

*IBM SPSS Missing Values 22*

**IBM**

**Hinweis**

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 23 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

## **Kapitel 1. Missing Values (Fehlende Werte) - Einleitung . . . . . 1**

## **Kapitel 2. Analyse fehlender Werte . . . . . 3**

Anzeigen der Muster fehlender Werte . . . . .	4
Anzeigen deskriptiver Statistiken für fehlende Werte	5
Schätzen von Statistiken und Imputieren fehlender Werte. . . . .	6
EM-Schätzung: Optionen . . . . .	7
Optionen für die Regressionsschätzung . . . . .	8
Vorhergesagte Variablen und Prädiktorvariablen . . . . .	9
Zusätzliche Funktionen beim Befehl MVA . . . . .	9

## **Kapitel 3. Multiple Imputation . . . . . 11**

Analysieren von Mustern. . . . .	12
----------------------------------	----

Imputieren fehlender Datenwerte . . . . .	13
Methode . . . . .	14
Nebenbedingungen. . . . .	15
Ausgabe . . . . .	16
Zusätzliche Funktionen beim Befehl MULTIPLE IMPUTATION . . . . .	17
Arbeiten mit Daten aus multipler Imputation . . . . .	17
Analysieren von Daten multipler Imputation . . . . .	18
Optionen für multiple Imputation . . . . .	21

## **Bemerkungen. . . . . 23**

Marken. . . . .	25
-----------------	----

## **Index . . . . . 27**



---

## Kapitel 1. Missing Values (Fehlende Werte) - Einleitung

Fälle mit fehlenden Werten stellen eine Herausforderung dar, da typische Modellverfahren diese Fälle einfach von der Analyse ausschließen. Wenn es wenige fehlende Werte (grob geschätzt weniger als 5 % der Gesamtzahl an Fällen) gibt und diese Werte als zufällig fehlend betrachtet werden, also das Fehlen eines Werts nicht von anderen Werten abhängt, dann ist die typische Methode des listenweisen Ausschlusses relativ sicher. Die Option "Missing Values+" kann Ihnen helfen zu bestimmen, ob der listenweise Ausschluss ausreichend ist, und stellt andernfalls Methoden zur Handhabung fehlender Werte bereit.

Die Analyse fehlender Werte im Vergleich zu Verfahren multipler Imputation

Die Option "Missing Values" stellt zwei Arten von Verfahren für die Handhabung fehlender Werte bereit:

- Die Verfahren der multiplen Imputation ermöglichen die Analyse von Mustern fehlender Daten und zielen auf eine eventuelle multiple Imputation der fehlenden Werte ab. Es werden mehrere Versionen des Datasets erzeugt, von denen jede ein eigenes Set an imputierten Werten enthält. Wenn statistische Analysen durchgeführt werden, werden die Parameterschätzungen für alle imputierten Datasets gesammelt. Sie stellen Schätzungen bereit, die im Allgemeinen genauer als die einzelner Imputationen sind.
- Die Analyse fehlender Werte stellt ein geringfügig anderes Set an beschreibenden Tools für die Analyse fehlender Daten bereit (insbesondere den MCAR-Test von Little) und umfasst eine Vielzahl einfacher Imputationsmethoden. Beachten Sie, dass die multiple Imputation im Allgemeinen als der einzelnen Imputation überlegen betrachtet wird.

Aufgaben fehlender Werte

Sie können mit der Analyse fehlender Wert anhand der folgenden grundlegenden Schritte beginnen:

1. **Untersuchen Sie das Fehlen.** Verwenden Sie die Analyse fehlender Werte und die Analyse von Mustern, um die Muster der fehlenden Werte in Ihren Daten zu untersuchen und zu bestimmen, ob eine multiple Imputation erforderlich ist.
2. **Imputieren Sie fehlende Werte.** Verwenden Sie "Fehlende Datenwerte imputieren", um imputierte fehlende Werte zu multiplizieren.
3. **Analysieren Sie die "vollständigen Daten".** Verwenden Sie ein Verfahren, das Daten der multiplen Imputation unterstützt. Informationen zur Analyse von Datasets der multiplen Imputation und eine Liste der Verfahren, die diese Daten unterstützen, finden Sie unter „Analysieren von Daten multipler Imputation“ auf Seite 18.



---

## Kapitel 2. Analyse fehlender Werte

Die Prozedur "Analyse fehlender Werte" dient primär drei Funktionen:

- Beschreiben des Musters fehlender Daten. Wo befinden sich die fehlenden Daten? Welches Ausmaß weisen sie auf? Tendieren Variablenpaare dazu, fehlenden Werte in mehreren Fällen aufzuweisen? Sind die Datenwerte extrem? Fehlen wahllos Werte?
- Schätzen der Mittelwerte, Standardabweichung, Kovarianzen und Korrelationen für verschiedene Methoden für fehlende Werte: listenweise Methode, paarweise Methode, Regressionsmethode oder EM-Methode (Maximierung des Erwartungswerts). Bei der paarweisen Methode werden auch die Häufigkeiten der paarweise vollständigen Fälle angezeigt.
- Füllt (imputierte) fehlende Werte mit geschätzten Werten mithilfe von Regressions- oder EM-Methoden. Multiple Imputation wird in der Regel jedoch als Methode betrachtet, die die genaueren Ergebnisse liefert.

Die Analyse fehlender Werte unterstützt Sie beim Umgang mit Problemen, die durch unvollständige Daten verursacht werden. Wenn Fälle mit fehlenden Werten sich systematisch von Fällen ohne fehlende Werte unterscheiden, können die Ergebnisse irreführend sein. Fehlende Daten können außerdem die Genauigkeit der berechneten Statistiken beeinträchtigen, da weniger Informationen vorliegen als ursprünglich geplant. Ein weiteres Problem ist die Annahme hinter vielen statistischen Prozeduren, dass alle Fälle vollständig sind. Fehlende Werte können den erforderlichen theoretischen Ansatz verkomplizieren.

**Beispiel.** Bei der Auswertung einer Leukämiebehandlung werden verschiedene Variablen gemessen. Es sind jedoch nicht alle Messwerte für alle Patienten verfügbar. Die Muster der fehlenden Daten werden angezeigt, tabellarisch dargestellt und für zufällig befunden. Eine EM-Analyse wird für die Schätzung der Mittelwerte, Korrelationen und Kovarianzen verwendet. Sie dient außerdem dazu, um festzustellen, ob die Daten in völlig zufälliger Weise fehlen. Die fehlenden Werte werden dann durch abgeleitete (imputierte) Werte ersetzt und zur weiteren Analyse in einer neuen Datendatei gespeichert.

**Statistiken.** Univariate Statistiken, einschließlich der Anzahl nicht fehlender Werte, dem Mittelwert, der Standardabweichung, der Anzahl fehlender Werte und der Anzahl von Extremwerten. Geschätzte Mittelwerte, Kovarianz- und Korrelationsmatrix unter Verwendung der listenweisen, paarweisen, EM- oder Regressionsmethode. MCAR-Test nach Little mit EM-Ergebnissen. Auswertung der Mittelwerte nach verschiedenen Methoden. Für Gruppen, die durch fehlende gegenüber nicht fehlende Werte definiert sind: *t*-Tests. Für alle Variablen: Muster der fehlenden Werte angezeigt nach Fällen und Variablen.

Erläuterung der Daten

**Daten.** Die Daten können kategorial oder quantitativ (metrisch oder stetig) sein. Die Berechnung von Statistiken und das Imputieren (Vorschreiben) fehlender Daten ist jedoch nur für die quantitativen Variablen möglich. Bei allen Variablen müssen die fehlenden Werte, die nicht als systemdefiniert fehlend codiert sind, als benutzerdefiniert fehlend definiert werden. Wenn beispielsweise für eine Frage in einem Fragebogen die Antwort *Ich weiß nicht* als 5 codiert ist und Sie diese als fehlend behandeln möchten, muss für diese Frage 5 als benutzerdefinierter fehlender Wert codiert werden.

**Häufigkeitsgewichtungen.** Häufigkeitsgewichtungen (Replikation) werden von dieser Prozedur berücksichtigt. Fälle mit einer negativen oder nullwertigen Replikationsgewichtung werden ignoriert. Nicht ganzzahligen Gewichtungen werden gekürzt.

**Annahmen.** Listenweisen, paarweisen und Regressionsschätzungen liegt die Annahme zugrunde, dass das Muster der fehlenden Werte nicht von den Datenwerten abhängt. Diese Bedingung ist als **völlig zufällig fehlend** oder MCAR (Missing Completely At Random) bekannt. Daher ergeben alle Schätzmethoden (einschließlich der EM-Methode) bei MCAR-Daten konsistente und unverzerrte Schätzungen der Kor-

relationen und Kovarianzen. Die Verletzung der MCAR-Annahme kann dazu führen, dass von der listenweisen, paarweisen bzw. Regressionsmethode verzerrte Schätzungen generiert werden. Wenn es sich nicht um MCAR-Daten handelt, muss die EM-Schätzung verwendet werden.

