

IBM SPSS Forecasting 29



Uwaga

Przed użyciem tych informacji i produktu, którego one dotyczą, przeczytaj informacje znajdujące się w sekcji “Uwagi” na stronie 53.

Informacje o produkcie

Niniejsze wydanie dotyczy wersji 29, wydania 0, modyfikacji 1 produktu IBM® SPSS Statistics oraz wszystkich kolejnych wydań i modyfikacji, dopóki nie zostanie to określone inaczej w nowych wydaniach.

© **Copyright International Business Machines Corporation .**

Spis treści

Rozdział 1. Prognozowanie.....	1
Wprowadzenie do szeregów czasowych.....	1
Dane szeregu czasowego.....	1
Przekształcenia danych.....	2
Okresy szacowania i walidacji.....	2
Tworzenie modeli i generowanie prognoz.....	3
Kreator modeli szeregów czasowych.....	3
Określanie opcji automatycznego doboru modelu.....	5
Niestandardowy model wygładzania wykładniczego.....	6
Modele ARIMA użytkownika.....	7
Wyniki.....	10
Zapisywanie predykcji modelu i specyfikacji modelu.....	12
Opcje.....	13
Dodatkowe właściwości komendy TSMODEL.....	14
Zastosuj modele szeregów czasowych.....	14
Wyniki.....	16
Zapisywanie predykcji modelu i specyfikacji modelu.....	18
Opcje.....	19
Dodatkowe właściwości komendy TSAPPLY.....	19
Dekompozycja sezonowa.....	20
Dekompozycja sezonowa: Zapisz.....	20
Dodatkowe właściwości komendy SEASON.....	21
Wykresy analizy spektralnej.....	21
Dodatkowe właściwości komendy SPECTRA.....	23
Modele przyczynowe szeregów czasowych.....	23
Aby uzyskać model przyczynowy szeregów czasowych.....	24
Szeregi czasowe dla modelowania	25
Obserwacje	26
Przedział czasowy analizy	27
Agregacja i rozkład	27
Brakujące wartości	28
Ogólne opcje danych	29
Ogólne opcje budowania	29
Prezentacja szeregów	29
Opcje wyników	30
Okres estymacji	32
Prognoza	33
Zapisz	33
Wyjście interaktywne	34
Stosowanie modelowania przyczynowego szeregów czasowych.....	35
Stosowanie modelowania przyczynowego szeregów czasowych.....	35
Prognozowanie modelem przyczynowym szeregów czasowych.....	35
Scenariusze modelowania przyczynowego szeregów czasowych.....	41
Miary dobroci dopasowania.....	47
Typy wartości odstających.....	48
Przewodnik po Plotkach ACF/PACF.....	48
Uwagi.....	53
Znaki towarowe.....	54

Indeks.....57

Rozdział 1. Prognozowanie

Następujące funkcje prognozowania są dostępne w SPSS Statistics produktu Professional Edition oraz w opcji Forecasting.

Wprowadzenie do szeregów czasowych

Szereg czasowy jest zbiorem obserwacji uzyskanych w wyniku regularnych pomiarów jednej zmiennej w pewnym okresie. Na przykład w szeregu danych dotyczących zapasów obserwacje mogą przedstawiać dzienne poziomy zapasów w okresie kilku miesięcy. Szereg odzwierciedlający udział produktu w rynku produktu może składać się z danych zbieranych co tydzień przez kilka lat. Szereg całkowitych wielkości sprzedaży może składać się z obserwacji dokonywanych raz w miesiącu przez wiele lat. W każdym z tych przykładów niektóre zmienne były obserwowane w regularnych, znanych odstępach czasu przez określony czas. Z tego względu postać danych typowego szeregu czasowego jest pojedynczą sekwencją lub listą obserwacji odzwierciedlających pomiary wykonywane w regularnych odstępach czasu.

Tabela 1. Szereg czasowy danych o zapasach

Czas	Tydzień	Dzień	Poziom zapasów
t ₁	1	poniedziałek	160
t ₂	1	wtorek	135
t ₃	1	Wednesday	129
t ₄	1	Thursday	122
t ₅	1	Friday	108
t ₆	2	Monday	150
		...	
t ₆₀	12	Friday	120

Jednym z najważniejszych powodów przeprowadzenia analizy szeregów czasowych jest dążenie do przewidzenia przyszłych wartości szeregu. Model szeregu, który wyjaśniał poprzednie wartości, może również przewidywać, czy kilka następnych wartości będzie większych czy mniejszych (i jak bardzo). Zdolność do takich predykcji jest ważna w każdym rodzaju działalności gospodarczej lub naukowej.

Dane szeregu czasowego

Dane kolumnowe

Każda zmienna szeregu czasowego zawiera dane dla jednego szeregu czasowego. Struktura ta jest tradycyjną strukturą danych szeregów czasowych, używaną w procedurze Kreator modeli szeregów czasowych, w procedurze Dekompozycja sezonowa oraz w procedurze Wykresy analizy spektralnej. Na przykład, aby zdefiniować szereg czasowy w Edytorze danych, kliknij kartę **Widok zmiennych** i wprowadź nazwę zmiennej w dowolnym pustym wierszu. Każda obserwacja w szeregu czasowym odpowiada obserwacji (wierszowi w Edytorze danych).

Po otwarciu arkusza kalkulacyjnego zawierającego dane szeregu czasowego każdy szereg powinien mieścić się w kolumnie arkusza. W przypadku, gdy użytkownik ma już arkusz kalkulacyjny z szeregami czasowymi rozmieszczonymi w wierszach, można otworzyć go mimo to i użyć opcji Transponuj w menu Dane w celu zamiany wierszy w kolumny.

Dane wielowymiarowe

W przypadku danych wielowymiarowych każda zmienna szeregu czasowego zawiera dane dla wielu szeregów czasowych. Oddzielne szeregi czasowe w ramach danej zmiennej są wówczas

identyfikowane na podstawie zestawu wartości zmiennych jakościowych, zwanych także zmiennymi wymiarów.

Na przykład dane sprzedaży dla różnych regionów i marek mogą być przechowywane w jednej zmiennej *sales*, tak że wymiarami w tym przypadku będą *region* i *brand*. Każda kombinacja *region* i *brand* określa konkretny szereg czasowy dla zmiennej *sales*. Na przykład w poniższej tabeli rekordy o wartości „north” dla wymiaru *region* oraz „brandX” dla wymiaru *brand* definiują jeden szereg czasowy.

data	region	brand	sprzedaż
01/01/2014	north	brandX	82350
01/01/2014	north	brandY	86380
01/01/2014	south	brandX	91375
01/01/2014	south	brandY	70320
01/02/2014	north	brandX	83275
01/02/2014	north	brandY	85260
01/02/2014	south	brandX	94760
01/02/2014	south	brandY	69870

Uwaga: Dane zaimportowane z kostek OLAP, na przykład z programu IBM Cognos TM1, są reprezentowane jako dane wielowymiarowe.

Przekształcenia danych

W analizie szeregów czasowych użyteczna jest pewna liczba procedur przekształcania danych udostępnianych w systemie podstawowym. Przekształcenia te mają zastosowanie tylko do danych kolumnowych, gdzie każda zmienna szeregów czasowych zawiera dane pojedynczego szeregu czasowego.

- Procedura Zdefiniuj daty (w menu Dane) generuje zmienne daty, które są używane do określania okresowości oraz do rozróżniania między okrasami historycznymi, walidacyjnymi i prognozowanymi. Algorytmy prognozowania są przeznaczone do pracy ze zmiennymi utworzonymi przez procedurę Zdefiniuj daty.
- Procedura Utwórz szereg czasowy (w menu Przekształcenia) tworzy nowe zmienne szeregów czasowych jako funkcje istniejących zmiennych szeregów czasowych. Zawiera funkcje, które na podstawie sąsiednich obserwacji dokonują wygładzania, uśredniania i różnicowania.
- Procedura Zastąp brakujące wartości (w menu Przekształcenia) zastępuje braki danych zdefiniowane i systemowe braki danych wartościami szacowanymi przy wykorzystaniu jednej z kilku metod. Braki danych na początku lub na końcu szeregu nie stanowią problemu, po prostu zmniejszają użyteczną długość szeregu. Luki wewnątrz szeregu (*osadzone* braki danych) mogą stanowić znacznie poważniejszy problem.

Szczegółowe informacje na temat przekształceń danych na analizy potrzeby szeregów czasowych zawiera *Podręcznik użytkownika systemu podstawowego*.

Okresy szacowania i walidacji

Często przydatne jest podzielenie szeregu czasowego na okres *oszacowania* lub *historyczny* i okres *walidacji*. Model jest opracowywany na podstawie obserwacji w okresie oszacowania (okresie historycznym), a następnie testowany w okresie walidacji. Wymuszając na modelu predykcję punktów, które są już znane (punktów w okresie walidacji), użytkownik uzyskuje pojęcie o jakości prognoz modelu.

Obserwacje w okresie walidacji są zwykle określane jako warianty kontrolne, ponieważ służą do kontroli, a nie są wykorzystywane w budowaniu modelu. Po upewnieniu się, że jakość predykcji modelu jest

zadowalająca, można ponownie zdefiniować okres oszacowania, aby uwzględnić warianty kontrolne, a następnie zbudować model finalny.

Tworzenie modeli i generowanie prognoz

Moduł dodatkowy Forecasting udostępnia następujące procedury do tworzenia modeli i generowania prognoz:

- Procedura [“Kreator modeli szeregów czasowych”](#) na stronie 3 tworzy modele dla szeregów czasowych i generuje prognozy. Zawiera funkcję Automatyczny dobór modelu, która automatycznie wybiera najlepszy model dla każdego z szeregów czasowych użytkownika. Dla doświadczonych analityków, którzy chcieliby mieć większą kontrolę nad procedurą, dostępne są także narzędzia do tworzenia modeli użytkownika.
- Procedura [“Zastosuj modele szeregów czasowych”](#) na stronie 14 stosuje istniejące modele szeregów czasowych, utworzone przez procedurę Modelowanie szeregów czasowych, do aktywnego zbioru danych. Można w ten sposób uzyskać prognozy dla szeregów, dla których dostępne są nowe lub zmienione dane, bez konieczności tworzenia modeli od nowa. Jeśli są powody, by podejrzewać, że model uległ zmianie, można ponownie go utworzyć za pomocą procedury Modelowanie szeregów czasowych.
- Procedura [“Modele przyczynowe szeregów czasowych”](#) na stronie 23 buduje autoregresywny model każdego szeregu i automatycznie wybiera najlepsze zmienne wejściowe, które z przewidywanym szeregiem łączą relacje przyczynowe. Procedura generuje wyniki interaktywne, których można użyć do eksplorowania relacji przyczynowych. Procedura może również generować prognozy, wykrywać wartości odstające i wybrać szereg z największym prawdopodobieństwem powodujący powstanie wartości odstającej.
- Procedura [“Prognozowanie modelem przyczynowym szeregów czasowych”](#) na stronie 35 stosuje model przyczynowy szeregów czasowych do aktywnego zbioru danych. Za pomocą tej procedury można uzyskać prognozy dla szeregów, dla których dostępne są nowe lub zmienione dane, bez konieczności ponownego tworzenia modeli. Może ona także posłużyć do określenia szeregu, który z największym prawdopodobieństwem powoduje powstanie wartości odstających wykrytych przez procedurę Modele przyczynowe szeregów czasowych.

Kreator modeli szeregów czasowych

Procedura Kreator modeli szeregów czasowych estymuje wygładzanie wykładnicze, modele autoregresyjnej średniej ruchomej (ARIMA) jednej zmiennej i modele ARIMA wielu zmiennych (lub modele funkcji przeniesienia) dla szeregów czasowych, a następnie generuje prognozy. W skład procedury wchodzi funkcja Automatyczny dobór modelu, która próbuje automatycznie zidentyfikować i oszacować najlepiej pasujący model ARIMA lub model wygładzania wykładniczego dla jednej lub więcej zależnych od siebie serii zmiennych, eliminując w ten sposób konieczność identyfikacji odpowiedniego modelu na drodze prób i błędów. Zamiast tego można określić niestandardowy model ARIMA lub model wygładzania wykładniczego.

Przykład. Jesteś menedżerem ds. produktów odpowiedzialnym za utworzenie prognozy na kolejne miesiące dla sprzedaży jednostkowej oraz przychodu dla 100 odrębnych produktów, ale masz niewielkie doświadczenie (lub w ogóle nie masz doświadczenia) w modelowaniu szeregów czasowych. Dostępne dane o sprzedaży jednostkowej dla wszystkich 100 produktów są zapisane w arkuszu kalkulacyjnym programu Excel. Po otwarciu arkusza kalkulacyjnego w programie IBM SPSS Statistics, można użyć modułu automatycznego doboru modelu i zlecić wykonanie prognoz na kolejny miesiąc. Moduł automatycznego doboru modelu znajduje najlepszy model sprzedaży jednostkowej każdego produktu i wykorzystuje ten model do prognozowania. Moduł automatycznego doboru modelu może obsługiwać wiele szeregów czasowych, więc aby uzyskać prognozy dla wszystkich produktów, wystarczy tylko raz uruchomić procedurę. Wybranie opcji zapisywania prognoz w aktywnym zbiorze danych powoduje, że wyniki można łatwo wyeksportować z powrotem do programu Excel.

Statystyki. Miara dobroci dopasowania: stacjonarny R -kwadrat, R -kwadrat (R^2), pierwiastek z błędu średniokwadratowego (RMSE), średni bezwzględny błąd (MAE), średni bezwzględny błąd procentowy (MAPE), maksymalny bezwzględny błąd (MaxAE), maksymalny bezwzględny błąd procentowy (MaxAPE),

znormalizowane bayesowskie kryterium informacyjne (BIC). Reszty: funkcja autokorelacji, funkcja autokorelacji cząstkowej, Ljung-Box Q . W przypadku modeli ARIMA: rzędy modelu ARIMA dla zmiennych zależnych, rzędy funkcji funkcji przeniesienia dla zmiennych niezależnych i wartości odstające oszacowań. A także oszacowania parametru wygładzania dla modeli wygładzania wykładniczego.

Wykresy. Wykresy podsumowujące dla wszystkich modeli: histogramy stacjonarnego R -kwadrat, R -kwadrat (R^2), pierwiastek z błędu średniokwadratowego (RMSE), średni bezwzględny błąd (ŚBB), średni bezwzględny błąd procentowy (MAPE), maksymalny bezwzględny błąd (MaxAE), maksymalny bezwzględny błąd procentowy (MaxAPE), znormalizowane bayesowskie kryterium informacyjne (BIC); wykresy skrzynkowe autokorelacji reszt i autokorelacji cząstkowej. Wyniki dla poszczególnych modeli: wartości prognozy, wartości dopasowania, wartości obserwowane, górna i dolna granica ufności, autokorelacje reszt i autokorelacje cząstkowe.

Wymagania dotyczące danych kreatora modeli szeregów czasowych

Dane. Zmienna zależna i zmienne niezależne powinny być numeryczne.

Założenia. Zmienna zależna i wszystkie zmienne niezależne są traktowane jako szeregi czasowe, co oznacza, że każda obserwacja reprezentuje punkt czasowy, przy czym kolejne obserwacje są rozdzielane stałym przedziałem czasowym.

- **Stacjonarność.** W przypadku niestandardowych modeli ARIMA modelowane szeregi czasowe powinny być stacjonarne. Najbardziej skutecznym sposobem przekształcenia szeregów niestacjonarnych na stacjonarne jest zastosowanie transformacja różnicowa – dostępna w oknie dialogowym Utwórz szereg czasowy.
- **Prognozy.** W przypadku tworzenia prognoz z użyciem modeli ze zmiennymi niezależnymi (predyktor) aktywny zestaw danych powinien zawierać wartości tych zmiennych dla wszystkich obserwacji z okresu prognozy. Ponadto, zmienne niezależne nie powinny zawierać żadnych braków danych w okresie estymacji.

Definiowanie dat

Chociaż nie jest to wymagane, zaleca się użycie okna dialogowego Definiuj datę i czas, aby określić datę powiązaną z pierwszą obserwacją oraz przedział czasowy pomiędzy kolejnymi obserwacjami. Należy to zrobić przed użyciem kreatora modeli szeregów czasowych i wyników z zestawu zmiennych, które tworzą etykiety dat powiązanych z każdą obserwacją. Pozwala to również ustawić zakładaną okresowość dat – przykładowo, okresowość wynosi 12, jeśli przedział czasu pomiędzy kolejnymi obserwacjami to jeden miesiąc. Taka okresowość jest wymagana, jeśli mają zostać utworzone modele sezonowe. Jeśli nie są wymagane modele okresowe ani etykiety wyników, można pominąć okno dialogowe Definiuj datę i czas. Etykietami powiązanymi z poszczególnymi obserwacjami są po prostu numery obserwacji.

Aby użyć kreatora modeli szeregów czasowych

1. Z menu wybierz:

Analizuj > Prognozowanie > Utwórz modele tradycyjne...

2. Na karcie Zmienne wybierz co najmniej jedną zmienną zależną do modelowania.
3. Z okna rozwijanego Metoda wybierz metodę modelowania. Aby wybrać modelowanie automatyczne, pozostaw wybraną metodę domyślną **Automatyczny dobór modelu**. Wówczas automatyczny dobór modelu określi najbardziej dopasowany model dla każdej zmiennej zależnej.

Aby utworzyć prognozy:

4. Kliknij kartę **Opcje**.
5. Określ okres prognozy. Spowoduje to utworzenie wykresu, który będzie obejmował prognozy i obserwowane wartości.

Opcjonalnie można wykonać następujące kroki:

- Wybierz co najmniej jedną zmienną niezależną. Zmienne niezależne są traktowane podobnie jak predyktory w analizie regresji, jednak są opcjonalne. Mogą być uwzględniane w modelach ARIMA, ale nie mają zastosowania w modelach, które nie są modelami wygładzania wykładniczego. Jeśli jako

metoda modelowania wybrana zostanie opcja **Automatyczny dobór modelu** i uwzględnione zostaną zmienne niezależne, uwzględniane będą tylko modele ARIMA.

- Kliknij opcję **Kryteria**, aby określić szczegóły modelowania.
- Zapisz predykcje, przedziały ufności i reszty modelu.
- Zapisz oszacowane modele w formacie XML. Zapisane modele można zastosować do nowych lub skorygowanych danych, aby uzyskać zaktualizowane prognozy bez ponownego budowania modeli.
- Uzyskaj statystyki podsumowujące dla wszystkich oszacowanych modeli.
- Określ funkcje przeniesienia zmiennych niezależnych w niestandardowych modelach ARIMA.
- Włącz automatyczne wykrywanie wartości odstających.
- Utwórz model specyficznych punktów czasowych jako wartości odstających dla niestandardowych modeli ARIMA.

Metody modelowania

Dostępne metody modelowania to:

Automatyczny dobór modelu. Funkcja Automatyczny dobór modelu automatycznie znajduje najlepiej pasujący model dla każdego szeregu zależnego. Jeśli określone są zmienne niezależne (predykcyjne), funkcja Automatyczny dobór modelu wybiera do włączenia do modeli ARIMA te, które mają statystycznie istotny związek z szeregiem zależnym. Zmienne modelu są przekształcane odpowiednio do potrzeb, za pomocą transformacji różnicującej i/lub pierwiastka kwadratowego lub logarytmu naturalnego. Domyślnie funkcja Automatyczny dobór modelu bierze pod uwagę modele wygładzania wykładniczego i modele ARIMA. Można jednak ograniczyć działanie funkcji Automatyczny dobór modelu do wyszukiwania tylko modeli ARIMA lub tylko modeli z wygładzania wykładniczego. Można także zdecydować o automatycznym wykrywaniu wartości odstających.

Wygładzanie wykładnicze. Ta opcja umożliwia określenie niestandardowego modelu wygładzania wykładniczego. Można wybrać spośród szeregu modeli wygładzania wykładniczego, różniących się między sobą sposobem traktowania trendów i sezonowości.

ARIMA. Ta opcja umożliwia określenie niestandardowego modelu ARIMA. Wiąże się to z wyraźnym określeniem autoregresywnych i średnich ruchomych zleceń, a także stopnia zróżnicowania. Można dołączyć niezależne (predykcyjne) zmienne i zdefiniować funkcje przeniesienia dla każdej lub wszystkich z nich. Można również określić automatyczne wykrywanie wartości odstających lub określić wyraźny zestaw wartości odstających.

Okres szacowania i okres prognozy

Okres szacowania. Okres szacowania definiuje zestaw obserwacji używanych do określenia modelu. Domyślnie, okres szacowania obejmuje wszystkie obserwacje w aktywnym zestawie danych. Aby ustawić okres szacowania, wybierz opcję **Na podstawie czasu lub zakresu obserwacji** w oknie dialogowym Wybierz obserwacje. W zależności od dostępnych danych okres szacowania, z jakiego korzysta procedura, może różnić się w zależności od zmiennej zależnej i dlatego będzie inny od wyświetlanej wartości. Dla zadanej zmiennej zależnej rzeczywistym okresem szacowania jest okres pozostały po wyeliminowaniu ciągłych braków danych zmiennej, występujących na początku lub na końcu danego okresu szacowania.

Okres prognozy. Okres prognozy rozpoczyna się od pierwszej obserwacji po okresie szacowania i domyślnie trwa do ostatniej obserwacji w aktywnym zestawie danych. Koniec okresu prognozy można ustawić na karcie Opcje.

Określanie opcji automatycznego doboru modelu

Automatyczny dobór modelu udostępnia opcje ograniczające zbiór modeli kandydackich, określające sposób traktowania wartości odstających i uwzględniania zmiennych zdarzeń.

Wybór modelu i specyfikacja zdarzenia

Karta Model umożliwia określenie typów modeli branych pod uwagę w metodzie Automatyczny dobór modelu oraz określenie zmiennych zdarzeń.

Typ modelu. Dostępne są następujące opcje:

- **Wszystkie modele.** Funkcja automatycznego doboru modeli uwzględnia zarówno modele ARIMA, jak i modele wygładzania wykładniczego.
- **Tylko modele wygładzania wykładniczego.** Automatyczny dobór modeli uwzględnia tylko modele wygładzania wykładniczego.
- **Tylko modele ARIMA.** Automatyczny dobór modeli uwzględnia tylko modele ARIMA.

Automatyczny dobór modelu uwzględnia modele sezonowe. Ta opcja jest aktywna tylko wówczas, jeśli dla aktywnego zbioru danych zdefiniowano okresowość. Po wybraniu (zaznaczeniu) tej opcji funkcja Automatyczny dobór modelu bierze pod uwagę zarówno sezonowe jak i niesezone modele. W przypadku niezaznaczenia tej opcji automatyczny dobór modeli uwzględnia tylko modele niesezone.

Bieżąca okresowość. Wskazuje okresowość (jeśli istnieje) aktualnie zdefiniowaną dla aktywnego zbioru danych. Bieżąca okresowość jest podawana w postaci liczby całkowitej – na przykład 12 dla okresu rocznego, przy czym każdy przypadek reprezentuje miesiąc. Jeśli nie ustawiono okresowości, widoczna jest wartość *Brak*. W przypadku modeli sezonowych wymagana jest okresowość. Okresowość można ustawić w oknie Zdefiniuj daty.

Zdarzenia. Wybierz dowolne zmienne niezależne, które mają być traktowane jak zmienne zdarzeń. W przypadku zmiennych zdarzeń obserwacje o wartości 1 wskazują czas, w którym oczekuje się, że zdarzenie będzie miało wpływ na szereg zmiennych zależnych. Wartości inne niż 1 oznaczają brak wpływu.

Sposób postępowania z wartościami odstającymi podczas automatycznego doboru

Karta Wartości odstające umożliwia wybór opcji automatycznego wykrywania wartości odstających oraz typów wykrywanych wartości odstających.

Automatycznie wykryj wartości odstające. Domyślnie automatyczne wykrywanie wartości odstających nie jest przeprowadzane. Wybierz (zaznacz) tę opcję, aby automatycznie wykrywać wartości odstające, a następnie wybierz jeden lub więcej spośród poniższych typów wartości odstających:

- Addytywne
- Przesunięcie poziomu
- Innowacyjne
- Przejściowe
- Sezonowo addytywne
- Trend lokalny
- Fragment dodatkowy

Niestandardowy model wygładzania wykładniczego

Typ modelu. Modele wygładzania wykładniczego¹ są klasyfikowane jako sezonowe lub niesezone. Modele sezonowe są dostępne tylko wtedy, gdy określono okresowość dla aktywnego zbioru danych (patrz sekcja „Bieżąca okresowość” poniżej).

- *Simple.* Ten model jest odpowiedni dla serii, w której nie ma trendu ani sezonowości. Jedynym parametrem wygładzania jest poziom. Wygasające wygładzanie wykładnicze jest najbardziej podobne do modelu ARIMA z 0 rzędów autoregresji, 1 rzędem różnic i 1 rzędem ruchomej średniej.
- *Trend liniowy Holta.* Ten model jest odpowiedni dla szeregów, w których występuje trend liniowy i nie ma sezonowości. Parametry wygładzania to poziom i trend; nie są one ograniczone swoimi wartościami. Model Holta jest bardziej ogólny niż model Browna, ale w przypadku dłuższych szeregów przeliczanie zajmuje więcej czasu. Wygładzanie wykładnicze Holta jest najbardziej podobne do modelu ARIMA z 0 rzędów autoregresji, 2 rzędami różnic i 2 rzędami ruchomej średniej.

¹ Gardner, E. S. 1985. Exponential smoothing: The state of the art. *Journal of Forecasting*, 4, 1-28.

- *Trend liniowy Browna*. Ten model jest odpowiedni dla szeregów, w których występuje trend liniowy i nie ma sezonowości. Parametry wygładzania to poziom i trend; zakłada się, że są one równe. Zatem model Browna jest specjalnym przypadkiem modelu Holta. Wygładzanie wykładnicze Browna jest podobne do modelu ARIMA z zerowym rzędem autoregresji, dwoma rzędami różnic oraz dwoma rzędami ruchomych średnich ze współczynnikiem drugiego rzędu ruchomej średniej równym kwadratowi połowy współczynnika pierwszego rzędu.
- *Trend wygasający*. Model ten jest odpowiedni dla szeregów z trendem liniowym, który umiera na zewnątrz i bez sezonowości. Parametry wygładzania to: poziom, trend i trend gasnący. Wygasające wygładzanie wykładnicze jest najbardziej podobne do modelu ARIMA z 1 rzędem autoregresji, 1 rzędem różnic i 2 rzędami ruchomej średniej.
- *Prosty sezonowy*. Ten model jest odpowiedni dla serii bez trendu i efektu sezonowego, który jest stały w czasie. Parametry wygładzania to: poziom i sezon. Proste sezonowe wygładzanie wykładnicze jest podobne do modelu ARIMA z zerowym rzędem autoregresji, jednym rzędem różnic oraz jednym rzędem różnic sezonowych i rzędami 1, p oraz p+1 średniej ruchomej, gdzie p jest liczbą okresów w przedziale sezonowym (dla danych miesięcznych p=12).
- *Addytywny model Wintersa*. Model ten jest odpowiedni dla szeregów o trendach liniowych i sezonowych, które nie zależą od poziomu serii. Parametry wygładzania to: poziom, trend i sezon. Wygładzanie wykładnicze addytywnego modelu Wintersa jest podobne do modelu ARIMA z zerowym rzędem autoregresji, jednym rzędem różnic oraz jednym rzędem różnic sezonowych i p+1 rzędami średniej ruchomej, gdzie p jest liczbą okresów w przedziale sezonowym (dla danych miesięcznych p=12).
- *Multiplikatywny model Wintersa*. Model ten jest odpowiedni dla szeregów z trendem liniowym i efektem sezonowym, który zależy od poziomu szeregu. Parametry wygładzania to: poziom, trend i sezon. Wygładzanie wykładnicze multiplikatywnego modelu Wintersa nie jest podobne do modelu ARIMA.

