

IBM SPSS Regression 29

IBM

Uwaga

Przed użyciem tych informacji i produktu, którego one dotyczą, przeczytaj informacje znajdujące się w sekcji “Uwagi” na stronie 51.

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. Regresja.....	1
Wybór procedury dla binarnej regresji logistycznej.....	1
Regresja logistyczna.....	2
Regresja logistyczna: Filtrowanie.....	3
Regresja logistyczna: Metody wyboru zmiennych.....	3
Regresja logistyczna: Definiuj zmienne jakościowe.....	4
Regresja logistyczna: Zapisz zmienne wynikowe.....	4
Regresja logistyczna: Opcje.....	5
Dodatkowe właściwości komendy LOGISTIC REGRESSION.....	5
Wielomianowa regresja logistyczna.....	6
Wielomianowa regresja logistyczna.....	7
Wielomianowa regresja logistyczna: Kategoria odniesienia.....	7
Wielomianowa regresja logistyczna: Statystyki.....	8
Wielomianowa regresja logistyczna: Kryteria.....	8
Wielomianowa regresja logistyczna: Opcje.....	9
Wielomianowa regresja logistyczna: Zapisz.....	9
Dodatkowe właściwości komendy NOMREG.....	10
Regresja probit.....	10
Analiza Probit: Definiuj zakres.....	11
Analiza Probit: Opcje.....	11
Dodatkowe właściwości komendy PROBIT.....	12
Regresja kwantylowa.....	12
Regresja kwantylowa: Kryteria.....	13
Regresja kwantylowa: Model.....	14
Regresja kwantylowa: Prezentacja.....	16
Regresja kwantylowa: Zapisz.....	17
Regresja kwantylowa: Eksportuj.....	17
Regresja nieliniowa.....	18
Logika warunkowa (regresja nieliniowa).....	19
Regresja nieliniowa: Parametry.....	19
Typowe modele regresji nieliniowej.....	20
Regresja nieliniowa: Funkcja straty.....	20
Regresja nieliniowa: Ograniczenia parametrów.....	21
Regresja nieliniowa: Zapisz zmienne wynikowe.....	21
Regresja nieliniowa: Opcje.....	21
Interpretowanie wyników regresji nieliniowej.....	22
Dodatkowe właściwości komendy NLR.....	22
Estymacja wag.....	22
Szacowanie wag: Opcje.....	24
Dodatkowe właściwości komendy WNK.....	24
Procedura Podwójna MNK.....	24
Regresja podwójną metodą najmniejszych kwadratów: Opcje.....	25
Dodatkowe właściwości komendy 2SLS.....	25
Schematy kodowania zmiennych jakościowych.....	25
Odchylenie.....	25
Proste.....	26
Helmerta.....	26
Różnica.....	27
Wielomianowe.....	27
Powtórzone.....	28
Specjalne.....	28

Wskaźnik.....	29
Regresja grzbietowa jądra.....	29
Parametry jądra.....	30
Regresja grzbietowa jądra: Opcje.....	30
Modele czasu przyspieszonego błędu parametrycznego.....	31
Modelowe modele czasu niepowodzenia w trybie przyspieszonym: Kryteria.....	32
Modelowe modele czasu niepowodzenia w trybie przyspieszonym: Model.....	33
Modelowe modele czasu niepowodzenia w trybie przyspieszonym: szacowanie.....	33
Modelowe modele czasu niepowodzenia w trybie przyspieszonym: Drukowanie.....	35
Modelowe modele czasu niepowodzenia w trybie przyspieszonym: predykcja.....	35
Modelowe modele czasu awarii z akceleracją parametryczną: wykres.....	36
Modele czasu przyspieszonego niepowodzenia parametrycznego: Eksport.....	37
Czas przeżycia AFT-definiowanie zdarzeń dla zmiennych statusu.....	37
Modelowanie parametrycznych przyspieszonych modeli czasu: Wybierz kategorię.....	37
Modele parametrycznych współużytkowanych modeli Frailty.....	38
Modele parametryczne współużytkowane frailty: kryteria.....	39
Modele parametryczne współużytkowane Frailty: Model.....	40
Modele współużytkowanych ramek parametrycznych: szacowanie.....	40
Modele parametryczne współużytkowane szandarowe: drukowanie.....	42
Modele sparametryzowane współużytkowane parametryczne: Predict.....	42
Modele parametryczne współużytkowane frailty: Plot.....	43
Modele sparametryzowanych współużytkowanych modeli: Eksport.....	44
Modele parametryczne współużytkowane Frailty: Definiuj zdarzenia.....	44
Modele parametryczne współużytkowane frailty-przykłady.....	45
Parametryczne współużytkowane modele Frailty-studium przypadku dla danych rekurencyjnych.....	47
Uwagi.....	51
Znaki towarowe.....	52
Indeks.....	55

Rozdział 1. Regresja

Następujące funkcje dotyczące regresji są dostępne w SPSS Statistics edycji Standard Edition lub module opcjonalnym Regression.

Wybór procedury dla binarnej regresji logistycznej

Modele binarnej regresji logistycznej mogą być dopasowywane za pomocą procedury regresji logistycznej oraz procedury wielomianowej regresji logistycznej. Każda z tych procedur zawiera opcje, które nie są dostępne w drugiej. Ważna różnica teoretyczna polega na tym, że procedura regresji logistycznej tworzy wszystkie predykcje, reszty, statystyki wpływu oraz testy dobroci dopasowania, wykorzystując dane na poziomie poszczególnych obserwacji, niezależnie od tego, w jaki sposób dane są wprowadzane i czy liczba schematów współzmiennych jest mniejsza niż całkowita liczba obserwacji; natomiast procedura wielomianowej regresji logistycznej wewnątrznie agreguje obserwacje w celu tworzenia podpopulacji z identycznymi schematami współzmiennych dla predyktorów, tworząc predykcje, reszty i testy dobroci dopasowania w oparciu o te podpopulacje. Jeśli wszystkie predyktory są zmiennymi jakościowymi lub predyktory ciągłe przyjmują tylko ograniczoną liczbę wartości — tak, że dla każdego odrębnego schematu współzmiennych występuje kilka obserwacji — to zastosowanie podpopulacji może generować prawidłowe testy dobroci dopasowania i użyteczne reszty, w odróżnieniu od analizy na poziomie pojedynczych obserwacji.

Regresja logistyczna

Udostępnia następujące unikalne funkcje:

- Test dobroci dopasowania Hosmera-Lemesowej dla modelu
- Analizy krokowe
- Kontrasty definiujące parametryzację modelu
- Alternatywne punkty podziału dla klasyfikacji
- Wykresy klasyfikacyjne
- Dopasowanie modelu utworzonego według jednego zbioru obserwacji do zachowanego testującego zbioru obserwacji
- Zapisywanie predykcji, reszt i statystyk wpływu

Wielomianowa regresja logistyczna

Udostępnia następujące unikalne funkcje:

- Test chi-kwadrat Pearsona i test odchyłań chi-kwadrat dobroci dopasowania modelu
- Określanie podpopulacji do grupowania danych dla celów badania dobroci dopasowania
- Tworzenie listy liczebności, przewidywanych liczebności i reszt z podziałem na podpopulacje
- Korekta oszacowań wariancji z uwzględnieniem nadmiernego rozproszenia
- Macierz kowariancji oszacowań parametrów
- Testy liniowych kombinacji parametrów
- Jawne określanie modeli zagnieżdżonych
- Dopasowanie 1-1 dopasowanych warunkowych modeli regresji logistycznej utworzonych na podstawie różniących się zmiennych

Uwagi:

- Obie te procedury dopasowują model dla danych binarnych, który jest uogólnionym modelem liniowym z rozkładem dwumianowym i funkcją łączenia logit. Jeśli do analizy dostępnych danych bardziej odpowiednie jest użycie innej funkcji łączenia, należy skorzystać z procedury Uogólnione modele liniowe.

- Jeśli dane zawierają powtarzane pomiary danych binarnych lub rekordy, które są w inny sposób ze sobą skorelowane, należy wziąć pod uwagę procedury Uogólnione liniowe modele mieszane lub Uogólnione równania estymujące.

Regresja logistyczna

Regresja logistyczna jest przydatna w sytuacjach, w których wymagane jest przewidywanie obecności lub braku cechy bądź wyniku na podstawie wartości zestawu predyktorów. Jest podobna do modelu regresji liniowej, ale nadaje się dla modeli, w których zmienna zależna jest dychotomiczna. Współczynniki regresji logistycznej mogą być używane do oszacowania ilorazów szans dla każdej zmiennej niezależnej w modelu. Regresja logistyczna ma zastosowanie do szerszego zakresu sytuacji badawczych niż analiza dyskryminacyjna.

Przykład. Jakie cechy stylu życia są czynnikami ryzyka dla choroby wieńcowej serca? Na podstawie próby pacjentów przebadanych pod względem palenia tytoniu, diety, aktywności fizycznej, spożycia alkoholu i statusu choroby wieńcowej można zbudować model wykorzystujący cztery zmienne stylu życia, aby przewidywać obecność lub brak choroby wieńcowej w próbie pacjentów. Model może być następnie używany do szacowania ilorazów szans dla każdego czynnika w celu określenia, na przykład, o ile bardziej prawdopodobne jest, że choroba wieńcowa rozwinie się u osób palących niż u niepalących.

Statystyki. Dla każdej analizy: łączna liczba obserwacji, wybrane obserwacje, ważne obserwacje. Dla każdej zmiennej jakościowej: kodowanie parametrów. Dla każdego kroku: zmienne wprowadzone lub usunięte, historia iteracji, -2 logarytm wiarygodności, dobroć dopasowania, statystyka dobroci dopasowania Hosmera-Lemeshowa, chi-kwadrat modelu, chi-kwadrat poprawy, tabela klasyfikacji, korelacja między zoiennymi, tabela zaobserwowanych grup i przewidzianych prawdopodobieństw, reszta chi-kwadrat. Dla każdej zmiennej w równaniu: współczynnik (B), błąd standardowy B , statystyka Walda, szacowany iloraz szans ($\exp(B)$), przedział ufności dla $\exp(B)$, logarytm wiarygodności w wypadku usunięcia składnika z modelu. Dla każdej zmiennej, która nie występuje w równaniu: statystyka oceny. Dla każdej obserwacji: zaobserwowana grupa, przewidywane prawdopodobieństwo, przewidywana grupa, reszta, reszta standaryzowana.

Metody. Modele można szacować, wprowadzając zmienne blokowo lub korzystając z dowolnej z następujących metod krokowych: warunkowej postępującej, postępującej LR, Walda postępującej, warunkowej wstecznej, wstecznej LR lub wstecznej Walda.

Wymagania dotyczące danych dla regresji logistycznej

Dane. Zmienna zależna powinna być dychotomiczna. Zmienne niezależne mogą być przedziałowe lub jakościowe; w przypadku zmiennych jakościowych wymagane jest kodowanie przy użyciu wartości sztucznych lub wskaźników (w procedurze dostępna jest opcja automatycznego przekodowywania zmiennych jakościowych).