Der EM-Schätzung liegt die Annahme zugrunde, dass das Muster der fehlenden Daten nur mit den beobachteten Daten zusammenhängt. Diese Bedingung wird **zufällig fehlend** oder MCAR ("missing at random") genannt. Aufgrund dieser Annahme können die Schätzungen unter Verwendung der verfügbaren Information korrigiert werden. So kann es beispielsweise in einer Studie über Bildung und Einkommen vorkommen, dass bei Personen mit niedrigerer Bildung eine höhere Anzahl fehlende Einkommenswerte vorliegt. In diesem Fall handelt es sich um MAR-Daten, nicht um MCAR-Daten. Anders ausgedrückt: Bei MAR hängt die Wahrscheinlichkeit, dass ein Einkommen angegeben wird, vom Bildungsniveau der betreffenden Person ab. Die Wahrscheinlichkeit kann abhängig von der Bildung, nicht jedoch abhängig vom Einkommen *innerhalb des betreffenden Bildungsniveaus* schwanken. Wenn die Wahrscheinlichkeit, dass ein Einkommen angegeben wird auch in Abhängigkeit vom Einkommen innerhalb der einzelnen Bildungsniveaus schwankt (wenn beispielsweise Personen mit hohem Einkommen ihr Einkommen nicht angeben), handelt es sich weder um MCAR-Daten noch um MAR-Daten. Dies ist eine ungewöhnliche Situation, bei deren Eintreten keine der Methoden angemessen ist.

**Verwandte Prozeduren.** Listenweise und paarweise Schätzungen können in vielen Prozeduren verwendet werden. Mit der linearen Regression und der Faktorenanalyse könne fehlende Werte durch die Mittelwerte ersetzt werden. Im Zusatzmodul "Forecasting" sind verschiedene Methoden verfügbar, um fehlende Werte in Zeitreihen zu ersetzen.

So berechnen Sie eine Analyse fehlender Werte:

1. Wählen Sie die folgenden Befehle aus den Menüs aus:  
**Analysieren > Analyse fehlender Werte...**
2. Wählen Sie mindestens eine quantitative (metrische) Variable zur Schätzung der Statistiken und der optionalen Imputation fehlender Werte aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie kategoriale Variablen (numerisch oder Zeichenfolge) aus und geben Sie für die Anzahl der Kategorien eine Grenze (**Maximalzahl der Kategorien**) ein.
- Klicken Sie auf **Muster** zur tabellarischen Darstellung der Muster fehlender Daten. Weitere Informationen finden Sie im Thema „Anzeigen der Muster fehlender Werte“.
- Klicken Sie auf **Deskriptive Statistik** zur Anzeige deskriptiver Statistiken fehlender Werte. Weitere Informationen finden Sie im Thema „Anzeigen deskriptiver Statistiken für fehlende Werte“ auf Seite 5.
- Wählen Sie eine Methode zur Schätzung der Statistiken (Mittelwerte, Kovarianzen und Korrelationen) und optionalen Imputation fehlender Werte aus. Weitere Informationen finden Sie im Thema „Schätzen von Statistiken und Imputieren fehlender Werte“ auf Seite 6.
- Wenn Sie "EM" oder "Regression" auswählen, klicken Sie auf **Variablen**, um das Subset anzugeben, das für die Schätzung verwendet wird. Weitere Informationen finden Sie im Thema „Vorhergesagte Variablen und Prädiktorvariablen“ auf Seite 9.
- Wählen Sie eine Variable für die Fallbeschriftung aus. Diese Variable dient zur Beschriftung von Fällen in Mustertabellen, die einzelne Fälle anzeigen.

---

## Anzeigen der Muster fehlender Werte

Sie können verschiedene Tabellen anzeigen lassen, die die Muster und das Ausmaß der fehlenden Daten zeigen. Mit diesen Tabellen können Sie Antworten auf folgende Fragen finden:

- Wo befinden sich fehlende Daten?
- Tendieren Variablenpaare dazu, fehlende Werte in einzelnen Fällen aufzuweisen?
- Sind Datenwerte extrem?



## Anzeigen

Für die Anzeige von Mustern fehlender Daten stehen drei Tabellentypen zur Verfügung.

**Fälle in Tabellen.** Die Muster fehlender Daten in den Analysevariablen werden in Tabellenform dargestellt, wobei für jedes Muster auch die Häufigkeiten angegeben werden. Mit **Variable sortieren nach dem Muster fehlender Werte** können Sie angeben, ob Häufigkeiten (Anzahl) und Variablen nach der Ähnlichkeit der Muster sortiert werden sollen. Mit **Muster weglassen bei weniger als n % der Fälle** können Sie Muster ausschließen, die nur selten vorkommen..

**Fälle mit fehlenden Werten.** Für die einzelnen Analysevariablen werden jeweils die einzelnen Fälle mit einem fehlenden Wert oder einem Extremwert tabellarisch dargestellt. Mit **Variable sortieren nach dem Muster fehlender Werte** können Sie angeben, ob Häufigkeiten (Anzahl) und Variablen nach der Ähnlichkeit der Muster sortiert werden sollen.

**Alle Fälle.** Die einzelnen Fälle werden tabellarisch dargestellt, und fehlende Werte und Extremwerte werden für jede Variable angegeben. Die Fälle werden in der Reihenfolge aufgeführt, in der sie in der Daten-datei auftreten, sofern unter **Sortieren nach** keine Variable angegeben wurde.

In den Tabellen, die einzelne Fälle anzeigen, werden folgende Symbole verwendet:

- + . Extrem hoher Wert
- . Extrem niedriger Wert
- S. Systemdefiniert fehlender Wert
- A. Erster Typ des benutzerdefiniert fehlenden Werts
- B. Zweiter Typ des benutzerdefiniert fehlenden Werts
- C. Dritter Typ des benutzerdefiniert fehlenden Werts

## Variablen

Sie können weitere Informationen für die in die Analyse aufgenommenen Variablen anzeigen. Die Variablen, die Sie unter **Zusätzliche Informationen für** hinzufügen, werden einzeln in der Tabelle der fehlenden Muster angezeigt. Bei quantitativen (metrischen) Variablen wird der Mittelwert und bei kategorialen Variablen wird die Anzahl der Fälle aufgeführt, die das Muster in jeder Kategorie aufweisen.

- **Sortieren nach.** Die Fälle werden entsprechend der aufsteigenden oder absteigenden Reihenfolge der Werte der angegebenen Variablen aufgeführt. Diese Option ist nur für **Alle Fälle** verfügbar.

So geben Sie Muster fehlender Werte an:

1. Wählen Sie im Dialogfeld "Analyse fehlender Werte" die Variablen aus, für die Muster fehlender Werte angezeigt werden sollen.
2. Klicken Sie auf **Muster**.
3. Wählen Sie die anzuzeigenden Mustertabellen aus.

---

## Anzeigen deskriptiver Statistiken für fehlende Werte

### Univariate Statistiken

Univariate Statistiken können zur Ermittlung des allgemeinen Ausmaßes der fehlenden Daten beitragen. Für jede Variable werden folgende Daten angezeigt:

- Anzahl nicht fehlender Werte

- Anzahl und Prozentsatz fehlender Werte

Für quantitative (metrische) Variablen werden außerdem folgende Daten angezeigt:

- Mittelwert
- Standardabweichung
- Anzahl extrem hoher und niedriger Werte

Statistik für Indikatorvariablen

Für jede Variable wird eine Indikatorvariable erstellt. Diese kategoriale Variable gibt an, ob die Variable für einen einzelnen Fall vorhanden ist oder fehlt. Die Indikatorvariablen werden verwendet, um die Tabellen mit Nichtübereinstimmungen, *T*-Tests und Häufigkeiten zu erstellen.

**Prozent der nicht übereinstimmenden Variablen.** Für jedes Variablenpaar wird der Prozentsatz von Fällen angezeigt, in denen eine Variable einen fehlenden Wert und die andere Variable einen nicht fehlenden Wert aufweist. Jedes Diagonalelement in der Tabelle enthält den Prozentsatz von fehlenden Werten für eine einzelne Variable.

**T-Test für Gruppen, die durch Indikatorvariablen gebildet werden.** Für jede quantitative Variable werden die Mittelwerte von zwei Gruppen mithilfe der Student-*T*-Statistik verglichen. Die Gruppen geben an, ob eine Variable vorhanden ist oder fehlt. Es werden die *T*-Statistik, Freiheitsgrade, Häufigkeiten von fehlenden und nicht fehlenden Werten sowie die Mittelwerte der beiden Gruppen angezeigt. Außerdem können Sie alle zweiseitigen Wahrscheinlichkeiten anzeigen, die der *T*-Statistik zugeordnet sind. Wenn Ihre Analyse zu mehreren Tests führt, dürfen Sie diese Wahrscheinlichkeiten nicht für Signifikanztests verwenden. Die Wahrscheinlichkeiten sind nur geeignet, wenn nur ein einziger Test berechnet wird.

**Kreuztabellen kategorialer und Indikatorvariablen.** Für jede kategoriale Variable wird eine Tabelle angezeigt. In der Tabelle werden für jede Kategorie die Häufigkeit und der Prozentsatz von nicht fehlenden Werten für die anderen Variablen angezeigt. Außerdem werden die Prozentsätze für jeden Typ von fehlenden Werten angezeigt.