Bieżąca okresowość. Wskazuje okresowość (jeśli istnieje) aktualnie zdefiniowaną dla aktywnego zbioru danych. Bieżąca okresowość jest podawana w postaci liczby całkowitej – na przykład 12 dla okresu rocznego, przy czym każdy przypadek reprezentuje miesiąc. Jeśli nie ustawiono okresowości, widoczna jest wartość *Brak*. W przypadku modeli sezonowych wymagana jest okresowość. Okresowość można ustawić w oknie Zdefiniuj daty.

Transformacja zmiennej zależnej. Istnieje możliwość określenia transformacji wykonywanej dla każdej zmiennej zależnej przed jej zamodelowaniem.

- **Brak.** Nie jest wykonywana żadna transformacja.
- **1/pierwiastek kwadratowy.** Przekształcenie pierwiastkiem kwadratowym.
- **Logarytm naturalny.** Transformacja z użyciem logarytmu naturalnego.

Modele ARIMA użytkownika

Program do modelowania szeregów czasowych umożliwia tworzenie niestandardowych modeli niesezonowych lub sezonowych ARIMA (Autoregressive Integrated Moving Average), zwanych również Box-Jenkins.² modele – z ustalonym zestawem zmiennych predyktorów lub bez niego. Można zdefiniować funkcje przeniesienia dla dowolnych lub wszystkich predyktorów i nakazać automatyczne wykrywanie wartości odstających albo wprost zdefiniować zbiór wartości odstających.

- W modelu jawnie uwzględnione są wszystkie zmienne niezależne (predyktory) określone na karcie Zmienne. Jest to inny sposób działania niż w przypadku automatycznego doboru modelu, w którym zmienne niezależne są uwzględniane tylko jeśli istnieje statystycznie znacząca relacja między nimi a zmienną zależną.

Określanie niestandardowych modeli ARIMA

Karta Model umożliwia określenie struktury niestandardowego modelu ARIMA.

² Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.

Rzędy ARIMA. Należy wprowadzić wartości dla różnych składników ARIMA modelu do odpowiednich komórek siatki Struktura. Wszystkie wartości muszą być nieujemnymi liczbami całkowitymi. W przypadku składników autoregresja i średnia ruchoma wartość ta reprezentuje rząd maksymalny. W modelu zostaną uwzględnione wszystkie niższe rzędy dodatnie. Na przykład, jeśli zostanie określona wartość 2, model będzie obejmował rzędy 2 i 1. Komórki w kolumnie Sezonowe są włączone tylko wtedy, gdy dla aktywnego zbioru danych zdefiniowano okresowość (patrz poniżej "Bieżąca okresowość").

- **Autoregresja (p).** Liczba rzędów autoregresji w modelu. Rzędy autoregresji określają, które z poprzednich wartości są używane do przewidywania bieżących wartości. Na przykład rząd autoregresji 2 oznacza, że do przewidywania bieżącej wartości zostanie użyta wartość dwu okresów czasu szeregu w przeszłości.
- **Różnica (d).** Określa rząd różnicowania stosowany względem szeregu przed oszacowaniem modeli. Różnicowanie jest niezbędne w przypadku obecności trendów (szeregi z trendami są zwykle niestacjonarne, zaś modelowanie ARIMA zakłada stacjonarność) i służy do usuwania ich wpływu. Rząd różnicowania odpowiada stopniowi trendu szeregu — różnicowanie pierwszego rzędu jest uwzględniane w przypadku trendów liniowych, różnicowanie drugiego rzędu w przypadku trendów kwadratowych itd.
- **Średnia ruchoma (q).** Liczba rzędów średniej ruchomej w modelu. Rzędy średniej ruchomej określają, w jaki sposób odchylenia od średniej szeregu dla poprzednich wartości są używane do przewidywania bieżących wartości. Na przykład rzędy średniej ruchomej 1 i 2 oznaczają, że odchylenia od wartości średniej szeregu z każdego z dwu ostatnich okresów czasu będą uwzględniane podczas przewidywania bieżących wartości szeregu.

Rzędy sezonowości. Autoregresja sezonowa, średnia ruchoma i różnicowanie odgrywają takie same role, jak ich niesezonowe odpowiedniki. W przypadku rzędów sezonowości na bieżące wartości szeregu wpływają jednak poprzednie wartości szeregu, rozdzielone jednym lub większą liczbą okresów sezonowości. Na przykład w przypadku danych miesięcznych (okres sezonowości 12) rząd sezonowości 1 oznacza, że na bieżącą wartość szeregu wpływa 12 okresów wartości szeregu poprzedzających okres bieżący. Rząd sezonowości 1, w przypadku danych miesięcznych jest taki sam, jak w przypadku rzędu niesezonowości wynoszącego 12.

Bieżąca okresowość. Wskazuje okresowość (jeśli istnieje) aktualnie zdefiniowaną dla aktywnego zbioru danych. Bieżąca okresowość jest podawana w postaci liczby całkowitej — na przykład 12 dla okresu rocznego, przy czym każdy przypadek reprezentuje miesiąc. Jeśli nie ustawiono okresowości, widoczna jest wartość *Brak*. W przypadku modeli sezonowych wymagana jest okresowość. Okresowość można ustawić w oknie Zdefiniuj daty.

Transformacja zmiennej zależnej. Istnieje możliwość określenia transformacji wykonywanej dla każdej zmiennej zależnej przed jej zamodelowaniem.

- **Brak.** Nie jest wykonywana żadna transformacja.
- **1/pierwiastek kwadratowy.** Przekształcenie pierwiastkiem kwadratowym.
- **Logarytm naturalny.** Transformacja z użyciem logarytmu naturalnego.

Dołącz stałe do modelu. Włączenie stałej jest standardowe, chyba że użytkownik ma pewność, że ogólna wartość średniej serii wynosi 0. Wykluczanie stałej jest zalecane, gdy stosowane jest różnicowanie.

Funkcje przenoszenia w niestandardowych modelach ARIMA

Karta Funkcje przenoszenia (dostępna tylko wtedy, gdy określono zmienne niezależne) umożliwia zdefiniowanie funkcji przenoszenia dla dowolnej lub wszystkich zmiennych niezależnych określonych na karcie Zmienne. Funkcje przenoszenia umożliwiają określenie sposobu, w jaki poprzednie wartości zmiennych niezależnych (predykcyjnych) są używane do przewidywania przyszłych wartości szeregu zmiennych zależnych.

Polecenia wykonania funkcji przenoszenia. Należy wprowadzić wartości dla różnych składników funkcji przenoszenia do odpowiednich komórek siatki Struktura. Wszystkie wartości muszą być nieujemnymi liczbami całkowitymi. W przypadku składników, takich jak licznik i mianownik wartość ta reprezentuje rząd maksymalny. W modelu zostaną uwzględnione wszystkie niższe rzędy dodatnie. Ponadto w przypadku składników typu licznik zawsze uwzględniany jest rząd 0. Na przykład, jeśli zostanie określona wartość 2 dla licznika, model będzie obejmował rzędy 2, 1 i 0. Jeśli dla mianownika

zostanie określona wartość 3, model będzie obejmował rzędy 3, 2 i 1. Komórki w kolumnie Sezonowe są włączone tylko wtedy, gdy dla aktywnego zbioru danych zdefiniowano okresowość (patrz poniżej "Bieżąca okresowość").

- **Licznik.** Rząd licznika funkcji przenoszenia. Określa, które z poprzednich wartości wybranych niezależnych (predykcyjnych) serii są używane do przewidywania bieżących wartości szeregu zmiennych zależnych. Na przykład rząd licznika równy 1 oznacza, że wartość szeregu niezależnego jeden okres wstecz — a także bieżąca wartość szeregu niezależnego — służą do przewidywania bieżącej wartości każdego szeregu zależnego.
- **Mianownik.** Rząd mianownika funkcji przenoszenia. Określa, jak odchylenia od średniej szeregu, dla poprzednich wartości wybranych szeregów niezależnych (predyktora) są używane do przewidywania bieżących wartości szeregu zależnego. Na przykład rząd mianownika 1 oznacza, że odchylenia od wartości średniej szeregu niezależnego jeden okres czasu wstecz zostaną uwzględnione podczas przewidywania bieżącej wartości każdego szeregu zależnego.
- **Różnica.** Określa rząd różnicowania stosowany względem wybranego szeregu niezależnego (predykcyjnego) przed oszacowaniem modeli. Różnicowanie jest niezbędne w przypadku obecności trendów i służy do usuwania ich wpływu.

Rzędy sezonowości. Składniki, takie jak licznik, mianownik i różnicowanie odgrywają takie same role, jak ich niesezonowe odpowiedniki. W przypadku rzędów sezonowości na bieżące wartości szeregu wpływają jednak poprzednie wartości szeregu, rozdzielone jednym lub większą liczbą okresów sezonowości. Na przykład w przypadku danych miesięcznych (okres sezonowości 12) rząd sezonowości 1 oznacza, że na bieżącą wartość szeregu wpływa 12 okresów wartości szeregu poprzedzających okres bieżący. Rząd sezonowości 1, w przypadku danych miesięcznych jest taki sam, jak w przypadku rzędu niesezonowości wynoszącego 12.

Bieżąca okresowość. Wskazuje okresowość (jeśli istnieje) aktualnie zdefiniowaną dla aktywnego zbioru danych. Bieżąca okresowość jest podawana w postaci liczby całkowitej — na przykład 12 dla okresu rocznego, przy czym każdy przypadek reprezentuje miesiąc. Jeśli nie ustawiono okresowości, widoczna jest wartość *Brak*. W przypadku modeli sezonowych wymagana jest okresowość. Okresowość można ustawić w oknie Zdefiniuj daty.

Opóźnienie. Ustawienie opóźnienia powoduje, że wpływ niezależnej zmiennej jest opóźniony o określoną liczbę przedziałów. Na przykład, jeśli dla opóźnienia ustawiono wartość 5, wartość zmiennej niezależnej w czasie t nie wpływa na prognozę do czasu upłynięcia pięciu okresów ($t + 5$).

Transformacja. Specyfikacja funkcji przenoszenia dla zestawu zmiennych niezależnych obejmuje także opcjonalną transformację do wykonania na tych zmiennych.

- **Brak.** Nie jest wykonywana żadna transformacja.
- **1/pierwiastek kwadratowy.** Przekształcenie pierwiastkiem kwadratowym.
- **Logarytm naturalny.** Transformacja z użyciem logarytmu naturalnego.

Wartości odstające w modelach ARIMA użytkownika

Karta Wartości odstające udostępnia następujące opcje do obsługi wartości odstających³: wykryj je automatycznie, określ konkretne punkty jako wartości odstające albo nie wykrywaj ich ani nie modeluj.

Nie wykrywaj wartości odstających ani ich nie modeluj. Domyślnie wartości odstające nie są ani wykrywane, ani modelowane. Wybór tej opcji pozwala dezaktywować wszelkie operacje wykrywania lub modelowania wartości odstających.

Automatycznie wykryj wartości odstające. Wybór tej opcji pozwala automatycznie wykrywać wartości odstające i wybrać jeden lub więcej spośród poniższych typów wartości odstających:

- Addytywne
- Przesunięcie poziome
- Innowacyjne

³ Pena, D., G. C. Tiao, i R. S. Tsay, eds. 2001. *A course in time series analysis*. New York: John Wiley and Sons.

- Przejściowe
- Sezonowo addytywne
- Trend lokalny
- Fragment dodatkowy

Model określa punkty czasu jako wartości odstające. Wybierz tę opcję, aby określić specyficzne punkty czasowe jako wartości odstające. Dla każdej wartości odstającej użyj osobnego wiersza w tabeli definicji wartości odstających. Wprowadź wartości dla wszystkich komórek w danym wierszu.

- **Typ.** Typ wartości odstającej. Obsługiwane są następujące typy: addytywna (domyślna), przesunięcie poziomu, innowacyjna, przejściowa, sezonowo addytywna i trend lokalny.

Uwaga 1: jeśli dla aktywnego zbioru danych nie zdefiniowano żadnej specyfikacji daty, tabela definicji wartości odstających zawiera pojedynczą kolumnę *Obserwacja*. Aby określić wartość odstającą, wprowadź liczbę wiersza (widoczną w edytorze danych) odpowiedniej obserwacji.

Uwaga 2: kolumna *Cykl* (jeśli jest obecna) w tabeli definicji wartości odstających odpowiada wartości zmiennej *CYCLE_* w aktywnym zbiorze danych.

Wyniki

Możliwe jest zarówno uzyskanie wyników obliczonych z poszczególnych modeli, jak i ze wszystkich modeli łącznie. Wyniki z poszczególnych modeli można ograniczyć do zbioru najlepiej lub najgorzej dopasowanych modeli na podstawie kryteriów określonych przez użytkownika.

Tabele statystyk i prognoz

Karta Statystyki udostępnia opcje służące do wyświetlania tabel wyników modelowania.

Wyświetl miary dopasowania, statystykę Ljung-Boxa oraz liczbę wartości odstających dla modelu.

Wybierz (zaznacz) tę opcję, aby wyświetlić tabelę zawierającą wybrane miary dopasowania, wartości statystyki Ljunga-Boxa i liczbę wartości odstających dla każdego oszacowanego modelu.

Miary dopasowania. Można wybrać jeden lub więcej z następujących elementów, które mają być uwzględnione w tabeli zawierającej miary dopasowania dla każdego oszacowanego modelu:

- Stacjonarny *R*-kwadrat
- *R*-kwadrat
- Pierwiastek błędu średniokwadratowego
- Średni bezwzględny błąd procentowy
- Średni błąd bezwzględny
- Maksymalny bezwzględny błąd procentowy
- Maksymalny błąd bezwzględny
- Znormalizowane bayesowskie kryterium informacyjne

Statystyki porównawcze modeli. Ta grupa opcji steruje sposobem wyświetlania tabeli zawierającej statystyki obliczone na wszystkich oszacowanych modelach. Każda opcja powoduje wygenerowanie odrębnej tabeli. Można wybrać jedną lub więcej z następujących opcji:

- **Dobroć dopasowania.** Tabela statystyk podsumowujących i percentyli dla stacjonarnego *R*-kwadratu, *R*-kwadratu, pierwiastka błędu średniokwadratowego, średniego bezwzględnego błędu procentowego, średniego bezwzględnego błędu, maksymalnego bezwzględnego błędu procentowego, maksymalnego błędu bezwzględnego i znormalizowanego bayesowskiego kryterium informacyjnego.
- **Funkcja autokorelacji reszt (ACF).** Tabela statystyk podsumowujących i percentyli dla autokorelacji reszt we wszystkich oszacowanych modelach.
- **Funkcja autokorelacji cząstkowych reszt (PACF).** Tabela statystyk podsumowujących i percentyli dla autokorelacji cząstkowych reszt we wszystkich oszacowanych modelach.

Statystyki dla poszczególnych modeli. Ta grupa opcji steruje sposobem wyświetlania tabeli zawierających szczegółowe informacje dla każdego oszacowanego modelu. Każda opcja powoduje wygenerowanie odrębnej tabeli. Można wybrać jedną lub więcej z następujących opcji:

- **Oceny parametrów.** Wyświetla tabelę oszacowań parametrów dla każdego oszacowanego modelu. Dla modeli wygładzania wykładniczego i modeli ARIMA są wyświetlane odrębne tabele. Jeśli obecne są wartości odstające, oszacowania parametrów są również wyświetlane w odrębnej tabeli.
- **Funkcja autokorelacji reszt (ACF).** Wyświetla tabelę autokorelacji reszt według opóźnienia dla każdego oszacowanego modelu. Tabela ta obejmuje przedziały ufności dla autokorelacji.
- **Funkcja autokorelacji cząstkowych reszt (PACF).** Wyświetla tabelę autokorelacji cząstkowych reszt według opóźnienia dla każdego oszacowanego modelu. Tabela ta obejmuje przedziały ufności dla autokorelacji cząstkowych.

Pokaż prognozy. Wyświetla tabelę prognoz modelu i przedziałów ufności dla każdego oszacowanego modelu. Okres prognozy jest konfigurowany na karcie Opcje.

Wykresy

Karta Wykresy udostępnia opcje służące do wyświetlania wykresów utworzonych na podstawie wyników modelowania.

Wykresy porównawcze modeli

Ta grupa opcji steruje sposobem wyświetlania wykresów zawierających statystyki obliczone na wszystkich oszacowanych modelach. Każda opcja powoduje wygenerowanie odrębnego wykresu. Można wybrać jedną lub więcej z następujących opcji:

- Stacjonarny R -kwadrat
- R -kwadrat
- Pierwiastek błędu średniokwadratowego
- Średni bezwzględny błąd procentowy
- Średni błąd bezwzględny
- Maksymalny bezwzględny błąd procentowy
- Maksymalny błąd bezwzględny
- Znormalizowane bayesowskie kryterium informacyjne
- Funkcja autokorelacji reszt (ACF)
- Funkcja autokorelacji cząstkowych reszt (PACF)

Wykresy dla poszczególnych modeli

Szeregi. Wybierz (zaznacz) tę opcję, aby wygenerować wykresy na podstawie wartości przewidywanych dla każdego oszacowanego modelu. Można wybrać jeden lub więcej z następujących elementów, które mają być uwzględnione na wykresie:

- **Wartości obserwowane.** Wartości obserwowane dla szeregu zmiennych zależnych.
- **Prognozy.** Wartości przewidywane przez model dla okresu prognozy.
- **Wartości dopasowania.** Wartości przewidywane przez model dla okresu estymacji.
- **Przedziały ufności dla prognoz.** Przedziały ufności dla okresu prognozy.
- **Przedziały ufności dla wartości dopasowania.** Przedziały ufności dla okresu estymacji.

Funkcja autokorelacji reszt (ACF). Wyświetla wykres autokorelacji reszt dla każdego oszacowanego modelu.

Funkcja autokorelacji cząstkowych reszt (PACF). Wyświetla wykres autokorelacji reszt cząstkowych dla każdego oszacowanego modelu.

Ograniczanie wyników do najlepiej lub najgorzej dopasowanych modeli

Karta Filtr wyników udostępnia opcje służące do ograniczania wyników tabelarycznych i wyników na wykresie do podzbioru oszacowanych modeli. Wyniki można ograniczyć do najlepiej dopasowanych i/lub najgorzej dopasowanych modeli odpowiednio do podanych kryteriów dopasowania. Domyślnie w danych wyjściowych zawarte są wszystkie oszacowane modele.

Najlepiej dopasowane modele. Wybierz (zaznacz) tę opcję, aby uwzględnić najlepiej dopasowane modele w wynikach. Wybierz miarę dobroci dopasowania i określ liczbę modeli do uwzględnienia. Wybranie tej opcji nie uniemożliwia wybrania modeli najgorzej dopasowanych. W takim przypadku wyniki będą zawierać modele najgorzej i najlepiej dopasowane.

- **Ustalona liczba modeli.** Określa, że wyniki są wyświetlane dla n najlepiej dopasowanych modeli. Jeśli liczba jest większa niż liczba oszacowanych modeli, wyświetlane są wszystkie modele.
- **Procent całkowitej liczby modeli.** Określa, które wyniki są wyświetlane dla modeli z wartościami dobroci dopasowania w górnych $n\%$ we wszystkich oszacowanych modelach.

Modele najgorzej dopasowane. Wybierz (zaznacz) tę opcję, aby uwzględnić najgorzej dopasowane modele w wynikach. Wybierz miarę dobroci dopasowania i określ liczbę modeli do uwzględnienia. Wybranie tej opcji nie uniemożliwia wybrania modeli najlepiej dopasowanych. W takim przypadku wyniki będą zawierać modele najgorzej i najlepiej dopasowane.

- **Ustalona liczba modeli.** Określa, że wyniki są wyświetlane dla n najgorzej dopasowanych modeli. Jeśli liczba jest większa niż liczba oszacowanych modeli, wyświetlane są wszystkie modele.
- **Procent całkowitej liczby modeli.** Określa, że wyniki są wyświetlane dla modeli z wartościami dobroci dopasowania w dolnych $n\%$ we wszystkich oszacowanych modelach.

Miara dobroci dopasowania. Wybierz miarę dobroci dopasowania, która będzie używana do filtrowania modeli. Ustawienie domyślne to stacjonarny R -kwadrat.

Zapisywanie predykcji modelu i specyfikacji modelu

Karta Zapisz umożliwia zapisywanie predykcji modelu jako nowych zmiennych w aktywnym zbiorze danych oraz zapisywanie specyfikacji modelu do zewnętrznego pliku w formacie XML.

Zapisz zmiennę. Predykcje modelu, przedziały ufności i reszty można zapisać jako nowe zmienne w aktywnym zbiorze danych. Dla każdego szeregu zmiennych zależnych generowany jest odrębny zbiór nowych zmiennych, a każda nowa zmienna zawiera wartości zarówno dla przedziałów estymacji, jak i prognozy. Nowe obserwacje są dodawane, jeśli okres prognozy jest dłuższy niż długość szeregu zmiennych zależnych. Wybierz pole wyboru Zapisz dla wszystkich nowych zmiennych, które chcesz zapisać. Domyślnie żadna nowa zmienna nie jest zapisywana.

- **Wartości przewidywane.** Wartości przewidywane przez model.
- **Dolne przedziały ufności.** Dolne przedziały ufności dla wartości przewidywanych.
- **Górne przedziały ufności.** Górne przedziały ufności dla wartości przewidywanych.
- **Reszty szumu.** Reszty modelu. Gdy wykonywane są przekształcenia zmiennej zależnej (na przykład logarytm naturalny), są to reszty dla przekształconych szeregów.
- **Prefiks nazwy zmiennej.** Określ prefiksy, które mają być używane dla nazw nowych zmiennych lub pozostaw prefiksy domyślne. Nazwy zmiennych składają się z powiązanej zmiennej zależnej i identyfikatora modelu. Nazwa zmiennej jest w razie potrzeby uzupełniana, aby nie dochodziło do kolizji między nazwami. Prefiks musi być zgodny z regułami nazewnictwa zmiennych.

Eksportuj plik modelu. Specyfikacje modelu dla wszystkich oszacowanych modeli są eksportowane do określonego pliku w formacie XML. Zapisanych modeli można używać do generowania zaktualizowanych prognoz.

- **Plik XML.** Specyfikacje modelu są zapisywane w pliku XML, którego można używać z aplikacjami IBM SPSS.
- **Plik PMML.** Specyfikacje modelu są zapisywane w pliku XML zgodnym z PMML, którego można używać z aplikacjami zgodnymi z PMML, w tym z aplikacjami IBM SPSS.

Opcje

Karta Opcje umożliwia ustawienie okresu prognozy, określenie sposobu postępowania z brakującymi wartościami, ustawienie szerokości przedziału ufności, określenie niestandardowego prefiksu dla identyfikatorów modelu oraz ustawienie liczby przesunięć wyświetlanych dla autokorelacji.

Okres prognozy. Okres prognozy rozpoczyna się zawsze od pierwszej obserwacji po zakończeniu okresu estymacji (zbiór obserwacji używanych do określenia modelu) i trwa do ostatniej obserwacji w aktywnym zbiorze danych lub daty określonej przez użytkownika. Domyślnie koniec okresu estymacji jest określany przez ostatnią obserwację w aktywnym zbiorze danych. Można jednak zmienić to ustawienie w oknie dialogowym Wybierz obserwacje, wybierając opcję **Na podstawie czasu lub zakresu obserwacji**.

- **Od pierwszej obserwacji po końcu okresu estymacji do ostatniej obserwacji w aktywnym zbiorze danych.** Wybierz tę opcję, gdy okres estymacji kończy się przed ostatnią obserwacją w aktywnym zbiorze danych, a prognozy mają obejmować okres do ostatniej obserwacji. Opcja ta jest zazwyczaj używana do tworzenia prognoz na okres wstrzymana, co umożliwia porównanie predykcji modelu z podzbiorem wartości rzeczywistych.
- **Od pierwszej obserwacji po końcu okresu estymacji do określonej daty.** Wybierz tę opcję, aby wyraźnie określić zakończenie okresu prognozy. Opcja ta jest zazwyczaj używana do tworzenia prognoz poza koniec szeregów rzeczywistych. Wprowadź wartości dla wszystkich komórek w tabeli Dane.

Jeśli dla aktywnego zbioru danych nie zdefiniowano żadnej daty, tabela Dane zawiera pojedynczą kolumnę *Obserwacja*. Aby określić koniec okresu prognozy, wprowadź numer wiersza (wyświetlany w Edytorze danych) odpowiedniej obserwacji.

Kolumna *Cykl* (jeśli jest obecna) w tabeli definicji Dane odpowiada wartości zmiennej *CYCLE_* w aktywnym zbiorze danych.

Braki danych zdefiniowane przez użytkownika. Te opcje określają sposób postępowania z brakami danych zdefiniowanymi przez użytkownika.

- **Traktuj jako nieprawidłowe.** Braki danych użytkownika są traktowane jak systemowe braki danych.
- **Traktuj jako ważne.** Braki danych zdefiniowane przez użytkownika są traktowane jako ważne wartości.

Zasady traktowania braków danych. Następujące reguły obowiązują podczas postępowania z brakami danych (w tym z systemowymi brakami danych oraz brakami danych użytkownika traktowanymi jako wartości nieważne) podczas procedury modelowania:

- Obserwacje z brakami danych zmiennej zależnej, które występują w okresie estymacji, są uwzględniane w modelu. Sposób postępowania z brakami danych zależy od metody estymacji.
- Jeśli dla zmiennej niezależnej istnieją braki danych w okresie estymacji, generowane jest ostrzeżenie. W przypadku automatycznego doboru modelu modele zawierające zmienną niezależną są szacowane bez zmiennej. W przypadku niestandardowych modeli ARIMA modele zawierające zmienną nie są szacowane.
- Jeśli w okresie prognozy występuje jakakolwiek zmienna niezależna z brakami danych, procedura generuje ostrzeżenie i wykonuje prognozę w możliwie najszerszym okresie.

Szerokość przedziału ufności (%). Przedziały ufności są obliczane dla predykcji modelu i autokorelacji reszt. Można określić dowolną wartość dodatnią mniejszą od 100. Domyślnie ustawiony jest 95% przedział ufności.

Przedrostek dla identyfikatorów modeli w wynikach. Każda zmienna zależna określona na karcie Zmienne powoduje wygenerowanie odrębnego oszacowanego modelu. Modele są rozróżniane dzięki unikalnym nazwom zawierającym prefiks, który można konfigurować, oraz sufiks w postaci liczby całkowitej. Można wprowadzić prefiks lub pozostawić domyślną wartość opcji *Model*.

aksymalna liczba przesunięć przedstawiana w wynikach ACF i PACF. Można określić maksymalną liczbę przesunięć wyświetlanych w tabelach i wykresach autokorelacji i autokorelacji cząstkowych.

Dodatkowe właściwości komendy TSMODEL

Istnieje możliwość dostosowania modelowania szeregów czasowych poprzez wklejenie zaznaczeń do okna składni i edycję wynikowej składni komend TSMODEL. Język składni komend umożliwia:

- Określenie okresu sezonowego danych (przy użyciu słowa kluczowego SEASONLENGTH w opcji komendy AUXILIARY). Powoduje to zastąpienie bieżącej okresowości (jeśli istniała) dla aktywnego zbioru danych.
- Określenie nienastępujących po sobie opóźnień dla składników niestandardowych modeli ARIMA i funkcji przeniesienia (przy użyciu opcji komend ARIMA i TRANSFERFUNCTION). Na przykład można określić niestandardowy model ARIMA z autoregresyjnymi opóźnieniami zleceń 1, 3 i 6; lub funkcję przeniesienia z opóźnieniami licznika zleceń 2, 5 i 8.
- Podać więcej niż jeden zestaw specyfikacji modelowania (na przykład metodę modelowania, rzędy ARIMA, zmienne niezależne itd.) dla pojedynczego uruchomienia procedury Konstruktor modeli szeregów czasowych (przy użyciu opcji komendy MODEL).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Zastosuj modele szeregów czasowych

Procedura Zastosuj modele szeregów czasowych ładuje istniejące szeregi czasowe z pliku zewnętrznego i stosuje je do aktywnego zbioru danych. Za pomocą tej procedury można uzyskać prognozy dla serii, dla których dostępne są nowe lub zmienione dane, bez konieczności przebudowywania modeli. Models are generated using the Time Series Modeler procedure.