Założenia. Regresja logistyczna nie opiera się na założeniach co do rozkładu w tym samym sensie, jak analiza dyskryminacyjna. Jednak rozwiązanie może być bardziej stabilne, jeśli predyktory charakteryzują się wielowymiarowymi rozkładami normalnymi. Ponadto, podobnie jak w przypadku innych form regresji, wielowspółliniowość między predyktorami może prowadzić do obciążonych oszacowań i zawyżonych błędów standardowych. Procedura jest najefektywniejsza, kiedy przynależność do grupy jest zmienną prawdziwie kategorialną. Jeśli przynależność do grupy opiera się na wartościach zmiennej ciągłej (na przykład wysoki a niski iloraz inteligencji), należy rozważyć możliwość wykorzystania regresji liniowej, tak aby skorzystać z bogatszych informacji oferowanych przez samą zmienną ilościową.

Procedury pokrewne. Procedura Wykres rozrzutu służy do monitorowania danych pod kątem wielowspółliniowości. Jeśli spełnione są założenia dotyczące wielowymiarowego rozkładu normalnego oraz równości macierzy wariancji-kowariancji, możliwe jest szybsze uzyskanie rozwiązania za pomocą procedury Analiza dyskryminacyjna. Jeśli wszystkie predyktory są zmiennymi jakościowymi, można również użyć procedury log-liniowej. Jeśli zmienna zależna jest zmienną ciągłą, należy użyć regresji liniowej. W celu sporządzenia wykresu prawdopodobieństw zapisanych za pomocą procedury regresji logistycznej możliwe jest wykorzystanie procedury krzywej ROC.

Uzyskiwanie analizy metodą regresji logistycznej

1. Z menu wybierz:

Analiza > Regresja > Dwumianowa regresja logistyczna...

- Wybierz jedną dychotomiczną zmienną zależną. Ta zmienna może być numeryczna lub łańcuchowa.
- Wybierz co najmniej jedną współzmienną. Aby uwzględnić składniki interakcji, zaznacz wszystkie zmienne uczestniczące w interakcji i wybierz pozycję **>a*b>**.

Aby wprowadzić zmienne w grupach (**blokach**), wybierz współzmiennie dla bloku, a następnie kliknij przycisk **Dalej**, aby określić nowy blok. Powtarzaj te czynności do czasu, aż zostaną określone wszystkie bloki.

Opcjonalnie można wybrać obserwacje do analizy. Wybierz zmienną filtrującą i wprowadź kryteria reguły.

Regresja logistyczna: Filtrowanie

W oszacowaniu modelu uwzględniane są obserwacje zdefiniowane przez regułę filtrowania. Na przykład jeśli wybrane zostaną zmienna, **jest równe** oraz wartość 5, to w oszacowaniu modelu uwzględnione zostaną tylko te obserwacje, w których wybrana zmienna ma wartość równą 5.

Statystyki i wyniki klasyfikacji są generowane zarówno dla wybranych, jak i dla niewybranych obserwacji. W ten sposób uzyskujemy mechanizm klasyfikacji nowych obserwacji na podstawie już istniejących danych lub podzbiory danych na podzbiory szkoleniowe i testowe w celu przeprowadzenia sprawdzenia na wygenerowanym modelu.

Regresja logistyczna: Metody wyboru zmiennych

Wybór metody pozwala na określenie w jaki sposób zmienne niezależne będą wprowadzane do analizy. Korzystając z różnych metod, dla jednego zbioru zmiennych można skonstruować wiele modeli regresji.

- Wprowadź.** Procedura doboru zmiennych, w której wszystkie zmienne z bloku są jednocześnie wprowadzane do analizy.
- Selekcja postępująca (warunkowa).** Metoda selekcji krokowej z testowaniem wprowadzania, opartym na istotności statystyk ocen oraz z testowaniem usuwania, bazującym na prawdopodobieństwie statystyki ilorazu wiarygodności opartej na warunkowych ocenach parametrów.
- Selekcja postępująca (iloraz wiarygodności).** Metoda selekcji krokowej z testowaniem wprowadzania, opartym na istotności statystyk ocen oraz z testowaniem usuwania, bazującym na prawdopodobieństwie statystyki ilorazu wiarygodności opartej na maksymalnych częściowych ocenach wiarygodności.
- Selekcja postępująca (statystyka Walda).** Metoda selekcji krokowej z testowaniem wprowadzania, opartym na istotności statystyk ocen oraz z testowaniem usuwania, bazującym na prawdopodobieństwie statystyki Walda.
- Eliminacja wsteczna (warunkowa).** Wybór krokowy wsteczny. Sprawdzenie, czy dana zmienna podlega usunięciu, opiera się na prawdopodobieństwie ilorazu wiarygodności wyliczonego na podstawie ocen parametrów warunkowych.
- Eliminacja wsteczna (iloraz wiarygodności).** Wybór krokowy wsteczny. Sprawdzenie, czy dana zmienna podlega usunięciu, opiera się na prawdopodobieństwie ilorazu wiarygodności wyliczonego na podstawie ocen maksymalnych wiarygodności częściowych.
- Eliminacja wsteczna (Wald).** Wybór krokowy wsteczny. Sprawdzenie, czy dana zmienna podlega usunięciu, opiera się na prawdopodobieństwie statystyki Walda.

Wartości istotności wyników zależą od dopasowania pojedynczego modelu. Z tego względu, jeśli stosowana jest metoda krokowa wartości istotności są zazwyczaj niepoprawne.

Wszystkie wybrane zmienne niezależne są dodawane do jednego modelu regresji. Można jednak również określić różne metody wprowadzania dla różnych podzbiorów zmiennych. Na przykład jeden blok zmiennych można wprowadzić do modelu regresji przy użyciu selekcji krokowej, a drugi za pomocą selekcji postępującej. Aby do modelu regresji dodać drugi blok zmiennych, kliknij przycisk **Dalej**.

Regresja logistyczna: Definiuj zmienne jakościowe

Można szczegółowo określić sposób traktowania zmiennych jakościowych przez procedurę Regresja logistyczna:

Współzmiennie. Zawiera listę wszystkich współzmiennych określonych w głównym oknie dialogowym – albo bezpośrednio, albo jako części interakcji, na dowolnej warstwie. Jeśli niektóre z tych zmiennych są łańcuchowe lub jakościowe, można użyć ich tylko jako współzmiennych jakościowych.

Współzmiennie jakościowe. Lista zmiennych określonych jako jakościowe. Przy każdej zmiennej w nawiasach wskazano sposób kodowania kontrastów. Zmienne łańcuchowe (oznaczone symbolem < po nazwie) są już wymienione na liście współzmiennych jakościowych. Wybierz dowolne inne współzmiennie jakościowe z listy Współzmiennie i przenieś je na listę Współzmiennie jakościowe.

Zmień kontrast. Umożliwia zmianę metody kontrastu. Dostępne są następujące metody kontrastu:

- **Wskaźnik.** Kontrasty wskazują na obecność lub nieobecność kategorii. Kategoria odniesienia jest przedstawiona w macierzy kontrastów jako rząd zer.
- **Prosty.** Każda kategoria predyktora z wyjątkiem kategorii odniesienia jest porównywana z kategorią odniesienia.
- **Różnica.** Każda kategoria predyktora z wyjątkiem pierwszej kategorii jest porównywana ze średnim efektem poprzednich kategorii. Znane również jako odwrotne kontrasty Helmerta.
- **Helmerta.** Każda kategoria predyktora z wyjątkiem ostatniej kategorii jest porównywana ze średnim efektem następujących kategorii.
- **Powtórzonej.** Każda kategoria predyktora z wyjątkiem ostatniej kategorii jest porównywana z następną kategorią.
- **Wielomianowy.** Kontrasty wielomianowe ortogonalne. Zakłada się, że kategorie są równo rozłożone. Kontrasty wielomianowe są dostępne wyłącznie dla zmiennych liczbowych.
- **Odchylenie.** Każda kategoria predyktora z wyjątkiem kategorii odniesienia jest porównywana z efektem łącznym.

W przypadku wybrania opcji **Odchylenie**, **Proste** lub **Wskaźnik**, jako kategorię odniesienia należy wybrać **Pierwsza** albo **Ostatnia**. Należy zauważyć, że metoda zmienia się dopiero po kliknięciu przycisku **Zmień**.

Współzmiennie łańcuchowe muszą być współzmiennymi jakościowymi. Aby usunąć zmienną łańcuchową z listy Współzmiennie jakościowe, należy usunąć wszystkie składniki zawierające tę zmienną z listy Współzmiennie w głównym oknie dialogowym.

Regresja logistyczna: Zapisz zmienne wynikowe

Wyniki regresji logistycznej można zapisać jako nowe zmienne w aktywnym zbiorze danych:

Wartości przewidywane. Zapisuje wartości przewidywane przez model. Dostępne opcje to Prawdopodobieństwa i Przynależność do grupy.

- *Prawdopodobieństwa.* Dla każdego przypadku zapisuje przewidywane prawdopodobieństwo wystąpienia zdarzenia. W tabeli w wynikach wyświetlana jest nazwa i zawartość wszelkich nowych zmiennych. „Zdarzenie” jest kategorią zmiennej zależnej o wyższej wartości; na przykład, jeśli zmienna zależna przyjmuje wartości 0 i 1, zapisywane jest przewidywane prawdopodobieństwo kategorii 1.
- *Przewidywana przynależność do grupy.* Grupa z największym prawdopodobieństwem posterioru, oparta na wynikach dyskryminacyjnych. Jest to grupa, do której klasyfikowana jest obserwacja na podstawie predykcji modelu.

Statystyki wpływu. Zapisuje wartości statystyk, które mierzą wpływ obserwacji na przewidywane wartości. Dostępne są następujące opcje: Cooka, Wartości wpływu i DfBeta.

- *Cook.* Statystyka logistyczna analogowa statystyki wpływu Cooka. Miara stopnia, w jakim zmieniłyby się wskaźniki reszt dla wszystkich obserwacji przy wykluczeniu poszczególnych obserwacji z obliczeń współczynników regresji.
- *Wartość wpływu.* Względny wpływ każdej obserwacji na dopasowanie modelu.

- *DfBetas*. Różnica w wartości beta jest zmianą współczynnika regresji, który wynika z wyłączenia danej obserwacji. Wartość tej różnicy jest obliczana dla wszystkich składników modelu, łącznie ze stałą (wyrazem wolnym).

Reszty. Powoduje zapisanie reszt. Dostępne opcje to Niestandaryzowane, Logit, Studentyzowane, Standaryzowane i Odchylenie.

- *Niestandaryzowane reszty*. Różnica pomiędzy wartością empiryczną, a wartością przewidywaną przez model.
- *Reszta logit*. Reszta dla przypadku, jeśli jest ona przewidywana w skali logit. Reszta logit to reszta podzielona przez przewidywane prawdopodobieństwo razy jeden minus przewidywane prawdopodobieństwo.
- *Reszta studentyzowana*. Zmiana w odchyleniu modelu w wyniku usunięcia obserwacji z analizy.
- *Standaryzowane reszty*. Reszta podzielona przez oszacowanie jego odchylenia standardowego. Standaryzowane reszty, znane także jako reszty Pearsona, mają średnią arytmetyczną 0 oraz odchylenie standardowe 1.
- *Deviance*. Reszty zależne od odchylenia modelu.