**Variablen weglassen, die in weniger als n % der Fälle fehlen.** Um die Tabellen zu verkleinern, können Sie die Statistiken weglassen, die nur für eine kleine Anzahl von Fällen berechnet werden.

So zeigen Sie deskriptive Statistiken an:

1. Wählen Sie im Dialogfeld "Analyse fehlender Werte" die Variablen aus, für die deskriptive Statistiken fehlender Werte angezeigt werden sollen.
2. Klicken Sie auf **Deskriptive Statistik**.
3. Wählen Sie die anzuzeigende deskriptive Statistik aus.

---

## Schätzen von Statistiken und Imputieren fehlender Werte

Sie können Mittelwerte, Standardabweichung, Kovarianzen und Korrelationen unter Verwendung der listenweisen Methode (nur vollständige Fälle), der paarweisen Methode, der EM-Methode (Maximierung des Erwartungswerts) bzw. der Regressionsmethode schätzen. Außerdem können Sie auswählen, dass die fehlenden Werte imputiert (vorgeschrieben) werden sollen, d. h. dass Ersatzwerte geschätzt werden sollen. Beachten Sie, dass multiple Imputation im Allgemeinen bei der Lösung des Problems fehlender Werte der einfachen Imputation überlegen ist. Der MCAR-Test von Little ist nach wie vor hilfreich bei der Bestimmung, ob eine Imputation erforderlich ist.

Listenweise Methode

Bei dieser Methode werden nur vollständige Fälle verwendet. Wenn eine der Analysevariablen fehlende Werte aufweist, wird der betreffende Fall aus den Berechnungen ausgeschlossen.

## Paarweise Methode

Bei dieser Methode werden Paare von Analysevariablen betrachtet und ein Fall wird nur verwendet, wenn er für beide Variablen nicht fehlende Werte aufweist. Häufigkeiten, Mittelwerte und Standardabweichungen werden für jedes Paar gesondert berechnet. Da andere fehlende Werte im Fall ignoriert werden, sind die für zwei Variablen berechneten Korrelationen und Kovarianzen nicht von Werten abhängig die in anderen Variablen fehlen.

## EM-Methode

Bei dieser Methode wird von einer Verteilung für die teilweise fehlenden Daten ausgegangen und die Schlussfolgerungen (Inferenzen) beruhen auf der Likelihood bei dieser Verteilung. Jede Iteration besteht aus einem E-Schritt und einem M-Schritt. Im E-Schritt wird die bedingte Erwartung der "fehlenden" Daten ermittelt, die auf den beobachteten Werten und den aktuellen Schätzungen der Parameter beruht. Anschließend werden die "fehlenden" Daten durch diese Erwartungen ersetzt. Im M-Schritt werden Maximum-Likelihood-Schätzungen der Parameter so berechnet, wie wenn die fehlenden Daten ergänzt worden wären. "Fehlend" steht in Anführungszeichen, da die fehlenden Werte nicht direkt ergänzt werden. Stattdessen, werden bei der Log-Likelihood Funktionen dieser Werte verwendet.

Die Chi-Quadrat-Statistik nach Roderick J. A. Little, die dazu dient zu testen, ob Werte in völlig zufälliger Weise fehlen (MCAR - Missing Completely At Random) ist als Fußnote zu den EM-Matrizen abgedruckt. Bei diesem Test besagt die Nullhypothese, dass die Daten völlig zufällig fehlen, und der  $p$ -Wert ist auf dem Niveau 0,05 signifikant. Wenn der Wert weniger als 0,05 beträgt, fehlen die Werte nicht völlig zufällig. Die Daten fehlen möglicherweise zufällig (MAR - Missing At Random) oder fehlen nicht zufällig (NMAR - Not Missing At Random). Sie können nicht von einer der Eigenschaften ausgehen, sondern müssen die Daten analysieren, um zu ermitteln, in welcher Form sie fehlen.

## Regressionsmethode (Faktorenanalyse)

Diese Methode berechnet Schätzungen für die mehrfach lineare Regression und verfügt über Optionen zur Erweiterung der Schätzungen durch Zufallskomponenten. Zu jedem vorhergesagten Wert kann das Verfahren ein Residuum aus einem zufällig ausgewählten vollständigen Fall, eine normale Zufallsabweichung oder eine Zufallsabweichung (anhand der Quadratwurzel der Residualvarianz (residual mean square) aus der  $t$ -Verteilung hinzufügen.

## EM-Schätzung: Optionen

Beim EM-Verfahren werden unter Verwendung eines iterativen Prozesses die Mittelwerte, die Kovarianzmatrix und die Korrelation der quantitativen (metrischen) Variablen mit fehlenden Werte geschätzt.

**Verteilung.** EM erstellt Schlussfolgerungen (Inferenzen) anhand der für die jeweilige Verteilung geltenden Likelihood. Standardmäßig wird eine Normalverteilung angenommen. Wenn Sie wissen, dass die Flanken der Verteilung länger sind als die einer Normalverteilung, können Sie anfordern, dass die Prozedur die Likelihood-Funktion aus einer Student- $T$ -Verteilung mit  $n$  Freiheitsgraden erstellt. Die gemischte Normalverteilung führt ebenfalls zu einer Verteilung mit längeren Flanken. Geben Sie die Quotienten der Standardabweichungen der gemischten Normalverteilung und das Mischungsverhältnis der beiden Verteilungen an. Bei der gemischten Normalverteilung wird davon ausgegangen, dass nur die Standardabweichungen der Verteilungen unterschiedlich sind. Die Mittelwerte müssen übereinstimmen.

**Maximalzahl der Iterationen.** Legt die maximale Anzahl der Iterationen zur Schätzung der wahren Kovarianz fest. Die Prozedur wird beendet, wenn diese Anzahl der Iterationen erreicht wurde, auch wenn die Schätzungen nicht konvergiert haben.

**Vervollständigte Daten speichern.** Sie können ein Dataset mit den imputierten Werten anstelle der fehlenden Werte speichern. Beachten Sie jedoch, dass kovarianzbasierte Statistiken, die die imputierten Werte

verwenden, zu einer zu niedrigen Schätzung der zugehörigen Parameterwerte führen. Der Grad der Unterschätzung ist proportional zu der Anzahl der Fälle die gemeinsam unbeobachtet sind.

So legen Sie EN-Optionen fest:

1. Wählen Sie im Dialogfeld "Analyse fehlender Werte" die Variablen aus, für die fehlende Werte mithilfe der EM-Methode geschätzt werden sollen.
2. Aktivieren Sie im Gruppenfeld "Schätzung" die Option **EM**.
3. Klicken Sie auf die Schaltfläche **Variablen**, um die vorhergesagten Variablen und die Prädiktorvariablen anzugeben. Weitere Informationen finden Sie im Thema „Vorhergesagte Variablen und Prädiktorvariablen“ auf Seite 9.
4. Klicken Sie auf **EM**.
5. Wählen Sie die gewünschten EM-Optionen aus.

## Optionen für die Regressionsschätzung

Bei der Regressionsmethode werden fehlende Werte unter Verwendung der mehrfachen linearen Regression geschätzt. Es werden die Mittelwerte, die Kovarianzmatrix und die Korrelationsmatrix der vorhergesagten Variablen angezeigt.

**Anpassung der Schätzung.** Bei der Regression kann den Regressionsschätzungen eine Zufallskomponente hinzugefügt werden. Sie können Residuen, normale Variaten, Student-T-Variaten oder keine Anpassung auswählen.

- *Residuen.* Es werden Fehlerterme zufällig aus den beobachteten Residuen vollständiger Fälle ausgewählt und den Regressionsschätzungen hinzugefügt.
- *Normale Variaten.* Fehlerterme werden beliebig aus einer Verteilung mit dem Erwartungswert 0 und einer Standardabweichung gleich der Quadratwurzel der mittleren Quadratsumme des Regressionsfehlerterms gezogen.
- *Student-T-Variaten.* Fehlerterme werden beliebig aus der t-Verteilung gezogen und anhand des Effektivwertfehlers (RMSE, Root Mean Squared Error, Wurzel des mittleren quadratischen Fehlers) skaliert.

**Maximale Anzahl der Prädiktoren.** Legt eine Obergrenze für die Anzahl der (unabhängigen) Prädiktorvariablen fest, die bei der Schätzung verwendet werden.

**Vervollständigte Daten speichern.** Schreibt ein Dataset in der aktuellen Sitzung oder eine externe Datendatei im IBM® SPSS Statistics-Format. Dabei werden die fehlenden Werte durch die Werte ersetzt, die bei der Regression geschätzt wurden.

So legen Sie die Regressionsoptionen fest:

1. Wählen Sie im Dialogfeld "Analyse fehlender Werte" die Variablen aus, für die fehlende Werte mithilfe der Regressionsmethode geschätzt werden sollen.
2. Aktivieren Sie im Gruppenfeld "Schätzung" die Option **Regression**.
3. Klicken Sie auf die Schaltfläche **Variablen**, um die vorhergesagten Variablen und die Prädiktorvariablen anzugeben. Weitere Informationen finden Sie im Thema „Vorhergesagte Variablen und Prädiktorvariablen“ auf Seite 9.
4. Klicken Sie auf **Regression**.
5. Wählen Sie die gewünschten Regressionsoptionen aus.