Przykład. Kierownik ds. inwentaryzacji w dużym sklepie detalicznym odpowiada za każdy z 5000 produktów. Skorzystał on z automatycznego doboru modelu w celu utworzenia modeli, które przewidują sprzedaż każdego produktu na trzy miesiące w przód. Magazyn danych jest odświeżany co miesiąc o rzeczywiste dane sprzedaży, które będą używane do wygenerowania prognoz aktualizowanych co miesiąc. Procedura Kreator modeli szeregów czasowych umożliwia wykonanie tego przy użyciu oryginalnych modeli, które są po prostu ponownie estymowane w celu uwzględnienia nowych danych.

Statystyki. Miara dobroci dopasowania: stacjonarny R -kwadrat, R -kwadrat (R^2), pierwiastek z błędu średniokwadratowego (RMSE), średni bezwzględny błąd (MAE), średni bezwzględny błąd procentowy (MAPE), maksymalny bezwzględny błąd (MaxAE), maksymalny bezwzględny błąd procentowy (MaxAPE), znormalizowane bayesowskie kryterium informacyjne (BIC). Reszty: funkcja autokorelacji, funkcja autokorelacji częściowej, statystyki Q Ljunga-Boxa.

Wykresy. Wykresy podsumowujące dla wszystkich modeli: histogramy stacjonarnego R -kwadrat, R -kwadrat (R^2), pierwiastek z błędu średniokwadratowego (RMSE), średni bezwzględny błąd (ŚBB), średni bezwzględny błąd procentowy (MAPE), maksymalny bezwzględny błąd (MaxAE), maksymalny bezwzględny błąd procentowy (MaxAPE), znormalizowane bayesowskie kryterium informacyjne (BIC); wykresy skrzynkowe autokorelacji reszt i autokorelacji częściowej. Wyniki dla poszczególnych modeli: wartości prognozy, wartości dopasowania, wartości obserwowane, górna i dolna granica ufności, autokorelacje reszt i autokorelacje częściowe.

Zagadnienia dotyczące danych z procedury Zastosuj modele szeregów czasowych

Dane. Zmienne (zależne i niezależne), do których będą stosowane modele, powinny być numeryczne.

Założenia. Modele są stosowane do zmiennych w aktywnym zbiorze danych o tych samych nazwach, co zmienne określone w modelu. Wszystkie takie zmienne są traktowane jako szeregi czasowe, co oznacza, że każda obserwacja reprezentuje punkt czasowy, przy czym kolejne obserwacje są rozdzielane stałym przedziałem czasowym.

- **Prognozy.** W przypadku tworzenia prognoz z użyciem modeli ze zmiennymi niezależnymi (predyktor) aktywny zestaw danych powinien zawierać wartości tych zmiennych dla wszystkich obserwacji z okresu prognozy. Jeśli parametry modelu zostaną poddane ponownej estymacji, zmienne niezależne nie powinny zawierać żadnych brakujących wartości w okresie estymacji.

Definiowanie dat

Procedura Kreator modeli szeregów czasowych wymaga, aby okresowość, jeśli jest określona, aktywnego zbioru danych była zgodna z okresowością modeli, które mają zostać zastosowane. Jeśli prognozowanie będzie wykonywane na tym samym zbiorze danych (prawdopodobnie z nowymi lub uaktualnionymi danymi), który był używany do budowy modelu, warunek ten będzie spełniony. Jeśli dla aktywnego zbioru danych nie istnieje żadna okresowość, można przejść do okna dialogowego Zdefiniuj daty w celu utworzenia okresowości. Jeśli jednak podczas tworzenia modeli nie określono okresowości, aktywny zbiór danych również nie powinien jej posiadać.

Zastosowanie modeli

1. Wybierz z menu następującą opcję:

Analizuj > Prognozowanie > Zastosuj modele tradycyjne...

2. Enter the file specification for a model file or click **Browse** and select a model file (model files are created with the Time Series Modeler procedure).

Opcjonalnie można wykonać następujące czynności:

- Ponowna estymacja parametrów modelu przy użyciu danych w aktywnym zbiorze danych. Prognozy są tworzone na podstawie parametrów poddanych ponownej estymacji.
- Zapisz predykcje, przedziały ufności i reszty modelu.
- Zapisz modele po ponownej estymacji w formacie XML.

Parametry modelu i miary dobroci dopasowania

Władaj z pliku modelu. Prognozy są tworzone przy użyciu parametrów modelu z pliku modelu bez konieczności ponownej estymacji tych parametrów. Miary dobroci dopasowania wyświetlane w wynikach i używane do filtrowania modeli (najlepszego lub najgorszego dopasowania) pochodzą z pliku modelu i odzwierciedlają dane używane podczas opracowywania (lub ostatniej aktualizacji) każdego modelu. Jeśli używana jest ta opcja, prognozy nie biorą pod uwagę danych historycznych (dla zmiennych zależnych i zmiennych niezależnych) w aktywnym zbiorze danych. Jeśli dane historyczne mają wpływać na prognozy, należy wybrać opcję **Estymuj ponownie na podstawie danych**. Ponadto, w prognozach nie są brane pod uwagę wartości szeregów zmiennych zależnych w okresie prognozy. Wartości zmiennych niezależnych w okresie prognozy są jednak brane pod uwagę. Jeśli istnieje więcej aktualnych wartości szeregów zależnych i mają one zostać uwzględnione w prognozach, należy wykonać ponowną estymację, dostosowując okres estymacji, aby uwzględnić te wartości.

Estymuj ponownie na podstawie danych. Parametry modelu są estymowane ponownie przy użyciu danych w aktywnym zbiorze danych. Ponowna estymacja parametrów modelu nie wpływa na strukturę modelu. Na przykład model ARIMA(1,0,1) nie zostanie zmieniony, ale parametry autoregresji i średnich ruchomych zostaną poddane ponownej estymacji. Ponowna estymacja nie spowoduje wykrycia nowych wartości odstających. Jeśli istnieją wartości odstające, zawsze pochodzą one z pliku modelu.

- **Okres szacowania.** Okres szacowania definiuje zestaw obserwacji używanych do ponownej estymacji parametrów modelu. Domyślnie, okres szacowania obejmuje wszystkie obserwacje w aktywnym zestawie danych. Aby ustawić okres szacowania, wybierz opcję **Na podstawie czasu lub zakresu obserwacji** w oknie dialogowym Wybierz obserwacje. W zależności od dostępnych danych okres szacowania, z jakiego korzysta procedura, może różnić się w zależności od modelu i dlatego będzie inny od wyświetlanej wartości. Dla zadanego modelu rzeczywistym okresem estymacji jest okres pozostały po wyeliminowaniu ciągłych braków danych zmiennej, ze zmiennej zależnej modelu, występujących na początku lub na końcu danego okresu estymacji.

Okres prognozowania

Okres prognozy dla każdego modelu zawsze zaczyna się wraz z pierwszą obserwacją po zakończeniu okresu estymacji i kończy się wraz z ostatnią obserwacją w aktywnym zbiorze danych lub z datą określoną przez użytkownika. Jeśli parametry nie zostaną poddane ponownej estymacji (jest to ustawienie domyślne), okres estymacji dla każdego modelu to zestaw obserwacji używanych podczas opracowywania (lub ostatniej aktualizacji) modelu.

- **Od pierwszej obserwacji po końcu okresu estymacji do ostatniej obserwacji w aktywnym zbiorze danych.** Wybierz tę opcję, gdy okres estymacji kończy się przed ostatnią obserwacją w aktywnym zbiorze danych, a prognozy mają obejmować okres do ostatniej obserwacji.

- **Od pierwszej obserwacji po końcu okresu estymacji do określonej daty.** Wybierz tę opcję, aby wyraźnie określić zakończenie okresu prognozy. Wprowadź wartości dla wszystkich komórek w tabeli Dane.

Jeśli dla aktywnego zbioru danych nie zdefiniowano żadnej daty, tabela Dane zawiera pojedynczą kolumnę *Obserwacja*. Aby określić koniec okresu prognozy, wprowadź numer wiersza (wyświetlany w Edytorze danych) odpowiedniej obserwacji.

Kolumna *Cykl* (jeśli jest obecna) w tabeli definicji Dane odpowiada wartości zmiennej *CYCLE_* w aktywnym zbiorze danych.

Wyniki

Możliwe jest zarówno uzyskanie wyników z poszczególnych modeli, jak i ze wszystkich modeli łącznie. Wyniki z poszczególnych modeli można ograniczyć do zbioru najlepiej lub najgorzej dopasowanych modeli na podstawie kryteriów określonych przez użytkownika.

Tabele statystyk i prognoz

Karta Statystyki udostępnia opcje służące do wyświetlania tabel statystyk dopasowania modeli, parametrów modeli, funkcji autokorelacji i prognoz. Wyświetlane są wartości miar dopasowania, wartości statystyki Ljunga-Boxa i parametry z pliku modelu, które odzwierciedlają wartości używane podczas opracowywania każdego modelu (lub ostatniej aktualizacji modelu), o ile nie wykonano ponownej estymacji parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele). Informacje o wartościach odstających zawsze pochodzą z pliku modelu.

Wyświetl miary dopasowania, statystykę Ljung-Boxa oraz liczbę wartości odstających dla modelu.

Wybierz (zaznacz) tę opcję, aby wyświetlić tabelę zawierającą wybrane miary dopasowania, wartości statystyki Ljunga-Boxa i liczbę wartości odstających dla każdego modelu.

Miary dopasowania. Można wybrać jeden lub więcej z następujących elementów, które mają być uwzględnione w tabeli zawierającej miary dopasowania dla każdego modelu:

- Stacjonarny *R*-kwadrat
- *R*-kwadrat
- Pierwiastek błędu średniokwadratowego
- Średni bezwzględny błąd procentowy
- Średni błąd bezwzględny
- Maksymalny bezwzględny błąd procentowy
- Maksymalny błąd bezwzględny
- Znormalizowane bayesowskie kryterium informacyjne

Statystyki porównawcze modeli. Ta grupa opcji umożliwia sterowanie wyświetlaniem tabel zawierających statystyki ze wszystkich modeli. Każda opcja powoduje wygenerowanie odrębnej tabeli. Można wybrać jedną lub więcej z następujących opcji:

- **Dobroć dopasowania.** Tabela statystyk podsumowujących i percentyli dla stacjonarnego *R*-kwadratu, *R*-kwadratu, pierwiastka błędu średniokwadratowego, średniego bezwzględnego błędu procentowego, średniego bezwzględnego błędu, maksymalnego bezwzględnego błędu procentowego, maksymalnego błędu bezwzględnego i znormalizowanego bayesowskiego kryterium informacyjnego.
- **Funkcja autokorelacji reszt (ACF).** Tabela statystyk podsumowujących i percentyli dla autokorelacji reszt we wszystkich oszacowanych modelach. Tabela ta jest dostępna tylko wtedy, gdy wykonano ponowną estymację parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele).
- **Funkcja autokorelacji cząstkowych reszt (PACF).** Tabela statystyk podsumowujących i percentyli dla autokorelacji cząstkowych reszt we wszystkich oszacowanych modelach. Tabela ta jest dostępna tylko wtedy, gdy wykonano ponowną estymację parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele).

Statystyki dla poszczególnych modeli. Ta grupa opcji steruje sposobem wyświetlania tabel zawierających szczegółowe informacje dla każdego modelu. Każda opcja powoduje wygenerowanie odrębnej tabeli. Można wybrać jedną lub więcej z następujących opcji:

- **Oceny parametrów.** Wyświetla tabelę oszacowań parametrów dla każdego modelu. Dla modeli wyładzania wykładniczego i modeli ARIMA są wyświetlane odrębne tabele. Jeśli obecne są wartości odstające, oszacowania parametrów są również wyświetlane w odrębnej tabeli.
- **Funkcja autokorelacji reszt (ACF).** Wyświetla tabelę autokorelacji reszt według opóźnienia dla każdego oszacowanego modelu. Tabela ta obejmuje przedziały ufności dla autokorelacji. Tabela ta jest dostępna tylko wtedy, gdy wykonano ponowną estymację parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele).
- **Funkcja autokorelacji cząstkowych reszt (PACF).** Wyświetla tabelę autokorelacji cząstkowych reszt według opóźnienia dla każdego oszacowanego modelu. Tabela ta obejmuje przedziały ufności dla autokorelacji cząstkowych. Tabela ta jest dostępna tylko wtedy, gdy wykonano ponowną estymację parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele).

Pokaż prognozy. Wyświetla tabelę prognoz modelu i przedziałów ufności dla każdego modelu.

Wykresy

Karta Wykresy zawiera opcje wyświetlania wykresów statystyk dopasowania modelu, funkcji autokorelacji oraz wartości serii (w tym prognoz).

Wykresy porównawcze modeli

Ta grupa opcji umożliwia sterowanie wyświetlaniem wykresów zawierających statystyki z wszystkich modeli. Wyświetlane są wartości z pliku modelu, które odzwierciedlają wartości używane podczas opracowywania każdego modelu (lub ostatniej aktualizacji modelu), o ile nie wykonano ponownej estymacji parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele). Ponadto wykresy autokorelacji są dostępne tylko wtedy, gdy wykonano ponową estymację parametrów modelu. Każda opcja powoduje wygenerowanie odrębnego wykresu. Można wybrać jedną lub więcej z następujących opcji:

- Stacjonarny R -kwadrat
- R -kwadrat
- Pierwiastek błędu średniokwadratowego
- Średni bezwzględny błąd procentowy
- Średni błąd bezwzględny
- Maksymalny bezwzględny błąd procentowy
- Maksymalny błąd bezwzględny
- Znormalizowane bayesowskie kryterium informacyjne
- Funkcja autokorelacji reszt (ACF)
- Funkcja autokorelacji cząstkowych reszt (PACF)

Wykresy dla poszczególnych modeli

Szeregi. Wybierz (zaznacz) tę opcję, aby wygenerować wykresy na podstawie wartości przewidywanych dla każdego modelu. Wartości obserwowane, wartości dopasowania, przedziały ufności dla wartości dopasowania i autokorelacje są dostępne tylko wtedy, gdy wykonano ponową estymację parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele). Można wybrać jeden lub więcej z następujących elementów, które mają być uwzględnione na wykresie:

- **Wartości obserwowane.** Wartości obserwowane dla szeregu zmiennych zależnych.
- **Prognozy.** Wartości przewidywane przez model dla okresu prognozy.
- **Wartości dopasowania.** Wartości przewidywane przez model dla okresu estymacji.
- **Przedziały ufności dla prognoz.** Przedziały ufności dla okresu prognozy.
- **Przedziały ufności dla wartości dopasowania.** Przedziały ufności dla okresu estymacji.

Funkcja autokorelacji reszt (ACF). Wyświetla wykres autokorelacji reszt dla każdego oszacowanego modelu.

Funkcja autokorelacji cząstkowych reszt (PACF). Wyświetla wykres autokorelacji reszt cząstkowych dla każdego oszacowanego modelu.

Ograniczanie wyników do najlepiej lub najgorzej dopasowanych modeli

Karta Filtr wyników udostępnia opcje służące do ograniczania wyników tabelarycznych i wyników na wykresie do podzbioru modeli. Wyniki można ograniczyć do najlepiej dopasowanych i/lub najgorzej dopasowanych modeli odpowiednio do podanych kryteriów dopasowania. Domyślnie w danych wyjściowych zawarte są wszystkie modele. Używane do filtrowania modeli wartości miar dopasowania pochodzą z pliku modelu i odzwierciedlają wartości używane podczas opracowywania każdego modelu (lub ostatniej aktualizacji modelu), o ile nie wykonano ponownej estymacji parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele).

Najlepiej dopasowane modele. Wybierz (zaznacz) tę opcję, aby uwzględnić najlepiej dopasowane modele w wynikach. Wybierz miarę dobroci dopasowania i określ liczbę modeli do uwzględnienia. Wybranie tej opcji nie uniemożliwia wybrania modeli najgorzej dopasowanych. W takim przypadku wyniki będą zawierać modele najgorzej i najlepiej dopasowane.

- **Ustalona liczba modeli.** Określa, że wyniki są wyświetlane dla n najlepiej dopasowanych modeli. Jeśli liczba jest większa niż łączna liczba modeli, wyświetlane są wszystkie modele.
- **Procent całkowitej liczby modeli.** Określa, że wyniki są wyświetlane dla modeli z wartościami dobroci dopasowania w górnych $n\%$ we wszystkich oszacowanych modelach.

Modele najgorzej dopasowane. Wybierz (zaznacz) tę opcję, aby uwzględnić najgorzej dopasowane modele w wynikach. Wybierz miarę dobroci dopasowania i określ liczbę modeli do uwzględnienia. Wybranie tej opcji nie uniemożliwia wybrania modeli najlepiej dopasowanych. W takim przypadku wyniki będą zawierać modele najgorzej i najlepiej dopasowane.

- **Ustalona liczba modeli.** Określa, że wyniki są wyświetlane dla n najgorzej dopasowanych modeli. Jeśli liczba jest większa niż łączna liczba modeli, wyświetlane są wszystkie modele.
- **Procent całkowitej liczby modeli.** Określa, że wyniki są wyświetlane dla modeli z wartościami dobroci dopasowania w dolnych $n\%$ we wszystkich modelach.

Miara dobroci dopasowania. Wybierz miarę dobroci dopasowania, która będzie używana do filtrowania modeli. Ustawienie domyślne to stacjonarny R -kwadrat.

Zapisywanie predykcji modelu i specyfikacji modelu

Karta Zapisz umożliwia zapisywanie predykcji modelu jako nowych zmiennych w aktywnym zbiorze danych oraz zapisywanie specyfikacji modelu do zewnętrznego pliku w formacie XML.

Zapisz zmienne. Predykcje modelu, przedziały ufności i reszty można zapisać jako nowe zmienne w aktywnym zbiorze danych. Każdy model powoduje wygenerowanie własnego zestawu nowych zmiennych. Nowe obserwacje są dodawane, jeśli okres prognozy jest dłuższy niż długość szeregu zmiennych zależnych powiązanych z modelem. Wartości przewidywane i przedziały ufności są tworzone tylko dla okresu prognozy, o ile parametry modelu nie zostaną poddane ponownej estymacji (**Estymuj ponownie na podstawie danych** na karcie Modele). Wybierz pole wyboru Zapisz dla wszystkich nowych zmiennych, które chcesz zapisać. Domyślnie żadna nowa zmienna nie jest zapisywana.

- **Wartości przewidywane.** Wartości przewidywane przez model.
- **Dolne przedziały ufności.** Dolne przedziały ufności dla wartości przewidywanych.
- **Górne przedziały ufności.** Górne przedziały ufności dla wartości przewidywanych.
- **Reszty szumu.** Reszty modelu. Gdy wykonywane są przekształcenia zmiennej zależnej (na przykład logarytm naturalny), są to reszty dla przekształconych szeregów. Opcja ta jest dostępna tylko wtedy, gdy wykonano ponowną estymację parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele).

- **Prefiks nazwy zmiennej.** Określ prefiksy, które mają być używane dla nazw nowych zmiennych lub pozostaw prefiksy domyślne. Nazwy zmiennych składają się z powiązanej zmiennej zależnej i identyfikatora modelu. Nazwa zmiennej jest w razie potrzeby uzupełniana, aby nie dochodziło do kolizji między nazwami. Prefiks musi być zgodny z regułami nazewnictwa zmiennych.

Eksportuj plik modelu Specyfikacje modelu, zawierające parametry po ponownej estymacji i statystyki dopasowania, są eksportowane do określonego pliku w formacie XML. Opcja ta jest dostępna tylko wtedy, gdy wykonano ponowną estymację parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele).

- **Plik XML.** Specyfikacje modelu są zapisywane w pliku XML, którego można używać z aplikacjami IBM SPSS.
- **Plik PMML.** Specyfikacje modelu są zapisywane w pliku XML zgodnym z PMML, którego można używać z aplikacjami zgodnymi z PMML, w tym z aplikacjami IBM SPSS.

Opcje

Karta Opcje umożliwia określenie sposobu postępowania z brakami danych, skonfigurowanie szerokości przedziału ufności oraz skonfigurowanie liczby przesunięć wyświetlanych dla autokorelacji.

Braki danych zdefiniowane przez użytkownika. Te opcje określają sposób postępowania z brakami danych zdefiniowanymi przez użytkownika.

- **Traktuj jako nieprawidłowe.** Braki danych użytkownika są traktowane jak systemowe braki danych.
- **Traktuj jako ważne.** Braki danych zdefiniowane przez użytkownika są traktowane jako ważne wartości.

Zasady traktowania braków danych. Następujące reguły obowiązują podczas postępowania z brakami danych (w tym z systemowymi brakami danych oraz brakami danych zdefiniowanymi przez użytkownika traktowanymi jako wartości nieważne):

- Obserwacje z brakami danych zmiennej zależnej, które występują w okresie estymacji, są uwzględniane w modelu. Sposób postępowania z brakami danych zależy od metody estymacji.
- W przypadku modeli ARIMA ostrzeżenie jest generowane, jeśli w okresie estymacji istnieje predyktor z brakami danych. Modele, które zawierają ten predyktor, nie są poddawane ponownej estymacji.
- Jeśli w okresie prognozy występuje jakakolwiek zmienna niezależna z brakami danych, procedura generuje ostrzeżenie i wykonuje prognozę w możliwie najszerszym okresie.

Szerokość przedziału ufności (%). Przedziały ufności są obliczane dla predykcji modelu i autokorelacji reszt. Można określić dowolną wartość dodatnią mniejszą od 100. Domyślnie ustawiony jest 95% przedział ufności.

aksymalna liczba przesunięć przedstawiana w wynikach ACF i PACF. Można określić maksymalną liczbę przesunięć wyświetlanych w tabelach i wykresach autokorelacji i autokorelacji cząstkowych. Opcja ta jest dostępna tylko wtedy, gdy wykonano ponowną estymację parametrów modelu (**Estymuj ponownie na podstawie danych** na karcie Modele).

Dodatkowe właściwości komendy TSAPPLY

Dodatkowe funkcje są dostępne po wklejeniu zaznaczeń do okna składni i edycji wynikowej składni komend TSAPPLY. Język składni komend umożliwia:

- Określenie, że tylko podzbiór modeli w pliku modelu ma być zastosowany do aktywnego zbioru danych (przy użyciu słów kluczowych DROP i KEEP w opcji komendy MODEL).
- Zastosowanie modeli z dwóch lub więcej plików modeli do danych (przy użyciu opcji komendy MODEL). Na przykład jeden plik modelu może zawierać modele serii, które reprezentują sprzedaż jednostkową, a inny może zawierać modele serii, które reprezentują przychody.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Dekompozycja sezonowa

Procedura Dekompozycja sezonowa umożliwia dekomponowanie szeregu na składnik sezonowy, połączony składnik trendu i cyklu oraz składnik „błędu”. Procedura ta jest implementacją metody Census Method I, zwanej także „metodą stosunku do średniej ruchomej”.

Przykład. Naukowiec jest zainteresowany analizą miesięcznych wyników pomiaru stężenia ozonu w konkretnej stacji meteorologicznej. Celem jest ustalenie, czy w danych występują jakiegokolwiek trendy. W celu ujawnienia rzeczywistych trendów naukowiec najpierw musi wziąć pod uwagę sezonową zmienność odczytów. Może zastosować procedurę Dekompozycja sezonowa w celu usunięcia systematycznych wahań sezonowych. Następnie przeprowadzi analizę trendów na szeregu skorygowanym sezonowo.

Statystyki. Zestaw czynników sezonowych.

Wymagania dotyczące danych do dekompozycji sezonowej

Dane. Zmienne powinny być numeryczne.

Założenia. Zmienne nie powinny zawierać żadnych osadzonych braków danych. Musi być zdefiniowany co najmniej jeden okresowy składnik daty.

Szacowanie czynników sezonowych

1. Z menu wybierz:

Analiza > Prognozowanie > Dekompozycja sezonowa...

2. Wybierz co najmniej jedną zmienną z listy dostępnych i przenieś je na listę Zmienne. Należy pamiętać, że lista obejmuje tylko zmienne numeryczne.

Typ modelu. Procedura dekompozycji sezonowej oferuje dwa różne podejścia do modelowania czynników sezonowych: multiplikatywne i addytywne.

- *Multiplikatywne.* Składnik sezonowy jest czynnikiem, przez który mnoży się szereg skorygowany sezonowo, otrzymując w wyniku pierwotne wartości szeregu. W rezultacie składniki sezonowe, które są proporcjonalne do całkowitego poziomu szeregu. Obserwacje bez zmienności sezonowej mają składnik sezonowości równy 1.
- *Addytywne.* Poprawki sezonowe są dodawane do danych korygowanych czasowo w celu uzyskania obserwowanych wartości. Poprawka taka ma na celu eliminację z danych efektu sezonowości w celu ujawnienia innych istotnych charakterystyk, które mogą być „maskowane” przez składnik sezonowy. W rezultacie moduł Trendy dokonuje oceny składników sezonowych, które nie zależą od całkowitego poziomu szeregów danych. Obserwacje bez zmienności sezonowej mają składnik sezonowości równy 0.

Wagi przy średnich ruchomych. Opcje Wagi przy średnich ruchomych umożliwiają określenie sposobu traktowania szeregu podczas obliczania średnich ruchomych. Opcje te są dostępne tylko wtedy, gdy okresowość serii jest parzysta. Jeśli okresowość jest nieparzysta, wszystkie punkty mają równe wagi.

- *Dla wszystkich punktów równe.* Średnie ruchome są obliczane z zakresem równym okresowości i we wszystkich punktach ważonych jednakowo. Ta metoda jest zawsze stosowana, jeśli cykliczność jest nieparzysta.
- *Granice przedziałów ważone przez 0,5.* Ruchome średnie szeregów z parzystą okresowością są obliczane z rozpiętością przedziału równą okresowości plus 1 i z granicami rozpiętości przedziałów ważonymi przez 0,5.

Opcjonalnie wykonaj następujące czynności:

- Kliknij przycisk **Zapisz**, aby określić, w jaki sposób mają być zapisywane nowe zmienne.

Dekompozycja sezonowa: Zapisz

Utwórz zmienne. Umożliwia wybranie sposobu traktowania nowych zmiennych.

- *Dodaj do pliku.* Nowa seria utworzona przez dekompozycję sezonową jest zapisywana jako zwykłe zmienne w aktywnym zbiorze danych. Nazwy zmiennych składają się z trzyliterowego przedrostka, podkreślenia oraz numeru.
- *Zastąp istniejące.* Nowa seria utworzona przez Sezonową Dekompozycję jest zapisywana jako zmienne tymczasowe w aktywnym zbiorze danych. Jednocześnie upuszczone zostają wszystkie istniejące tymczasowe zmienne, utworzone przez procedury Prognoza. Nazwy zmiennych składają się z trzyliterowego przedrostka, znaku (#) oraz numeru.
- *Nie twórz.* Nowe szeregi nie zostają dodane do aktywnego zbioru danych.