Eksportuj informacje o modelu do pliku XML. Parametr szacuje i (opcjonalnie) i kowariancje są eksportowane do określonego pliku w formacie XML (PMML). Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. .

Regresja logistyczna: Opcje

Można określić opcje analizy regresji logistycznej:

Statystyki i wykresy. Pozwala na zlecenie obliczenia statystyk i utworzenia wykresów. Dostępne opcje to Wykresy klasyfikacji, dobroci dopasowania Hosmera-Lemeshowa, Lista reszt obserwacjami, Korelacje oszacowań, Przebieg iteracji i Przedział ufności dla $\exp(B)$. Wybierz jedną z opcji w grupie Wyświetl, aby wyświetlić statystyki i wykresy Na każdym kroku albo tylko dla modelu finalnego, w ostatnim kroku.

- *Statystyka dobroci dopasowania Hosmera-Lemeshowa*. Ta statystyka dobroci dopasowania jest bardziej wytrzymała niż tradycyjna statystyka dobroci dopasowania używana w regresji logistycznej, szczególnie w przypadku modeli o ciągłych współzmiennych i badaniach o małych rozmiarach próby. Polega ona na grupowaniu obserwacji w decyle ryzyka i porównywanie w ramach każdego decylu prawdopodobieństwa obserwowanego z prawdopodobieństwem oczekiwanym.

Prawdopodobieństwo w metodzie krokowej. Umożliwia określanie kryteriów, według których zmienne są wprowadzane do równania i z niego usuwane. Istnieje możliwość określenia kryteriów wprowadzania lub usuwania zmiennych.

- *Prawdopodobieństwo w metodzie krokowej*. Zmienna jest wprowadzana do modelu, jeśli prawdopodobieństwo jej statystyki wyniku jest mniejsze od wartości pozycji i jest usuwane, jeśli prawdopodobieństwo jest większe niż wartość usunięcia. Aby zmienić ustawienia domyślne, należy wpisać dodatnie wartości wprowadzenia i usunięcia. Wartość wprowadzenia musi być mniejsza niż wartość usunięcia.

Punkt odcięcia klasyfikacji. Umożliwia określenie punktu podziału na potrzeby klasyfikowania obserwacji. Obserwacje z wartościami przewidywanymi, które przekraczają punkt podziału klasyfikacji, są klasyfikowane jako dodatnie, a wartości poniżej punktu odcięcia są klasyfikowane jako ujemne. W celu zmiany ustawienia domyślnego należy wprowadzić wartość zakresu od 0,01 do 0,99.

Maksymalna liczba powtórzeń. Umożliwia zmianę maksymalnej liczby powtórzeń modelu przed zakończeniem.

Dołącz stałe do modelu. Umożliwia wskazanie, czy model powinien zawierać stały składnik. Gdy ta opcja jest wyłączona, składnik stały jest równy 0.

Dodatkowe właściwości komendy LOGISTIC REGRESSION

Język składni komend umożliwia również:

- Identyfikowanie wyników obserwacji na podstawie wartości lub etykiet zmiennej.
- Określanie odstępów między raportami z iteracji. Zamiast prezentować oszacowania parametrów po każdej iteracji, można zażądać oszacowań parametrów co n iteracji.
- Zmiana kryteriów zakończenia iteracji i sprawdzenie nadmiarowości.
- Określenie listy zmiennych dla list obserwacji.
- Zaoszczędzenie pamięci dzięki przechowywaniu danych każdej grupy podzbiorów w zewnętrznym pliku roboczym podczas przetwarzania.

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

Wielomianowa regresja logistyczna

Wielomianowa regresja logistyczna jest przydatna w sytuacjach, w których można klasyfikować obiekty na podstawie wartości zbioru predyktorów. Ten typ regresji jest podobny do regresji logistycznej, ale ma charakter bardziej ogólny, ponieważ zmienna zależna nie jest ograniczona do dwóch kategorii.

Przykład. Aby skuteczniej promować filmy, studia filmowe chcą przewidzieć rodzaj filmu, który widzowie kin prawdopodobnie będą chcieli oglądać. Wykonując wielomianową regresję logistyczną, studio może określić siłę wpływu wieku, płci i statusu związku (ma/nie ma partnera) na rodzaj preferowanych filmów. Następnie studio może skierować kampanię promującą film do grupy osób, które prawdopodobnie będą nim zainteresowane.

Statystyki. Przebieg iteracji, współczynniki parametrów, asymptotyczne macierze kowariancji i korelacji, testy ilorazu wiarygodności dla efektów modelu i cząstkowych, -2 logarytmy wiarygodności. Chi-kwadrat Pearsona i dobroć dopasowania chi-kwadrat. Coxa i Snella, R^2 Nagelkerke'a i McFaddena. Klasyfikacja: obserwowane a przewidywane częstości według kategorii odpowiedzi. Tabela krzyżowa: częstości obserwowane i przewidywane (z resztami) oraz proporcje według schematu współzmiennych i kategorii odpowiedzi.

Metody. Wielomianowy model logitowy jest dopasowywany do pełnego modelu czynnikowego lub modelu określonego przez użytkownika. Oszacowanie parametrów realizowane jest za pomocą iteracyjnego algorytmu maksymalnej wiarygodności.

Wymagania dotyczące danych dla wielomianowej regresji logistycznej

Dane. Zmienna zależna powinna być jakościowa. Zmienne niezależne mogą być czynnikami lub współzmiennymi. Co do zasady czynniki powinny być zmiennymi jakościowymi, a współzmiennie powinny być zmiennymi ilościowymi.

Założenia. Zakłada się, że iloraz szans dla każdej z dwóch kategorii jest niezależny od wszystkich innych kategorii odpowiedzi. Na przykład, jeśli na rynek wprowadzany jest nowy produkt, to zakłada się, że będzie to miało wpływ w równym stopniu, proporcjonalnie, na udziały w rynku wszystkich innych produktów. Ponadto, biorąc pod uwagę schemat współzmiennych, przyjmuje się, że odpowiedzi są niezależnymi zmiennymi wielowymiarowymi.

Wykonywanie wielomianowej regresji logistycznej

1. Wybierz z menu następującą opcję:

Analizuj > Regresja > Wielomianowa regresja logistyczna...

2. Wybierz jedną zmienną zależną.

3. Czynniki są opcjonalne i mogą być liczbowe albo jakościowe.

4. Współzmiennie są opcjonalne, ale muszą być liczbowe, jeśli zostaną określone.

Wielomianowa regresja logistyczna

Domyślnie wielomianowa regresja logistyczna tworzy model z głównymi efektami czynników i współzmiennych, ale w tym oknie dialogowym można określić model użytkownika lub zażądać krokowej selekcji modelu.

Określ model. Model efektów głównych zawiera efekty główne współzmiennych i czynników, ale nie zawiera efektów interakcji. Pełny model czynnikowy obejmuje efekty główne wszystkich czynników oraz interakcje między wszystkimi czynnikami. Nie obejmuje interakcji współzmiennych. Można utworzyć model użytkownika, aby określić podzbiory interakcji między czynnikami lub interakcji między współzmiennymi albo zażądać krokowej selekcji składników modelu.

Czynniki i współzmiennie. Lista zawiera czynniki i współzmiennie.

Składniki wymuszone. Składniki dodane do listy wymuszonych są zawsze uwzględniane w modelu.

Składniki krokowe. Składniki dodane do listy selekcji krokowej są uwzględniane w modelu zgodnie z jedną z następujących metod selekcji krokowej, wybraną przez użytkownika:

- **Wprowadzanie postępujące.** W tej metodzie na początku model nie zawiera żadnych składników krokowych. W każdym kroku najbardziej znaczący składnik krokowy jest dodawany do modelu — do momentu, aż żaden ze składników krokowych będących poza modelem nie miałby już statystycznie istotnego wkładu w model, gdyby został do niego dodany.
- **Eliminacja wsteczna.** Ta metoda rozpoczyna się od wprowadzenia do modelu wszystkich terminów określonych na liście selekcji krokowej. W każdym kroku najmniej znaczący składnik krokowy jest usuwany z modelu — do momentu, aż wszystkie składniki pozostałe w modelu będą miały statystycznie istotny wkład w model.
- **Krokowa postępująca.** Ta metoda rozpoczyna się od modelu, który zostałby wybrany przy użyciu metody wprowadzania postępującego. Od tego momentu algorytm naprzemiennie stosuje eliminację wsteczną składników krokowych w modelu i wprowadzanie postępujące składników pozostawionych poza modelem. Ten proces jest kontynuowany do momentu, gdy żadne składniki nie spełniają już kryteriów wprowadzenia albo usunięcia.
- **Krokowa wsteczna.** Ta metoda rozpoczyna się od modelu, który zostałby wybrany metodą eliminacji wstecznej. Od tego momentu algorytm naprzemiennie stosuje wprowadzanie postępujące składników pozostawionych poza modelem i eliminację wsteczną składników krokowych w modelu. Ten proces jest kontynuowany do momentu, gdy żadne składniki nie spełniają już kryteriów wprowadzenia albo usunięcia.

Uwzględnij wyraz wolny w modelu. Umożliwia uwzględnienie lub wykluczenie wyrazu wolnego w modelu.

Budowanie składników

Dla wybranych czynników i współzmiennych:

Interakcja. Dla wszystkich wybranych zmiennych tworzy składnik interakcji najwyższego rzędu.

Efekty główne. Dla każdej wybranej zmiennej tworzy składnik efektów głównych.

Wszystkie 2 rzędu. Tworzy wszystkie możliwe interakcje drugiego rzędu wybranych zmiennych.

Wszystkie 3 rzędu. Tworzy wszystkie możliwe interakcje trzeciego rzędu wybranych zmiennych.

Wszystkie 4 rzędu. Tworzy wszystkie możliwe interakcje czwartego rzędu wybranych zmiennych.

Wszystkie 5 rzędu. Tworzy wszystkie możliwe interakcje piątego rzędu wybranych zmiennych.

Wielomianowa regresja logistyczna: Kategoria odniesienia

Domyślnie procedura Wielomianowa regresja logistyczna wybiera ostatnią kategorię jako kategorię odniesienia. To okno dialogowe umożliwia wybór kategorii odniesienia, a także kolejności kategorii.

Kategoria odniesienia. Określ pierwszą, ostatnią lub własną kategorię.

Uporządkowanie kategorii. W porządku rosnącym najniższa wartość definiuje pierwszą kategorię, a najwyższa wartość definiuje ostatnią. W porządku malejącym najwyższa wartość definiuje pierwszą kategorię, a najniższa wartość — ostatnią.

Wielomianowa regresja logistyczna: Statystyki

Dla wielomianowej regresji logistycznej można określić następujące statystyki:

Podsumowanie przetwarzania obserwacji. Ta tabela zawiera informacje na temat określonych zmiennych jakościowych.

Model. Statystyki dla całego modelu.