## Vorhergesagte Variablen und Prädiktorvariablen

Standardmäßig werden alle quantitativen Variablen für EM und Regressions-schätzung verwendet. Falls erforderlich, können Sie bestimmte Variablen als vorhergesagte Variablen bzw. Prädiktorvariablen in den Schätzungen auswählen. Eine Variable kann prinzipiell in beiden Listen enthalten sein, es gibt jedoch Situationen, in denen es sinnvoll ist, die Verwendung einer Variablen einzuschränken. So vermeiden es einige Analytiker, die Werte von Ergebnisvariablen zu schätzen. Außerdem kann es sinnvoll sein, für verschiedene Schätzungen auch unterschiedliche Variablen zu verwenden und die Prozedur mehrmals auszuführen. Wenn Ihnen beispielsweise ein Set von Items vorliegt, bei denen es sich um die Bewertungen des Pflegepersonals handelt, und ein weiteres Set mit den Bewertungen der Ärzteschaft, kann es sinnvoll sein, eine Ausführung zur Schätzung der fehlenden Items für das Pflegepersonal und eine weitere Ausführung für die Schätzungen der Items der Ärzteschaft durchzuführen.

Bei Verwendung der Regressionsmethode ist noch ein weiterer Faktor zu berücksichtigen. Bei der mehrfachen Regression kann die Verwendung eines großen Subsets unabhängiger Variablen zu schlechteren vorhergesagten Werten führen als eine kleinere Untergruppe. Daher muss eine Variable mindestens ein  $F$  für die Aufnahme von 4,0 erreichen, um verwendet zu werden. Dieser Grenzwert kann über die Syntax geändert werden.

So geben Sie vorhergesagte Variablen und Prädiktorvariablen an:

1. Wählen Sie im Dialogfeld "Analyse fehlender Werte" die Variablen aus, für die fehlende Werte mithilfe der Regressionsmethode geschätzt werden sollen.
2. Aktivieren Sie im Gruppenfeld "Schätzung" die Option **EM** oder **Regression**.
3. Klicken Sie auf **Variablen**.
4. Wenn Sie nur bestimmte und nicht alle Variablen als vorhergesagte Variablen und Prädiktorvariablen verwenden möchten, aktivieren Sie **Variablen auswählen** und verschieben Sie die Variablen in die entsprechende(n) Liste(n).

---

## Zusätzliche Funktionen beim Befehl MVA

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Schlüsselwort **DESCRIBE** in den Unterbefehlen **MPATTERN**, **DPATTERN** und **TPATTERN** können Sie separate deskriptive Variablen für Muster fehlender Werte, Datenmuster und Muster in Tabellen festlegen.
- Mit dem Unterbefehl **DPATTERN** können Sie mehrere Sortiervariablen für die Tabelle der Datenmuster festlegen.
- Mit dem Unterbefehl **DPATTERN** können Sie mehrere Sortiervariablen für die Datenmuster festlegen.
- Mit dem Unterbefehl **EM** können Sie die Toleranz und Konvergenz festlegen.
- Mit dem Unterbefehl **REGRESSION** können Sie die Toleranz und den  $F$ -Wert für die Aufnahme festlegen.
- Mit den Unterbefehlen **EM** und **REGRESSION** können Sie verschiedene Variablenlisten für das EM-Verfahren und die Regression festlegen.
- Für **TTESTS**, **TABULATE** und **MISMATCH** können Sie unterschiedliche Prozentsätze für das Unterdrücken von angezeigten Fällen festlegen.

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.



## Kapitel 3. Multiple Imputation

Der Zweck der multiplen Imputation ist die Erzeugung möglicher Werte für fehlende Werte, um so verschiedene "vollständige" Sets an Daten zu erzeugen. Analyseverfahren, die mit Datensets aus multipler Imputation arbeiten, erzeugen Ausgaben für jedes "vollständige" Dataset sowie eine gemeinsame Ausgabe, die schätzt, welche Ergebnisse entstanden wären, wenn das Originaldataset keine fehlenden Werte besitzte. Diese zusammengefassten Ergebnisse sind in der Regel genauer als die, die durch einfache Imputationsmethoden entstehen.

Multiple Imputation - Datenüberlegungen












**Analysevariablen.** Die Analysevariablen können wie folgt gestaltet sein:

- *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.
- *Metrisch.* 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).

Bei der Prozedur wird davon ausgegangen, dass allen Variablen das richtige Messniveau zugewiesen wurde. Sie können das Messniveau für eine Variable jedoch vorübergehend ändern. Klicken Sie hierzu mit der rechten Maustaste auf die Variable in der Liste der Quellvariablen und wählen Sie das gewünschte Messniveau im Popup-Menü aus. Weitere Informationen zum dauerhaften Ändern des Messniveaus einer Variablen

Messniveau und Datentyp sind durch ein Symbol neben der jeweiligen Variablen in der Variablenliste gekennzeichnet:

Tabelle 1. Messniveausymbole

	Numerisch	Zeichenfolge	Datum	Zeit
Metrisch (stetig)		entfällt		
Ordinal				
Nominal				

**Häufigkeitsgewichtungen.** Häufigkeitsgewichtungen (Replikation) werden von dieser Prozedur berücksichtigt. Fälle mit einer negativen oder nullwertigen Replikationsgewichtung werden ignoriert. Nicht ganzzahlige Gewichtungen werden auf die nächste Ganzzahl gerundet.

**Analysegewichtung.** Analysegewichtungen (Regression oder Stichprobe) werden in Zusammenfassungen von fehlenden Werten und in passende Imputationsmodelle integriert. Fälle mit einer negativen oder nullwertigen Analysegewichtung werden ausgeschlossen.

**Komplexe Stichproben.** Das Verfahren der multiplen Imputation ist nicht explizit für Schichten, Cluster oder andere komplexe Stichprobenstrukturen gedacht, es kann jedoch endgültige Stichprobengewichtungen in Form der Analysegewichtungsvariablen akzeptieren. Beachten Sie auch, dass Prozeduren für komplexe Stichproben nicht automatisch mehrere imputierte Datasets analysieren. Eine komplette Liste der Prozeduren, die Pooling unterstützen, finden Sie unter „Analysieren von Daten multipler Imputation“ auf Seite 18.

**Fehlende Werte.** Sowohl benutzer- als auch systemdefiniert fehlende Werte werden als ungültige Werte behandelt. Beide Arten von fehlenden Werten werden ersetzt, wenn Werte imputiert werden, und beide Arten werden als ungültige Werte von als Einflussfaktoren in Imputationsmodellen verwendeten Variablen behandelt. Benutzer- und systemdefiniert fehlende Werte werden auch bei Fehlanalysen als fehlende Werte behandelt.

**Replikation von Ergebnissen (Fehlende Datenwerte imputieren).** Wenn Sie Ihre Imputation exakt reproduzieren möchten, müssen Sie nicht nur dieselben Einstellungen für die Prozedur, sondern auch denselben Initialisierungswert für den Zufallszahlengenerator, dieselbe Datenreihenfolge und dieselbe Variablenreihenfolge verwenden.

- **Generierung von Zufallszahlen.** Die Prozedur verwendet Zufallszahlengenerierung bei der Berechnung der imputierten Werte. Um zu einem späteren Zeitpunkt dieselben randomisierten Ergebnisse zu reproduzieren, müssen Sie vor jeder Ausführung der Prozedur "Fehlende Datenwerte imputieren" denselben Initialisierungswert für den Zufallszahlengenerator verwenden.
- **Fallreihenfolge.** Werte werden in der Fallreihenfolge imputiert.
- **Reihenfolge der Variablen.** Die Imputationsmethode der vollständig konditionalen Spezifikation imputiert Werte in der Reihenfolge der Liste der Analysevariablen.

Für multiple Imputation stehen zwei spezielle Dialogfelder zur Verfügung.

- **Muster analysieren** stellt deskriptive Messungen der Muster der fehlenden Werte in den Daten bereit und eignet sich als Untersuchungsschritt vor der Imputation.
- **Fehlende Datenwerte imputieren** wird verwendet, um multiple Imputationen zu erzeugen. Die vollständigen Datasets können mit Prozeduren analysiert werden, die Datasets mit multipler Imputation unterstützen. Informationen zur Analyse von Datasets der multiplen Imputation und eine Liste der Verfahren, die diese Daten unterstützen, finden Sie unter „Analysieren von Daten multipler Imputation“ auf Seite 18.

---

## Analysieren von Mustern

"Muster analysieren" stellt deskriptive Messungen der Muster der fehlenden Werte in den Daten bereit und eignet sich als Untersuchungsschritt vor der Imputation.

**Beispiel.** Ein Telekommunikationsanbieter möchte einen besseren Einblick in die Servicenutzungsmuster in seiner Kundendatenbank gewinnen. Er verfügt über die vollständigen Daten der von seinen Kunden genutzten Services, jedoch fehlen in den demografischen Informationen, die das Unternehmen gesammelt hat, einige Werte. Eine Analyse der Muster von fehlenden Werten kann helfen, die nächsten Schritte für die Imputation zu bestimmen. Weitere Informationen finden Sie im Thema .