Nazwy nowych zmiennych

Procedura dekompozycji sezonowej tworzy cztery nowe zmienne (szeregi), z następującymi przedrostkami trzyliterowymi, dla każdego określonego szeregu:

SAF. *Czynniki korygujące wahania sezonowe.* Wartości odzwierciedlają wpływ każdego okresu na poziomie szeregu.

SAS. *Szereg skorygowany o wahania sezonowe.* Są to wartości uzyskane po usunięciu sezonowej zmienności szeregu.

STC. *Wygładzone składniki cyklu i trendu.* Wartości te odzwierciedlają trend i cykliczność w szeregu.

ERR. *Wartości reszt lub „błędów”.* Wartości, które pozostają po usunięciu z szeregu składników sezonowych, trendów i cyklu.

Dodatkowe właściwości komendy SEASON

Język składni komend umożliwia również:

- Określenie okresowości w komendzie SEASON zamiast wybierania jednej z możliwości dostępnych w procedurze Definiuj datę i czas.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Wykresy analizy spektralnej

Procedura Wykresy analizy spektralnej służy do identyfikowania okresowego działania w szeregu czasowym. Zamiast analizować zmienność od jednego punktu do następnego, bada ona zmienność szeregu jako całości, wykrywając składniki okresowe o różnych częstotliwościach. Szereg wygładzony obejmuje silniejsze składniki okresowe o mniejszych częstotliwości; zmienność losowa (biały szum) rozdziela siłę składników pomiędzy wszystkie częstotliwości.

Szeregi, które obejmują braki danych, nie mogą być analizowane z użyciem tej procedury.

Przykład. Szybkość budowania nowych domów jest ważnym wskaźnikiem statusu finansowego. Dane rozpoczęcia budowy zwykle wykazują silną zależność od składników związanych z sezonowością. Jednak w danych istnieją dłuższe cykle, które analitycy muszą brać pod uwagę podczas oceny bieżących wartości.

Statystyki. Transformacje sinusoidalne i cosinusoidalne, wartość periodogramu i spektralne oszacowania gęstości dla każdego składnika częstotliwości lub okresu. W przypadku wybrania analizy dwóch zmiennych: części rzeczywiste i urojone periodogramu krzyżowego, gęstość współspektralna, spektrum kwadraturowe, wzmocnienie, kwadrat koherencji i widmo fazowe dla każdej składowej częstotliwości lub okresu.

Wykresy. W przypadku analizy jednej zmiennej i dwóch zmiennych: periodogram i gęstość spektralna. W przypadku analizy dwóch zmiennych: kwadrat koherencji, spektrum kwadraturowe, amplituda krzyżowa, gęstość współspektralna, widmo fazowe i wzmocnienie.

Wymagania dotyczące danych dla wykresów analizy spektralnej

Dane. Zmienne powinny być numeryczne.

Założenia. Zmienne nie powinny zawierać żadnych osadzonych braków danych. Szereg czasowy do przeanalizowania powinien być stacjonarny, a żadna średnia niezerowa nie powinna być odejmowana od szeregu.

- *Stacjonarne.* Warunek, który musi zostać spełniony przez serię czasową, do której pasuje model ARIMA. Przykładowo, szereg MA może być stacjonarny, natomiast szeregi AR i ARMA nie muszą. Stacjonarność szeregu czasowego oznacza niezmienność jego średniej i wariancji w czasie.

Uzyskiwanie analizy spektralnej

1. Z menu wybierz:

Analiza > Szeregi czasowe > Analiza spektralna...

2. Wybierz co najmniej jedną zmienną z dostępnej listy i przenieś wybrane zmienne na listę Zmienne. Należy pamiętać, że lista obejmuje tylko zmienne numeryczne.
3. Wybierz jedną z opcji Okno spektralne, aby określić sposób wygładzenia periodogramu w celu uzyskania oszacowania gęstości spektralnej. Dostępne opcje wygładzenia to: Tukey-Hamming, Tukey, Parzen, Bartlett, Daniell (jednostka) i Brak.
 - *Tukey-Hamming.* Wagi są następujące: $W_k = .54D_p(2\pi f_k) + .23D_p(2\pi f_k + \pi/p) + .23D_p(2\pi f_k - \pi/p)$, dla $k = 0, \dots, p$, gdzie p jest całkowitą częścią połowy zakresu, a D_p jest jądrem Dirichleta rzędu p .
 - *Tukey.* Wagi są następujące: $W_k = 0.5D_p(2\pi f_k) + 0.25D_p(2\pi f_k + \pi/p) + 0.25D_p(2\pi f_k - \pi/p)$, dla $k = 0, \dots, p$, gdzie p jest całkowitą częścią połowy zakresu, a D_p jest jądrem Dirichleta rzędu p .
 - *Parzen.* Wagi są następujące: $W_k = 1/p(2 + \cos(2\pi f_k)) (F[p/2](2\pi f_k))^{**2}$, dla $k = 0, \dots, p$, gdzie p jest całkowitą częścią połowy zakresu, a $F[p/2]$ jest jądrem Fejera rzędu $p/2$.
 - *Bartlett.* Kształt okna spektralnego, dla którego wagi górnej połowy okna są obliczane jako $W_k = F_p(2\pi f_k)$, dla $k = 0, \dots, p$, gdzie p jest całkowitą częścią połowy zakresu, a F_p jest jądrem Fejera rzędu p . Dolna połowa jest symetryczna z górną połową.
 - *Daniell (jednostka).* Kształt okna spektralnego, dla którego wszystkie wagi są równe 1.
 - *Brak.* Brak wygładzania. Po wybraniu tej opcji ocena gęstości spektralnej będzie taka sama jak periodogramu.

Rozpiętość. Zakres następujących po sobie wartości, dla których wykonywana jest wygładzanie. Zwykle przyjmuje się nieparzystą liczbę całkowitą. Im większa rozpiętość, tym wykres gęstości spektralnej jest bardziej wygładzony.

Centruj zmienne. Dopasowuje szereg, aby otrzymać średnią równą 0 przed obliczaniem spektrum i aby usunąć duży składnik, który może być powiązany ze średnią szeregu.

Analiza dwóch zmiennych - pierwsza z każdą następną. Jeśli wybrano co najmniej dwie zmienne, można wybrać tę opcję, aby utworzyć żądanie analizy spektralnej dwóch zmiennych.

- Pierwsza zmienna z listy Zmienne jest traktowana jako zmienna niezależna, a wszystkie pozostałe zmienne są traktowane treated jako zmienne zależne.
- Każdy szereg po pierwszym jest analizowany względem pierwszego szeregu, niezależnie od nazwy pozostałych szeregów. Analiza jednej zmiennej poszczególnych szeregów również jest wykonywana.

Wykres. Periodogram i gęstość spektralna to opcje dostępne zarówno dla analizy jednej zmiennej, jak i dwóch zmiennych. Pozostałe opcje są dostępne tylko dla analizy dwóch zmiennych.

- *Periodogram.* Niewygładzona fabuła amplitudy spektralnej (wykreślona na skali logarytmicznej) względem częstotliwości lub okresu. Zmienność o niskiej częstotliwości charakteryzuje szeregi wygładzone. Równomierny rozrzut zmienności wśród wszystkich częstotliwości wskazuje na „biały szum”.
- *Koherencja kwadratowa.* Wynik korzyści dwóch szeregów.
- *Kwadratura spektrum.* Wyimaginowana część periodogramu krzyżowego, która jest miarą korelacji składowych częstotliwości poza-fazowych dwóch szeregów czasowych. Komponenty są przesunięte w fazie o $\pi/2$ radianów.
- *Amplituda krzyżowa.* Pierwiastek kwadratowy sumy kwadratów różnic między wartościami dla pozycji.

- *Gęstość spektralna*. Periodogram wygładzony w taki sposób, aby wyeliminować nieregularną zmienność.
- *Gęstość kospektralna*. Rzeczywista część periodogramu krzyżowego, będąca miarą korelacji częstotliwości w fazie komponentów dwóch szeregów czasowych.
- *Spektrum fazy*. Miara zakresu, w jakim każdy komponent częstotliwości jednego szeregu wyprzedza lub opóźnia inny komponent.
- *Korzyści*. Iloraz rozdzielenia amplitudy krzyżowej przez gęstość spektralną dla jednej z serii. Każdy z dwóch szeregów ma własną wartość korzyści.

Według częstotliwości. Wszystkie wykresy kreślone są według częstotliwości, od 0 (stała lub średnia) do 0,5 (cykl dwóch obserwacji).

Według okresu. Wszystkie wykresy są tworzone według okresu, od 2 (termin dla cyklu dwóch obserwacji) do okresu równego liczbie obserwacji (wartości stałej lub średniej). Okres wyświetlany jest na skali logarymicznej.

Dodatkowe właściwości komendy SPECTRA

Język składni komend umożliwia również:

- Zapisanie obliczonych zmiennych analizy spektralnej do aktywnego zbioru danych w celu późniejszego użycia.
- Określenie niestandardowych wag okna spektralnego.
- Utworzenie wykresów na podstawie częstotliwości i okresu.
- Wydrukowanie pełnej listy wszystkich wartości widocznych na wykresie.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Modele przyczynowe szeregów czasowych

Modelowanie przyczynowe szeregów czasowych jest próbą wykrycia kluczowych zależności przyczynowych w danych o szeregach czasowych. W procesie modelowania przyczynowego szeregów czasowych użytkownik określa zbiór szeregów przewidywanych i zbiór potencjalnych zmiennych wejściowych dla tych szeregów przewidywanych. Następnie procedura buduje autoregresywny model każdego szeregu przewidywanego i uwzględnia tylko te zmienne wejściowe, które z przewidywanym szeregiem łączą relacje przyczynowe. Ta strategia różni się od tradycyjnego modelowania szeregów czasowych, w którym użytkownik musi jawnie określić predyktory dla przewidywanego szeregu. Ponieważ w modelowaniu przyczynowym szeregów czasowych zazwyczaj buduje się modele dla wielu powiązanych szeregów czasowych, wynik tego modelowania nazywamy *systemem modelu*.

W kontekście modelowania przyczynowego szeregów czasowych termin *przyczynowe* odnosi się do przyczynowości w sensie Grangera. Szereg czasowy X jest uznawany za „przyczynę w sensie Grangera” innego szeregu czasowego Y, jeśli regresja dla Y przy przeszłych wartościach szeregów X i Y zwraca lepszy model dla Y niż regresja tylko przy przeszłych wartościach dla Y.

Przykłady

Osoby podejmujące decyzje biznesowe mogą korzystać z modelowania przyczynowego szeregów czasowych do odkrywania relacji przyczynowych w dużych zestawach metryk w oparciu o czas opisujących procesy biznesowe. Efektem analizy może być zidentyfikowanie możliwych do zmiany danych wejściowych mających największy wpływ na kluczowe wskaźniki wydajności.

Menedżerowie dużych systemów IT mogą korzystać z modelowania przyczynowego szeregów czasowych do wykrywania anomalii w dużych zbiorach powiązanych ze sobą metryk operacyjnych. Model przyczynowy umożliwi wówczas więcej niż tylko wykrywanie anomalii i odkrywanie najbardziej prawdopodobnych ich przyczyn.

Wymagania dotyczące zmiennych

Wymagana jest co najmniej jedna zmienna przewidywana. Domyślnie zmienne o predefiniowanej roli Brak nie są używane.

Struktura danych

Modelowanie przyczynowe szeregów czasowych obsługuje dwa rodzaje struktur danych.

Dane kolumnowe

W przypadku danych kolumnowych każda zmienna szeregu czasowego zawiera dane dla jednego szeregu czasowego. Struktura ta jest tradycyjną strukturą danych szeregów czasowych, stosowaną przez kreator modeli szeregów czasowych.

Dane wielowymiarowe

W przypadku danych wielowymiarowych każda zmienna szeregu czasowego zawiera dane dla wielu szeregów czasowych. Oddzielne szeregi czasowe w ramach danej zmiennej są wówczas identyfikowane na podstawie zestawu wartości zmiennych jakościowych, zwanych także zmiennymi *wymiarów*. Na przykład dane sprzedaży dla dwu różnych kanałów sprzedaży (detalicznego i internetowego) mogą być przechowywane w jednej zmiennej *sales*. Zmienna wymiarowa o nazwie *channel* i wartościach 'retail' oraz 'web' określa rekordy powiązane z każdym z tych dwu kanałów sprzedaży.

Aby uzyskać model przyczynowy szeregów czasowych

Ta funkcja wymaga opcji Statistics Forecasting.

Z menu wybierz jedną z opcji:

Analizuj > Prognozowanie > Utwórz modele przyczynowe szeregów czasowych...

1. Jeśli obserwacje są zdefiniowane przez zmienną daty/czasu, określ tę zmienną.
2. W przypadku danych wielowymiarowych określ zmienne wymiarów, które identyfikują szereg czasowy.
 - Kolejność określona w zmiennych wymiaru definiuje kolejność, w jakiej będą wyświetlane wszystkie następnne okna dialogowe i wyniki. Kolejność zmiennych wymiarów można zmieniać za pomocą przycisków strzałek w górę i w dół.
 - Można określić podzbiory wartości wymiarów, które ograniczą analizę do konkretnych wartości zmiennych wymiarów. Na przykład, jeśli używane są wymiary regionu i marki, można ograniczyć analizę do konkretnego regionu. Podzbiory wymiarów mają zastosowanie do wszystkich zmiennych metryk użytych w analizie.
 - Można także zmodyfikować analizę, określając wartości wymiarów według ról na poziomie zmiennej metryki. Na przykład, jeśli istnieje wymiar kanału sprzedaży (z wartościami 'stacjonarny' i 'internetowy') oraz metryki sprzedaży i reklamy w tych kanałach, można określić reklamę WWW jako wejście dla handlu stacjonarnego i internetowego. Domyślnie ten typ modyfikacji jest dozwolony i ograniczony do wyboru maksymalnej określonej liczby odrębnych wartości (domyślnie 250) dla każdej zmiennej wymiaru.
3. Kliknij przycisk **Kontynuuj**.

Uwaga: Kroki 1, 2 i 3 nie mają zastosowania, jeśli aktywny zbiór danych ma specyfikację daty. Specyfikacje dat tworzy się w oknie dialogowym Definiowanie daty i czasu lub za pomocą komendy DATE.
4. Kliknij przycisk **Zmienne**, aby określić szeregi czasowe, które mają zostać uwzględnione w modelu, a także celu określenia, w jaki sposób są zdefiniowane obserwacje. Co najmniej jedna zmienna musi być określona jako przewidywana albo jako jednocześnie wejściowa i przewidywana.
5. Kliknij opcję **Specyfikacja danych**, aby określić opcjonalne ustawienia obejmujące przedział czasowy analizy, ustawienia agregacji i rozkładu oraz sposób postępowania z brakami danych.
6. Kliknij **Opcje tworzenia**, aby zdefiniować okres estymacji, zawartość wyników oraz ustawienia tworzenia, takie jak maksymalna liczba zmiennych wejściowych na jedną zmienną przewidywaną.

7. Kliknij **Opcje modelu**, aby zlecić prognozy, zapisać predykcje i wyeksportować system modelu do pliku zewnętrznego.
8. Kliknij przycisk **Uruchom**, aby uruchomić procedurę.

Szeregi czasowe dla modelowania

Na karcie Zmienne należy użyć ustawień **Szeregi czasowe**, aby określić szeregi, jakie będą uwzględniane w systemie modelu.

W przypadku danych kolumnowych pojęcie *szeregi* ma takie samo znaczenie jak pojęcie *zmienna*. W przypadku danych wielowymiarowych zmienne, które zawierają szeregi czasowe, są nazywane zmiennymi *metryk*. Szeregi czasowe, dla danych wielowymiarowych, są definiowane przez zmienną metryk i wartość dla każdej zmiennej wymiaru. Przedstawione poniżej rozważania mają zastosowanie do danych kolumnowych oraz do danych wielowymiarowych.

- Szeregi określane jako potencjalne zmienne wejściowe oraz jako przewidywane i wejściowe są uwzględniane w celu dołączenia do modelu dla każdej zmiennej przewidywanej. Model dla każdej zmiennej przewidywanej zawsze uwzględni wartości opóźnione samej zmiennej przewidywanej.
- Szeregi określane jako wymuszone zmienne wejściowe zawsze są uwzględniane w modelu dla każdej zmiennej przewidywanej.
- Co najmniej jeden szereg musi być określony jako przewidywany lub jako przewidywany i wejściowy.
- Jeśli zaznaczono opcję **Użyj wstępnie zdefiniowanych ról**, zmienne z rolą Zmienna wejściowa są ustawiane jako potencjalne zmienne wejściowe. Żadna wstępnie zdefiniowana rola nie jest odwzorowywana na potencjalną zmienną wejściową.

Dane wielowymiarowe

W przypadku danych wielowymiarowych zmienne metryk i powiązane role są określane w siatce, w której każdy wiersz określa pojedynczą metrykę lub rolę. Domyślnie system modelu obejmuje szeregi dla wszystkich kombinacji zmiennych wymiarów dla każdego wiersza w siatce. Jeśli na przykład dostępne są wymiary dla *region* i *brand*, domyślnie określenie metryki *sales* jako przewidywanej oznacza, że istnieje osobny szereg przewidywany sprzedaży dla każdej kombinacji wartości *region* i *brand*.

Dla każdego wiersza w siatce można dostosować zestaw wartości dowolnej zmiennej wymiaru, klikając przycisk wielokropka danego wymiaru. Ta czynność otworzy podrzędne okno dialogowe Wybierz wartości wymiarów. Wiersze siatki można również dodawać, usuwać lub kopiować.

W kolumnie **Liczebność szeregu** wyświetlana jest liczba zestawów wartości wymiarów, jakie są aktualnie określone dla powiązanej metryki. Wyświetlana wartość może być większa niż rzeczywista liczba szeregów (jeden szereg na zestaw). Taka sytuacja ma miejsce, jeśli niektóre z określonych kombinacji wartości wymiarów nie odpowiadają szeregom objętym przez powiązaną metrykę.

Wybierz wartości wymiarów

W przypadku danych wielowymiarowych można dostosować analizę poprzez określenie, które wartości wymiarów mają zastosowanie do konkretnej zmiennej metryki dla konkretnej roli. Na przykład, jeśli *sales* jest zmienną metryki a *channel* jest wymiarem z wartościami 'retail' i 'web', można określić, że 'web' *sales* jest zmienną wejściową, a 'retail' *sales* jest zmienną przewidywaną. Można również określić podzbiory wymiarów, jakie będą miały zastosowanie do wszystkich zmiennych metryk użytych w analizie. Na przykład, jeśli *region* jest zmienną wymiaru, która określa region geograficzny, wówczas można ograniczyć analizę do konkretnych regionów.

Wszystkie wartości

Ta opcja określa, że uwzględniane są wszystkie wartości bieżącej zmiennej wymiaru. Jest to opcja domyślna.

Wybierz wartości do uwzględnienia lub wykluczenia

Tej opcji należy użyć, aby określić zbiory wartości dla bieżącej zmiennej wymiaru. Jeśli dla opcji **Tryb** wybrana jest wartość **Uwzględnij**, uwzględniane będą tylko wartości określone na liście **Wybrane**

wartości. Jeśli dla opcji **Tryb** wybrana jest wartość **Wyklucz**, wówczas uwzględniane są wszystkie wartości inne niż te określone na liście **Wybrane wartości**.

Wartości, z których można dokonywać wyboru, można filtrować. Wartości, które spełniają warunki filtrowania, są wyświetlane na karcie **Dopasowane**, a wartości, które nie spełniają warunków filtrowania, są wyświetlane na karcie **Niedopasowane** w postaci listy **Niewybrane wartości**. Karta **Wszystkie** zawiera listę wszystkich niewybranych wartości, niezależnie od warunków filtrowania.

- Podczas określania filtru można użyć gwiazdki (*) jako symbolu wieloznacznego.
- Aby wyczyścić bieżący filtr, dla poszukiwanego terminu w oknie dialogowym Filtruj wyświetlane wartości należy wprowadzić pustą wartość.

Obserwacje

Na karcie Zmienne należy użyć ustawień **Obserwacje**, aby określić zmienne definiujące obserwacje.

Uwaga: Jeśli aktywny zbiór danych zawiera specyfikację daty, wówczas obserwacje są definiowane przez te specyfikacje daty i nie można ich zmodyfikować w ramach procedury modelowania przyczynowego szeregów czasowych. Specyfikacje dat tworzy się w oknie dialogowym Definiowanie daty i czasu lub za pomocą komendy DATE.

Obserwacje definiowane jako data/czas

Można określić, czy obserwacje będą definiowane zmienną w formacie daty, czasu lub daty/czasu, czy też przez zmienną łańcuchową, która reprezentuje datę/czas. Zmienne łańcuchowe mogą prezentować datę w formacie RRRR-MM-DD, a czas w formacie GG:MM:SS lub datę/czas w formacie RRRR-MM-DD GG:MM:SS. W reprezentacji łańcuchowej zera początkowe mogą być pominięte. Przykładowo łańcuch 2014-9-01 odpowiada zapisowi 2014-09-01.

Oprócz zmiennej, która definiuje obserwacje, należy wybrać odpowiedni przedział czasowy, w którym będą opisywane obserwacje. W zależności od określonego przedziału czasowego można również dokonać innych ustawień, takich jak przedział między obserwacjami (przyrost) lub liczba dni w tygodniu. Poniższe stwierdzenia mają zastosowanie do przedziałów czasowych:

- Wartości **Nieregularny** należy użyć, jeśli obserwacje są nierównomiernie rozmieszczone w czasie, na przykład są wykonywane w chwili, gdy następuje przetwarzanie zamówienia sprzedaży. Jeśli wybrana zostanie opcja **Nieregularny**, należy określić przedział czasowy, jaki będzie używany do analizy, wybierając ustawienia **Przedział czasowy** na karcie Specyfikacja danych.
- Jeśli obserwacje reprezentują datę i czas, a przedział czasowy to godziny, minuty lub sekundy, wówczas należy wybrać opcję **Godziny dziennie**, **Minuty dziennie** lub **Sekundy dziennie**. Jeśli obserwacje reprezentują czas (trwanie) bez odniesienia do daty, a przedział czasowy to godziny, minuty lub sekundy, wówczas należy użyć opcji **Godziny (nieokresowo)**, **Minuty (nieokresowo)** lub **Sekundy (nieokresowo)**.
- Na podstawie wybranego przedziału czasowego procedura może wykryć brakujące obserwacje. Wykrywanie brakujących obserwacji jest konieczne, ponieważ procedura zakłada, że wszystkie obserwacje są równomiernie rozłożone w czasie i że nie ma brakujących obserwacji. Na przykład, jeśli przedziałem czasu są dni, a po dacie 2014-10-27 występuje 2014-10-29, istnieje brakująca obserwacja dla 2014-10-28. Dla wszystkich brakujących obserwacji wprowadzane są wartości. Ustawienia obsługi braków danych można wprowadzić na karcie Specyfikacja danych.
- Określone przedziały czasowe umożliwiają procedurze wykrycie wielu obserwacji w jednym przedziale czasowym, które muszą być zagregowane i które są dopasowane do obserwacji dla granicy przedziału (np. pierwszy dzień miesiąca), dzięki czemu obserwacje będą równomiernie rozmieszczone. Na przykład, jeśli przedziałem czasu są miesiące, to zagregowane zostanie wiele dat z tego samego miesiąca. Ten typ agregacji jest nazywany *grupowaniem*. Domyślnie obserwacje są sumowane podczas grupowania. Można określić inną metodę grupowania, np. średnia z obserwacji; ustawienia **Agregacja i rozkład** można wprowadzić na karcie Specyfikacja danych.
- Przy niektórych przedziałach czasowych istnieją dodatkowe ustawienia umożliwiające zdefiniowanie odstępów w zwykle równomiernie rozmieszczonych przedziałach. Na przykład, jeśli przedziałem

czasu są dni, ale istotne są tylko dni robocze, można określić pięciodniowy tydzień rozpoczynający się od poniedziałku.

Obserwacje zdefiniowane jako okresy lub okresy cykliczne

Obserwacje mogą być zdefiniowane przez co najmniej jedną zmienną całkowitą, która reprezentuje okresy lub powtarzające się cyklicznie okresy, aż do dowolnej liczby poziomów cyklicznych. Taka struktura pozwala opisać serie obserwacji, które nie pasują do jednego ze standardowych przedziałów czasowych. Przykładowo, rok fiskalny trwający tylko 10 miesięcy może być opisany przez zmienną cyklu, która reprezentuje lata, i zmienną okresu, która reprezentuje miesiące, przy czym długość jednego cyklu wynosi 10.

Zmienne, które określają okresy cykliczne, definiują hierarchię poziomów okresowości, w której najniższy poziom jest definiowany przez zmienną **Okres**. Następny wyższy poziom jest określany przez zmienną cyklu z poziomem 1, po której następuje zmienna cyklu z poziomem 2 itd. Wartości zmiennych dla każdego poziomu, z wyjątkiem najwyższego, muszą być okresowe w odniesieniu do kolejnego najwyższego poziomu. Wartości dla najwyższego poziomu nie mogą być okresowe. Na przykład, dla 10-miesięcznego roku fiskalnego miesiące występują okresowo w latach, ale lata nie są okresowe.

- Długość cyklu na poszczególnych poziomach stanowi okresowość dla kolejnego najniższego poziomu. W przykładzie dot. roku fiskalnego istnieje tylko jeden poziom cyklu, a długość cyklu wynosi 10, ponieważ kolejny najniższy poziom reprezentuje miesiące, a w określonym roku fiskalnym jest 10 miesięcy.
- Należy określić wartość początkową dla każdej zmiennej okresowej, która nie rozpoczyna się od 1. To ustawienie jest niezbędne do wykrywania braków danych. Na przykład, jeśli zmienna okresowa rozpoczyna się od 2, ale wartość początkowa jest określona jako 1, wówczas procedura zakłada, że istnieje brak danych dla pierwszego okresu w każdym cyklu dla tej zmiennej.

Obserwacje definiowane jako zgodne z kolejnością rekordów

W przypadku danych kolumnowych można określić, czy obserwacje są definiowane zgodnie z kolejnością rekordów, czyli pierwszy rekord reprezentuje pierwszą obserwację, drugi rekord reprezentuje drugą obserwację itd. Zakłada się wówczas, że rekordy reprezentują obserwacje, które są równomiernie rozłożone w czasie.

Przedział czasowy analizy

Przedział czasowy, który jest używany do analizy, może różnić się od przedziału czasowego dla obserwacji. Na przykład, jeśli przedział czasowy obserwacji to dni, jako przedział czasowy dla analizy można wybrać miesiące. Wówczas przed zbudowaniem modelu dane agregowane są z dziennych na miesięczne. Można również rozłożyć dane z dłuższego przedziału czasu na krótszy. Przykładowo, jeśli obserwacje są przeprowadzane kwartalnie, wówczas można rozłożyć dane z kwartalnych na miesięczne.

Opcje możliwe do wyboru dla przedziału czasowego, w jakim wykonywana jest analiza, zależą od sposobu zdefiniowania obserwacji oraz wyznaczonego dla nich przedziału czasowego. W szczególności, jeśli obserwacje są definiowane przez okresy cykliczne lub jeśli dla aktywnego zbioru danych zdefiniowana jest specyfikacja daty, wówczas obsługiwana jest tylko agregacja. W takim przypadku przedział czasowy dla analizy musi być większy od przedziału czasowego dla obserwacji lub mu równy.

Przedział czasowy dla analizy można określić w ustawieniach **Przedział czasowy** na karcie Specyfikacja danych. Metoda, którą dane są agregowane lub rozkładane, jest określana w ustawieniach **Agregacja i rozkład** na karcie Specyfikacja danych.