- **Pseudo R-kwadrat.** Prezentuje statystyki R^2 Coxa i Snella oraz Nagelkerke'a i McFaddena.
- **Podsumowanie kroku.** W tej tabeli podsumowane są efekty wprowadzone lub usunięte w każdym kroku metody krokowej. Nie jest ona generowana, jeśli w oknie dialogowym Model nie określono selekcji krokowej składników modelu.
- **Dopasowanie modelu.** Ta tabela zawiera porównanie modelu dopasowanego z modelem zawierającym tylko wyraz wolny i modelem pustym.
- **Kryteria informacyjne.** Tabela ta zawiera kryterium informacyjne Akaike (AIC) oraz kryterium informacyjne bayesowskie Schwarz (BIC).
- **Prawdopodobieństwa komórek.** Przedstawia tabelę częstości obserwowanych i oczekiwanych (z resztami) oraz proporcje według schematu współzmiennych i kategorii odpowiedzi.
- **Tabela klasyfikacji.** Przedstawia tabelę odpowiedzi zaobserwowanych wobec przewidywanych.
- **Chi-kwadrat statystyki dobroci dopasowania.** Przedstawia statystyki chi-kwadrat Pearsona oraz ilorazu wiarygodności. Statystyki są obliczane dla schematów współzmiennych, które zostały określone przez wszystkie czynniki i współzmiennne lub przez podzbiór czynników i współzmiennych zdefiniowany przez użytkownika.
- **Miary monotoniczności.** Przedstawia tabelę zawierającą informacje na temat liczby par zgodnych, par niezgodnych i par powiązanych. W tabeli widoczne są także Współczynnik Somers' D, Lambda Goodmana i Kruskala, Współczynnik Tau-a Kendalla oraz Indeks zgodności C.

Parametry. Statystyki związane z parametrami modelu.

- **Oszacowania.** Prezentuje oszacowania parametrów modelu z poziomem ufności określonym przez użytkownika.
- **Test ilorazu wiarygodności.** Przedstawia testy ilorazu wiarygodności dla efektów cząstkowych modelu. Test dla całego modelu jest przedstawiany automatycznie.
- **Korelacje asymptotyczne.** Przedstawia macierz korelacji oszacowań parametrów.
- **Kowariancje asymptotyczne.** Przedstawia macierz kowariancji oszacowań parametrów.

Definiuj podpopulacje. Umożliwia wybranie podzbioru czynników i współzmiennych w celu zdefiniowania schematów współzmiennych używanych przy wyznaczaniu prawdopodobieństwa komórek oraz w testach dobroci dopasowania.

Wielomianowa regresja logistyczna: Kryteria

Dla wielomianowej regresji logistycznej można określić następujące kryteria:

Iteracje. Umożliwia określenie maksymalnej liczby powtórzeń algorytmu, maksymalnej liczby kroków połowienia, tolerancji zbieżności dla zmian logarytmu wiarygodności i parametrów, częstości prezentowania postępu algorytmu iteracyjnego, iteracji, od której procedura powinna rozpocząć sprawdzanie kompletności lub quasi-kompletności podziału danych.

- **Zbieżność logarytmu wiarygodności.** Przyjmuje się, że zbieżność ma miejsce, gdy względna zmiana funkcji logarytmu wiarygodności jest mniejsza niż określona wartość. Kryterium nie jest stosowane, jeśli wartość wynosi 0. Podaj liczbę nieujemną.

- **Zbieżność parametru.** Przyjmuje się, że zbieżność ma miejsce, gdy bezwzględna zmiana oszacowania parametru jest mniejsza niż ta wartość. Kryterium nie jest stosowane, jeśli wartość wynosi 0.

Delta. Umożliwia określenie wartości nieujemnej mniejszej niż 1. Ta wartość jest dodawana do każdej pustej komórki tabeli krzyżowej kategorii odpowiedzi według wzoru współzmiennej. Pomaga to w ustabilizowaniu algorytmu i zapobiega odchyleniom w oszacowaniach.

Tolerancja osobliwości. Umożliwia określenie tolerancji wykrywania osobliwości.

Wielomianowa regresja logistyczna: Opcje

Dla wielomianowej regresji logistycznej można określić następujące opcje:

Skala rozproszenia. Można określić skalę rozproszenia, która będzie używana do korygowania oszacowania macierzy kowariancji parametrów. Opcja **Odchylenie** powoduje oszacowanie skali przy użyciu statystyki funkcji odchylenia (chi-kwadrat ilorazu wiarygodności). Opcja **Pearsona** powoduje oszacowanie skali przy użyciu statystyki chi-kwadrat Pearsona. Można także określić własną wartość skalującą. Musi to być dodatnia wartość liczbowa.

Opcje metod krokowych. Opcje te dają kontrolę nad kryteriami statystycznymi, gdy do budowania modelu używane są metody krokowe. Są one zignorowane, jeśli w oknie dialogowym Model nie określono selekcji krokowej składników modelu.

- **Prawdopodobieństwo wprowadzenia.** Jest to prawdopodobieństwo statystyki ilorazu wiarygodności, przy którym zmienna jest wprowadzana. Im większe jest określone prawdopodobieństwo, tym łatwiej zmienne wprowadzane są do modelu. To kryterium jest uwzględniane tylko wtedy, gdy wybrana jest metoda wprowadzania postępującego, selekcji krokowej postępującej lub selekcji krokowej wstecznej.
- **Test wprowadzenia.** Jest to metoda wprowadzania składników stosowana w metodach krokowych. Można wybrać test ilorazu wiarygodności albo test oceny. To kryterium jest uwzględniane tylko wtedy, gdy wybrana jest metoda wprowadzania postępującego, selekcji krokowej postępującej lub selekcji krokowej wstecznej.
- **Prawdopodobieństwo usunięcia.** Jest to prawdopodobieństwo statystyki ilorazu wiarygodności, przy którym zmienna jest usuwana. Im większe jest określone prawdopodobieństwo, tym łatwiej zmienne usuwane są z modelu. To kryterium jest uwzględniane tylko wtedy, gdy wybrana jest metoda eliminacji wstecznej, wprowadzania postępującego, selekcji krokowej postępującej lub selekcji krokowej wstecznej.
- **Test usunięcia.** Jest to metoda usuwania składników stosowana w metodach krokowych. Można wybrać test ilorazu wiarygodności albo test Walda. To kryterium jest uwzględniane tylko wtedy, gdy wybrana jest metoda eliminacji wstecznej, wprowadzania postępującego, selekcji krokowej postępującej lub selekcji krokowej wstecznej.
- **Minimum efektów krokowych w modelu.** W przypadku używania metody eliminacji wstecznej lub metody krokowej selekcji wstecznej określa minimalną liczbę składników do uwzględnienia w modelu. Wyraz wolny nie jest liczony jako składnik modelu.
- **Maksimum efektów krokowych w modelu.** W przypadku użycia metody wprowadzania postępującego lub selekcji krokowej postępującej określa maksymalną liczbę składników do uwzględnienia w modelu. Wyraz wolny nie jest liczony jako składnik modelu.
- **Wymuszaj hierarchiczne wprowadzanie i usuwanie składników.** Ta opcja umożliwia określenie, czy należy wprowadzić ograniczenia dotyczące uwzględniania składników modelu. W przypadku wybrania opcji hierarchicznej wymagane jest, aby dla każdego uwzględnianego składnika najpierw uwzględnić w modelu wszystkie składniki niższego rzędu, które stanowią jego części. Na przykład, jeśli obowiązuje wymagania hierarchiczności, wówczas składniki *Stan cywilny* i *Płeć* muszą już znajdować się w modelu, aby można było dodać interakcję *Stan cywilny * Płeć*. Trzy przetączniki decydują o roli współzmiennych w określaniu hierarchii.

Wielomianowa regresja logistyczna: Zapisz

Okno dialogowe Zapisz umożliwia zapisanie zmiennych w pliku roboczym i wyeksportowanie informacji o modelu do pliku zewnętrznego.

Zapisywane zmienne. Można zapisać następujące zmienne:

- **Estymowane prawdopodobieństwa reakcji.** Są to oszacowane prawdopodobieństwa klasyfikowania schematów czynników/współzmiennych do kategorii odpowiedzi. Liczba estymowanych prawdopodobieństw jest równa liczbie kategorii zmiennej odpowiedzi; zapisanych zostaje maksymalnie 25.
- **Przewidywana kategoria.** Jest to kategoria odpowiedzi z największym oczekiwanym prawdopodobieństwem schematu czynnika/współzmiennych.
- **Przewidywane prawdopodobieństwa kategorii.** Jest to maksymalna wartość szacowanych prawdopodobieństw odpowiedzi.
- **Prawdopodobieństwo rzeczywistej kategorii.** Jest to szacowane prawdopodobieństwo zaklasyfikowania schematu czynnika/współzmiennych do obserwowanej kategorii.

Eksportuj informacje o modelu do pliku XML. Parametr szacuje i (opcjonalnie) i kowariancje są eksportowane do określonego pliku w formacie XML (PMML). Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. .

Dodatkowe właściwości komendy NOMREG

Język składni komend umożliwia również:

- Określenie kategorii referencyjnej zmiennej zależnej.
- Uwzględnienie obserwacji z brakami danych użytkownika.
- Tworzenie własnych testów hipotezy poprzez określanie hipotez zerowych jako liniowych kombinacji parametrów.

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

Regresja probit

Ta procedura służy do pomiaru relacji między siłą bodźca a odsetkiem przypadków wykazujących określoną odpowiedź na bodziec. Jest przydatna w sytuacjach, w których mamy wyniki dychotomiczne, co do których podejrzewamy, że znajdują się pod wpływem poziomów pewnych zmiennych niezależnych lub są przez nie spowodowane. Szczególnie dobrze nadaje się do analizy danych eksperymentalnych. Procedura ta pozwoli oszacować siłę bodźca konieczną do wywołania pewnego odsetka odpowiedzi, takiego jak mediana dawki skutecznej.

Przykład. Jak skutecznie nowy pestycyd zwalcza mrówki i w jakim stężeniu najlepiej go stosować? Możemy przeprowadzić eksperyment, w którym poddamy mrówki działaniu pestycydu w różnych stężeniach, a następnie zarejestrujemy liczbę mrówek faktycznie zabitych i liczbę mrówek poddanych działaniu pestycydu. Stosując regresję Probit względem tych danych, można określić siłę zależności między stężeniem i skutecznością zabijania, a także wyznaczyć stężenie pestycydu, które będzie odpowiednie, gdy chcemy mieć pewność, że zginie co najmniej 95% mrówek.

Statystyki. Współczynniki regresji i błędy standardowe, wyraz wolny i błąd standardowy, chi-kwadrat dobroci dopasowania Pearsona, częstości obserwowane i częstości oczekiwane, przedziały ufności dla poziomów skutecznych zmiennych niezależnych. Wykresy: wykresy przekształconych odpowiedzi.

Ta procedura wykorzystuje algorytmy zaproponowane i zaimplementowane w NPSOL[®] przez Gill, Murray, Saunders & Wright, aby oszacować parametry modelu.

Wymagania dotyczące danych dla regresji Probit

Dane. Dla każdej wartości zmiennej niezależnej (lub każdej kombinacji wartości wielu zmiennych niezależnych) zmienna odpowiedzi powinna być liczbą obserwacji z wartościami, które oznaczają interesującą odpowiedź, a zmienna częstości obserwowanych powinna być liczbą wszystkich obserwacji z tymi wartościami dla zmiennej niezależnej. Zmienna czynnika powinna być jakościowa i zakodowana jako liczby całkowite.