Wählen Sie die folgenden Befehle aus den Menüs aus:

**Analysieren > Multiple Imputation > Muster analysieren...**

1. Wählen Sie mindestens zwei Analysevariablen aus. Die Prozedur analysiert Muster fehlender Daten für diese Variablen.

Optionale Einstellungen



**Analysegewichtung.** Diese Variable enthält Analysegewichtungen (Regression oder Stichprobe). Das Verfahren integriert Analysegewichtungen in Zusammenfassungen fehlender Werte. Fälle mit einer negativen oder nullwertigen Analysegewichtung werden ausgeschlossen.

**Ausgabe.** Die folgende optionale Ausgabe ist verfügbar:

- **Zusammenfassung der fehlenden Werte.** Zeigt ein unterteiltes Kreisdiagramm an, das die Anzahl und die Prozentzahlen der Analysevariablen, Fälle oder einzelne Datenwerte enthält, die über einen oder mehrere fehlende Werte verfügen.
- **Muster fehlender Werte.** Zeigt tabulierte Muster fehlender Werte an. Jedes Muster entspricht einer Gruppe von Fällen mit dem gleichen Muster unvollständiger und vollständiger Daten bei Analysevariablen. Sie können diese Ausgabe verwenden, um zu bestimmen, welche monotone Imputationsmethode für Ihre Daten verwendet werden kann und in welchem Maße Ihre Daten einem monotonen Muster entsprechen. Die Prozedur ordnet Analysevariablen, um ein monotones Muster preiszugeben bzw. anzunähern. Wenn kein nicht monotones Muster nach der Neuordnung existiert, können Sie daraus schließen, dass die Daten ein monotonen Muster besitzen, wenn die Analysevariablen als solche geordnet sind.

- **Variablen mit der höchsten Häufigkeit an fehlenden Werten.** Zeigt eine Tabelle der Analysevariablen, sortiert nach Prozent der fehlenden Werte in absteigender Reihenfolge, an. Die Tabelle enthält deskriptive Statistiken (Mittelwert und Standardabweichung) für metrische Variablen.

Sie können die maximale Zahl an anzuzeigenden Variablen und den Mindestprozentsatz fehlender Werte für eine Variable, der dargestellt wird, steuern. Es wird das Set von Variablen angezeigt, die beiden Kriterien entspricht. Zum Beispiel verlangt das Einstellen der Maximalzahl von Variablen auf 50 und des Mindestprozentsatzes fehlender Werte auf 25, dass die Tabelle bis zu 50 Variablen anzeigt, die mindestens 25 % fehlende Werte besitzen. Wenn es 60 Analysevariablen gibt, aber nur 15 25 % oder mehr fehlende Werte haben, enthält die Ausgabe nur 15 Variablen.

---

## Imputieren fehlender Datenwerte

"Fehlende Datenwerte imputieren" wird verwendet, um multiple Imputationen zu erzeugen. Die vollständigen Datensets können mit Prozeduren analysiert werden, die Datensets mit multipler Imputation unterstützen. Informationen zur Analyse von Datensets der multiplen Imputation und eine Liste der Verfahren, die diese Daten unterstützen, finden Sie unter „Analysieren von Daten multipler Imputation“ auf Seite 18.

**Beispiel.** Ein Telekommunikationsanbieter möchte einen besseren Einblick in die Servicenutzungsmuster in seiner Kundendatenbank gewinnen. Er verfügt über die vollständigen Daten der von seinen Kunden genutzten Services, jedoch fehlen in den demografischen Informationen, die das Unternehmen gesammelt hat, einige Werte. Zudem fehlen diese Werte nicht völlig zufällig, daher wird das Dataset mithilfe multipler Imputation vervollständigt. Weitere Informationen finden Sie im Thema .

Wählen Sie die folgenden Befehle aus den Menüs aus:

### Analysieren > Multiple Imputation > Fehlende Datenwerte imputieren...

1. Wählen Sie mindestens zwei Variablen im Imputationsmodell aus. Die Prozedur imputiert mehrere Werte für fehlende Daten für diese Variablen.
2. Die Anzahl der zu berechnenden Imputationen. Standardmäßig ist dieser Wert 5.
3. Geben Sie ein Dataset oder eine Datendatei im IBM SPSS Statistics-Format an, in das die imputierten Daten geschrieben werden sollen.

Das Ausgabedataset besteht aus den ursprünglichen Falldaten mit fehlenden Daten plus einem Set von Fällen mit imputierten Werten für jede Imputation. Wenn beispielsweise das ursprüngliche Dataset 100 Fälle enthält und Sie haben fünf Imputationen, umfasst das Ausgabedataset 600 Fälle. Alle Variablen im Eingabedataset sind im Ausgabedataset enthalten. Wörterbucheigenschaften (Namen, Be-

schriftungen etc.) von bestehenden Variablen werden in das neue Dataset kopiert. Die Datei enthält auch eine neue Variable, *Imputation\_*, eine numerische Variable, die die Imputation angibt (0 für Originaldaten, 1..n für Fälle mit imputierten Werten).

Die Prozedur definiert automatisch die Variable *Imputation\_* als Teilungsvariable, wenn das Ausgabedataset erstellt wird. Wenn bei Ausführung der Prozedur Aufteilungen wirksam sind, enthält das Ausgabedataset ein Set an Imputationen für jede Kombination von Werten von ausgeteilten Variablen.

## Optionale Einstellungen

**Analysegewichtung.** Diese Variable enthält Analysegewichtungen (Regression oder Stichprobe). Die Prozedur umfasst Analysegewichtungen in Regressions- und Klassifizierungsmodellen, die verwendet werden, um fehlende Werte zu imputieren. Analysegewichtungen werden auch in Zusammenfassungen imputierter Werte verwendet, zum Beispiel Mittelwert, Standardabweichung und Standardfehler. Fälle mit einer negativen oder nullwertigen Analysegewichtung werden ausgeschlossen.

## 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.

## Methode

Die Registerkarte "Methode" gibt an, wie fehlende Werte einschließlich der verwendeten Modelltypen imputiert werden. Kategoriale Prädiktoren sind als Indicator (Dummy) codiert.

**Imputationsmethode.** Die Methode **Automatisch** scannt die Daten und verwendet die monotone Methode, wenn die Daten ein monotoneres Muster fehlender Werte zeigen. Andernfalls wird die vollständig konditionale Spezifikation verwendet. Wenn Sie sich sicher sind, welche Methode Sie verwenden wollen, können Sie sie als eine Methode unter **Benutzerdefiniert** angeben.

- **Vollständig konditionale Spezifikation.** Dies ist eine iterative Markov Chain Monte Carlo (MCMC) Methode, die verwendet werden kann, wenn das Muster fehlender Daten willkürlich (monoton oder nicht monoton) ist.

Für jede Iteration und jede Variable in der in der Variablenliste angegebenen Reihenfolge passt die Methode der vollständig konditionalen Spezifikation ein univariates (einzelne abhängige Variable) Modell mit allen anderen Variablen im Modell als Prädiktoren an und imputiert dann die fehlenden Werte für die anzupassende Variable. Die Methode wird fortgesetzt, bis die maximale Zahl an Iterationen erreicht ist, und die imputierten Werte in der maximalen Iteration werden in das imputierte Dataset gespeichert.

**Maximalzahl der Iterationen.** Gibt die Anzahl der Iterationen oder Schritte an, die die von der Methode der vollständig konditionalen Spezifikation verwendete Markov-Kette durchläuft. Wenn die Methode der vollständig konditionalen Spezifikation automatisch gewählt wurde, verwendet sie die Standardzahl von 10 Iterationen. Wenn Sie die vollständig konditionale Spezifikation explizit wählen, können Sie eine benutzerdefinierte Zahl an Iterationen angeben. Sie müssen gegebenenfalls die Anzahl

der Iterationen erhöhen, wenn die Markov-Kette nicht konvergiert. Auf der Registerkarte "Ausgabe" können Sie den Iterationsverlaufdaten der vollständig konditionalen Spezifikation speichern und sie als Diagramm ausgeben, um die Konvergenz zu beurteilen.

- **Monoton.** Dies ist eine nicht iterative Methode, die nur verwendet werden kann, wenn die Daten ein monotones Muster fehlender Werte haben. Ein monotones Muster existiert, wenn Sie die Variablen so ordnen können, dass alle vorhergehenden Variablen auch nicht fehlende Werte haben, wenn eine Variable einen nicht fehlenden Wert hat. Wenn Sie dies als **benutzerdefinierte** Methode angeben, stellen Sie sicher, die Variablen in der Liste in einer Reihenfolge anzugeben, die ein monotones Muster aufweist.

Für jede Variable in der monotonen Reihenfolge passt die monotone Methode ein univariates (einzelne abhängige Variable) Modell mit allen vorhergehenden Variablen im Modell als Prädiktoren an und imputiert dann die fehlenden Werte für die anzupassende Variable. Diese imputierten Werte werden in das imputierte Dataset gespeichert.

**Zweiwegeinteraktionen aufnehmen.** Wenn die Imputationsmethode automatisch gewählt wird, enthält das Imputationsmodell für jede Variable eine Konstante und Haupteffekte für Prädiktorvariablen. Wenn eine bestimmte Methode gewählt wird, können Sie optional alle möglichen Zweiwegeinteraktionen in die kategorialen Prädiktorvariablen aufnehmen.