Agregacja i rozkład

Funkcje agregacji

Jeśli przedział czasowy użyty dla analizy jest dłuższy niż przedział czasowy dla obserwacji, dane wejściowe zostają zregulowane. Przykładowo agregacja jest przeprowadzana, kiedy przedział czasowy dla obserwacji to dni, a przedział czasowy dla analizy to miesiące. Dostępne są następujące funkcje agregacji: średnia, suma, dominanta, wartość minimalna lub maksymalna.

Funkcje rozkładu

Jeśli przedział czasowy użyty dla analizy jest krótszy niż przedział czasowy dla obserwacji, dane wejściowe zostają rozłożone. Przykładowo rozkład jest przeprowadzany, kiedy przedział czasowy dla obserwacji to kwartały, a przedział czasowy dla analizy to miesiące. Dostępne są następujące funkcje rozkładu: średnia lub suma.

Funkcje grupujące

Grupowanie jest stosowane, kiedy obserwacje są definiowane przez datę/czas i w tym samym przedziale czasowym występuje wiele obserwacji. Na przykład, jeśli przedział czasowy dla obserwacji to miesiące, wówczas wiele dat z tego samego miesiąca jest grupowanych i tworzone jest ich powiązanie z miesiącem, w którym występują. Dostępne są następujące funkcje grupowania: średnia, suma, dominanta, wartość minimalna lub maksymalna. Grupowanie jest zawsze przeprowadzane, kiedy obserwacje są zdefiniowane przez datę/czas, a przedział czasowy dla obserwacji jest określony jako Nieregularny.

Uwaga: Chociaż grupowanie jest formą agregacji, jest przeprowadzane przed rozpoczęciem obsługi braków danych, podczas gdy formalna agregacja jest wykonywana po zakończeniu obsługi braków danych. Jeśli przedział czasowy dla obserwacji jest określony jako Nieregularny, agregacja jest wykonywana tylko za pomocą funkcji grupowania.

Agreguj obserwacje przekraczające granice dnia do dnia poprzedniego

Określa, czy obserwacje, których czas przekracza granicę dnia, są agregowane na wartości dla dnia poprzedniego. Przykładowo, dla obserwacji godzinowych trwających osiem godzin dziennie i rozpoczynających się o godzinie 20:00, to ustawienie określi, czy obserwacje od godziny 00:00 do 04:00 będą uwzględniane w zagregowanych wynikach dla poprzedniego dnia. To ustawienie ma zastosowanie tylko w przypadku, kiedy przedział czasowy dla obserwacji to Godziny dziennie, Minuty dziennie lub Sekundy dziennie, a przedział czasowy dla analizy to Dni.

Ustawienia niestandardowe dla określonych zmiennych

Funkcje agregacji, rozkładu i grupowania dla zmiennej można określić na podstawie zmiennej. Ustawienia te zastępują domyślne ustawienia funkcji agregacji, rozkładu i grupowania.

Brakujące wartości

Brakujące wartości w danych wejściowych są zastępowane przez wartość podstawianą. Dostępne są zastępujące metody zastępowania:

Interpolacja liniowa

Powoduje zastąpienie braków danych przy wykorzystaniu interpolacji liniowej. W interpolacji używana jest ostatnia ważna wartość przed brakiem danych oraz pierwsza ważna za brakiem. Jeśli pierwsza lub ostatnia obserwacja w szeregu zawiera brakujące wartości, wówczas używane są dwie najbliższe niebrakujące wartości na początku i na końcu serii.

Średnia szeregu

Zastępuje braki danych średnią obliczoną ze wszystkich obserwacji.

Rozpiętość sąsiednich punktów

Powoduje zastąpienie braków danych średnią z ważnych wartości sąsiednich. Rozpiętość sąsiednich punktów, to liczba poprawnych wartości występujących przed i po brakującej wartości, jakie są wykorzystywane do obliczenia średniej.

Rozpiętość sąsiednich punktów

Powoduje zastąpienie braków danych medianą ważnych wartości sąsiednich. Rozpiętość sąsiednich punktów, to liczba poprawnych wartości występujących przed brakującą wartością i po niej, jakie są wykorzystywane do obliczenia mediany.

Trend liniowy

Ta opcja wykorzystuje niebrakujące obserwacje w szeregu do dopasowania prostego modelu regresji liniowej, który jest następnie używany w celu przypisania brakujących wartości.

Inne ustawienia:

Maksymalny procent braków danych (%)

Określa maksymalną wartość procentową braków danych, jaka jest dozwolona w szeregu. Szeregi z większą liczbą braków danych od określonego maksimum są wykluczane z analizy.

Brakujące wartości użytkownika

Ta opcja określa, czy braki danych użytkownika będą traktowane jako poprawne dane i czy będą uwzględniane w szeregu. Domyślnie braki danych użytkownika są wykluczane i są traktowane jako systemowe braki danych, które później są zastępowane.

Ogólne opcje danych

Maksymalna liczba odrębnych wartości na zmienną wymiaru

To ustawienie dotyczy danych wielowymiarowych i określa maksymalną liczbę odrębnych wartości, jaka jest dozwolona dla dowolnej zmiennej wymiaru. Domyślnie ograniczenie jest ustawione na 10000, ale wartość tę można zmienić na dowolnie dużą liczbę.

Ogólne opcje budowania

Szerokość przedziału ufności (%)

To ustawienie decyduje o przedziałach ufności dla prognoz i parametrów modelu. Można określić dowolną wartość dodatnią mniejszą od 100. Domyślnie ustawiony jest 95% przedział ufności.

Maksymalna liczba zmiennych wejściowych na każdą zmienną przewidywaną

To ustawienie określa maksymalną liczbę zmiennych wejściowych, jaka jest dozwolona w modelu dla każdej zmiennej przewidywanej. Można określić liczbę całkowitą z zakresu od 1 do 20. Model dla każdej zmiennej przewidywanej zawsze uwzględnia wartości opóźnione samej zmiennej przewidywanej; ustawienie tej wartości na 1 spowoduje, że tylko zmienna wejściowa jest samą zmienną przewidywaną.

Tolerancja modelu

To ustawienie kontroluje proces iteracyjny, jaki jest stosowany do określenia najlepszego zestawu zmiennych wejściowych dla każdej zmiennej przewidywanej. Można określić dowolną wartość większą od zera. Domyślną wartością jest 0,001.

Próg (%) dla wartości odstającej

Obserwacja jest oznaczana jako odstająca, jeśli obliczone wg modelu prawdopodobieństwo, że jest odstająca, przekracza wyznaczony próg. Można określić wartość z zakresu od 50 do 100.

Liczba opóźnień dla każdej zmiennej wejściowej

To ustawienie określa liczbę opóźnień dla każdej zmiennej wejściowej w modelu dla każdej zmiennej przewidywanej. Domyślnie liczba opóźnień jest określana automatycznie na podstawie przedziału czasowego używanego do analizy. Przykładowo, jeśli przedział czasowy to miesiące (z przyrostem o jeden miesiąc), wówczas liczba opóźnień wynosi 12. Opcjonalnie można jawnie określić liczbę opóźnień. Podana wartość musi być liczbą całkowitą z zakresu od 1 do 20.

Prezentacja szeregów

Te opcje określają szeregi (przewidywane lub wejściowe), dla których wyświetlany jest wynik. Zawartość wyniku dla określonego szeregu jest określana przez ustawienia **Opcje wyników**.

Przedstaw zmiennie przewidywane w powiązaniu z modelami o najlepszym dopasowaniu

Domyślnie wynik jest wyświetlany dla zmiennych przewidywanych, które są powiązane z 10 najlepszymi modelami, określonymi na podstawie wartości R-kwadrat. Można określić inną stałą liczbę najlepszych modeli lub podać odstępek tych modeli. Można również wybrać jedną z następujących miar dopasowania:

R-kwadrat

Miara dobroci dopasowania modelu liniowego, czasami nazywana współczynnikiem determinacji. Jest to część zmienności w zmiennej przewidywanej wyjaśniona przez model. Przyjmuje wartość z zakresu od 0 do 1. Małe wartości wskazują, że model nie pasuje dobrze do danych.

Pierwiastek z procentowego błędu średniokwadratowego

Miara wielkości przewidywanych wartości modelu różni się od obserwowanych wartości szeregu. Jest niezależna od używanych jednostek i tym samym może być używana do porównywania szeregów używających różnych jednostek.

Pierwiastek błędu średniokwadratowego

Pierwiastek kwadratowy obliczany z przeciętnego odchylenia kwadratowego. Mierzy, jak bardzo szereg zależny odbiega od poziomu przewidywanego przez model; miara wyrażona w jednostkach używanych przez szereg zależny.

BIC

Bayesowskie kryterium informacyjne. Miara wyboru i porównywania modeli w oparciu o -2 zredukowane logarytm wiarygodności. Mniejsze wartości oznaczają lepszy model. Wartość BIC „karze” także modele przeparametryzowane (na przykład złożone modele z dużą liczbą danych wejściowych), jednak silniej niż miara AIC.

AIC

Kryterium informacyjne Akaike. Miara wyboru i porównywania modeli w oparciu o -2 zredukowane logarytm wiarygodności. Mniejsze wartości oznaczają lepszy model. Wartość AIC „karze” modele przeparametryzowane (na przykład złożone modele z dużą liczbą danych wejściowych).

Określ poszczególne szeregi

Można określić poszczególne szeregi, dla których potrzebny jest wynik.

- W przypadku danych kolumnowych należy określić zmienne, które zawierają wymagane szeregi. Kolejność określonych zmiennych definiuje kolejność, w jakiej są wyświetlane w wyniku.
- W przypadku danych wielowymiarowych należy określić konkretny szereg, dodając wpis do siatki dla zmiennej metryki, która zawiera szereg. Następnie należy określić wartości zmiennych wymiaru, które definiują szereg.
 - Można wprowadzić wartość dla każdej zmiennej wymiaru bezpośrednio do siatki lub wybrać ją z listy dostępnych wartości wymiarów. Aby wybrać wartość z listy, należy kliknąć przycisk wielokropka w komórce wybranego wymiaru. Ta czynność otworzy podrzędne okno dialogowe Wybierz zmienną wymiaru.
 - Listę wartości wymiaru w podrzędnym oknie dialogowym Wybierz zmienną wymiaru można przeszukać, klikając ikonę lornetki i wprowadzając poszukiwany termin. Spacje są traktowane jako część poszukiwanego terminu. Gwiazdka (*) w poszukiwanym terminie nie oznacza symbolu wieloznacznego.
 - Kolejność szeregów w siatce definiuje kolejność, w jakiej są wyświetlane w wyniku.

Wynik dla danych kolumnowych i danych wielowymiarowych jest ograniczony do 30 szeregów. Ten limit uwzględnia poszczególne szeregi (wejściowe i przewidywane) określone przez użytkownika i przewidywane powiązane z najlepszymi modelami. Szeregi określone przez użytkownika mają pierwszeństwo w stosunku do przewidywanych powiązanych z najlepszymi modelami.

Opcje wyników

Te opcje określają zawartość wyniku. Opcje z grupy **Wyniki dla zmiennych przewidywanych** generują wynik dla zmiennych przewidywanych, które są powiązane z najlepszymi modelami w ustawieniach **Prezentacja szeregów**. Opcje z grupy **Wyniki dla szeregów czasowych** generują wyniki dla poszczególnych szeregów, które zostały określone w ustawieniach **Prezentacja szeregów**.

Ogólny model systemu

Wyświetla graficzną reprezentację relacji przyczynowych pomiędzy szeregami w systemie modelu. Tabele statystyk dopasowania modelu i wartości odstających dla wyświetlanych przewidywanych są uwzględniane jako część element wyniku. Jeśli ta opcja zostanie wybrana w grupie **Wyniki dla szeregów czasowych**, tworzony jest osobny element wyniku dla każdego pojedynczego szeregu, jaki został określony w ustawieniach **Prezentacja szeregów**.

Relacje przyczynowe pomiędzy szeregami mają przypisany poziom istotności, gdzie niższy poziom istotności oznacza bardziej istotny związek. Istnieje możliwość ukrycia relacji, których poziom istotności jest większy od określonej wartości.

Statystyki dopasowania modelu i wartości odstające

Tabele statystyk dopasowania modelu i wartości odstających dla szeregów przewidywanych, jakie zostały wybrane do wyświetlenia. Tabele te zawierają te same informacje, co tabele w wizualizacji Ogólny system modelu. Obsługują one wszystkie standardowe funkcje przedstawiania i edytowania tabel.

Efekty i parametry modelu

Tabele testów efektów modelu i parametrów modelu dla szeregów przewidywanych, jakie zostały wybrane do wyświetlenia. Testy efektów modelu obejmują statystykę F i powiązaną wartość istotności dla każdej zmiennej wejściowej uwzględnianej w modelu.

Diagram wpływu

Przedstawia graficzną reprezentację relacji przyczynowych pomiędzy szeregami będącymi przedmiotem zainteresowania a innymi szeregami, na które one wpływają lub które mają na nie wpływ. Szeregi, które wpływają na szeregi będące przedmiotem zainteresowania, są nazywane *przyczynami*. Wybór opcji **Efekty** powoduje wygenerowanie diagramu wpływu, który jest inicjowany w celu wyświetlenia efektów. Wybór opcji **Przyczyny** powoduje wygenerowanie diagramu wpływu, który jest inicjowany w celu wyświetlenia przyczyn. Wybór opcji **Przyczyny i skutki** powoduje wygenerowanie dwóch osobnych diagramów wpływu; jeden jest inicjowany dla przyczyn, a drugi dla efektów. Przyczyny i efekty można w sposób interaktywny przełączać w elemencie wyniku, który wyświetla diagram wpływu.

Można określić liczbę poziomów przyczyn i efektów do wyświetlenia, gdzie pierwszy poziom stanowi szereg będący przedmiotem zainteresowania. Każdy dodatkowy poziom przedstawia bardziej pośrednie przyczyny lub efekty szeregu będącego przedmiotem zainteresowania. Przykładowo, trzeci poziom wyświetlanych efektów składa się z szeregów z drugiego poziomu stanowiących bezpośrednio dane wejściowe. Na szeregi z trzeciego poziomu szeregi będące przedmiotem zainteresowania mają więc pośredni wpływ, ponieważ szeregi te stanowią bezpośrednio dane wejściowe dla szeregów z drugiego poziomu.

Wykres szeregu

Wykresy obserwowanych i przewidywanych wartości dla szeregów przewidywanych, które zostały wybrane do wyświetlenia. Jeśli żądanie obejmowało prognozy, wykres przedstawia również wartości prognozowane i przedziały ufności dla prognoz.

Wykres reszt

Wykresy reszt modelu dla szeregów przewidywanych, jakie zostały wybrane do wyświetlenia.

Najlepsze zmienne wejściowe

Wykresy poszczególnych wyświetlanych przewidywanych w czasie wraz z 3 najlepszymi zmiennymi wejściowymi dla przewidywanych. Najlepsze zmienne wejściowe to zmienne z najniższą wartością istotności. W celu dostosowania różnych skal zmiennych wejściowych i przewidywanych na osi y prezentowane są oceny z z każdego szeregu.

Tabela prognoz

Tabele zawierające przewidziane wartości i przedziały ufności dla tych prognoz określone dla przewidywanych szeregów, jakie zostały wybrane do wyświetlenia.

Analiza podstawowych przyczyn wartości odstających

Określa, które szeregi z największym prawdopodobieństwem będą powodowały powstawanie poszczególnych wartości odstających w szeregu będącym przedmiotem zainteresowania. Analiza podstawowych przyczyn wartości odstających jest przeprowadzana dla każdego szeregu przewidywanego, który jest uwzględniony na liście poszczególnych szeregów w ustawieniach **Prezentacja szeregów**.

Wyniki

Interaktywna tabela i wykres wartości odstających

Tabela i wykres odstających wartości oraz podstawowych przyczyn powstawania tych wartości dla każdego szeregu będącego przedmiotem zainteresowania. Każda wartość

jest zapisywana w tabeli w osobnym wierszu. Wykres ten przedstawia diagramu wpływu. Wybór wiersza w tabeli powoduje wyróżnienie ścieżki (na diagramie wpływu) prowadzącej od szeregu będącego przedmiotem zainteresowania do szeregu, który z największym prawdopodobieństwem był przyczyną powstania powiązanej wartości odstającej.

Tabela przestawna wartości odstających

Tabela odstających wartości oraz głównych przyczyn powstawania tych wartości dla każdego szeregu będącego przedmiotem zainteresowania. Ta tabela zawiera te same informacje, co tabela na ekranie interaktywnym. Tabela ta obsługuje wszystkie standardowe funkcje przestawiania i edytowania tabel.

Poziomy przyczynowe

Użytkownik może określić liczbę poziomów, jakie będą uwzględnione w wyszukiwaniu podstawowych przyczyn. Informacje na temat poziomów, jakie są tutaj używane, są takie same, jak w przypadku diagramów wpływu.

Dopasowanie wszystkich modeli

Histogram dopasowania modelu dla wszystkich modeli i wybranych statystyk dopasowania. Dostępne są następujące statystyki dopasowania:

R-kwadrat

Miara dobroci dopasowania modelu liniowego, czasami nazywana współczynnikiem determinacji. Jest to część zmienności w zmiennej przewidywanej wyjaśniona przez model. Przyjmuje wartość z zakresu od 0 do 1. Małe wartości wskazują, że model nie pasuje dobrze do danych.

Pierwiastek z procentowego błędu średniokwadratowego

Miara wielkości przewidywanych wartości modelu różni się od obserwowanych wartości szeregu. Jest niezależna od używanych jednostek i tym samym może być używana do porównywania szeregów używających różnych jednostek.

Pierwiastek błędu średniokwadratowego

Pierwiastek kwadratowy obliczany z przeciętnego odchylenia kwadratowego. Mierzy, jak bardzo szereg zależny odbiega od poziomu przewidywanego przez model; miara wyrażona w jednostkach używanych przez szereg zależny.

BIC

Bayesowskie kryterium informacyjne. Miara wyboru i porównywania modeli w oparciu o -2 zredukowane logarytm wiarygodności. Mniejsze wartości oznaczają lepszy model. Wartość BIC „karze” także modele przeparametryzowane (na przykład złożone modele z dużą liczbą danych wejściowych), jednak silniej niż miara AIC.

AIC

Kryterium informacyjne Akaike. Miara wyboru i porównywania modeli w oparciu o -2 zredukowane logarytm wiarygodności. Mniejsze wartości oznaczają lepszy model. Wartość AIC „karze” modele przeparametryzowane (na przykład złożone modele z dużą liczbą danych wejściowych).

Wartości odstające w czasie

Wykres słupkowy liczby wartości odstających dla wszystkich zmiennych przewidywanych i każdego przedziału czasowego w okresie oszacowania.

Transformacje szeregów

Tabela wszystkich transformacji, jakie zostały przeprowadzone dla szeregu w systemie modelu. Możliwe transformacje to podstawianie brakujących wartości, agregacja i rozkład.

Okres estymacji

Domyślnie okres estymacji zaczyna się od czasu z najwcześniejszą obserwacją, a kończy w czasie z najpóźniejszą obserwacją we wszystkich szeregach.

Wyznaczony przez czas początkowy i końcowy

Można określić datę rozpoczęcia i zakończenia okresu estymacji lub można określić tylko datę rozpoczęcia lub tylko datę zakończenia. Jeśli rozpoczęcie lub zakończenie okresu estymacji zostanie pominięte, użyta będzie wartość domyślna.

- Jeśli obserwacje są zdefiniowane przez określenie zmiennej daty/czasu, wówczas wartości rozpoczęcia i zakończenia okresu należy wprowadzić w takim samym formacie, jaki został użyty dla zmiennej daty/czasu.
- W przypadku obserwacji definiowanych na podstawie okresów cyklicznych należy określić wartość dla każdej zmiennej okresu cyklicznego. Każda zmienna jest wyświetlana w osobnej kolumnie.
- Jeśli w efekcie aktywnego zbioru danych znajduje się specyfikacja daty, wówczas należy określić wartość dla każdego składnika (np. Miesiąc) specyfikacji daty. Każdy składnik jest wyświetlany w osobnej kolumnie.
- Jeśli obserwacje są definiowane przez kolejność rekordów, rozpoczęcie i zakończenie okresu estymacji jest definiowane przez liczbę wierszy (wyświetlaną w edytorze danych) dla danej obserwacji.

Wyznaczony przez najwcześniejszy lub najpóźniejszy przedział czasowy

Definiuje okres estymacji jako określoną liczbę przedziałów czasowych, która rozpoczyna się od najwcześniejszego przedziału czasowego lub kończy na najpóźniejszym przedziale czasowym określonym w danych, z opcjonalnym przesunięciem. W tym kontekście przedział czasowy odnosi się do przedziału czasowego dla analizy. Przykładowo, założmy, że obserwacje są przeprowadzane miesięcznie, ale przedział czasowy dla analizy to kwartały. Określenie wartości **Najpóźniejszy** i wartości 24 dla opcji **Liczba przedziałów czasowych** będzie oznaczało ostatnie 24 kwartały.

Opcjonalnie można wykluczyć określoną liczbę przedziałów czasowych. Przykładowo, określenie ostatnich 24 przedziałów czasowych i 1 do wykluczenia oznacza, że okres oszacowania składa się z 24 przedziałów, które poprzedzają ostatni.

Prognoza

Opcja **Rozszerz rekordy na przedziały z przyszłości** pozwala ustawić liczbę przedziałów czasowych do prognozowania poza koniec okresu estymacji. Przedział czasowy jest w tym przypadku przedziałem czasowym dla analizy, określonym na karcie Specyfikacja danych. Z chwilą wywołania prognoz następuje automatyczna budowa modeli autoregresji dla szeregów wejściowych niebędących jednocześnie wartościami przewidywanymi. Modele te są następnie używane do generowania wartości dla tych szeregów wejściowych w okresie prognozy.

Zapisz

Opcje miejsca docelowego

Można zapisać zarówno transformacje danych (takie jak agregacja lub podstawienie brakujących wartości), jak i nowe zmienne (określone w ustawieniach **Zapisz zmienne przewidywane**) do pliku danych IBM SPSS Statistics lub nowego zbioru danych w bieżącej sesji. Wartości daty/czasu w zapisanych danych są wyrównane do początków poszczególnych przedziałów czasowych, np. pierwszych dni miesiąca, i reprezentują przedział czasu analizy systemu modelu. Wszelkie nowe zmienne można zapisywać w aktywnym zbiorze danych pod warunkiem, że obserwacje są zdefiniowane przez specyfikację daty lub kolejność rekordów, a dane nie są zagregowane.

Zapisz zmienne przewidywane

Można zapisać predykcje modelu, przedziały ufności i reszty jako nowe zmienne. Dla każdej zmiennej przewidywanej wybranej do zapisania generowany jest odrębny zbiór nowych zmiennych, a każda nowa zmienna zawiera wartości zarówno dla przedziałów estymacji, jak i prognozy. W przypadku wartości przewidywanych, przedziałów ufności i reszt szumu można określić trzon nazwy, który będzie przedrostkiem nazw nowych zmiennych. Cała nazwa zmiennej jest złączeniem trzonu nazwy oraz nazwy zmiennej zawierającej szereg przewidywany. Trzon nazwy musi być zgodny z regułami nazewnictwa zmiennych. Nazwa zmiennej jest w razie potrzeby uzupełniana, aby nie dochodziło do kolizji między nazwami.

Oznacz obserwacje zawierające prognozę

Tworzy zmienną wskazującą, czy rekord zawiera dane prognozy. Można określić nazwę zmiennej. Nazwa domyślna to *ForecastIndicator*.

Zapisywane zmienne przewidywane

Określa, czy nowe zmienne tworzone będą dla wszystkich szeregów przewidywanych w systemie modelu, czy tylko dla szeregów przewidywanych określonych w ustawieniach **Prezentacja szeregów**.

Eksportuj system modelu

Zapisuje system modelu w skompresowanym pliku archiwum (.zip). Pliku systemu modelu może użyć procedura Prognozowanie modelem przyczynowym szeregów czasowych, aby uzyskać aktualizowane prognozy lub wygenerować dostępny wynik. Plik może być również używany przez procedurę Scenariusze modeli przyczynowych szeregów czasowych do uruchamiania analizy scenariuszy.

Wyjście interaktywne

Wynik modelowania przyczynowego szeregów czasowych zawiera interaktywne obiekty wynikowe. Interaktywne właściwości są dostępne po aktywowaniu (dwukrotnym kliknięciu) obiektu wykresu w oknie wynikowym.

Ogólny model systemu

Wyświetla relacje przyczynowe między szeregami w systemie modelu. Wszystkie linie łączące konkretną zmienną przewidywaną z jej zmiennymi wejściowymi mają ten sam kolor. Grubość linii oznacza istotność związku przyczynowego, przy czym grubsze linie oznaczają bardziej istotny związek. Zmienne wejściowe, które nie są jednocześnie przewidywanymi, oznaczone są czarnym kwadratem.

- Istnieje możliwość wyświetlenia relacji najlepszych modeli, określonego szeregu, wszystkich szeregów lub modeli bez danych wejściowych. Najlepsze modele są tymi, które spełniają kryteria najlepiej dopasowanych modeli wprowadzone na stronie **Szereg do wyświetlenia**.
- Można wygenerować diagramy wpływu dla jednego lub wielu szeregów, wybierając nazwę szeregu w tabeli, klikając ją prawym przyciskiem myszy i wybierając z menu kontekstowego polecenie **Utwórz Diagram wpływu**.
- Istnieje możliwość ukrycia relacji przyczynowych, których poziom istotności jest większy od określonej wartości. Mniejszy poziom istotności oznacza bardziej znaczącą relację przyczynową.
- Można wyświetlić relacje dla konkretnego szeregu, wybierając w tabeli nazwę szeregu, klikając ją prawym przyciskiem i wybierając z menu kontekstowego polecenie **Podświetl relacje dla szeregu**.

Diagram wpływu

Przedstawia graficzną reprezentację relacji przyczynowych pomiędzy szeregami będącymi przedmiotem zainteresowania a innymi szeregami, na które one wpływają lub które mają na nie wpływ. Szeregi, które wpływają na szeregi będące przedmiotem zainteresowania, są nazywane *przyczynami*.

- Analizowany szereg można zmienić, wpisując nazwę nowego potrzebnego szeregu. Dwukrotne kliknięcie dowolnego węzła na diagramie wpływu zmieni analizowany szereg na ten, który jest powiązany z danym węzłem.
- Istnieje możliwość przetęczenia obrazu między przyczynami i skutkami oraz zmiany liczby wyświetlanych poziomów przyczyn lub skutków.
- Jednokrotne kliknięcie dowolnego węzła wyświetla szczegółowy diagram kolejności dla szeregu powiązanego z węzłem.

Analiza podstawowych przyczyn wartości odstających

Określa, które szeregi z największym prawdopodobieństwem będą powodowały powstawanie poszczególnych wartości odstających w szeregu będącym przedmiotem zainteresowania.

- Podstawową przyczynę dowolnej wartości odstającej można określić, wybierając jej wiersz w tabeli Wartości odstające. Można ją także wyświetlić, klikając ikonę wartości odstającej w wykresie sekwencji.
- Jednokrotne kliknięcie dowolnego węzła wyświetla szczegółowy diagram kolejności dla szeregu powiązanego z węzłem.

Ogólna jakość modelu

Histogram dopasowania modelu dla wszystkich modeli, w szczególności dla statystyki dopasowania. Kliknięcie słupka na wykresie słupkowym filtruje wykres punktowy tak, aby wyświetlane były wyłącznie modele powiązane z wybranym słupkiem. Model dla określonej serii docelowej na wykresie punktowym można znaleźć, określając nazwę szeregu.

Rozkład wartości odstających

Wykres słupkowy liczby wartości odstających dla wszystkich zmiennych przewidywanych i każdego przedziału czasowego w okresie oszacowania. Kliknięcie słupka na wykresie słupkowym filtruje wykres punktowy tak, aby wyświetlane były wyłącznie wartości odstające powiązane z wybranym słupkiem.