Założenia. Obserwacje powinny być niezależne. Jeśli mamy dużą liczbę wartości zmiennych niezależnych względem liczby obserwacji, tak jak np. w typowym badaniu obserwacyjnym, statystyki chi-kwadrat i dobroci dopasowania mogą być nieważne.

Procedury pokrewne. Analiza Probit jest ściśle związana z regresją logistyczną; w istocie w wypadku wybrania przekształcenia logit opisywana tutaj procedura obliczy regresję logistyczną. Zasadniczo analiza Probit jest odpowiednia do analizy zaplanowanych eksperymentów, natomiast regresja logistyczna jest bardziej odpowiednia do badań obserwacyjnych. Różnice w wynikach odzwierciedlają te różnice w zastosowaniach. Procedura analizy Probit podaje oszacowania wartości skutecznych przy różnych wskaźnikach odpowiedzi (w tym medianę dawki skutecznej), podczas gdy procedura regresji logistycznej podaje oszacowania ilorazów szans dla zmiennych niezależnych.

Uzyskiwanie analizy regresji Probit

1. Wybierz z menu następujące opcje:

Analiza > Regresja > Probit ...

2. Wybierz zmienną częstości odpowiedzi. Zmienna ta określa liczbę obserwacji wykazujących odpowiedź (reakcję) na bodziec testowy. Wartości tej zmiennej nie mogą być ujemne.
3. Wybierz zmienną częstości obserwowanych. Ta zmienna wskazuje liczbę obserwacji, wobec których zastosowano bodziec. Wartości tej zmiennej nie mogą być ujemne i nie mogą być mniejsze niż wartości zmiennej częstości odpowiedzi dla każdej obserwacji.

Opcjonalnie można wybrać zmienną czynnika. Jeśli ta zmienna zostanie wybrana, zdefiniuj zakresy dla grup w polu **Definiuj zakres**.

4. Wybierz co najmniej jedną współzmienną. Zmienna ta zawiera poziom bodźca zastosowanego do każdej obserwacji. Aby przekształcić współzmienną, wybierz przekształcenie z listy rozwijanej **Przekształcenia**. Jeśli nie zostanie zastosowana żadne przekształcenie, a istnieje grupa kontrolna, to grupa kontrolna jest uwzględniana w analizie.

5. Wybierz model **Probit** albo **Logit**.

Model Probit

Stosuje transformację probitową (odwrotność funkcji skumulowanego standardowego rozkładu normalnego) do proporcji odpowiedzi.

Model logitowy

Umożliwia zastosowanie transformacji logitowej (logarytmy szans) do proporcji odpowiedzi.

Analiza Probit: Definiuj zakres

Pozwala to na określenie poziomów zmiennej czynnika, która będzie analizowana. Poziomy czynnika muszą być zakodowane jako kolejne liczby całkowite, a analizowane będą wszystkie poziomy w podanym zakresie.

Analiza Probit: Opcje

Dla analizy probitowej można określić opcje:

Statystyki. Umożliwia zlecenie obliczenia następujących opcjonalnych statystyk: Częstości, Względna moc medialna, Test równoległości i Granice ufności.

- *Względna moc medialna.* Wyświetla współczynnik potencji mediany dla każdej pary poziomów czynników. Pokazuje także 95-procentowe przedziały ufności dla każdej względnej mocy medialnej. Względne moce medialne nie są dostępne, gdy nie ma zmiennej czynnika lub gdy jest więcej niż jedna współzmienna.
- *Test równoległości.* Sprawdzenie hipotezy, że wszystkie poziomy czynników mają wspólny gradient.
- *Granice ufności.* Przedziały ufności dla dawki czynnika, wymaganej do uzyskania określonego prawdopodobieństwa reakcji.

Statystyki Przedziały ufności i względna siła medialna są niedostępne, jeśli wybrano więcej niż jedną współzmienną. Statystyki Względna moc medialna i Test równoległości są dostępne tylko wtedy, gdy wybrano zmienną czynnika.

Proporcja reakcji naturalnej. Pozwala na określenie naturalnej stopy reakcji (proporcji reakcji naturalnej), nawet w przypadku braku bodźca. Dostępne alternatywy to: Brak, Oblicz na podstawie danych i Wartość.

- *Oblicz na podstawie danych.* Oszacowanie naturalnego wskaźnika odpowiedzi z przykładowych danych. Dane powinny zawierać obserwację reprezentującą poziom kontrolny, dla którego wartość współzmiennych wynosi 0. Probit szacuje naturalny wskaźnik odpowiedzi, używając proporcji odpowiedzi dla poziomu kontrolnego jako wartości początkowej.
- *Wartość.* Ustawia naturalną szybkość odpowiedzi w modelu (wybierz tę pozycję, gdy znasz wcześniej naturalny wskaźnik odpowiedzi). Należy wprowadzić naturalną proporcję odpowiedzi (musi być mniejsza od 1). Jeśli na przykład odpowiedź występuje w 10% czasu, gdy bodziec wynosi 0, należy wpisać wartość 0,10.

Kryteria. Umożliwia określenie parametrów iteracyjnego algorytmu estymacji parametrów. Można przestawić wartości domyślne opcji Maksymalna liczba iteracji, Limit kroku i Optymalizacja tolerancji.

Dodatkowe właściwości komendy PROBIT

Język składni komend umożliwia również:

- Zażądanie analizy zarówno dla modelu probit, jak i modelu logit.
- Kontrolowanie sposobu postępowania z brakami danych.
- Przekształcenie współzmiennych według podstaw innych niż 10 lub logarytm naturalny.

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

Regresja kwantylowa

Regresja jest metodą statystyczną, która jest szeroko stosowana w modelowaniu ilościowym. Wielokrotna regresja liniowa jest podstawowym i standardowym podejściem, w którym badacze używają wartości kilku zmiennych w celu wyjaśnienia lub przewidzenia wartości średnich wyniku ilościowego. Jednakże w wielu okolicznościach jesteśmy bardziej zainteresowani medianą lub arbitralnym kwantylem wyniku ilościowego.

Regresja kwantylowa modeluje relację między zbiorem predyktorów (zmiennych niezależnych) i konkretnych centyli (lub kwantyli) zmiennej przewidywanej (zależnej), najczęściej medianę. Ma dwie główne przewagi nad zwykłą regresją metodą najmniejszych kwadratów:

- W regresji kwantylowej nie przyjmuje się żadnych założeń dotyczących rozkładu zmiennej przewidywanej.
- Regresja kwantylowa jest odporna na wpływy obserwacji odstających

Regresja kwantylowa jest szeroko stosowana w takich branżach, jak ekologii, opieka zdrowotna i ekonometria.

Przykład

Jaki jest związek między całkowitym dochodem gospodarstwa domowego a odsetkiem dochodów wydatkowanych na żywność? Prawo Engela w ekonomii głosi, że w miarę wzrostu dochodów, część dochodów wydatkowanych na żywność spada, nawet jeśli wydatki na żywność bezwzględnie rosną. Stosując regresję kwantylową do tych danych, można określić, jakie wydatki na żywność ponosi 90% rodzin (dla 100 rodzin z danym dochodem), gdy nie interesują nas średnie wydatki na żywność.

Statystyki

Regresja kwantylowa, metoda Simplex, algorytm optymalizacji nieliniowej Frischa-Newtona metodą punktów wewnętrznych, Barrodale i Roberts, Bofinger, Hall Sheather, szerokość pasma, poziom istotności, operacje na macierzach, kryterium zbieżności, wagi regresji, składnik stałej, zmienna przewidywana, reszty predykcji, tabele, wykres predykcji, oszacowania parametrów, macierz kowariancji, macierz korelacji, wartości obserwowane, przedział ufności.

W tej procedurze wykorzystano algorytmy zaproponowane przez Koenker, R. W. i Bassett, G. W. (1978). Regression quantiles, *Econometrica*, 46, 33–50.

Wymagania dotyczące danych dla regresji kwantylowej

Dane

Wymagana jest jedna liczbowa zmienna zależna. Zmienna przewidywana musi być zmienną ilościową. Predyktory mogą być zmiennymi ilościowymi lub sztucznymi zmiennymi w przypadku predyktorów jakościowych. Aby można było uruchomić analizę, wymagany jest wyraz wolny lub co najmniej jeden predyktor.

Założenia

Regresja kwantylowa nie czyni założeń dotyczących rozkładu zmiennej przewidywanej i jest odporna na wpływ obserwacji odstających.

Procedury pokrewne

Analiza kwantylowa jest pokrewna regresji metodą najmniejszych kwadratów.

Uzyskiwanie analizy regresji kwantylowej

1. Z menu wybierz:

Analiza > Regresja > Kwantylowa...

Okno dialogowe umożliwia określenie zmiennej przewidywanej, czynników, kowariancji i zmiennych wagi do zastosowania w regresji kwantylowej. W oknie dialogowym dostępna jest także opcja oszczędzania pamięci na potrzeby złożonych analiz lub dużych zbiorów danych.

2. Wybierz liczbową zmienną przewidywaną. Do uruchomienia analizy wymagana jest tylko jedna zmienna przewidywana. Dozwolone są tylko zmienne numeryczne.

3. Opcjonalnie wybierz jedną lub więcej zmiennych czynnikowych. Zmienne ilościowe nie są dozwolone.

4. Opcjonalnie wybierz jedną lub więcej współzmiennych. Zmienne łańcuchowe nie są dozwolone.

Uwaga: Jeśli listy **Czynniki** i **Współzmienne** są puste, a w oknie dialogowym Model została wybrana opcja **Uwzględnij wyraz wolny w modelu**, zostanie wyświetlony następujący komunikat:

```
No effects have been specified. Therefore, an intercept only model will be fit.
Do you want to fit an intercept-only model?
```

5. Opcjonalnie wybierz zmienną wagi regresji. Zmienne łańcuchowe nie są dozwolone.

6. Opcjonalnie wybierz opcję **Zachowaj pamięć dla złożonych analiz lub dużych zestawów danych**. To ustawienie określa, czy dane mają być przechowywane w zewnętrznym pliku podczas przetwarzania. Włączenie tego ustawienia pozwala zaoszczędzić pamięć podczas wykonywania złożonych analiz lub analiz z dużymi zbiorami danych.

Regresja kwantylowa: Kryteria

Okno Kryteria udostępnia opcje dotyczące:

Kwantyl

Udostępnia opcje służące do określania kwantyli.

Określ pojedyncze kwantyle

Jeśli ta opcja jest wybrana, do wykonania analizy wymagana jest co najmniej jedna wartość. Dozwolonych jest wiele wartości, a każda wartość musi należeć do przedziału $[0, 1]$. Można określić wiele odległości rozdzielonych spacjami. Do pracy z wartościami na liście wartości kwantyli można używać przycisków **Dodaj**, **Zmień** i **Usuń**.

Wszystkie wartości muszą być unikalne (zduplikowane wartości nie są dozwolone). Wartość domyślna to 0, 5.