**Modelltyp für metrische Variablen.** Wenn die Imputationsmethode automatisch gewählt wird, wird lineare Regression als univariates Modell für metrische Variablen verwendet. Wenn eine bestimmte Methode gewählt wird, können Sie alternativ Predictive Mean Matching (PMM) als Modell für metrische Variablen wählen. PMM ist eine Variante der linearen Regression, die imputierte Werte, die durch das Regressionsmodell berechnet wurden, mit dem nächsten beobachteten Wert abgleicht.

Logistische Regression wird immer als univariates Modell für kategoriale Variablen verwendet. Unabhängig vom Modelltyp werden kategoriale Prädiktoren mit Indikatorcodierung (Dummy) gehandhabt.

**Toleranz für Prüfung auf Singularität.** Singuläre (bzw. nicht invertierbare) Matrizen weisen linear abhängige Spalten auf, die zu ernststen Problemen für den Schätzalgorithmus führen können. Auch annähernd singuläre Matrizen können zu schlechten Ergebnissen führen, daher behandelt die Prozedur eine Matrix, deren Determinante unter dem Toleranzwert liegt, als singulär. Geben Sie einen positiven Wert ein.

## Nebenbedingungen

Mithilfe der Registerkarte "Nebenbedingungen" können Sie die Rolle einer Variablen während der Imputation beschränken und den Bereich der imputierten Werte einer metrischen Variablen so einschränken, dass sie plausibel sind. Zusätzlich können Sie die Analyse auf Variablen mit weniger als einem maximalen Prozentsatz fehlender Werte einschränken.

**Daten für Variablenzusammenfassung durchsuchen.** Wenn Sie auf **Daten durchsuchen** klicken, zeigt die Liste Analysevariablen und jeweils den beobachteten Prozentwert für fehlend, Minimum und Maximum. Die Zusammenfassungen können auf allen Fällen oder auf einem Durchlauf der ersten  $n$  Fälle wie im Textfeld "Fälle" angegeben beruhen. Durch Klicken auf **Erneut durchsuchen** werden die Verteilungszusammenfassungen aktualisiert.

Nebenbedingungen definieren

- **Rolle.** Hierüber können Sie das Set der zu imputierenden und/oder als Prädiktoren zu behandelnden Variablen anpassen. Üblicherweise wird jede Analysevariable im Imputationsmodell sowohl als abhängige Variable als auch als Prädiktor betrachtet. Die **Rolle** kann verwendet werden, um die Imputation von Variablen, die Sie **Nur als Prädiktor verwenden** wollen, auszuschalten oder um Variablen von der Verwendung als Prädiktoren (**Nur imputieren**) auszuschließen und so das Vorhersagemodell kompakter zu machen. Dies ist die einzige Nebenbedingung, die für kategoriale Variablen oder für Variablen, die nur als Prädiktoren verwendet werden, angegeben werden kann.
- **Min und Max.** In diesen Spalten können Sie die minimal und maximal zulässigen imputierten Werte für metrische Variablen angeben. Wenn ein imputierter Wert außerhalb dieses Bereichs liegt, zieht das

Verfahren einen anderen Wert, bis es einen findet, der im Bereich liegt, oder bis die maximale Zahl an Ziehungen erreicht ist (siehe **Maximale Ziehungen** unten). Diese Spalten sind nur verfügbar, wenn **Lineare Regression** als Modelltyp für metrische Variablen auf der Registerkarte "Methode" ausgewählt ist.

- **Runden.** Einige Variablen können als metrische Variablen verwendet werden, haben aber Werte, die weiter natürlich beschränkt sein können, z. B. muss die Anzahl der Personen in einem Haushalt eine Ganzzahl sein und der in einem Geschäft ausgegebene Betrag kann keine Bruchteile von Cents umfassen. In dieser Spalte kann die kleinste zulässige Stückelung festgelegt werden. Beispiel: Um ganzzahlige Werte zu erhalten, geben Sie 1 als Rundungswert an; um Werte auf den nächsten Cent zu runden, geben Sie 0,01 an. Im Allgemeinen werden Werte auf das nächste ganzzahlige Vielfache des angegebenen Rundungswerts gerundet. Die folgende Tabelle zeigt, wie sich unterschiedliche Rundungswerte auf den imputierten Wert 6,64823 (vor der Rundung) auswirken.

Tabelle 2. Runden von Ergebnissen.

Rundungswert	Wert, auf den 6,64832 gerundet wird
10	10
E	7
0,25	6,75
0,1	6,6
0,01	6,65

**Variablen mit großen Mengen an fehlenden Daten ausschließen.** Normalerweise werden Analysevariablen imputiert und als Prädiktoren verwendet, unabhängig davon, wie viele fehlende Werte sie besitzen, vorausgesetzt, sie haben ausreichend Daten, um ein Imputationsmodell zu schätzen. Sie können Variablen ausschließen, die einen hohen Prozentsatz an fehlenden Werten haben. Wenn Sie zum Beispiel 50 als **Maximaler Prozentsatz fehlend** angeben, werden Analysevariablen, bei denen mehr als 50 % der Werte fehlen, nicht imputiert und sie werden auch nicht als Prädiktoren bei Imputationsmodellen verwendet.

**Maximale Ziehungen.** Wenn Minimum- oder Maximumwerte für imputierte Werte von metrischen Variablen angegeben werden (siehe **Min und Max** oben), versucht die Prozedur, Werte für einen Fall zu ziehen, bis ein Set an Werten gefunden ist, das innerhalb des angegebenen Bereichs liegt. Wenn mit der angegebenen Zahl an Ziehungen pro Fall kein Set an Werten gefunden wird, zieht die Prozedur ein anderes Set an Modellparametern und wiederholt den Prozess der Fallziehung. Ein Fehler tritt auf, wenn ein Set von Werten im Bereich nicht in der angegebenen Zahl von Fall- und Parameterziehungen gefunden wird.

Beachten Sie, dass höhere Werte eine längere Verarbeitungszeit bedeuten. Wenn die Prozedur lange dauert oder keine geeigneten Ziehungen findet, prüfen Sie die angegebenen Minimum- und Maximumwerte, um sicherzustellen, dass sie angemessen sind.

## Ausgabe

**Anzeigen.** Steuert die Anzeige der Ausgabe. Eine Gesamtimputationszusammenfassung wird immer angezeigt. Sie enthält Tabellen in Bezug auf die Imputationsspezifikationen, die Iterationen (für die Methode vollständiger konditionaler Spezifikation), die abhängigen imputierten Variablen, die abhängigen Variablen, die von der Imputation ausgeschlossen sind, und die Imputationssequenz. Wenn angegeben, werden auch die Nebenbedingungen für Analysevariablen angezeigt.

- **Imputationsmodell.** Zeigt das Imputationsmodell für abhängige Variablen und Prädiktoren an und enthält den univariaten Modelltyp, Modelleffekte und die Anzahl der imputierten Werte.
- **Deskriptive Statistiken.** Zeigt die deskriptive Statistik für abhängige Variablen an, für die Werte imputiert sind. Für metrische Variablen enthält die deskriptive Statistik Mittelwert, Anzahl, Standardabweichung, Minimum und Maximum für die Originaleingabedaten (vor der Imputation), imputierte Werte (durch Imputation) und vollständige Daten (Original- und imputierte Werte gemeinsam - durch Imputation). Für kategoriale Variablen enthält die deskriptive Statistik Anzahl und Prozent nach Kate-

gorie für die Originaleingabedaten (vor der Imputation), imputierte Werte (durch Imputation) und vollständige Daten (Original- und imputierte Werte gemeinsam - durch Imputation).

**Iterationsverlauf.** Wenn die Methode vollständiger konditionaler Spezifikation verwendet wird, können Sie ein Dataset anfordern, das die Iterationsverlaufsdaten für die Imputation nach vollständiger konditionaler Spezifikation enthält. Das Dataset enthält Mittelwerte und Standardabweichungen nach Iteration und Imputation für jede metrische abhängige Variable, für die Werte imputiert sind. Sie können die Daten als Diagramm darstellen, um die Beurteilung der Modellkonvergenz zu erleichtern. Weitere Informationen finden Sie im Thema .

---

## Zusätzliche Funktionen beim Befehl MULTIPLE IMPUTATION

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Geben Sie ein Subset von Variablen an, für die deskriptive Statistik angezeigt wird (Unterbefehl IMPUTATIONSUMMARIES).
- Geben Sie eine Analyse fehlender Muster und Imputation in einem einzigen Lauf der Prozedur an.
- Geben Sie die maximale Anzahl an Modellparametern an, die zulässig sind, wenn eine Variable imputiert wird (Schlüsselwort MAXMODELPARAM).

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

---

## Arbeiten mit Daten aus multipler Imputation

Wenn ein Dataset multipler Imputation (MI) erstellt wird, wird eine Variable mit dem Namen *Imputation\_* und der Variablenbeschriftung *Imputationszahl* hinzugefügt und das Dataset wird danach in aufsteigender Reihenfolge sortiert. Fälle aus dem Originaldataset haben einen Wert von 0. Fälle imputierter Werte sind von 1 bis  $M$  nummeriert, wobei  $M$  die Zahl der Imputationen ist.