Uwaga:

- Jeśli przy zapisywaniu dokumentu wynikowego, który zawiera interaktywne wyniki modelowania przyczynowego szeregów czasowych, mają być zachowane właściwości interaktywne, to w oknie dialogowym Zapisz wynik jako musi być zaznaczona opcja **Składaj wymagane informacje o modelu z dokumentem wynikowym**.
- Niektóre właściwości interaktywne wymagają, by aktywnym zbiorem danych były dane używane do zbudowania systemu modelu przyczynowego szeregów czasowych.

Stosowanie modelowania przyczynowego szeregów czasowych

Stosowanie modelowania przyczynowego szeregów czasowych

Dostępne są dwie procedury służące do stosowania modeli utworzonych za pomocą procedury Modelowanie przyczynowe szeregów czasowych. Obie procedury wymagają pliku systemu modelu, który można zapisać w ramach procedury Modelowanie przyczynowe szeregów czasowych.

Prognozowanie modelem przyczynowym szeregów czasowych

Za pomocą tej procedury można uzyskać prognozy dla szeregów, dla których dostępne są nowe lub zmienione dane, bez konieczności ponownego tworzenia modeli. Można także wygenerować wszelkie wyniki dostępne w procedurze Modelowanie przyczynowe szeregów czasowych.

Scenariusze modelowania przyczynowego szeregów czasowych

Za pomocą tej procedury można sprawdzić, jak konkretne wartości konkretnego szeregu czasowego w systemie modelu wpływają na wartości przewidywane szeregu czasowego związane z nim przyczynowo.

Prognozowanie modelem przyczynowym szeregów czasowych

Procedura Prognozowanie modelem przyczynowym szeregów czasowych ładuje plik systemu modeli utworzony przez procedurę Modelowanie przyczynowe szeregów czasowych i stosuje modele do aktywnego zbioru danych. Za pomocą tej procedury można uzyskać prognozy dla szeregów, dla których dostępne są nowe lub zmienione dane, bez konieczności ponownego tworzenia modeli. Można także wygenerować wszelkie wyniki dostępne w procedurze Modelowanie przyczynowe szeregów czasowych.

Założenia

- Struktura danych w aktywnym zbiorze danych, kolumnowa albo wielowymiarowa, musi być taka sama, jak struktura używana na etapie tworzenia systemu modeli. W przypadku danych wielowymiarowych zmienne wymiarów muszą być takie same, jak te, których użyto do utworzenia systemu modeli. Ponadto wartości wymiarów użyte do tworzenia systemu modeli muszą istnieć w aktywnym zbiorze danych.
- Modele są stosowane do zmiennych w aktywnym zbiorze danych na podstawie nazw, tj. do zmiennych o tych samych nazwach, co określone w systemie modeli.
- Zmienne, które definiowały obserwacje podczas tworzenia systemu modeli, muszą istnieć w aktywnym zbiorze danych. Zakłada się, że przedział czasowy między obserwacjami jest taki sam, jak podczas tworzenia modeli. Jeśli obserwacje zostały zdefiniowane przez specyfikację daty, to ta sama

specyfikacja daty musi istnieć w aktywnym zestawie danych. Specyfikacje dat tworzy się w oknie dialogowym Definiowanie daty i czasu lub za pomocą komendy DATE.

- Przedział czasowy analizy i wszelkie ustawienia agregacji, rozkładu i brakujących wartości są takie same, jak przy tworzeniu modeli.

Aby zastosować prognozowanie modelem przyczynowym szeregów czasowych

Ta funkcja wymaga opcji Statistics Forecasting.

Z menu wybierz jedną z opcji:

Analizuj > Prognozowanie > Zastosuj modele tradycyjne...

1. Wprowadź specyfikację pliku dla pliku systemu modelu lub kliknij przycisk **Przeglądaj** i wybierz plik systemu modelu. Pliki systemu modelu są tworzone za pomocą procedury modelowania przyczynowego szeregów czasowych.
2. Kliknij opcję ponownej estymacji modeli, tworzenia prognoz i i generowania wyników.
3. Kliknij przycisk **Kontynuuj**.
4. Określ, czy mają być używane istniejące parametry modelu, czy też parametry modelu mają być estymowane ponownie na podstawie danych w aktywnym zbiorze danych.
5. Określ, jak daleko w przyszłość ma wybiegać prognoza, lub wybierz opcję rezygnacji z prognozowania.
6. Kliknij przycisk **Opcje**, aby określić zawartość wyników.
7. Kliknij przycisk **Zapisz**, aby zapisać predykcje i wyeksportować zaktualizowany model do pliku zewnętrznego po ponownej estymacji parametrów modelu.
8. Kliknij przycisk **Uruchom**, aby uruchomić procedurę.

Parametry modelu i prognozy

Łaładuj z pliku modelu

Prognozy generowane są na podstawie parametrów modelu z pliku systemu modelu oraz danych z aktywnego zbioru danych, bez ponownego szacowania parametrów modelu. Miary dobroci dopasowania prezentowane w wynikach i używane do wyboru najlepiej dopasowanych modeli pochodzą z pliku systemu modelu. Miary dopasowania odzwierciedlają wówczas dane używane na etapie tworzenia (lub ostatniej modyfikacji) modelu. Ta opcja jest odpowiednia do generowania prognoz i wyników na podstawie danych użytych do stworzenia systemu modelu.

Estymuj ponownie na podstawie danych

Parametry modelu są estymowane ponownie przy użyciu danych w aktywnym zbiorze danych. Ponowna estymacja parametrów modelu nie wpływa na to, które dane wejściowe są uwzględniane w modelu dla każdej ze zmiennych przewidywanych. Ta opcja jest odpowiednia, gdy dostępne są dane spoza pierwotnego okresu estymacji i pożądanym jest wygenerowanie prognoz lub innych wyników z uwzględnieniem zaktualizowanych danych.

Wszystkie obserwacje

Określa, że okres estymacji zaczyna się od czasu z najwcześniejszą obserwacją, a kończy w czasie z najpóźniejszą obserwacją we wszystkich szeregach.

Wyznaczony przez czas początkowy i końcowy

Można określić datę rozpoczęcia i zakończenia okresu estymacji lub można określić tylko datę rozpoczęcia lub tylko datę zakończenia. Jeśli rozpoczęcie lub zakończenie okresu estymacji zostanie pominięte, użyta będzie wartość domyślna.

- Jeśli obserwacje są zdefiniowane przez określenie zmiennej daty/czasu, wówczas wartości rozpoczęcia i zakończenia okresu należy wprowadzić w takim samym formacie, jaki został użyty dla zmiennej daty/czasu.

- W przypadku obserwacji definiowanych na podstawie okresów cyklicznych należy określić wartość dla każdej zmiennej okresu cyklicznego. Każda zmienna jest wyświetlana w osobnej kolumnie.
- Jeśli w efekcie aktywnego zbioru danych znajduje się specyfikacja daty, wówczas należy określić wartość dla każdego składnika (np. Miesiąc) specyfikacji daty. Każdy składnik jest wyświetlany w osobnej kolumnie.
- Jeśli obserwacje są definiowane przez kolejność rekordów, rozpoczęcie i zakończenie okresu estymacji jest definiowane przez liczbę wierszy (wyświetlaną w edytorze danych) dla danej obserwacji.

Wyznaczony przez najwcześniejszy lub najpóźniejszy przedział czasowy

Definiuje okres estymacji jako określoną liczbę przedziałów czasowych, która rozpoczyna się od najwcześniejszego przedziału czasowego lub kończy na najpóźniejszym przedziale czasowym określonym w danych, z opcjonalnym przesunięciem. W tym kontekście przedział czasowy odnosi się do przedziału czasowego dla analizy. Przykładowo, założmy, że obserwacje są przeprowadzane miesięcznie, ale przedział czasowy dla analizy to kwartały. Określenie wartości **Najpóźniejszy** i wartości 24 dla opcji **Liczba przedziałów czasowych** będzie oznaczało ostatnie 24 kwartały.

Opcjonalnie można wykluczyć określoną liczbę przedziałów czasowych. Przykładowo, określenie ostatnich 24 przedziałów czasowych i 1 do wykluczenia oznacza, że okres oszacowania składa się z 24 przedziałów, które poprzedzają ostatni.

Rozszerz rekordy na przedziały z przyszłości

Pozwala ustawić liczbę przedziałów czasowych do prognozowania poza koniec okresu estymacji. Przedział czasowy jest w tym przypadku przedziałem czasowym dla analizy. Z chwilą wywołania prognoz następuje automatyczna budowa modeli autoregresji dla szeregów wejściowych niebędących jednocześnie wartościami przewidywanymi. Modele te są następnie używane do generowania wartości dla tych szeregów wejściowych w okresie prognozy w celu uzyskania prognoz dla zmiennych przewidywanych tych szeregów.

Opcje ogólne

Szerokość przedziału ufności (%)

To ustawienie decyduje o przedziałach ufności dla prognoz i parametrów modelu. Można określić dowolną wartość dodatnią mniejszą od 100. Domyślnie ustawiony jest 95% przedział ufności.

Próg (%) dla wartości odstającej

Obserwacja jest oznaczana jako odstająca, jeśli obliczone wg modelu prawdopodobieństwo, że jest odstająca, przekracza wyznaczony próg. Można określić wartość z zakresu od 50 do 100.

Prezentacja szeregów

Te opcje określają szeregi (przewidywane lub wejściowe), dla których wyświetlany jest wynik. Zawartość wyniku dla określonego szeregu jest określana przez ustawienia **Opcje wyników**.

Przedstaw zmienne przewidywane w powiązaniu z modelami o najlepszym dopasowaniu

Domyślnie wynik jest wyświetlany dla zmiennych przewidywanych, które są powiązane z 10 najlepszymi modelami, określonymi na podstawie wartości R-kwadrat. Można określić inną stałą liczbę najlepszych modeli lub podać odestek tych modeli. Można również wybrać jedną z następujących miar dopasowania:

R-kwadrat

Miara dobroci dopasowania modelu liniowego, czasami nazywana współczynnikiem determinacji. Jest to część zmienności w zmiennej przewidywanej wyjaśniona przez model. Przyjmuje wartość z zakresu od 0 do 1. Małe wartości wskazują, że model nie pasuje dobrze do danych.

Pierwiastek z procentowego błędu średniokwadratowego

Miara wielkości przewidywanych wartości modelu różni się od obserwowanych wartości szeregu. Jest niezależna od używanych jednostek i tym samym może być używana do porównywania szeregów używających różnych jednostek.

Pierwiastek błędu średniokwadratowego

Pierwiastek kwadratowy obliczany z przeciętnego odchylenia kwadratowego. Mierzy, jak bardzo szereg zależny odbiega od poziomu przewidywanego przez model; miara wyrażona w jednostkach używanych przez szereg zależny.

BIC

Bayesowskie kryterium informacyjne. Miara wyboru i porównywania modeli w oparciu o -2 zredukowane logarytm wiarygodności. Mniejsze wartości oznaczają lepszy model. Wartość BIC „karze” także modele przeparametryzowane (na przykład złożone modele z dużą liczbą danych wejściowych), jednak silniej niż miara AIC.

AIC

Kryterium informacyjne Akaike. Miara wyboru i porównywania modeli w oparciu o -2 zredukowane logarytm wiarygodności. Mniejsze wartości oznaczają lepszy model. Wartość AIC „karze” modele przeparametryzowane (na przykład złożone modele z dużą liczbą danych wejściowych).

Określ poszczególne szeregi

Można określić poszczególne szeregi, dla których potrzebny jest wynik.

- W przypadku danych kolumnowych należy określić zmienne, które zawierają wymagane szeregi. Kolejność określonych zmiennych definiuje kolejność, w jakiej są wyświetlane w wyniku.
- W przypadku danych wielowymiarowych należy określić konkretny szereg, dodając wpis do siatki dla zmiennej metryki, która zawiera szereg. Następnie należy określić wartości zmiennych wymiaru, które definiują szereg.
 - Można wprowadzić wartość dla każdej zmiennej wymiaru bezpośrednio do siatki lub wybrać ją z listy dostępnych wartości wymiarów. Aby wybrać wartość z listy, należy kliknąć przycisk wielokropka w komórce wybranego wymiaru. Ta czynność otworzy podrzędne okno dialogowe Wybierz zmienną wymiaru.
 - Listę wartości wymiaru w podrzędnym oknie dialogowym Wybierz zmienną wymiaru można przeszukać, klikając ikonę lornetki i wprowadzając poszukiwany termin. Spacje są traktowane jako część poszukiwanego terminu. Gwiazdka (*) w poszukiwanym terminie nie oznacza symbolu wieloznacznego.
 - Kolejność szeregów w siatce definiuje kolejność, w jakiej są wyświetlane w wyniku.

Wynik dla danych kolumnowych i danych wielowymiarowych jest ograniczony do 30 szeregów. Ten limit uwzględnia poszczególne szeregi (wejściowe i przewidywane) określone przez użytkownika i przewidywane powiązane z najlepszymi modelami. Szeregi określone przez użytkownika mają pierwszeństwo w stosunku do przewidywanych powiązanych z najlepszymi modelami.

Opcje wyników

Te opcje określają zawartość wyniku. Opcje z grupy **Wyniki dla zmiennych przewidywanych** generują wynik dla zmiennych przewidywanych, które są powiązane z najlepszymi modelami w ustawieniach **Prezentacja szeregów**. Opcje z grupy **Wyniki dla szeregów czasowych** generują wyniki dla poszczególnych szeregów, które zostały określone w ustawieniach **Prezentacja szeregów**.

Ogólny model systemu

Wyświetla graficzną reprezentację relacji przyczynowych pomiędzy szeregami w systemie modelu. Tabele statystyk dopasowania modelu i wartości odstających dla wyświetlanych przewidywanych są uwzględniane jako część element wyniku. Jeśli ta opcja zostanie wybrana w grupie **Wyniki dla szeregów czasowych**, tworzony jest osobny element wyniku dla każdego pojedynczego szeregu, jaki został określony w ustawieniach **Prezentacja szeregów**.

Relacje przyczynowe pomiędzy szeregami mają przypisany poziom istotności, gdzie niższy poziom istotności oznacza bardziej istotny związek. Istnieje możliwość ukrycia relacji, których poziom istotności jest większy od określonej wartości.

Statystyki dopasowania modelu i wartości odstające

Tabele statystyk dopasowania modelu i wartości odstających dla szeregów przewidywanych, jakie zostały wybrane do wyświetlenia. Tabele te zawierają te same informacje, co tabele w wizualizacji

Ogólny system modelu. Obsługują one wszystkie standardowe funkcje przestawiania i edytowania tabel.

Efekty i parametry modelu

Tabele testów efektów modelu i parametrów modelu dla szeregów przewidywanych, jakie zostały wybrane do wyświetlenia. Testy efektów modelu obejmują statystykę F i powiązaną wartość istotności dla każdej zmiennej wejściowej uwzględnianej w modelu.

Diagram wpływu

Przedstawia graficzną reprezentację relacji przyczynowych pomiędzy szeregami będącymi przedmiotem zainteresowania a innymi szeregami, na które one wpływają lub które mają na nie wpływ. Szeregi, które wpływają na szeregi będące przedmiotem zainteresowania, są nazywane *przyczynami*. Wybór opcji **Efekty** powoduje wygenerowanie diagramu wpływu, który jest inicjowany w celu wyświetlenia efektów. Wybór opcji **Przyczyny** powoduje wygenerowanie diagramu wpływu, który jest inicjowany w celu wyświetlenia przyczyn. Wybór opcji **Przyczyny i skutki** powoduje wygenerowanie dwóch osobnych diagramów wpływu; jeden jest inicjowany dla przyczyn, a drugi dla efektów. Przyczyny i efekty można w sposób interaktywny przełączać w elemencie wyniku, który wyświetla diagram wpływu.

Można określić liczbę poziomów przyczyn i efektów do wyświetlenia, gdzie pierwszy poziom stanowi szereg będący przedmiotem zainteresowania. Każdy dodatkowy poziom przedstawia bardziej pośrednie przyczyny lub efekty szeregu będącego przedmiotem zainteresowania. Przykładowo, trzeci poziom wyświetlanych efektów składa się z szeregów z drugiego poziomu stanowiących bezpośrednie dane wejściowe. Na szeregi z trzeciego poziomu szeregi będące przedmiotem zainteresowania mają więc pośredni wpływ, ponieważ szeregi te stanowią bezpośrednie dane wejściowe dla szeregów z drugiego poziomu.

Wykres szeregu

Wykresy obserwowanych i przewidywanych wartości dla szeregów przewidywanych, które zostały wybrane do wyświetlenia. Jeśli żądanie obejmowało prognozy, wykres przedstawia również wartości prognozowane i przedziały ufności dla prognoz.

Wykres reszt

Wykresy reszt modelu dla szeregów przewidywanych, jakie zostały wybrane do wyświetlenia.

Najlepsze zmienne wejściowe

Wykresy poszczególnych wyświetlanych przewidywanych w czasie wraz z 3 najlepszymi zmiennymi wejściowymi dla przewidywanych. Najlepsze zmienne wejściowe to zmienne z najniższą wartością istotności. W celu dostosowania różnych skal zmiennych wejściowych i przewidywanych na osi y prezentowane są oceny z z każdego szeregu.

Tabela prognoz

Tabele zawierające przewidziane wartości i przedziały ufności dla tych prognoz określane dla przewidywanych szeregów, jakie zostały wybrane do wyświetlenia.

Analiza podstawowych przyczyn wartości odstających

Określa, które szeregi z największym prawdopodobieństwem będą powodowały powstawanie poszczególnych wartości odstających w szeregu będącym przedmiotem zainteresowania. Analiza podstawowych przyczyn wartości odstających jest przeprowadzana dla każdego szeregu przewidywanego, który jest uwzględniony na liście poszczególnych szeregów w ustawieniach **Prezentacja szeregów**.

Wyniki

Interaktywna tabela i wykres wartości odstających

Tabela i wykres odstających wartości oraz podstawowych przyczyn powstawania tych wartości dla każdego szeregu będącego przedmiotem zainteresowania. Każda wartość jest zapisywana w tabeli w osobnym wierszu. Wykres ten przedstawia diagramu wpływu. Wybór wiersza w tabeli powoduje wyróżnienie ścieżki (na diagramie wpływu) prowadzącej od szeregu będącego przedmiotem zainteresowania do szeregu, który z największym prawdopodobieństwem był przyczyną powstania powiązanej wartości odstającej.

Tabela przestawna wartości odstających

Tabela odstających wartości oraz głównych przyczyn powstawania tych wartości dla każdego szeregu będącego przedmiotem zainteresowania. Ta tabela zawiera te same informacje, co tabela na ekranie interaktywnym. Tabela ta obsługuje wszystkie standardowe funkcje przestawiania i edytowania tabel.

Poziomy przyczynowe

Użytkownik może określić liczbę poziomów, jakie będą uwzględnione w wyszukiwaniu podstawowych przyczyn. Informacje na temat poziomów, jakie są tutaj używane, są takie same, jak w przypadku diagramów wpływu.

Dopasowanie wszystkich modeli

Histogram dopasowania modelu dla wszystkich modeli i wybranych statystyk dopasowania. Dostępne są następujące statystyki dopasowania:

R-kwadrat

Miara dobroci dopasowania modelu liniowego, czasami nazywana współczynnikiem determinacji. Jest to część zmienności w zmiennej przewidywanej wyjaśniona przez model. Przyjmuje wartość z zakresu od 0 do 1. Małe wartości wskazują, że model nie pasuje dobrze do danych.

Pierwiastek z procentowego błędu średniokwadratowego

Miara wielkości przewidywanych wartości modelu różni się od obserwowanych wartości szeregu. Jest niezależna od używanych jednostek i tym samym może być używana do porównywania szeregów używających różnych jednostek.

Pierwiastek błędu średniokwadratowego

Pierwiastek kwadratowy obliczany z przeciętnego odchylenia kwadratowego. Mierzy, jak bardzo szereg zależny odbiega od poziomu przewidywanego przez model; miara wyrażona w jednostkach używanych przez szereg zależny.

BIC

Bayesowskie kryterium informacyjne. Miara wyboru i porównywania modeli w oparciu o -2 zredukowane logarytm wiarygodności. Mniejsze wartości oznaczają lepszy model. Wartość BIC „karze” także modele przeparametryzowane (na przykład złożone modele z dużą liczbą danych wejściowych), jednak silniej niż miara AIC.

AIC

Kryterium informacyjne Akaike. Miara wyboru i porównywania modeli w oparciu o -2 zredukowane logarytm wiarygodności. Mniejsze wartości oznaczają lepszy model. Wartość AIC „karze” modele przeparametryzowane (na przykład złożone modele z dużą liczbą danych wejściowych).

Wartości odstające w czasie

Wykres słupkowy liczby wartości odstających dla wszystkich zmiennych przewidywanych i każdego przedziału czasowego w okresie oszacowania.

Transformacje szeregów

Tabela wszystkich transformacji, jakie zostały przeprowadzone dla szeregu w systemie modelu. Możliwe transformacje to podstawianie brakujących wartości, agregacja i rozkład.

Zapisz

Zapisz zmienne przewidywane

Można zapisać predykcje modelu, przedziały ufności i reszty jako nowe zmienne. Dla każdej zmiennej przewidywanej wybranej do zapisania generowany jest odrębny zbiór nowych zmiennych, a każda nowa zmienna zawiera wartości zarówno dla przedziałów estymacji, jak i prognozy. W przypadku wartości przewidywanych, przedziałów ufności i reszt szumu można określić trzon nazwy, który będzie przedrostkiem nazw nowych zmiennych. Cała nazwa zmiennej jest złączeniem trzonu nazwy oraz nazwy zmiennej zawierającej szereg przewidywany. Trzon nazwy musi być zgodny z regułami nazewnictwa zmiennych. Nazwa zmiennej jest w razie potrzeby uzupełniana, aby nie dochodziło do kolizji między nazwami.

Oznacz obserwacje zawierające prognozę

Tworzy zmienną wskazującą, czy rekord zawiera dane prognozy. Można określić nazwę zmiennej. Nazwa domyślna to *ForecastIndicator*.

Zapisywane zmienne przewidywane

Określa, czy nowe zmienne tworzone będą dla wszystkich szeregów przewidywanych w systemie modelu, czy tylko dla szeregów przewidywanych określonych w ustawieniach **Prezentacja szeregów**.

Opcje miejsca docelowego

Można zapisać zarówno transformacje danych (takie jak agregacja lub podstawienie brakujących wartości), jak i nowe zmienne (określone w ustawieniach **Zapisz zmienne przewidywane**) do pliku danych IBM SPSS Statistics lub nowego zbioru danych w bieżącej sesji. Wartości daty/ czasu w zapisanych danych są wyrównane do początków poszczególnych przedziałów czasowych, np. pierwszych dni miesiąca, i reprezentują przedział czasu analizy systemu modelu. Wszelkie nowe zmienne można zapisywać w aktywnym zbiorze danych pod warunkiem, że obserwacje są zdefiniowane przez specyfikację daty lub kolejność rekordów, a dane nie są zagregowane.

Eksportuj system modelu

Zapisuje system modelu w skompresowanym pliku archiwum (.zip). Plik systemu modelu można później wykorzystać w tej procedurze. Plik ten może być również używany przez procedurę Scenariusze modeli przyczynowych szeregów czasowych do uruchamiania analizy scenariuszy. Ta opcja jest dostępna tylko w przypadku ponownej estymacji parametrów modelu.

Scenariusze modelowania przyczynowego szeregów czasowych

Procedura Scenariusze modelowania przyczynowego szeregów czasowych uruchamia zdefiniowane przez użytkownika scenariusze dla systemu modelowania przyczynowego szeregów czasowych wraz z danymi z aktywnego zbioru danych. Zmienna *scenario* jest definiowana przez szereg czasowy będący szeregiem źródłowym (*root series*) oraz przez zestaw zdefiniowanych przez użytkownika wartości dla tego szeregu w podanym zakresie czasu. Podane wartości są następnie używane do generowania predykcji dla szeregów czasowych, na które wpływa szereg źródłowy. Procedura wymaga pliku systemu modelu utworzonego w procedurze Modelowanie przyczynowe szeregów czasowych. Zakłada się, że aktywny zbiór danych to te same dane, które były używane do utworzenia pliku systemu modelu.

Przykład

Korzystając z procedury modelowania przyczynowego szeregów czasowych, osoba odpowiedzialna za decyzje biznesowe zidentyfikowała kluczową metrykę wpływającą na pewną liczbę ważnych wskaźników wydajności. Metryka poddaje się kontroli, dlatego osoba odpowiedzialna za decyzje zamierza zbadać wpływ różnych zestawów wartości metryki na dane w następnym kwartale. Można to łatwo zrobić, wczytując plik systemu modelu do procedury modelowania przyczynowego szeregów czasowych i określając zestaw wartości dla kluczowej metryki.

Aby uruchomić scenariusze modelowania przyczynowego szeregów czasowych

Ta funkcja wymaga opcji Statistics Forecasting.

Z menu wybierz jedną z opcji:

Analizuj > Prognozowanie > Zastosuj modele tradycyjne...

1. Wprowadź specyfikację pliku dla pliku systemu modelu lub kliknij przycisk **Przeglądaj** i wybierz plik systemu modelu. Pliki systemu modelu są tworzone za pomocą procedury modelowania przyczynowego szeregów czasowych.
2. Kliknij opcję, aby uruchomić scenariusze.
3. Kliknij przycisk **Kontynuuj**.
4. Na karcie Scenariusze (w oknie dialogowym Scenariusze modelowania przyczynowego szeregów czasowych) kliknij opcję **Zdefiniuj okres scenariusza** i wskaż okres scenariusza.
5. W przypadku danych kolumnowych kliknij przycisk **Dodaj scenariusz**, aby zdefiniować każdy ze scenariuszy. W przypadku danych wielowymiarowych kliknij przycisk **Dodaj scenariusz**, aby

zdefiniować każdy ze scenariuszy, a następnie kliknij opcję **Dodaj grupę scenariuszy**, aby zdefiniować każdą z grup scenariuszy.

6. Kliknij **Opcje**, aby wskazać zawartość danych wynikowych oraz określić zasięg szeregu, na który będzie wpływał scenariusz.

7. Kliknij przycisk **Uruchom**, aby uruchomić procedurę.

Definiowanie okresu scenariusza

Okres scenariusza to okres, w którym użytkownik określa wartości używane do uruchamiania scenariuszy. Powinien on rozpoczynać się przed okresem estymacji lub po jego zakończeniu. Opcjonalnie można określić predykcję poza koniec okresu scenariusza. Domyślnie predykcje są generowane po zakończeniu okresu scenariusza. Wszystkie scenariusze korzystają z tego samego okresu scenariusza oraz ze specyfikacji określających, jak daleko ma sięgać predykcja.

Uwaga: Predykcje rozpoczynają się w pierwszym okresie czasu po rozpoczęciu okresu scenariusza. Na przykład, jeśli okres scenariusza rozpoczyna się z datą 2014-11-01, zaś przedział czasu jest mierzony w miesiącach, wówczas pierwsza predykcja ma datę 2014-12-01.