Określ kwantyle siatki

Gdy ta opcja jest wybrana, można wybrać siatkę kwantyli z przedziału od wartości **Początek** (wartość1) do wartości **Koniec** (wartość2) z przyrostem **Co** (wartość3). Dozwolony jest tylko

jeden poprawny zbiór [wartość1 DO wartość2 CO wartość3]. Musi on spełniać warunek $0 \leq \text{wartość1} \leq \text{wartość2} \leq 1$. Przypadki, w których $\text{wartość1} = \text{wartość2}$, są równoważne określeniu jednej wartości1, niezależnie od wartości3.

Metoda estymacji

Udostępnia opcje do określania metody estymacji modelu.

Automatycznie wybrane przez program

Umożliwia automatyczne wybranie odpowiedniej metody estymacji. To jest ustawienie domyślne.

Algorytm Simplex

Wywołuje algorytm Simplex, którego autorami są Barrodale i Roberts.

Nieliniowa optymalizacja Frischa-Newtona metodą punktów wewnętrznych

Wywołuje nieliniową optymalizację Frischa-Newtona metodą punktów wewnętrznych.

Po estymacji

Udostępnia opcje po estymacji wariancji-kowariancji oszacowań parametrów i przedziałów ufności dla przewidzianych wartości zmiennej przewidywanej.

Przyjmij, że obserwacje są IID

Gdy to ustawienie jest wybrane, przyjmuje się, że składniki błędów mają rozkład niezależny i identyczny. Gdy to ustawienie nie jest wybrane, czas obliczeń może znacznie wzrosnąć w przypadku dużych modeli. Domyślnie to ustawienie jest wybrane.

Typ szerokości pasma

Określa metodę szerokości pasma używaną do oszacowania macierzy wariancji-kowariancji estymacji parametrów (**Bofinger** lub **Hall-Sheather**). Ustawienie domyślne to **Bofinger**.

Metoda numeryczna

Dostępne są następujące opcje:

Tolerancja osobliwości

Określa wartość tolerancji dla manipulacji macierzowych w metodzie punktów wewnętrznych. Podana wartość musi być jedną wartością double (0, 10^{-3}), a wartością domyślną jest 10^{-12} .

Zbieżność

Określa kryterium zbieżności metody liczbowej. Podana wartość musi być jedną wartością double (0, 10^{-3}), a wartością domyślną jest 10^{-6} .

Maksymalna liczba iteracji

Określa maksymalną liczbę iteracji. Podana wartość musi być jedną dodatnią liczbą całkowitą. Wartość domyślna to 2000.

Braki danych

Udostępnia opcje określające sposób obsługi braków danych.

Wyklucz systemowe i zdefiniowane przez użytkownika braki danych

Jeśli ta opcja jest wybrana, wykluczane są zarówno braki danych użytkownika, jak i systemowe braki danych.

Zdefiniowane braki danych są traktowane jako ważne wartości

Po wybraniu tej opcji braki danych użytkownika są traktowane jako poprawne.

Przedział ufności (%)

Określa poziom istotności. Jeśli ta wartość jest określona, wartość musi być pojedynczą wartością typu double z zakresu od 0 do 100. Wartość domyślna wynosi 95.

Regresja kwantylowa: Model

Okno dialogowe Model zawiera opcje służące do określania efektów i wag używanych w modelu. W przypadku określenia ich przez sam model lub pominięcia model będzie zawierał wyraz wolny i wszystkie efekty główne z współzmiennymi na liście współzmiennych oraz czynnikami na liście czynników.

Określ efekty modelu

Model domyślny obejmuje tylko wyraz wolny, dlatego należy jawnie określić inne efekty modelu. Zamiast tego można utworzyć składniki zagnieżdżone lub niezagnieżdżone. Po wybraniu opcji **Buduj składniki** dla składników niezagnieżdżonych dostępne są następujące opcje efektów i interakcji.

Efekty główne

Dla każdej wybranej zmiennej tworzy składnik efektów głównych.

Interakcje

Dla wszystkich wybranych zmiennych tworzy składnik interakcji najwyższego rzędu.

Czynnikowy

Tworzy wszystkie możliwe interakcje i efekty główne wybranych zmiennych.

Wszystkie 2. rzędu

Tworzy wszystkie możliwe interakcje drugiego rzędu wybranych zmiennych.

Wszystkie 3. rzędu

Tworzy wszystkie możliwe interakcje trzeciego rzędu wybranych zmiennych.

Wszystkie 4. rzędu

Tworzy wszystkie możliwe interakcje czwartego rzędu wybranych zmiennych.

Wszystkie 5. rzędu

Tworzy wszystkie możliwe interakcje piątego rzędu wybranych zmiennych.

Jeśli wybrano opcję **Buduj zagnieżdżone składniki**, można zbudować składniki zagnieżdżone. Składniki zagnieżdżone są przydatne do modelowania efektu czynników lub współzmiennych, których wartości nie wchodzi w interakcje z poziomami innego czynnika. Na przykład sieć sklepów spożywczych może analizować zwyczaje zakupowe swoich klientów w kilku sklepach. Ponieważ każdy klient bywa regularnie tylko w jednym z tych sklepów, efekt *Klient* jest **zagnieżdżony w** efekcie *Lokalizacja sklepu*.

Ponadto można uwzględnić efekty interakcji, takie jak składniki wielomianowe z tą samą współzmienną, lub dodać wiele poziomów zagnieżdżenia do składnika zagnieżdżonego.

Uwagi dotyczące zagnieżdżenia:

- Chcąc uwzględnić efekt interakcji między dwoma czynnikami, należy użyć słowa kluczowego BY lub gwiazdki (*), aby dołączyć czynniki, które biorą udział w interakcji.
- Czynniki wewnątrz efektu interakcji muszą być różne.
- Użyj par nawiasów, aby włączyć efekt poprzez zagnieżdżenie jednego składnika w innym.
- Gdy obecna jest więcej niż jedna para nawiasów, każda para musi być zawarta (zagnieżdżona) w innej.
- Dozwolone jest wielokrotne zagnieżdżanie.
- Interakcje między efektami zagnieżdżonymi nie są obsługiwane.

Ograniczenia: W odniesieniu do składników zagnieżdżonych obowiązują następujące ograniczenia:

- Wszystkie czynniki w interakcji muszą być unikalne. A zatem, jeśli A jest czynnikiem, to niedozwolone jest określenie $A*A$.
- Wszystkie czynniki w efekcie zagnieżdżonym muszą być unikalne. A zatem, jeśli A jest czynnikiem, to niedozwolone jest określenie $A(A)$.
- Efekt nie może być zagnieżdżony w obrębie współzmiennej. A zatem, jeśli A jest czynnikiem, a X jest współzmienną, to określenie $A(X)$ jest niedozwolone.

Uwzględnij wyraz wolny w modelu

Jeśli ta opcja zostanie wybrana, wyraz wolny będzie uwzględniany w modelu. Jeśli ta opcja nie jest wybrana, do wykonania analizy wymagany jest co najmniej jeden predyktor. Ustawienie jest domyślnie włączone.

Regresja kwantylowa: Prezentacja

Okno dialogowe Wyświetl udostępnia ustawienia wyników i sterowania wykresami.

Drukowanie

Dostępne są następujące opcje wyników.

Oszacowania parametrów

Wyświetlane są oszacowania parametrów i odpowiednie statystyki testu oraz przedziały ufności. Opcjonalnie oprócz surowych oszacowań parametrów można wyświetlić wykładnicze oszacowania parametrów.

Macierz kowariancji szacowanych parametrów

Wyświetla macierz kowariancji szacowanych parametrów.

Macierz korelacji szacowanych parametrów

Wyświetla macierz korelacji oszacowanych parametrów.

Wykres i tabela

Dostępne są następujące opcje wykresów:

Wykreśl oszacowania parametrów dla

Można wybrać wykres oszacowania parametrów dla określonej liczby najistotniejszych efektów lub dla wszystkich efektów w modelu. Ustawienie liczby górnych (najistotniejszych) efektów określa liczbę kategorii lub kombinacji kategorii w efekcie mieszanym, która będzie kreślona w ramach interakcji jednej współzmiennnej i jednego lub dwóch czynników. Wartość musi być jedną dodatnią liczbą całkowitą (wartością domyślną jest 50).

Uwagi:

- Gdy podana liczba całkowita jest większa niż liczba kategorii lub kombinacji, wykresy predykcji są tworzone dla wszystkich efektów.
- To ustawienie jest poprawne tylko wtedy, gdy dla ustawienia **Wartości kwantyli** w oknie Kryteria określono wiele wartości. Gdy określona jest jedna wartość kwantyla, nie są tworzone żadne wykresy.

Wyświetl wykres przewidywanych wobec obserwowanych

Steruje tworzeniem wykresu wartości przewidywanych wobec obserwowanych. Gdy ta opcja jest włączona, tworzony jest pojedynczy wykres zawierający punkty (z różnymi kolorowymi kropkami reprezentującymi różne kwantyle). To ustawienie jest domyślnie wyłączone.

Przewiduj efekty w modelu

Jeśli to ustawienie jest włączone, dostępne są następujące opcje:

Wykreśl lub wyświetl tabelę x najwyższych efektów

Określ liczbę najwyższych efektów, których wykres predykcji lub tabela predykcji zostanie utworzona. Wartość 3 jest wartością domyślną.

Uwaga: Gdy podana liczba całkowita jest większa niż liczba poprawnych efektów w modelu, wykresy predykcji lub tabele predykcji są tworzone dla wszystkich efektów.

Wykreśl lub wyświetl tabelę efektów określonych przez użytkownika

Uwagi dotyczące poprawnych efektów:

- Efekt z jedną współzmienną (w tym dużą mocą samej współzmiennnej): utwórz jeden wykres zawierający przewidywane linie dla różnych kwantyli.
- Efekt z jednym czynnikiem: utwórz tabelę predykcji dla kategorii czynników dla różnych kwantyli.
- Efekt z interakcją dwóch czynników: dla każdego kwantyla utwórz tabelę predykcji dla kategorii dwóch czynników.
- Efekt interakcji jednej współzmiennnej i jednego lub dwóch czynników: dla każdego kwantyla utwórz wykres zawierający linie dla każdej kategorii lub kombinacji kategorii w ramach efektu interakcji.

- Maksymalna liczba kombinacji do wykreślenia zależy od wartości ustawienia **Wykreśl maksymalnie xx kateg. lub kombinac. kategorii w efekcie mieszanym**.

Efekty, które zostały przeniesione z listy **Efekty modelu** do listy **Linie predykcji**, są uwzględniane na wykresie. Wykresy nie są tworzone w przypadkach, gdy określone efekty są stałe (są usuwane z procesu budowania modelu).

Efekty, które zostały przeniesione z listy **Efekty modelu** do listy **Tabele predykcji** są uwzględniane przy tworzeniu tabel. Tabele nie są tworzone w przypadkach, gdy określone efekty są stałe (są usuwane z procesu budowania modelu).

Wykreśl maksymalnie xx kateg. lub kombinac. kategorii w efekcie mieszanym

Określa maksymalną liczbę kombinacji kategorii do wykreślenia. Wartością domyślną jest 10.