Wenn Sie ein Dataset öffnen, identifiziert das Vorhandensein der *Imputation\_* das Dataset als mögliches MI-Dataset.

Aktivieren eines MI-Datasets für die Analyse

Das Dataset muss mit der Option **Gruppen vergleichen** mit *Imputation\_* als Gruppierungsvariable aufgeteilt werden, um in Analysen als MI-Dataset behandelt zu werden. Sie können auch Aufteilungen bei anderen Variablen definieren.

Wählen Sie die folgenden Befehle aus den Menüs aus:

### Daten > Datei aufteilen...

1. Wählen Sie die Option **Gruppen vergleichen**.
2. Wählen Sie *Imputationszahl [Imputation\_]* als Variable, um Fälle danach zu gruppieren.

Alternativ wird die Datei, wenn Sie Markierungen einschalten (siehe unten), bei *Imputationszahl (Imputation\_)* geteilt.

Unterscheidung von imputierten Werten und beobachteten Werten

Sie können imputierte Werte von beobachteten Werten über die Zellenhintergrundfarbe, die Schriftart und den Fettdruck (für imputierte Werte) unterscheiden. Wenn Sie in der aktuellen Sitzung ein neues Dataset mit "Fehlende Werte ersetzen" erstellen, werden Markierungen standardmäßig eingeschaltet. Wenn Sie eine gespeicherte Datendatei öffnen, die Imputationen enthält, werden Markierungen ausgeschaltet.

Um die Markierungen einzuschalten, wählen Sie aus den Menüs im Dateneditor:

## Ansicht > Imputierte Daten markieren...

Alternativ können Sie Markierungen einschalten, indem Sie in der Datenansicht des Dateneditors auf die Schaltfläche zur Imputationsmarkierung rechts in der Bearbeitungsleiste klicken.

Wechseln zwischen Imputationen

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

**Bearbeiten > Gehe zu Imputation...**

2. Wählen Sie die Imputation (oder die Originaldaten) aus der Dropdown-Liste.

Alternativ können Sie in der Datenansicht des Dateneditors die Imputation aus der Dropdown-Liste in der Bearbeitungsleiste auswählen.

Die relative Fallposition wird bei der Auswahl der Imputationen beibehalten. Wenn es im Originaldataset 1.000 Fälle gibt, wird Fall 1.034, der 34. Fall in der ersten Imputation, oben im Raster angezeigt. Wenn Sie Imputation 2 in der Dropdown-Liste auswählen, würde Fall 2.034, der 34. Fall in Imputation 2 oben im Raster angezeigt werden. Wenn Sie **Originaldaten** in der Dropdown-Liste wählen, würde Fall 34 oben im Raster angezeigt werden. Auch die Spaltenposition wird beibehalten, wenn zwischen Imputationen gewechselt wird, sodass der Vergleich von Werten zwischen Imputationen erleichtert wird.

Transformieren und Bearbeiten imputierter Werte

Manchmal müssen Sie Transformationen an imputierten Daten durchführen. Zum Beispiel könnten Sie das Protokoll aller Werte einer Gehaltsvariablen nehmen und das Ergebnis in einer neuen Variablen speichern. Ein Wert, der über imputierte Daten berechnet wurde, wird als imputiert behandelt, wenn er sich von dem Wert, der mit den Originaldaten berechnet wurde, unterscheidet.

Wenn Sie einen imputierten Wert in einer Zelle des Dateneditors bearbeiten, wird die Zelle immer noch als imputiert behandelt. Es wird nicht empfohlen, imputierte Werte auf diese Weise zu bearbeiten.

---

## Analysieren von Daten multipler Imputation

Viele Prozeduren unterstützen das Pooling von Ergebnissen aus Analysen multipler imputierter Datensets. Wenn Imputationsmarkierungen eingeschaltet sind, wird ein Spezialsymbol neben den Prozeduren angezeigt, die Pooling unterstützen. Im Untermenü "Deskriptive Statistik" des Menüs "Analysieren" zum Beispiel unterstützen "Häufigkeiten", "Deskriptive Statistik", "Explorative Datenanalyse" und "Kreuztabellen" Pooling, während "Verhältnisskala", "P-P-Diagramme" und "Q-Q-Diagramme" kein Pooling unterstützen.

Sowohl die Tabellenausgabe als auch Modell-PMML unterstützen Pooling. Es gibt keine neue Prozedur für die Anforderung gepoolter Ausgabe. Stattdessen haben Sie über eine neue Registerkarte im Dialogfeld "Optionen" die Möglichkeit, die Ausgabe multipler Imputation zu steuern.

- **Pooling der Tabellenausgabe.** Standardmäßig werden die Ergebnisse, wenn Sie eine unterstützte Prozedur an einem MI-Dataset (Multiple Imputation) ausführen, automatisch für jede Imputation, die Originaldaten (nicht imputiert) und gepoolte (final) Ergebnisse erzeugt, die die Variation über die Imputationen berücksichtigen. Die gepoolten Statistiken unterscheiden sich je nach Prozedur.
- **Pooling von PMML.** Sie können auch gepoolte PMML von unterstützten Prozeduren erhalten, die PMML exportieren. Gepooltes PMML wird auf die gleiche Weise angefordert und wird statt nicht gepoolter PMML gespeichert.

Nicht unterstützte Prozeduren erzeugen entweder gepoolte Ausgabe oder gepoolte PMML-Dateien.

Pooling-Stufen

Die Ausgabe wird mittels einer von zwei Stufen gepoolt:

- **Naive Kombination.** Nur der gepoolte Parameter ist verfügbar.
- **Univariate Kombination.** Der gepoolte Parameter, sein Standardfehler, die Teststatistik und die effektiven Freiheitsgrade, der  $p$ -Wert, das Konfidenzintervall und die Pooling-Diagnose (Bruchteil der fehlenden Informationen, relative Effizienz, relativer Anstieg der Varianz) werden, wenn verfügbar, angezeigt.

Koeffizienten (Regression und Korrelation), Mittelwerte (und mittlere Differenzen) und Häufigkeiten werden typischerweise in Pools zusammengefasst. Wenn der Standardfehler der Statistik verfügbar ist, wird das univariate Pooling verwendet, andernfalls das naive Pooling.

Prozeduren, die Pooling unterstützen

Die folgenden Prozeduren unterstützen MI-Datasets mit den für jeden Ausgabeteil angegebenen Pooling-Stufen.

**Häufigkeiten.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Statistiken" unterstützt Mittelwerte bei univariatem Pooling (wenn auch der Standardfehler des Mittelwerts angefordert wird) und Gültiges-N und Fehlendes-N bei naivem Pooling.
- Die Tabelle "Häufigkeiten" unterstützt Häufigkeit bei naivem Pooling.

**Deskriptive Statistiken.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Deskriptive Statistiken" unterstützt Mittelwerte bei univariatem Pooling (wenn auch der Standardfehler des Mittelwerts angefordert wird) und N bei naivem Pooling.

**Kreuztabellen.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Kreuztabelle" unterstützt Anzahl bei naivem Pooling.

**Mittelwert.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Bericht" unterstützt Mittelwerte bei univariatem Pooling (wenn auch der Standardfehler des Mittelwerts angefordert wird) und N bei naivem Pooling.

**T-Test bei einer Stichprobe.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Statistik" unterstützt Mittelwert bei univariatem Pooling und N bei naivem Pooling.
- Die Tabelle "Test" unterstützt mittlere Differenz bei univariatem Pooling.

**T-Test bei unabhängigen Stichproben.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Gruppenstatistik" unterstützt Mittelwert bei univariatem Pooling und N bei naivem Pooling.
- Die Tabelle "Test" unterstützt mittlere Differenz bei univariatem Pooling.

**T-Test bei Stichproben mit paarigen Werten.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Statistik" unterstützt Mittelwerte bei univariatem Pooling und N bei naivem Pooling.
- Die Tabelle "Korrelationen" unterstützt Korrelationen und N bei naivem Pooling.
- Die Tabelle "Test" unterstützt Mittelwert bei univariatem Pooling.

**Einfaktorielle ANOVA.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Deskriptive Statistik" unterstützt Mittelwert bei univariatem Pooling und N bei naivem Pooling.
- Die Tabelle "Kontrasttests" unterstützt Kontrastwert bei univariatem Pooling.

**Lineare gemischte Modelle.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Deskriptive Statistik" unterstützt Mittelwert und N bei naivem Pooling.
- Die Tabelle "Schätzungen fester Effekte" unterstützt Schätzungen bei univariatem Pooling.

- Die Tabelle "Schätzungen von Kovarianzparametern" unterstützt Schätzungen bei univariatem Pooling.
- Die Tabelle "Geschätzte Randmittel: Schätzungen" unterstützt Mittelwert bei univariatem Pooling.
- Die Tabelle "Geschätzte Randmittel: Paarweise Vergleiche" unterstützt Mittelwertdifferenz bei univariatem Pooling.

**Verallgemeinerte lineare Modelle und verallgemeinerte Schätzungsgleichungen.** Diese Prozeduren unterstützen gepooltes PMML.