Wyznaczony przez początek, koniec i czas predykcji

- Jeśli obserwacje są zdefiniowane przez określenie zmiennej daty/czasu, wówczas wartości rozpoczęcia, zakończenia i predykcji należy wprowadzić w takim samym formacie, jaki został użyty dla zmiennej daty/czasu. Wartości zmiennych daty/czasu są dopasowane do rozpoczęcia powiązanego przedziału czasowego. Na przykład, jeśli przedział czasowy dla analizy to miesiące, wówczas wartość 10/10/2014 jest korygowana na 10/01/2014, co oznacza rozpoczęcie miesiąca.
- W przypadku obserwacji definiowanych na podstawie okresów cyklicznych należy określić wartość dla każdej zmiennej okresu cyklicznego. Każda zmienna jest wyświetlana w osobnej kolumnie.
- Jeśli w efekcie aktywnego zbioru danych znajduje się specyfikacja daty, wówczas należy określić wartość dla każdego składnika (np. Miesiąc) specyfikacji daty. Każdy składnik jest wyświetlany w osobnej kolumnie.
- Jeśli obserwacje są definiowane przez kolejność rekordów, rozpoczęcie, zakończenie i predykcja są definiowane przez liczbę wierszy (wyświetlaną w edytorze danych) dla danej obserwacji.

Wyznaczony przez przedziały zależne od końca okresu estymacji

Definiuje rozpoczęcie i zakończenie w odniesieniu do liczby przedziałów czasowych odpowiadających końcowi okresu estymacji, gdzie przedział czasowy jest przedziałem czasowym dla analizy. Koniec okresu estymacji jest zdefiniowany jako przedział czasowy 0. Przedziały czasowe przed końcem okresu estymacji mają wartości ujemne, a przedziały czasowe po końcu okresu estymacji mają wartości dodatnie. Można również określić liczbę przedziałów dla predykcji po zakończeniu okresu scenariusza. Domyślną wartością jest 0.

Założmy na przykład, że przedział czasowy dla analizy to miesiące, jako początek przedziału określono wartość 1, wartość 3 dla zakończenia przedziału oraz wartość 1 dla określenia, jak daleko poza koniec okresu scenariusza ma sięgać predykcja. Okres scenariusza jest wówczas ustawiony na 3 miesiące po zakończeniu okresu estymacji. Predykcje są generowane dla drugiego i trzeciego miesiąca okresu scenariusza i dla jeszcze 1 miesiąca poza koniec okresu scenariusza.

Dodawanie scenariuszy i grup scenariuszy

Na karcie Scenariusze określone są scenariusze, jakie mają zostać uruchomione. Aby zdefiniować scenariusz, konieczne jest zdefiniowanie najpierw okresu scenariusza poprzez kliknięcie przycisku **Zdefiniuj okres scenariusza**. Scenariusze i grupy scenariuszy (dotyczy tylko danych wielowymiarowych) są tworzone poprzez kliknięcie powiązanego przycisku **Dodaj scenariusz** lub **Dodaj grupę scenariuszy**. Wybierając określony scenariusz lub grupę scenariuszy w powiązanej siatce, można je edytować, utworzyć kopię lub usunąć.

Dane kolumnowe

W kolumnie **Zmienna źródłowa** w siatce określone są zmienne szeregów czasowych, których wartości są zastępowane przez wartości scenariusza. W kolumnie **Wartości scenariusza** wyświetlane są określone wartości scenariusza w kolejności od najwcześniejszych do najpóźniejszych. Jeśli wartości scenariusza są zdefiniowane przez wyrażenie, wówczas w kolumnie wyświetlane jest wyrażenie.

Dane wielowymiarowe

Indywidualne scenariusze

W każdym wierszu w siatce Indywidualne scenariusze określone są szeregi czasowe, których wartości są zastępowane przez określone wartości scenariusza. Szereg jest definiowany przez kombinację zmiennej, jaka została określona w kolumnie **Metryka źródłowa** oraz określonej wartości dla każdej zmiennej wymiaru. Zawartość kolumny **Wartości scenariusza** jest taka sama, jak w przypadku danych kolumnowych.

Grupy scenariuszy

Grupa scenariusza definiuje zestaw scenariuszy, które zostały utworzone w oparciu o jedną zmienną metryki źródłowej i wiele zbiorów wartości wymiaru. Każdy zbiór wartości wymiaru (jedna wartość dla zmiennej wymiaru) dla określonej zmiennej metryki definiuje szereg czasowy. Pojedynczy scenariusz jest wówczas generowany dla każdego szeregu czasowego, którego wartości są następnie zastępowane przez wartości scenariusza. Wartości scenariusza dla grupy scenariuszy są określone przez wyrażenie, które jest następnie stosowane do każdego szeregu czasowego w grupie.

W kolumnie **Liczba szeregów** wyświetlana jest liczba zbiorów wartości wymiarów, które są powiązane z grupą scenariuszy. Wyświetlana wartość może być większa niż rzeczywista liczba szeregów, które są powiązane z grupą scenariusza (jeden szereg dla zbioru). Taka sytuacja ma miejsce, jeśli niektóre z określonych kombinacji wartości wymiarów nie odpowiadają szeregom zawartym w metryce źródłowej dla grupy.

Przykładem grupy scenariusza może być zmienna metryki *advertising* oraz dwie zmienne wymiarów *region* i *brand*. Można zdefiniować grupę scenariusza dla zmiennej *advertising* stanowiącej metrykę źródłową i obejmuje wszystkie kombinacje z innych *region* i *brand*. Następnie można określić wartość $\text{advertising} \times 1.2$ jako wyrażenie do zbadania efektu zwiększenia wartości *advertising* o 20 procent dla wszystkich szeregów czasowych, jakie są powiązane ze zmienną *advertising*. Jeśli istnieją 4 wartości *region* i 2 wartości *brand*, wówczas istnieje 8 takich szeregów czasowych i 8 scenariuszy zdefiniowanych przez grupę.

Definicja scenariusza

Ustawienia dla definiowania scenariusza zależą od tego, czy dane są danymi kolumnowymi czy wielowymiarowymi.

Szereg źródłowy

Określa szereg początkowy dla scenariusza. Każdy scenariusz jest oparty na jednym szeregu początkowym. W przypadku danych kolumnowych wybierana jest zmienna, która definiuje szereg początkowy. W przypadku danych wielowymiarowych należy określić szereg początkowy, dodając wpis do siatki dla zmiennej metryki, która zawiera szereg. Następnie należy określić wartości zmiennych wymiaru, które definiują szereg źródłowy. Następujące opcje mają zastosowanie podczas określania wartości wymiarów:

- Można wprowadzić wartość dla każdej zmiennej wymiaru bezpośrednio do siatki lub wybrać ją z listy dostępnych wartości wymiarów. Aby wybrać wartość z listy, należy kliknąć przycisk wielokropka w komórce wybranego wymiaru. Ta czynność otworzy podrzędne okno dialogowe Wybierz zmienną wymiaru.
- Listę wartości wymiaru w podrzędnym oknie dialogowym Wybierz zmienną wymiaru można przeszukać, klikając ikonę lornetki i wprowadzając poszukiwany termin. Spacje są traktowane jako część poszukiwanego terminu. Gwiazdka (*) w poszukiwanym terminie nie oznacza symbolu wieloznacznego.

Określ docelowe zmienne przewidywane

Tej opcji należy użyć, jeśli znane są konkretne zmienne przewidywane, na które wpływają szeregi czasowe oraz użytkownik zamierza zbadać efekty tylko dla tych zmiennych przewidywanych. Domyślnie zmienne przewidywane, na które wpływa szereg początkowy, są ustalane automatycznie. Można określić zasięg szeregu, na który ma wpływ scenariusz, używając opcji na karcie Opcje.

W przypadku danych kolumnowych należy wybrać dowolne zmienne przewidywane. W przypadku danych wielowymiarowych należy określić szereg przewidywany, dodając wpis do siatki dla przewidywanej zmiennej metryki, która zawiera szereg. Domyślnie uwzględniane są wszystkie szeregi, które są zawarte w określonej zmiennej metryki. Można dostosować zbiór uwzględnionych szeregów poprzez dostosowanie wartości dla co najmniej jednej zmiennej wymiaru. Aby dostosować uwzględnione wartości wymiaru, należy kliknąć przycisk wielokropka dla wybranego wymiaru. Ta czynność otworzy okno dialogowe Wybierz wartości wymiarów.

W kolumnie **Liczebność szeregu** (dla danych wielowymiarowych) wyświetlana jest liczba zestawów wartości wymiarów, jakie są aktualnie określone dla powiązanej metryki przewidywanej. Wyświetlana wartość może być większa niż rzeczywista liczba szeregów przewidywanych (jeden szereg na zestaw). Taka sytuacja ma miejsce, jeśli niektóre z określonych kombinacji wartości wymiarów nie odpowiadają szeregom objętym przez powiązaną metrykę przewidywaną.

Scenariusz 1

Każdy scenariusz musi mieć unikalny identyfikator. Identyfikator jest wyświetlany w wyniku, który jest powiązany ze scenariuszem. Nie ma żadnych ograniczeń, poza unikalnością, dla wartości identyfikatora.

Określ wartości scenariusza dla szeregu źródłowego

Tej opcji należy użyć, aby określić jawne wartości dla szeregów początkowych w okresie scenariusza. Należy określić wartość liczbową dla każdego przedziału czasowego, jaki został wymieniony w siatce. Można uzyskać wartości szeregów początkowych (rzeczywistych i przewidywanych) dla każdego przedziału z okresu scenariusza, klikając opcje **Odczyt**, **Prognoza** lub **Odczyt\Prognoza**.

Określ wyrażenie dla wartości scenariusza dla szeregu źródłowego

Można zdefiniować wyrażenie do obliczania wartości dla szeregów początkowych w okresie scenariusza. Można wprowadzić wyrażenie bezpośrednio lub kliknąć przycisk kalkulatora i utworzyć wyrażenie za pomocą konstruktora wyrażeń wartości scenariusza.

- Wyrażenie może zawierać dowolną zmienną przewidywaną lub przewidywaną w systemie modelu.
- Jeśli okres scenariusza będzie rozszerzony poza istniejące dane, wówczas wyrażenie jest stosowane do wartości prognozowanych dla zmiennych z wyrażenia.
- W przypadku danych wielowymiarowych każda zmienna w wyrażeniu określa szereg czasowy, który jest definiowany przez zmienną oraz wartości wymiaru, jakie zostały określone dla metryki źródłowej. Są to szeregi czasowe, które są używane do oszacowania wyrażenia.

Założmy na przykład, że zmienną źródłową jest *advertising*, a wyrażenie to $advertising * 1.2$. Wartości dla *advertising*, jakie zostały użyte w scenariuszu, reprezentują 20 procent wzrostu wobec istniejących wartości.

Uwaga: Scenariusze są tworzone poprzez kliknięcie **Dodaj scenariusz** na karcie Scenariusze.

Konstruktor wyrażeń wartości scenariusza

Użyj opcji Scenariusze modelowania przyczynowego szeregów czasowych do utworzenia wyrażenia, które służy do obliczania wartości scenariusza dla pojedynczego scenariusza lub grupy scenariuszy. Aby utworzyć wyrażenie, można w pole **Wyrażenie** wkleić składniki lub wpisać je w polu **Wyrażenie** bezpośrednio.

- Wyrażenie może zawierać dowolną zmienną przewidywaną lub przewidywaną w systemie modelu.
- Możesz wkleić zmienne poprzez wybranie grupy z grupy z listy grup Funkcji i dwukrotne kliknięcie funkcji lub zmiennej na liście Funkcji (lub wybrać funkcję lub zmienną i kliknąć przycisk ze strzałką znajdujący się obok listy funkcji grupy). Wprowadź wszystkie parametry oznaczone znakami zapytania. Funkcja grupy o nazwie **Wszystkie** zapewnia listę wszystkich dostępnych funkcji. Krótki opis aktualnie wybranej funkcji jest wyświetlony w zarezerwowanym obszarze okna dialogowego.

- Stałe łańcuchowe muszą być ujęte w cudzysłów.
- Jeśli wartości zawierają dziesiętne, do wskazania dziesiętnej należy użyć kropki (.).

Aby uzyskać dostęp do opcji Konstruktor wyrażeń wartości scenariusza, należy kliknąć przycisk kalkulatora w oknie dialogowym Definicja scenariusza lub Definicja grupy scenariuszy.

Wybierz wartości wymiarów

W przypadku danych wielowymiarowych można dostosować wartości wymiarów, które definiują zmienne przewidywane, na które wpływa scenariusz lub grupa scenariuszy. Można również dostosować wartości wymiarów, które definiują zbiór szeregów początkowych dla grupy scenariusza.

Wszystkie wartości

Ta opcja określa, że uwzględniane są wszystkie wartości bieżącej zmiennej wymiaru. Jest to opcja domyślna.

Wybierz wartości

Tej opcji należy użyć, aby określić zbiory wartości dla bieżącej zmiennej wymiaru. Wartości, z których można dokonywać wyboru, można filtrować. Wartości, które spełniają warunki filtrowania, są wyświetlane na karcie **Dopasowane**, a wartości, które nie spełniają warunków filtrowania, są wyświetlane na karcie **Niedopasowane** w postaci listy **Niewybrane wartości**. Karta **Wszystkie** zawiera listę wszystkich niewybranych wartości, niezależnie od warunków filtrowania.

- Podczas określania filtru można użyć gwiazdki (*) jako symbolu wieloznacznego.
- Aby wyczyścić bieżący filtr, dla poszukiwanego terminu w oknie dialogowym Filtruj wyświetlane wartości należy wprowadzić pustą wartość.

Aby dostosować wartości wymiaru dla zmienionych zmiennych przewidywanych:

1. W oknie dialogowym Definicja scenariusza lub Definicja grupy scenariuszy należy wybrać metrykę przewidywaną, dla której wartości wymiaru mają zostać dostosowane.
2. Należy kliknąć przycisk wielokropka w kolumnie wymiaru, jaki ma być dostosowany.

Aby dostosować wartości wymiaru dla szeregu początkowego w grupie scenariuszy:

1. W oknie dialogowym Definicja grupy scenariuszy należy kliknąć przycisk wielokropka (w siatce szeregów początkowych) dla wymiaru, jaki ma być dostosowany.

Definicja grupy scenariuszy

Szereg początkowy

Określa zbiór szeregów początkowych dla grupy scenariuszy. Pojedynczy scenariusz jest generowany dla każdego szeregu czasowego w zbiorze. Należy określić szereg początkowy, dodając wpis do siatki dla zmiennej metryki, która zawiera wybrany szereg. Następnie należy określić wartości zmiennych wymiaru, które definiują zbiór. Domyślnie uwzględniane są wszystkie szeregi, które są zawarte w określonej zmiennej metryki źródłowej. Można dostosować zbiór uwzględnionych szeregów poprzez dostosowanie wartości dla co najmniej jednej zmiennej wymiaru. Aby dostosować uwzględnione wartości wymiaru, należy kliknąć przycisk wielokropka dla wybranego wymiaru. Ta czynność otworzy okno dialogowe Wybierz wartości wymiarów.

W kolumnie **Liczebność szeregu** wyświetlana jest liczba zestawów wartości wymiarów, jakie są aktualnie uwzględniane dla powiązanej metryki źródłowej. Wyświetlana wartość może być większa niż rzeczywista liczba szeregów dla grupy scenariusza (jeden szereg dla zbioru). Taka sytuacja ma miejsce, jeśli niektóre z określonych kombinacji wartości wymiarów nie odpowiadają szeregom objętym przez metrykę źródłową.

Określ zmieniony szereg przewidywany

Tej opcji należy użyć, jeśli znane są konkretne zmienne przewidywane, na które wpływa zbiór szeregów początkowych oraz użytkownik zamierza zbadać efekty tylko dla tych zmiennych przewidywanych. Domyślnie zmienne przewidywane, na które wpływa każdy szereg początkowy, są ustalane automatycznie. Można określić zasięg szeregu, na który ma wpływ każdy pojedynczy scenariusz, używając opcji na karcie Opcje.

Należy określić szereg przewidywany, dodając wpis do siatki dla przewidywanej zmiennej metryki, która zawiera szereg. Domyślnie uwzględniane są wszystkie szeregi, które są zawarte w określonej zmiennej metryki. Można dostosować zbiór uwzględnionych szeregów poprzez dostosowanie wartości dla co najmniej jednej zmiennej wymiaru. Aby dostosować uwzględnione wartości wymiaru, należy kliknąć przycisk wielokropka dla wybranego wymiaru. Ta czynność otworzy okno dialogowe Wybierz wartości wymiarów.

W kolumnie **Liczebność szeregu** wyświetlana jest liczba zestawów wartości wymiarów, jakie są aktualnie określone dla powiązanej metryki przewidywanej. Wyświetlana wartość może być większa niż rzeczywista liczba szeregów przewidywanych (jeden szereg na zestaw). Taka sytuacja ma miejsce, jeśli niektóre z określonych kombinacji wartości wymiarów nie odpowiadają szeregom objętym przez powiązaną metrykę przewidywaną.

Przedrostek identyfikatora scenariusza

Każda grupa scenariuszy musi mieć unikalny przedrostek. Przedrostek służy do budowania identyfikatora, jaki jest wyświetlany w wyniku powiązaniem z poszczególnymi pojedynczymi scenariuszami w grupie scenariuszy. Identyfikator dla pojedynczego scenariusza składa się z przedrostka, po którym znajduje się znak podkreślenia, a następnie wartość zmiennej wymiaru, która identyfikuje szereg początkowy. Wartości wymiarów są oddzielane znakami podkreślenia. Nie ma żadnych ograniczeń, poza unikalnością, dla wartości przedrostka.

Wyrażenie dla wartości scenariusza dla szeregu początkowego

Wartości scenariusza dla grupy scenariuszy są określane za pomocą wyrażenia, które jest następnie używane do wyliczenia wartości dla każdego szeregu początkowego w grupie. Można wprowadzić wyrażenie bezpośrednio lub kliknąć przycisk kalkulatora i utworzyć wyrażenie za pomocą konstruktora wyrażeń wartości scenariusza.

- Wyrażenie może zawierać dowolną zmienną przewidywaną lub przewidywaną w systemie modelu.
- Jeśli okres scenariusza będzie rozszerzony poza istniejące dane, wówczas wyrażenie jest stosowane do wartości prognozowanych dla zmiennych z wyrażenia.
- Dla każdego szeregu początkowego w grupie zmienne w wyrażeniu określają szeregi czasowe, które są definiowane przez te zmienne oraz wartości wymiaru, które definiują szereg początkowy. Są to szeregi czasowe, które są używane do oszacowania wyrażenia. Przykładowo, jeśli szereg początkowy jest definiowany przez wyrażenie `region= ' north '` oraz `brand= ' X '`, wówczas szeregi czasowe, używane w wyrażeniu są definiowane przez te same wartości wymiaru.

Załóżmy na przykład, że zmienna metryki źródłowej to *advertising* i dostępne są dwie zmienne wymiaru *region* i *brand*. Należy również założyć, że grupa scenariuszy zawiera również wszystkie kombinacje wartości zmiennych wymiarów. Następnie można określić wartość `advertising*1.2` jako wyrażenie do zbadania efektu zwiększenia wartości *advertising* o 20 procent dla wszystkich szeregów czasowych, jakie są powiązane ze zmienną *advertising*.

Uwaga: Grupy scenariuszy mają zastosowanie tylko do danych wielowymiarowych i tworzone są poprzez kliknięcie opcji **Dodaj grupę scenariuszy** na karcie Scenariusze.

Opcje

Maksymalny poziom dla zmienionych zmiennych przewidywanych

Określa maksymalną liczbę poziomów zmienionej zmiennej przewidywanej. Każdy kolejny poziom, maksymalnie do 5, obejmuje zmienne przewidywane, na które szereg początkowy ma bardziej pośredni wpływ. Szczególnie dotyczy to pierwszego poziomu obejmującego zmienne przewidywane, dla których szereg początkowy stanowi bezpośrednią wartość wejściową. Bezpośrednie dane wejściowe zmiennych przewidywanych na drugim poziomie stanowią zmienne przewidywane z pierwszego itd. Zwiększenie wartości tego ustawienia powoduje również zwiększenie złożoności obliczeń i może wpływać na wydajność.

Maksimum automatycznie wykrywanych zmiennych przewidywanych

Określa maksymalną liczbę zmiennych przewidywanych, które są automatycznie wykrywane dla każdego szeregu początkowego. Zwiększenie wartości tego ustawienia powoduje również zwiększenie złożoności obliczeń i może wpływać na wydajność.

Diagram wpływu

Wyświetla graficzną reprezentację relacji przyczynowych pomiędzy szeregiem początkowym dla każdego scenariusza a szeregiem przewidywanym, na który ma on wpływ. Tabele dla wartości scenariusz i wartościami przewidywanymi dla zmienionych zmiennych przewidywanych są uwzględniane jako część elementu wyniku. Graf obejmuje wykresy przewidywanych wartości zmienionych zmiennych przewidywanych. Jednokrotne kliknięcie dowolnego węzła w diagramie wpływu otwiera szczegółowy diagram kolejności dla szeregu powiązanego z węzłem. Dla każdego scenariusza generowany jest osobny diagram wpływu.

Wykresy szeregów

Generuje wykresy szeregów przewidywanych wartości dla wszystkich zmienionych zmiennych przewidywanych w scenariuszu.

Tabele prognoz i scenariuszy

Tabele przewidywanych wartości i wartości scenariusza dla każdego scenariusza. Tabele te zawierają te same informacje, co tabele w diagramie wpływu. Obsługują one wszystkie standardowe funkcje przestawiania i edytowania tabel.

Uwzględnij przedziały ufności na wykresach i w tabelach

Określa, czy przedziały ufności dla predykcji scenariusza są uwzględniane na wykresie i w wyniku.

Szerokość przedziału ufności (%)

To ustawienie kontroluje przedziału ufności dla predykcji scenariusza. Można określić dowolną wartość dodatnią mniejszą od 100. Domyślnie ustawiony jest 95% przedział ufności.

Miary dobroci dopasowania

W tej sekcji znajdują się definicje miar dobroci dopasowania stosowanych w modelowaniu szeregów czasowych.

- *Stacjonarność r-kwadrat*. Miara, która porównuje stacjonarną część modelu z prostym modelem średniej. Miara ta jest używana zamiast zwykłego r-kwadrat, gdy istnieje trend lub wzorec sezonowy. Stacjonarność r-kwadrat może być ujemna w zakresie od nieskończoności ujemnej do 1. Wartości ujemne oznaczają, że rozważany model jest gorszy od modelu bazowego. Wartości dodatnie oznaczają, że brany pod uwagę model jest lepszy niż model bazowy.
- *R-kwadrat*. Oszacowanie proporcji całkowitej zmienności w serii, która została wyjaśniona przez model. Miara ta jest przydatna, w przypadku szeregów stacjonarnych. R-kwadrat może być ujemny w zakresie od nieskończoności ujemnej do 1. Wartości ujemne oznaczają, że rozważany model jest gorszy od modelu bazowego. Wartości dodatnie oznaczają, że brany pod uwagę model jest lepszy niż model bazowy.
- *PPOK*. Pierwiastek błędu średniokwadratowego. Pierwiastek kwadratowy obliczany z przeciętnego odchylenia kwadratowego. Mierzy, jak bardzo szereg zależny odbiega od poziomu przewidywanego przez model; miara wyrażona w jednostkach używanych przez szereg zależny.
- *MAPE*. Średni Bezwzględny Błąd Procentowy. Mierzy jak bardzo szereg zależny odbiega od poziomu przewidywanego przez model. Jest niezależny od używanych jednostek i tym samym może być używany do porównywania szeregów używających różnych jednostek.
- *ŚBB*. Średni błąd bezwzględny. Mierzy jak bardzo szereg odbiega od poziomu przewidywanego przez model. ŚBB jest zgłaszany w pierwotnych jednostkach szeregu.
- *MaxAPE*. Maksymalny bezwzględny błąd procentowy. Największy prognozowany błąd, wyrażony jako procent. Miara ta jest pomocna w wyobrażeniu najgorszego możliwego scenariusza prognozy.
- *MaxAE*. Maksymalny błąd bezwzględny. Największy prognozowany błąd, wyrażony w jednostkach używanych przez szereg. Podobnie jak w przypadku MaksPBB, błąd ten jest pomocny w wyobrażeniu najgorszego możliwego scenariusza prognozy. Maksymalny błąd bezwzględny oraz maksymalny procentowy błąd bezwzględny mogą wystąpić w różnych miejscach szeregu, np. gdy błąd bezwzględny dużego szeregu jest w niewielkim stopniu większy od błędu bezwzględnego wartości małego szeregu. W takim wypadku maksymalny błąd bezwzględny wystąpi dla wartości dużego szeregu, a maksymalny procentowy błąd bezwzględny wystąpi dla wartości małego szeregu.

- *Znormalizowane BKI*. Znormalizowane Bayesowskie Kryterium Informacyjne. Ogólna miara całkowitego dopasowania modelu, która ma na celu uwzględnienie złożoności modelu. Jest to ocena oparta na średni błąd kwadratowy, zawierająca karę za liczbę parametrów w modelu oraz długość szeregu. Kara powoduje usunięcie przewagi modeli z większą liczbą parametrów, ułatwiając porównanie statystyki w różnych modelach tego samego szeregu.

Typy wartości odstających

W tej sekcji znajdują się definicje typów wartości odstających używanych w modelowaniu szeregów czasowych.

- *Addytyw*. Obserwacja odstająca, która wpływa na pojedynczą obserwację. Przykładowo błąd kodowania danych można zidentyfikować jako addytywną obserwację odstającą.
- *Przesunięcie poziomu*. Obserwacja odstająca przenosi wszystkie obserwacje za pomocą stałej, zaczynając od konkretnego punktu serii. Zmiana poziomu może być przyczyną zmiany zasad.
- *Innowacyjna*. Obserwacja odstająca, która działa jako uzupełnienie poziomu hałasu w określonym punkcie serii. W przypadku szeregów stacjonarnych innowacyjna obserwacja odstająca ma wpływ na kilka obserwacji. W przypadku szeregów niestacjonarnych innowacyjna obserwacja odstająca może mieć wpływ na wszystkie obserwacje rozpoczynające się w danym punkcie szeregu.
- *Przejęciowy*. Obserwacja odstająca, której wpływ maleje wykładniczo do 0.
- *Poprawka sezonowa*. Obserwacja odstająca, która wpływa na określoną obserwację i wszystkie kolejne obserwacje oddzielone od niej przez jeden lub więcej okresów sezonowych. Wpływ na wszystkie obserwacje jest równy. Sezonowa addytywna obserwacja odstająca może wystąpić, jeśli, rozpoczynając od danego roku, sprzedaż w styczniu każdego roku rośnie.
- *Trend lokalny*. Obserwacja odstająca rozpoczynająca trend lokalny w określonym punkcie szeregu.
- *Połączenie addytywów*. Grupa dwóch lub większej liczby kolejnych wartości odstających. Wybranie tego typu obserwacji grupy skutkuje wykryciem pojedynczych addytywnych obserwacji odstających oprócz ich grup.

Przewodnik po Płotkach ACF/PACF

Przedstawione tu wykresy są to czyste lub teoretyczne procesy ARIMA. Poniżej przedstawiono ogólne wytyczne dotyczące identyfikowania procesu:

- Seria Niestacjonarne posiada ACF, która pozostaje znacząca przez pół tuzina lub więcej opóźnień, a nie szybko spadając do 0. Należy zmienić taką serię, dopóki nie będzie ona stacjonarna, zanim będzie można zidentyfikować proces.
- Procesy autoregresyjne mają wykładniczo zmniejszający się ACF i kolce w pierwszej lub większej części PACF. Liczba kolców wskazuje kolejność autoregresji.
- Przenoszenie średnich procesów ma kolce w pierwszej lub większej liczby opóźnień w ACF i wykładniczo zmniejszającej się PACF. Liczba kolców wskazuje kolejność średniej ruchomej.
- Procesy mieszane (ARMA) zwykle wykazują spadki wykładnicze zarówno w ACF, jak i w PACF.