Regresja kwantylowa: Zapisz

Okno dialogowe Zapisz udostępnia opcje oceny modelu.

Przewidywana wartość odpowiedzi

Gdy ta opcja jest zaznaczona, oceniane są wartości zmiennej przewidywanej.

Reszta

Gdy ta opcja jest zaznaczona, oceniane są reszty predykcji.

Dolna granica przedziału predykcji

Gdy ta opcja jest zaznaczona, oceniane są dolne granice przedziałów predykcji.

Górna granica przedziału predykcji

Gdy ta opcja jest zaznaczona, oceniane są górne granice przedziałów predykcji.

Uwaga: Dla każdej opcji zapisywania można określić nazwę zmiennej. Jeśli określono trzon nazwy, musi on być poprawną nazwą zmiennej. Po określeniu wielu wartości dla ustawienia **Wartości kwantyli** w oknie dialogowym Kryteria używany jest trzon nazwy, po którym następuje znak podkreślenia „_” i zroszumiaty dla człowieka przyrostek określający kwantyl.

Regresja kwantylowa: Eksportuj

Okno dialogowe Eksportuj udostępnia opcje określające, które statystyki są eksportowane, sposób eksportowania statystyk (zewnętrzny plik danych lub zbiory danych) oraz sposób obsługi danych podczas przetwarzania (normalne przetwarzanie lub wstrzymywanie w zewnętrznym pliku tymczasowym na czas przetwarzania).

Oceny parametrów macierzy kowariancji

Jeśli ta opcja jest wybrana, włączone są opcje zapisu macierzy kowariancji oszacowań parametrów w zewnętrznym pliku danych lub wcześniej zadeklarowanym zbiorze danych.

Oceny parametrów macierzy korelacji

Jeśli ta opcja jest wybrana, włączone są opcje zapisu macierzy korelacji oszacowań parametrów w zewnętrznym pliku danych lub wcześniej zadeklarowanym zbiorze danych.

Macierz korelacji/kowariancji zostanie zapisana w jednym zbiorze danych lub zewnętrznym pliku w obecności wielu kwantyli regresji

W przypadku obecności wielu kwantyli ta opcja przetacza zapisywanie macierzy kowariancji/korelacji do jednego albo do wielu zbiorów danych lub zewnętrznych plików danych. Gdy ta opcja nie jest włączona, macierze są zapisywane w pojedynczym zewnętrznym pliku danych lub zbiorze danych. Gdy ta opcja jest włączona, macierze są zapisywane w wielu zewnętrznych plikach danych lub zbiorach danych. To ustawienie odnosi skutek tylko wtedy, gdy dla ustawienia **Wartości kwantyli** w oknie dialogowym Kryteria określono wiele wartości.

Uwaga: Ta opcja jest dostępna tylko wtedy, gdy wybrana jest opcja **Macierz kowariancji oszacowań parametrów** lub **Macierz korelacji oszacowań parametrów**.

Eksportuj informacje o modelu do pliku XML

Po wybraniu udostępnia opcje eksportowania informacji o modelu do pliku XML o wskazanej nazwie i położeniu.

Eksportuj jako XML

Po wybraniu opcji **Eksportuj informacje o modelu do pliku XML** można wybrać opcję eksportowania zarówno oszacowań parametrów, jak i macierzy kowariancji, albo tylko oszacowań parametrów. Domyślne ustawienie to **Oszacowania parametrów i macierz kowariancji**.

Konwencje dotyczące nazw plików

- Jeśli dla ustawienia **Wartości kwantyli** w oknie dialogowym Kryteria określona jest pojedyncza wartość, nazwy pliku danych zewnętrznych lub zbiory danych zawierają odpowiednio słowa `savefile` lub `dataset`.
- Jeśli dla ustawienia **Wartości kwantyli** w oknie dialogowym Kryteria określono wiele wartości, każdy kwantyl jest zapisywany w jednym zewnętrznym pliku danych lub zbiorze danych.
- Do nazwy pliku danych lub nazwy zbioru danych automatycznie dodawany jest znak podkreślenia „_”, po którym następuje zrostwiek dla człowieka przyrostek kwantyla. Na przykład, jeśli w polu **Wartości kwantyli** zostaną określone wartości 0,25, 0,50 i 0,75, do nazw plików danych (przed rozszerzeniem `.sav`) zostaną dołączone przyrostki `_25`, `_50` i `_75`.
- Dla każdego przyrostka kwantyla można podać dodatkowe cyfry (jeśli jest to konieczne).
- Zero wiodące ani kropka dziesiętna z pola **Wartości kwantyli** nie są używane w przyrostku.
- Gdy w polu **Wartości kwantylu** stosowana jest notacja naukowa, jest ona przed przeniesieniem do przyrostka przekształcana na wartość dziesiętną.

Regresja nieliniowa

Regresja nieliniowa jest metodą wyszukiwania nieliniowego modelu relacji między zmienną zależną a zbiorem zmiennych niezależnych. W odróżnieniu od tradycyjnej regresji liniowej, która ograniczona jest do estymacji modeli liniowych, regresja nieliniowa może szacować modele z arbitralnymi relacjami pomiędzy zmiennymi niezależnymi a zmiennymi zależnymi. Realizuje się to przy użyciu iteracyjnych algorytmów estymacji. Należy zauważyć, że ta procedura nie jest potrzebna w przypadku prostych modeli wielomianowych w postaci $Y = A + BX^2$. Definiując $W = X^2$, otrzymujemy prosty model liniowy, $Y = A + BW$, który można oszacować za pomocą tradycyjnych metod, takich jak regresja liniowa.

Przykład. Czy liczba ludności może być przewidywana na podstawie czasu? Z wykresu rozrzutu wynika, że prawdopodobnie istnieje silna relacja między liczbą ludności a czasem, ale relacja ta jest nieliniowa, dlatego wymaga zastosowania specjalnych metod estymacji w postaci regresji nieliniowej. Poprzez ustanowienie odpowiedniego równania, takiego jak logistyczny model wzrostu liczby ludności, możemy uzyskać dobre oszacowanie modelu, co pozwoli nam na predykcję liczby ludności w okresach, w których faktycznie nie została ona zmierzona.

Statystyki. Dla każdej iteracji: oszacowania parametrów i resztowa suma kwadratów. Dla każdego modelu: suma kwadratów dla regresji, reszta, nieskorygowana i skorygowana suma, oszacowania parametrów, asymptotyczne błędy standardowe oraz asymptotyczna macierz korelacji oszacowań parametrów.

Uwaga: Ograniczona regresja nieliniowa szacuje parametry modelu przy użyciu algorytmów, które zaproponowali i zaimplementowali w programie NPSOL[®] Gill, Murray, Saunders i Wright.

Wymagania dotyczące danych dla regresji nieliniowej

Dane. Zmienne zależne i niezależne powinny być zmiennymi ilościowymi. Zmienne jakościowe, takie jak wyznanie, dyscyplina studiów lub miejsce zamieszkania, muszą być zakodowane w zmiennych binarnych (sztucznych) lub innych zmiennych kontrastowych.

Założenia. Wyniki są ważne tylko wtedy, gdy określono funkcję dokładnie opisującą relację między zmiennymi zależnymi i niezależnymi. Ponadto bardzo ważny jest dobór dobrych wartości początkowych. Nawet jeśli określono poprawną formę funkcjonalną modelu, w przypadku użycia nieodpowiednich wartości początkowych model może nie być zbieżny lub można uzyskać rozwiązanie optymalne lokalnie, a nie globalnie.

Procedury pokrewne. Wiele modeli, które na pierwszy rzut oka wydają się być nieliniowe, można przekształcić w modele liniowe odpowiednie do analizy za pomocą regresji liniowej. W razie wątpliwości co do właściwego modelu procedura estymacji krzywej może pomóc w identyfikacji użytecznych relacji funkcjonalnych w danych.

Uzyskiwanie analizy metodą regresji nieliniowej

1. Z menu wybierz:

Analizuj > Regresja > Nieliniowa...

- Wybierz jedną liczbową zmienną zależną z listy zmiennych w aktywnym zbiorze danych.
- Aby zbudować wyrażenie modelu, wprowadź wyrażenie w polu **Wyrażenie modelu** lub wklej komponenty (zmienna, parametry, funkcje) do pola.
- Aby zidentyfikować parametry w modelu, kliknij opcję **Parametry**.

Model segmentowany (taki, który pobiera różne postacie w różnych częściach swojej dziedziny) musi być zdefiniowany za pomocą logiki warunkowej w ramach jednej instrukcji modelu.

Logika warunkowa (regresja nieliniowa)

Używając logiki warunkowej, można określić model posegmentowany. Aby użyć logiki warunkowej w wyrażeniu modelu lub funkcji straty, należy utworzyć sumę szeregu składników, po jednym dla każdego warunku. Każdy składnik powinien mieć postać wyrażenia logicznego (w nawiasach) pomnożonego przez wyrażenie, które powinno być wynikiem, gdy wyrażenie logiczne jest prawdziwe.

Rozważmy na przykład model segmentowany równy 0 dla $X \leq 0$, X dla $0 < X < 1$, and 1 for $X \geq 1$. Odpowiednie wyrażenie jest następujące:

$$(X \leq 0) * 0 + (X > 0 \ \& \ X < 1) * X + (X \geq 1) * 1.$$

Wszystkie wyrażenia logiczne w nawiasach przyjmują wartość 1 (prawda) albo 0 (fałsz). Zatem:

Jeśli $X \leq 0$, wyrażenie wynosi $1 * 0 + 0 * X + 0 * 1 = 0$.

Jeśli $0 < X < 1$, wyrażenie wynosi $0 * 0 + 1 * X + 0 * 1 = X$.

Jeśli $X \geq 1$, to zmniejsza się do $0 * 0 + 0 * X + 1 * 1 = 1$.

Można łatwo tworzyć bardziej skomplikowane przykłady, podstawiając różne wyrażenia logiczne i wyrażenia wynikowe. Należy pamiętać, że podwójne nierówności, takie jak $0 < X < 1$, must be written as compound expressions, such as $(X > 0 \ \& \ X < 1)$.

W wyrażeniach logicznych można używać zmiennych łańcuchowych:

$$(\text{city} = \text{'Warszawa'}) * \text{costliv} + (\text{city} = \text{'Lublin'}) * 0.59 * \text{costliv}$$

W ten sposób powstanie inne wyrażenie (z wartością kosztów życia, *costliv*) dla warszawiaków, a inne (59% tej wartości) dla mieszkańców Lublina. Stałe łańcuchowe muszą być ujęte w cudzysłów lub apostrofy, tak jak to pokazano w przykładzie.

Regresja nieliniowa: Parametry

Parametry to części modelu, które są estymowane przez procedurę regresji nieliniowej. Parametry mogą być stałymi addytywnymi, współczynnikami multiplikatywnymi, wykładnikami lub wartościami używanymi podczas wyznaczania wartości funkcji. Wszystkie zdefiniowane parametry zostaną wyświetlone (wraz z ich wartościami początkowymi) na liście Parametry w głównym oknie dialogowym.