- Die Tabelle "Informationen zu kategorialen Variablen" unterstützt N und Prozente bei naivem Pooling.
- Die Tabelle "Informationen zu stetigen Variablen" unterstützt N und Mittelwert bei naivem Pooling.
- Die Tabelle "Parameterschätzungen" unterstützt den Koeffizienten B bei univariatem Pooling.
- Die Tabelle "Geschätzte Randmittel: Schätzkoeffizienten" unterstützt Mittelwert bei naivem Pooling.
- Die Tabelle "Geschätzte Randmittel: Schätzungen" unterstützt Mittelwert bei univariatem Pooling.
- Die Tabelle "Geschätzte Randmittel: Paarweise Vergleiche" unterstützt Mittelwertdifferenz bei univariatem Pooling.

**Bivariate Korrelationen.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Deskriptive Statistik" unterstützt Mittelwert und N bei naivem Pooling.
- Die Tabelle "Korrelationen" unterstützt Korrelationen und N bei univariatem Pooling. Beachten Sie, dass Korrelationen vor dem Pooling mit der z-Transformation von Fisher transformiert und nach dem Pooling wieder rücktransformiert werden.

**Partielle Korrelation.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Deskriptive Statistik" unterstützt Mittelwert und N bei naivem Pooling.
- Die Tabelle "Korrelationen" unterstützt Korrelationen bei naivem Pooling.

**Lineare Regression.** Diese Prozedur unterstützt gepooltes PMML.

- Die Tabelle "Deskriptive Statistik" unterstützt Mittelwert und N bei naivem Pooling.
- Die Tabelle "Korrelationen" unterstützt Korrelationen und N bei naivem Pooling.
- Die Tabelle "Koeffizienten" unterstützt B bei univariatem Pooling und Korrelationen bei naivem Pooling.
- Die Tabelle "Korrelationskoeffizienten" unterstützt Korrelationen bei naivem Pooling.
- Die Tabelle "Residuenstatistik" unterstützt Mittelwert und N bei naivem Pooling.

**Binäre logistische Regression.** Diese Prozedur unterstützt gepooltes PMML.

- Die Tabelle "Variablen in der Gleichung" unterstützt B bei univariatem Pooling.

**Multinomiale logistische Regression.** Diese Prozedur unterstützt gepooltes PMML.

- Die Tabelle "Parameterschätzungen" unterstützt den Koeffizienten B bei univariatem Pooling.

**Ordinale Regression.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Parameterschätzungen" unterstützt den Koeffizienten B bei univariatem Pooling.

**Diskriminanzanalyse.** Diese Prozedur unterstützt gepooltes Modell-XML.

- Die Tabelle "Gruppenstatistik" unterstützt Mittelwert und Gültiges N bei naivem Pooling.
- Die Tabelle "Gepoolt innerhalb von Gruppenmatrizen" unterstützt Korrelationen bei naivem Pooling.
- Die Tabelle "Kanonische Diskriminanzfunktionskoeffizienten" unterstützt nicht standardisierte Koeffizienten bei naivem Pooling.
- Die Tabelle "Funktionen bei Gruppenmittelpunkten" unterstützt nicht standardisierte Koeffizienten bei naivem Pooling.
- Die Tabelle "Klassifizierungsfunktionskoeffizienten" unterstützt Koeffizienten bei naivem Pooling.



**Chi-Quadrat-Test.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Deskriptive Statistik" unterstützt Mittelwert und N bei naivem Pooling.
- Die Tabelle "Häufigkeiten" unterstützt Beobachtetes N bei naivem Pooling.

**Test auf Binomialverteilung.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Deskriptive Statistik" unterstützt Mittelwert und N bei naivem Pooling.
- Die Tabelle "Test" unterstützt N, beobachteter Anteil und Testanteil bei naivem Pooling.

**Sequenztest.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Deskriptive Statistik" unterstützt Mittelwert und N bei naivem Pooling.

**Kolmogorov-Smirnov-Test bei einer Stichprobe.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Deskriptive Statistik" unterstützt Mittelwert und N bei naivem Pooling.

**Test bei zwei unabhängigen Stichproben.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Ränge" unterstützt mittlerer Rang und N bei naivem Pooling.
- Die Tabelle "Häufigkeiten" unterstützt N bei naivem Pooling.

**Test bei mehreren unabhängigen Stichproben.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Ränge" unterstützt mittlerer Rang und N bei naivem Pooling.
- Die Tabelle "Häufigkeiten" unterstützt Anzahlen bei naivem Pooling.

**Tests bei zwei verbundenen Stichproben.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Ränge" unterstützt mittlerer Rang und N bei naivem Pooling.
- Die Tabelle "Häufigkeiten" unterstützt N bei naivem Pooling.

**Tests bei mehreren verbundenen Stichproben.** Die folgenden Funktionen werden unterstützt:

- Die Tabelle "Ränge" unterstützt mittlerer Rang bei naivem Pooling.

**Cox-Regression.** Diese Prozedur unterstützt gepooltes PMML.

- Die Tabelle "Variablen in der Gleichung" unterstützt B bei univariatem Pooling.
- Die Tabelle "Kovariate Mittelwerte" unterstützt Mittelwert bei naivem Pooling.

---

## Optionen für multiple Imputation

Die Registerkarte "Multiple Imputationen" steuert zwei Arten von Voreinstellungen für multiple Imputationen:

**Erscheinungsbild imputierter Daten.** Standardmäßig werden Zellen mit imputierten Daten mit einer anderen Hintergrundfarbe als Zellen mit nicht imputierten Daten angezeigt. Das Erscheinungsbild der imputierten Daten sollte es Ihnen erleichtern, durch ein Dataset zu blättern und diese Zellen zu finden. Sie können die Standardhintergrundfarbe für die Zellen und die Schriftfamilie ändern und imputierte Daten fett darstellen.

**Analyseausgabe.** Diese Gruppe steuert die Art der Viewer-Ausgabe, die erzeugt wird, wenn ein multiples imputiertes Dataset analysiert wird. Standardmäßig wird die Ausgabe für das Originaldataset (vor der Imputation) und für jedes der imputierten Datasets erzeugt. Zusätzlich werden finale zusammengefasste Ergebnisse für die Verfahren erzeugt, die das Pooling von imputierten Daten unterstützen. Bei univariatem Pooling wird auch die Pooling-Diagnose angezeigt. Sie können die Ausgaben, die Sie nicht sehen möchten, jedoch unterdrücken.

So stellen Sie die Optionen für multiple Imputation ein:

Wählen Sie die folgenden Befehle aus den Menüs aus:

**Bearbeiten > Optionen**

Klicken Sie auf die Registerkarte "Multiple Imputation".

---

## 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

## A

- Analyse fehlender Werte 3
  - deskriptive Statistiken 5
  - EM 7
  - Erwartungsmaximierung 9
  - Imputieren fehlender Werte 6
  - MCAR-Test 6
  - Methoden 6
  - Muster 4
  - Regression 8
  - Schätzen von Statistiken 6
  - zusätzliche Funktionen beim Befehl 9
- Analysieren von Mustern 12

## E

- EM
  - in "Analyse fehlender Werte" 7

## F

- Fälle in Tabellen
  - in "Analyse fehlender Werte" 4
- Fehlende Werte
  - univariate Statistiken 5

## H

- Häufigkeiten extremer Werte
  - in "Analyse fehlender Werte" 5
- Häufigkeitstabellen
  - in "Analyse fehlender Werte" 5

## I

- Imputieren fehlender Datenwerte 13
  - Ausgabe 16
  - Imputationsmethode 14
  - Nebenbedingungen 15
- Indikatorvariablen
  - in "Analyse fehlender Werte" 5
- Indikatorvariablen für fehlende Werte
  - in "Analyse fehlender Werte" 5
- Iterationsverlauf
  - in multipler Imputation 16

## K

- Korrelationen
  - in "Analyse fehlender Werte" 7, 8
- Kovarianz
  - in "Analyse fehlender Werte" 7, 8

## L

- Listenweiser Ausschluss
  - in "Analyse fehlender Werte" 3

## M

- MCAR-Test
  - in "Analyse fehlender Werte" 3
- MCAR-Test nach Little 6
  - in "Analyse fehlender Werte" 3
- Mittelwert
  - in "Analyse fehlender Werte" 5, 7, 8
- monotone Imputation
  - in multipler Imputation 14
- Multiple Imputation 11, 17, 18
  - Analysieren von Mustern 12
  - Imputieren fehlender Datenwerte 13

## N

- Nichtübereinstimmung
  - in "Analyse fehlender Werte" 5
- Normale Variaten
  - in "Analyse fehlender Werte" 8

## P

- Paarweiser Ausschluss
  - in "Analyse fehlender Werte" 3

## R

- Regression
  - in "Analyse fehlender Werte" 8
- Residuen
  - in "Analyse fehlender Werte" 8

## S

- Sortieren von Fällen
  - in "Analyse fehlender Werte" 4
- Standardabweichung
  - in "Analyse fehlender Werte" 5
- Student-T-Test
  - in "Analyse fehlender Werte" 8

## T

- t-Test
  - in "Analyse fehlender Werte" 5
- Tabellarische Darstellung von Kategorien
  - in "Analyse fehlender Werte" 5

## U

- Unvollständige Daten
  - siehe "Analyse fehlender Werte" 3

## V

- Vollständig konditionale Spezifikation
  - in multipler Imputation 14







