ermöglichen, wenden sich an folgende Adresse:

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

Die Bereitstellung dieser Informationen kann unter Umständen von bestimmten Bedingungen - in einigen Fällen auch von der Zahlung einer Gebühr - abhängig sein.

Die Lieferung des in diesem Dokument beschriebenen Lizenzprogramms sowie des zugehörigen Lizenzmaterials erfolgt auf der Basis der IBM Rahmenvereinbarung bzw. der Allgemeinen Geschäftsbedingungen von IBM, der IBM Internationalen Nutzungsbedingungen für Programmpakete oder einer äquivalenten Vereinbarung.

Alle in diesem Dokument enthaltenen Leistungsdaten stammen aus einer kontrollierten Umgebung. Die Ergebnisse, die in anderen Betriebsumgebungen erzielt werden, können daher erheblich von den hier erzielten Ergebnissen abweichen. Einige Daten stammen möglicherweise von Systemen, deren Entwicklung noch nicht abgeschlossen ist. Eine Gewährleistung, dass diese Daten auch in allgemein verfügbaren Systemen erzielt werden, kann nicht gegeben werden. Darüber hinaus wurden einige Daten unter Umständen durch Extrapolation berechnet. Die tatsächlichen Ergebnisse können davon abweichen. Benutzer dieses Dokuments sollten die entsprechenden Daten in ihrer spezifischen Umgebung prüfen.

Alle Informationen zu Produkten anderer Anbieter stammen von den Anbietern der aufgeführten Produkte, deren veröffentlichten Ankündigungen oder anderen allgemein verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann daher keine Aussagen zu Leistung, Kompatibilität oder anderen Merkmalen machen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter sind an den jeweiligen Anbieter zu richten.

Aussagen über Pläne und Absichten von IBM unterliegen Änderungen oder können zurückgenommen werden und repräsentieren nur die Ziele von IBM.

Diese Veröffentlichung enthält Beispiele für Daten und Berichte des alltäglichen Geschäftsablaufs. Sie sollen nur die Funktionen des Lizenzprogramms illustrieren und können Namen von Personen, Firmen, Marken oder Produkten enthalten. Alle diese Namen sind frei erfunden; Ähnlichkeiten mit tatsächlichen Namen und Adressen sind rein zufällig.

Kopien oder Teile der Beispielprogramme bzw. daraus abgeleiteter Code müssen folgenden Copyrightvermerk beinhalten:

Diese Veröffentlichung enthält Beispiele für Daten und Berichte des alltäglichen Geschäftsablaufs. Sie sollen nur die Funktionen des Lizenzprogramms illustrieren und können Namen von Personen, Firmen, Marken oder Produkten enthalten. Alle diese Namen sind frei erfunden; Ähnlichkeiten mit tatsächlichen Namen und Adressen sind rein zufällig.

Kopien oder Teile der Beispielprogramme bzw. daraus abgeleiteter Code müssen folgenden Copyrightvermerk beinhalten:

© (Name Ihrer Firma) (Jahr). Teile des vorliegenden Codes wurden aus Beispielprogrammen der IBM Corp. abgeleitet.

© Copyright IBM Corp. \_Jahr/Jahre angeben\_. Alle Rechte vorbehalten.



---

## Marken

IBM, das IBM Logo und ibm.com sind Marken oder eingetragene Marken der IBM Corporation in den USA und/oder anderen Ländern. Weitere Produkt- und Servicenamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM Marken finden Sie auf der Webseite „Copyright and trademark information“ unter [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Adobe, das Adobe-Logo, PostScript und das PostScript-Logo sind Marken oder eingetragene Marken der Adobe Systems Incorporated in den USA und/oder anderen Ländern.

Intel, das Intel-Logo, Intel Inside, das Intel Inside-Logo, Intel Centrino, das Intel Centrino-Logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium und Pentium sind Marken oder eingetragene Marken der Intel Corporation oder ihrer Tochtergesellschaften in den USA oder anderen Ländern.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Java und alle auf Java basierenden Marken und Logos sind Marken oder eingetragene Marken der Oracle Corporation und/oder ihrer verbundenen Unternehmen.



---

# Index

## C

Cluster 11  
Clusteranalyse 11  
Clusteranalyse (Direktmarketing) 11

## K

Kaufneigung 21  
Kontrollpakettest 25

## L

Logistische Regression (Direktmarketing) 21

## P

Profile über potenzielle Kunden (Option "Direktmarketing") 13

## R

RFM 3, 6, 7, 8  
    Klassierung 5  
    Kundendaten 4  
    Transaktionsdaten 3  
Rücklaufquoten nach Postleitzahlen 17