Na etapie identyfikacji nie trzeba się martwić o znak ACF lub PACF, ani o szybkość, z jaką wykładniczo maleje ACF lub PACF zbliża się do 0. Zależą one od znaku i rzeczywistej wartości współczynników AR i MA. W niektórych przypadkach jest to wykładniczo zmniejszający się zmienność ACF między wartościami dodatnimi i ujemnymi.

wykresy ACF i PACF z rzeczywistych danych nigdy nie są tak czyste, jak pokazywane są wykresy. Musisz nauczyć się wyłapać to, co jest niezbędne w każdej danej działce. Zawsze sprawdzaj ACF i PACF reszt, w przypadku, gdy identyfikacja jest błędna. Należy pamiętać, że:

- Procesy sezonowe pokazują te wzorce w opóźnieniach sezonowych (wielokrotności okresu sezonowego).

- Użytkownik jest uprawniony do traktowania nieistotnych wartości jako 0. Oznacza to, że można ignorować wartości, które leżą w przedziałach ufności na wykresach. Nie trzeba ich ignorować, szczególnie jeśli kontynuują one wzorzec statystycznie istotnych wartości.
- Autokorelacja okazjonalna będzie statystycznie istotna tylko przez przypadek. Można zignorować statystycznie istotną autokorelację, jeśli jest izolowana, najlepiej w dużym opóźnieniu, a jeśli nie występuje w sezonowym opóźnieniu.

Bardziej kompletną dyskusję na temat wykresów ACF i PACF można znaleźć w dowolnym tekście analizy ARIMA.

Tabela 3. ARIMA (0,0, 1), $q>0$

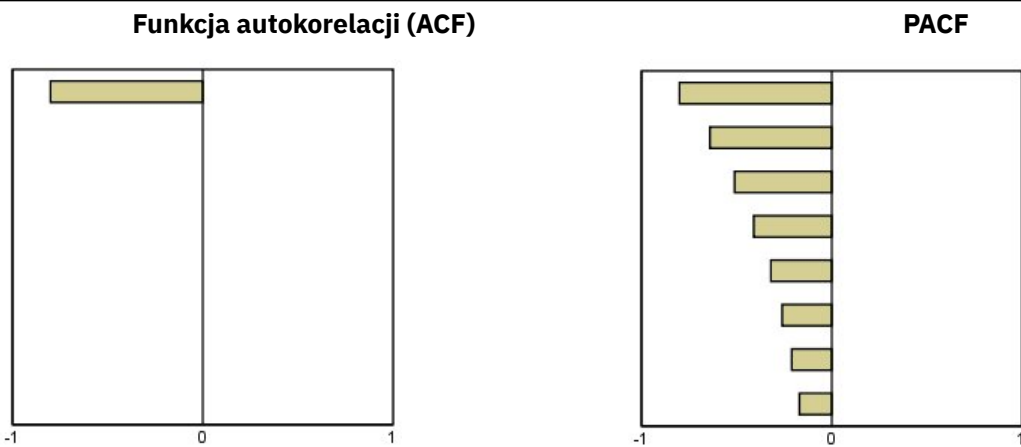
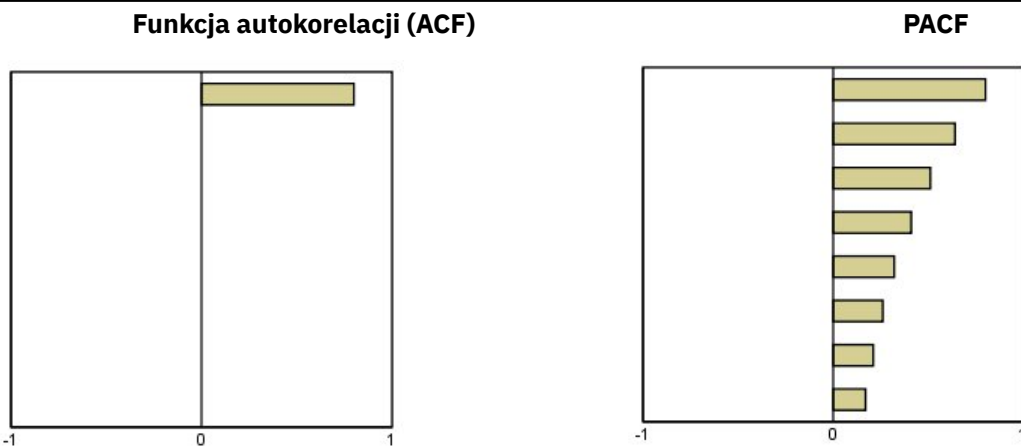
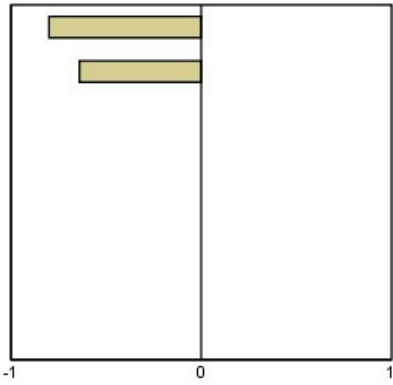


Tabela 4. ARIMA (0,0, 1), $q<0$



ARIMA (0,0, 2), $1_2 > 0$

Funkcja autokorelacji (ACF)



PACF

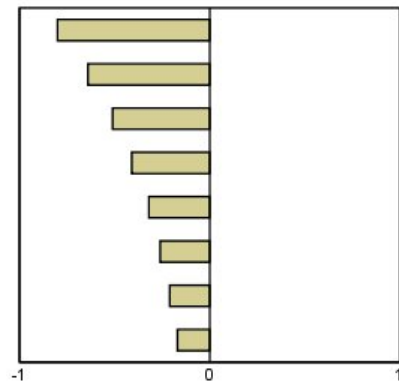
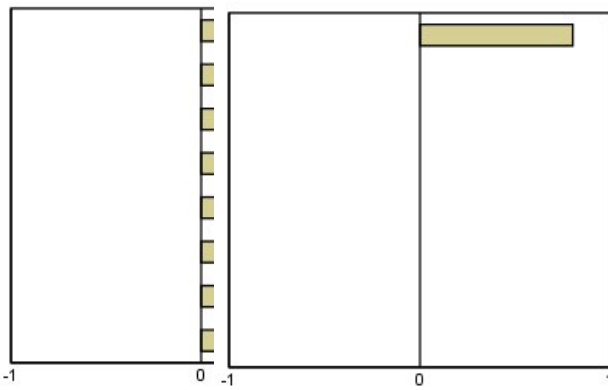


Tabela 5. ARIMA (1,0, 0), $f > 0$

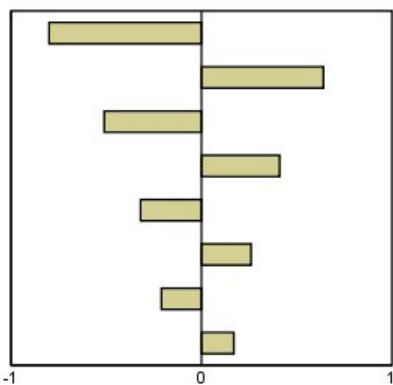
Funkcja autokorelacji (ACF)



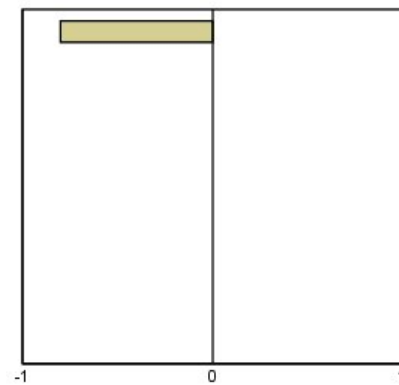
PACF

Tabela 6. ARIMA (1,0, 0), $f < 0$

Funkcja autokorelacji (ACF)

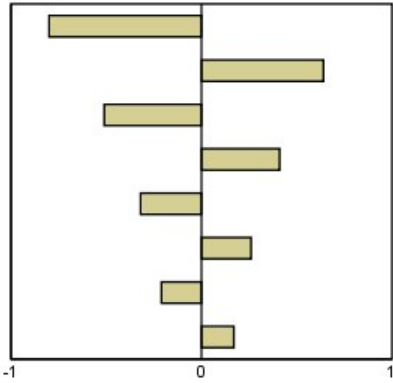


PACF

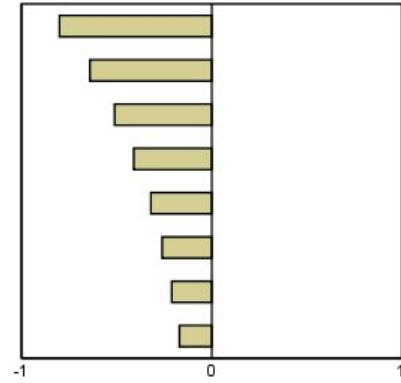


ARIMA (1,0, 1), $\theta < 0, \phi > 0$

Funkcja autokorelacji (ACF)

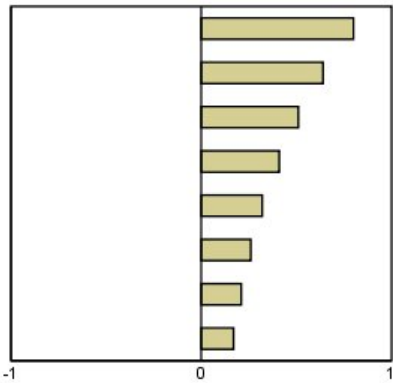


PACF



ARIMA (2,0,0), $\phi_2 > 0$

Funkcja autokorelacji (ACF)



PACF

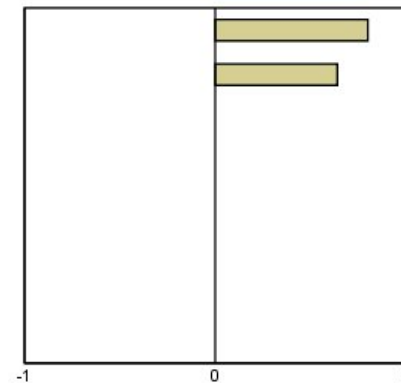
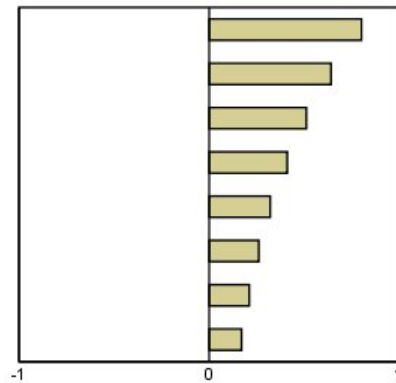


Tabela 7. ARIMA (0,1,0) (seria zintegrowana)

Funkcja autokorelacji (ACF)



Uwagi

Niniejsza publikacja została przygotowana z myślą o produktach i usługach oferowanych w Stanach Zjednoczonych. IBM może udostępniać ten materiał w innych językach. Jednakże w celu uzyskania dostępu do takiego materiału istnieje konieczność posiadania egzemplarza produktu w takim języku.

Produktów, usług lub opcji opisywanych w tym dokumencie IBM nie musi oferować we wszystkich krajach. Informacje o produktach i usługach dostępnych w danym kraju można uzyskać od lokalnego przedstawiciela IBM. Odwołanie do produktu, programu lub usługi IBM nie oznacza, że można użyć wyłącznie tego produktu, programu lub usługi IBM. Zamiast nich można zastosować ich odpowiednik funkcjonalny pod warunkiem że nie narusza to praw własności intelektualnej IBM. Jednakże cała odpowiedzialność za ocenę przydatności i sprawdzenie działania produktu, programu lub usługi pochodzących od producenta innego niż IBM spoczywa na użytkowniku.

IBM może posiadać patenty lub złożone wnioski patentowe na produkty, o których mowa w niniejszej publikacji. Przedstawienie tej publikacji nie daje żadnych uprawnień licencyjnych do tychże patentów. Pisemne zapytania w sprawie licencji można przesyłać na adres:

IBM Director of Licensing

IBM Corporation

*North Castle Drive, MD-NC119
Armonk, NY 10504-1785 U.S.A.*

Zapytania dotyczące zestawów znaków dwubajtowych (DBCS) należy kierować do lokalnych działów własności intelektualnej IBM (IBM Intellectual Property Department) lub wysłać je na piśmie na adres:

Intellectual Property Licensing

*Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokio 103-8510, Japonia*

INTERNATIONAL BUSINESS MACHINES CORPORATION DOSTARCZA TĘ PUBLIKACJĘ W STANIE, W JAKIM SIĘ ZNAJDUJE ("AS IS") BEZ UDZIELANIA JAKICHKOLWIEK GWARANCJI (W TYM TAKŻE RĘKOJMI), WYRAŻNYCH LUB DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU ORAZ GWARANCJI, ŻE PUBLIKACJA NIE NARUSZA PRAW STRON TRZECICH. Ustawodawstwa niektórych krajów nie dopuszczają zastrzeżeń dotyczących gwarancji wyraźnych lub domniemanych w odniesieniu do pewnych transakcji; w takiej sytuacji powyższe zdanie nie ma zastosowania.

Informacje zawarte w tej publikacji mogą zawierać nieścisłości techniczne lub błędy drukarskie. Informacje te są okresowo aktualizowane, a zmiany te zostaną uwzględnione w kolejnych wydaniach tej publikacji. IBM zastrzega sobie prawo do wprowadzania ulepszeń i/lub zmian w produktach i/lub programach opisanych w tej publikacji w dowolnym czasie, bez wcześniejszego powiadomienia.

Wszelkie wzmianki w tej publikacji na temat stron internetowych firm innych niż IBM zostały wprowadzone wyłącznie dla wygody użytkowników i w żadnym razie nie stanowią zachęty do ich odwiedzania. Materiały dostępne na tych stronach nie są częścią materiałów opracowanych dla tego produktu IBM, a użytkownik korzysta z nich na własną odpowiedzialność.

IBM ma prawo do używania i rozpowszechniania informacji przystanych przez użytkownika w dowolny sposób, jaki uzna za właściwy, bez żadnych zobowiązań wobec ich autora.

Licencjodawcy tego programu, którzy chcieliby uzyskać informacje na temat programu w celu: (i) umożliwienia wymiany informacji między niezależnie utworzonymi programami i innymi programami

(łącznie z opisywanym) oraz (ii) wykorzystywania wymienianych informacji, powinni skontaktować się z:

IBM Director of Licensing

IBM Corporation

*North Castle Drive, MD-NC119
Armonk, NY 10504-1785 U.S.A.*

Informacje takie mogą być udostępnione, o ile spełnione zostaną odpowiednie warunki, w tym, w niektórych przypadkach, zostanie uiszczona stosowna opłata.

Licencjonowany program opisany w niniejszej publikacji oraz wszystkie inne licencjonowane materiały dostępne dla tego programu są dostarczane przez IBM na warunkach określonych w Umowie IBM z Klientem, Międzynarodowej Umowie Licencyjnej IBM na Program lub w innych podobnych umowach zawartych między IBM i użytkownikami.

Dane dotyczące wydajności i cytowane przykłady zostały przedstawione jedynie w celu zobrazowania sytuacji. Faktyczne wyniki dotyczące wydajności mogą się różnić w zależności do konkretnych warunków konfiguracyjnych i operacyjnych.

Informacje dotyczące produktów innych podmiotów niż IBM zostały uzyskane od dostawców tych produktów, z ich publicznych ogłoszeń lub innych dostępnych publicznie źródeł. IBM nie testował tych produktów i nie może potwierdzić dokładności pomiarów wydajności, kompatybilności ani żadnych innych danych związanych z produktami firm innych niż IBM. Pytania dotyczące możliwości produktów firm innych niż IBM należy kierować do dostawców tych produktów.

Wszelkie stwierdzenia dotyczące przyszłych kierunków rozwoju i zamierzeń IBM mogą zostać zmienione lub wycofane bez powiadomienia.

Publikacja ta zawiera przykładowe dane i raporty używane w codziennych operacjach działalności gospodarczej. W celu kompleksowego zilustrowania tej działalności podane przykłady zawierają nazwy osób, firm i ich produktów. Wszystkie te nazwy/nazwiska są fikcyjne i jakiegokolwiek podobieństwo do istniejących nazw/nazwisk jest całkowicie przypadkowe.

LICENCJA W ZAKRESIE PRAW AUTORSKICH:

Niniejsza publikacja zawiera przykładowe aplikacje w kodzie źródłowym ilustrujące techniki programowania w różnych systemach operacyjnych. Użytkownik może kopiować, modyfikować i rozpowszechniać te programy przykładowe w dowolnej formie bez uiszczania opłat na rzecz IBM, w celu rozbudowy, użytkowania, handlowego lub w celu rozpowszechniania aplikacji zgodnych z aplikacyjnym interfejsem programowym dla tego systemu operacyjnego, dla którego napisane były programy przykładowe. Programy przykładowe nie zostały gruntownie przetestowane. IBM nie może zatem gwarantować ani sugerować niezawodności, użyteczności i funkcjonalności tych programów. Programy przykładowe są dostarczane w stanie, w jakim się znajdują ("AS IS"), bez jakichkolwiek gwarancji (rękojmię również wyłącza się). IBM nie ponosi odpowiedzialności za jakiegokolwiek szkody wynikające z używania programów przykładowych.

Każda kopia programu przykładowego lub jakiegokolwiek jego fragment, jak też jakiegokolwiek prace pochodne muszą zawierać następujące uwagi dotyczące praw autorskich:

© Copyright IBM Corp. 2021. Fragmenty tego kodu pochodzą z przykładowych programów produktu IBM Corp. Programy przykładowe.

© Copyright IBM Corp. 1989-2021. Wszelkie prawa zastrzeżone.

Znaki towarowe

IBM, logo IBM i ibm.com są znakami towarowymi lub zastrzeżonymi znakami towarowymi International Business Machines Corp., zarejestrowanymi w wielu systemach prawnych na całym świecie. Pozostałe nazwy produktów i usług mogą być znakami towarowymi IBM lub innych przedsiębiorstw. Aktualna lista znaków towarowych IBM dostępna jest w serwisie WWW, w sekcji "Copyright and trademark

information" (Informacje o prawach autorskich i znakach towarowych), pod adresem www.ibm.com/legal/copytrade.shtml.

Adobe, logo Adobe, PostScript oraz logo PostScript są znakami towarowymi lub zastrzeżonymi znakami towarowymi Adobe Systems Incorporated w Stanach Zjednoczonych i/lub w innych krajach.

Intel, logo Intel, Intel Inside, logo Intel Inside, Intel Centrino, logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium i Pentium są znakami towarowymi lub zastrzeżonymi znakami towarowymi Intel Corporation lub przedsiębiorstw podporządkowanych w Stanach Zjednoczonych i w innych krajach.

Linux jest zastrzeżonym znakiem towarowym Linusa Torvaldsa w Stanach Zjednoczonych i/lub w innych krajach.

Microsoft, Windows, Windows NT oraz logo Windows są znakami towarowymi Microsoft Corporation w Stanach Zjednoczonych i/lub w innych krajach.

UNIX jest zastrzeżonym znakiem towarowym Open Group w Stanach Zjednoczonych i w innych krajach.

Java oraz wszystkie znaki towarowe i logo dotyczące Java są znakami towarowymi firmy i jej firm zależnych.

Indeks

A

addytywna wartość odstająca
w kreatorze modeli szeregów czasowych [6](#), [9](#)
analiza harmoniczna [21](#)
analiza szeregów czasowych
modelowanie przyczynowe szeregów czasowych [23](#)

B

braki danych
w procedurze Zastosuj modele szeregów czasowych [19](#)
brakujące wartości
w kreatorze modeli szeregów czasowych [13](#)

D

dane historyczne
w kreatorze modeli szeregów czasowych [11](#)
w procedurze Zastosuj modele szeregów czasowych [17](#)
Dekompozycja sezonowa
modele [20](#)
obliczanie średnich ruchomych [20](#)
tworzenie zmiennych [20](#)
założenia [20](#)
zapisywanie zmiennych wynikowych [20](#)
dobroć dopasowania
definicje [47](#)
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)

F

funkcja autokorelacji
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
wykresy dla czystych procesów ARIMA [48](#)
Funkcja autokorelacji (ACF)
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
wykresy dla czystych procesów ARIMA [48](#)
funkcja autokorelacji cząstkowej
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
wykresy dla czystych procesów ARIMA [48](#)
Funkcja autokorelacji cząstkowej (PACF)
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
funkcje przenoszenia
opóźnienie [8](#)
rzędy licznika [8](#)

funkcje przenoszenia (*kontynuacja*)
rzędy mianownika [8](#)
rzędy różnicowania [8](#)
rzędy sezonowości [8](#)

I

innowacyjna wartość odstająca
w kreatorze modeli szeregów czasowych [6](#), [9](#)

K

Kreator modeli szeregów czasowych
ARIMA [3](#), [7](#)
brakujące wartości [13](#)
funkcja autokorelacji cząstkowych reszt [10](#), [11](#)
funkcja autokorelacji reszt [10](#), [11](#)
funkcje przenoszenia [8](#)
Moduł automatycznego doboru modelu [3](#)
najlepiej i najgorzej dopasowane modele [12](#)
nazwy modeli [13](#)
nazwy nowych zmiennych [12](#)
okres prognozy [3](#), [13](#)
okres szacowania [3](#)
okresowość [5–8](#)
parametry modelu [10](#)
prognozy [10](#), [11](#)
przedziały ufności [11](#), [13](#)
statystyka Boxa-Ljunga [10](#)
statystyka dobroci dopasowania [10](#), [11](#)
statystyki na wszystkich modelach [10](#), [11](#)
transformacja szeregu [6–8](#)
wartości dopasowania [11](#)
wartości odstające [6](#), [9](#)
wygładzanie wykładnicze [3](#), [6](#)
zapisywanie predykcji [12](#)
zapisywanie specyfikacji modelu w pliku XML [12](#)
zdarzenia [5](#)

M

maksymalny bezwzględny błąd
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
maksymalny bezwzględny błąd procentowy
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
Maksymalny bezwzględny błąd procentowy (MaxAPE)
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
Maksymalny błąd bezwzględny (MaxAE)
w kreatorze modeli szeregów czasowych [10](#), [11](#)

Maksymalny błąd bezwzględny (MaxAE) (*kontynuacja*)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
MaxAE [47](#)
MaxAPE [47](#)
model wygładzania wykładniczego Browna [6](#)
model wygładzania wykładniczego Holta [6](#)
model wygładzania wykładniczego Wintersa
addytywny [6](#)
multiplikatywny [6](#)
model wygładzania wykładniczego z wygasającym trendem [6](#)
modele
ARIMA [3](#)
Moduł automatycznego doboru modelu [3](#)
wygładzanie wykładnicze [3](#), [6](#)
modele ARIMA
funkcje przenoszenia [8](#)
wartości odstające [9](#)
modele wygładzania wykładniczego [3](#), [6](#)
modelowanie przyczynowe szeregów czasowych [23](#), [25–30](#),
[32–34](#)
Moduł automatycznego doboru modelu
ograniczanie obszaru modelu [5](#)
wartości odstające [6](#)

N

naturalna transformacja logarytmiczna
w kreatorze modeli szeregów czasowych [6–8](#)
nazwy modeli
w kreatorze modeli szeregów czasowych [13](#)
nazwy zmiennych
w kreatorze modeli szeregów czasowych [12](#)
w procedurze Zastosuj modele szeregów czasowych [18](#)

O

okres historyczny [2](#)
okres prognozy
w kreatorze modeli szeregów czasowych [3](#), [13](#)
w procedurze Zastosuj modele szeregów czasowych [14](#)
okres szacowania
w kreatorze modeli szeregów czasowych [3](#)
w procedurze Zastosuj modele szeregów czasowych [14](#)
okres walidacji [2](#)
okresowość
w kreatorze modeli szeregów czasowych [5–8](#)

P

PACF
wykresy dla czystych procesów ARIMA [48](#)
parametry modelu
w kreatorze modeli szeregów czasowych [10](#)
w procedurze Zastosuj modele szeregów czasowych [16](#)
pierwiastek błędu średniokwadratowego
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
Pierwiastek błędu średniokwadratowego (RMSE)
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)

ponowna estymacja parametrów modelu
w procedurze Zastosuj modele szeregów czasowych [14](#)
prognozowanie modelem przyczynowym szeregów
czasowych [35–38](#), [40](#)
prognozy
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
prosty model wygładzania wykładniczego [6](#)
przedziały ufności
w kreatorze modeli szeregów czasowych [11](#), [13](#)
w procedurze Zastosuj modele szeregów czasowych [17](#),
[19](#)
przejściowa wartość odstająca
w kreatorze modeli szeregów czasowych [6](#), [9](#)

R

R-kwadrat
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
R2 [47](#)
reszty
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)

S

scenariusze modelowania przyczynowego szeregów
czasowych [41–46](#)
sezonowa addytywna wartość odstająca
w kreatorze modeli szeregów czasowych [6](#), [9](#)
stacjonarny R-kwadrat
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
statystyka Boxa-Ljunga
w kreatorze modeli szeregów czasowych [10](#)
w procedurze Zastosuj modele szeregów czasowych [16](#)

Ś

średni bezwzględny błąd
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
średni bezwzględny błąd procentowy
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
Średni bezwzględny błąd procentowy (MAPE)
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)
Średni błąd bezwzględny (MAE)
w kreatorze modeli szeregów czasowych [10](#), [11](#)
w procedurze Zastosuj modele szeregów czasowych [16](#),
[17](#)

T

transformacja logarytmiczna
w kreatorze modeli szeregów czasowych [6–8](#)
transformacja pierwiastkiem kwadratowym
w kreatorze modeli szeregów czasowych [6–8](#)

W

warianty kontrolne [2](#)
wartości dopasowania
w kreatorze modeli szeregów czasowych [11](#)
w procedurze Zastosuj modele szeregów czasowych [17](#)
wartości odstające
definicje [48](#)
modele ARIMA [9](#)
Moduł automatycznego doboru modelu [6](#)
wartość odstająca w postaci przesunięcia poziomu
w kreatorze modeli szeregów czasowych [6, 9](#)
wartość odstająca w postaci trendu lokalnego
w kreatorze modeli szeregów czasowych [6, 9](#)
wiązka oddziaływania odstająca
w kreatorze modeli szeregów czasowych [6, 9](#)
Wykresy analizy spektralnej
analiza spektralna dwóch zmiennych [21](#)
okna spektralne [21](#)
wyśrodkowanie transformacji [21](#)
założenia [21](#)

X

XML
zapisywanie modeli szeregów czasowych w pliku XML
[12](#)
zapisywanie ponownie estymowanych modeli
w formacie XML [18](#)

Z

zapisywanie
nazwy nowych zmiennych [12, 18](#)
ponownie estymowane modele w formacie XML [18](#)
predykcje modelu [12, 18](#)
specyfikacje modelu w pliku XML [12](#)
Zastosuj modele szeregów czasowych
braki danych [19](#)
funkcja autokorelacji cząstkowych reszt [16, 17](#)
funkcja autokorelacji reszt [16, 17](#)
najlepiej i najgorzej dopasowane modele [18](#)
nazwy nowych zmiennych [18](#)
okres prognozy [14](#)
okres szacowania [14](#)
parametry modelu [16](#)
ponowna estymacja parametrów modelu [14](#)
prognozy [16, 17](#)
przedziały ufności [17, 19](#)
statystyka Boxa-Ljunga [16](#)
statystyka dobroci dopasowania [16, 17](#)
statystyki na wszystkich modelach [16, 17](#)
wartości dopasowania [17](#)
zapisywanie ponownie estymowanych modeli
w formacie XML [18](#)
zapisywanie predykcji [18](#)

zdarzenia

w kreatorze modeli szeregów czasowych [5](#)
znormalizowane BKI (bayesowskie kryterium informacyjne)
w kreatorze modeli szeregów czasowych [10, 11](#)
w procedurze Zastosuj modele szeregów czasowych [16, 17](#)