Name. Dla każdego parametru należy podać nazwę. Nazwa ta musi być poprawną nazwą zmiennej i musi być nazwą używaną w wyrażeniu modelu w głównym oknie dialogowym.

Wartość początkowa. Umożliwia określenie wartości początkowej parametru, najlepiej jak najbliższej spodziewanego ostatecznego rozwiązania. Nieodpowiednie wartości początkowe mogą spowodować brak zbieżności bądź uzyskanie rozwiązania, które jest lokalne (a nie globalne) lub jest fizycznie niemożliwe.

Użyj wartości początkowych z poprzedniej analizy. Jeśli z tego okna dialogowego została już wcześniej uruchomiona regresja nieliniowa, można wybrać tę opcję, aby jako początkowych wartości parametrów użyć ich wartości z poprzedniego uruchomienia. Umożliwia to kontynuowanie poszukiwań rozwiązania w sytuacji, gdy zbieżność następuje powoli. (Początkowe wartości będą nadal wyświetlane na liście Parametry w głównym oknie dialogowym).

Uwaga: Wybór ten jest zachowywany w tym oknie dialogowym w pozostałej części sesji. W przypadku zmiany modelu należy koniecznie usunąć zaznaczenie tej opcji.

Typowe modele regresji nieliniowej

W poniższej tabeli przedstawiono przykładowe komendy modelu dla wielu opisanych w literaturze modeli regresji nieliniowej. Losowo wybrany model prawdopodobnie nie będzie dobrze dopasowany do analizowanych danych. Potrzebne są odpowiednie wartości początkowe parametrów, a niektóre modele do osiągnięcia zbieżności wymagają ograniczeń.

Tabela 1. Przykładowe komendy modelu

Nazwa	Równanie modelu
Regresja asymptotyczna	$b1 + b2 * \exp(b3 * x)$
Regresja asymptotyczna	$b1 - (b2 * (b3 ** x))$
Gęstość	$(b1 + b2 * x) ** (-1 / b3)$
Gaussa	$b1 * (1 - b3 * \exp(-b2 * x ** 2))$
Gompertz	$b1 * \exp(-b2 * \exp(-b3 * x))$
Johnson-Schumacher	$b1 * \exp(-b2 / (x + b3))$
Zmodyfikowany logarytmiczny	$(b1 + b3 * x) ** b2$
Log-logistyczny	$b1 - \ln(1 + b2 * \exp(-b3 * x))$
Prawo Metcherlicha przyrostów mniej niż proporcjonalnych	$b1 + b2 * \exp(-b3 * x)$
Michaelisa Mentena	$b1 * x / (x + b2)$
Morgan-Mercer-Florin	$(b1 * b2 + b3 * x ** b4) / (b2 + x ** b4)$
Pearl-Reed	$b1 / (1 + b2 * \exp(-(b3 * x + b4 * x ** 2 + b5 * x ** 3)))$
Stosunek sześciątów	$(b1 + b2 * x + b3 * x ** 2 + b4 * x ** 3) / (b5 * x ** 3)$
Stosunek kwadratów	$(b1 + b2 * x + b3 * x ** 2) / (b4 * x ** 2)$
Richards	$b1 / ((1 + b3 * \exp(-b2 * x)) ** (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1 ** (1 - b4) - b2 * \exp(-b3 * x)) ** (1 / (1 - b4))$
Weibulla	$b1 - b2 * \exp(-b3 * x ** b4)$
Gęstość wyników	$(b1 + b2 * x + b3 * x ** 2) ** (-1)$

Regresja nieliniowa: Funkcja straty

Funkcja straty w regresji nieliniowej jest to funkcja, która jest minimalizowana przez algorytm. Należy wybrać opcję **Suma kwadratów reszt**, aby zminimalizować sumę kwadratów reszt, albo opcję **Funkcja straty zdefiniowana przez użytkownika**, aby zminimalizować inną funkcję.

Jeśli zostanie wybrana opcja **Funkcja straty zdefiniowana przez użytkownika**, należy zdefiniować funkcję straty, której suma (ze wszystkich obserwacji) powinna być minimalizowana przez wybór wartości parametrów.

- Większość funkcji strat obejmuje zmienną specjalną *RESID_*, która reprezentuje resztę. (Domyślna funkcja straty, którą jest suma kwadratów reszt, może zostać wprowadzona jawnie jako *RESID_**2*). Jeśli konieczne jest użycie wartości przewidywanej w funkcji straty, jest ona równa wartości zmiennej zależnej pomniejszonej o resztę.
- Można zdefiniować warunkową funkcję straty za pomocą logiki warunkowej.

Można wpisać wyrażenie w polu Funkcja straty zdefiniowana przez użytkownika albo wkleić komponenty wyrażenia do pola. Stałe łańcuchowe muszą być ujęte w cudzysłowy lub apostrofy, a stałe liczbowe muszą być wpisane w formacie amerykańskim, z kropką dziesiętną.

Regresja nieliniowa: Ograniczenia parametrów

Ograniczenie dotyczy dozwolonej wartości parametru podczas iteracyjnego poszukiwania rozwiązania. Wartości wyrażen liniowych są wyznaczone przed wykonaniem kroku, dlatego można użyć ograniczeń liniowych, aby zapobiec wykonywaniu kroków, które mogą spowodować wygenerowanie zbyt dużej wartości. Wartości wyrażen nieliniowych są wyznaczone po wykonaniu kroku.

Każde równanie lub nierówność musi zawierać następujące elementy:

- Wyrażenie, w którym występuje co najmniej jeden parametr modelu. Wpisz wyrażenie lub użyj bloku klawiszy, który umożliwi wklejanie liczb, operatorów lub nawiasów do wyrażenia. Można wpisać wymagane parametry razem z pozostałą częścią wyrażenia albo wkleić elementy z listy Parametry po lewej stronie. W ograniczeniu nie można używać zwykłych zmiennych.
- Jeden z trzech operatorów logicznych \leq , $=$, or \geq .
- Stałą liczbową, z którą porównywane jest wyrażenie za pomocą operatora logicznego. Wpisz stałą. Stałe liczbowe muszą być wpisywane w formacie amerykańskim, gdzie kropka jest separatorem dziesiętnym.

Regresja nieliniowa: Zapisz zmienne wynikowe

Istnieje możliwość zapisania szeregu zmiennych wynikowych w aktywnym pliku danych. Dostępne opcje to Reszty, Wartości przewidywane, Pochodne i Wartości funkcji straty. Zmienne te mogą zostać użyte w kolejnych analizach w celu przetestowania dopasowania modelu lub do zidentyfikowania problematycznych obserwacji.

- *Residuals*. Powoduje zapisanie reszty w zmiennej o nazwie *resid*.
- *Wartości przewidywane*. Powoduje zapisanie wartości przewidywanych w zmiennej o nazwie *pred_*.
- *Pochodne*. Jeden instrument pochodny jest zapisywany dla każdego parametru modelu. Nazwy pochodnych są tworzone przez dodanie przedrostka „d.” do pierwszych sześciu znaków nazw parametrów.
- *Wartości funkcji straty*. Ta opcja jest dostępna, jeśli użytkownik określi własną funkcję utraty danych. Do wartości funkcji strat przypisywana jest nazwa zmiennej *loss*.

Regresja nieliniowa: Opcje

Opcje umożliwiają wpływ na różne aspekty analizy metodą regresji nieliniowej:

Oceny metodą bootstrap. Metoda estymacji standardowego błędu statystyki z użyciem powtarzających się próbek z oryginalnego zestawu danych. Polega na doborze prób (z wymianą obserwacji) w celu uzyskania dużej liczby prób tej samej wielkości, co pierwotny zbiór danych. Dla każdej z takich prób przeprowadzane jest oszacowanie równania nieliniowego. Następnie obliczany jest błąd standardowy oszacowania każdego parametru jako odchylenie standardowe ocen metodą bootstrap. Za wartości początkowe każdej próby bootstrap przyjmowane są wartości parametrów pierwotnego zbioru danych. Wymaga to użycia algorytmu sekwencyjnego programowania kwadratowego.

Metoda oszacowania. Pozwala na wybór metody estymacji, o ile jest to możliwe. (Niektóre wybory w tym lub innych oknach dialogowych wymagają algorytmu sekwencyjnego programowania kwadratowego). Dostępne alternatywy to: Sekwencyjne programowanie kwadratowe i algorytm Levenberga-Marquardta.

- *Sekwencyjne programowanie kwadratowe*. Ta metoda jest dostępna dla modeli ograniczonych i nieograniczonych. Metoda sekwencyjnego programowania kwadratowego jest używana automatycznie, gdy określony zostanie model ograniczony, zdefiniowany przez użytkownika funkcją strat lub metodą bootstrap. Można wprowadzić nowe wartości w pola Maksymalna liczba iteracji i Ograniczenie kroku, można także zmienić wybór na listach rozwijanych Tolerancja optymalizacji, Precyzja funkcji i Nieskończona wielkość kroku.
- *Levenberg-Marquardt*. Jest to domyślny algorytm dla modeli nieograniczonych. Metoda Levenberga-Marquardta nie jest dostępna, gdy określony zostanie model ograniczony, zdefiniowany przez użytkownika funkcją strat lub metodą bootstrap. W polu Maksymalna liczba iteracji można wpisać nowe wartości, można także zmienić wybór na rozwijanych listach Zbieżność sumy kwadratów i Zbieżność parametru.

Interpretowanie wyników regresji nieliniowej

Problemy regresji nieliniowej często stwarzają trudności obliczeniowe:

- Wybór wartości początkowych dla parametrów wpływa na zbieżność. Należy wybrać rozsądne wartości początkowe, które, jeśli to możliwe, powinny być zbliżone do oczekiwanego ostatecznego rozwiązania.
- Czasami jeden z algorytmów działa lepiej niż inny w przypadku konkretnego problemu. W oknie dialogowym Opcje wybierz inny algorytm, jeśli jest dostępny. (Jeśli zostanie określona funkcja straty lub niektóre typy ograniczeń, nie będzie można użyć algorytmu Levenberga-Marquardta).
- Jeśli iteracje zatrzymały się tylko dlatego, że wykonano już maksymalną liczbę iteracji, model „finalny” prawdopodobnie nie jest dobrym rozwiązaniem. Wybierz opcję **Użyj wartości początkowych z poprzedniej analizy** w oknie dialogowym Parametry, aby kontynuować iteracje, lub — co będzie lepszym rozwiązaniem — wybierz inne wartości początkowe.
- Modele, które wymagają potęgowania dużych wartości lub podnoszenia wartości do dużej potęgi mogą powodować nadmiary lub niedomiary (liczby tak duże albo tak małe, że komputer nie będzie w stanie ich zapisać). Niekiedy można tego uniknąć, wybierając odpowiednie wartości początkowe lub narzucając ograniczenia parametrów.

Dodatkowe właściwości komendy NLR

Język składni komend umożliwia również:

- Nazwa pliku, z którego mają być odczytywane wartości początkowe dla oszacowań parametrów.
- Określ więcej niż jedną instrukcję modelu i funkcję straty. Ułatwia to określenie modelu posegmentowanego.
- Określanie własnych pochodnych zamiast używania tych, które zostały obliczone przez program.
- Określenie liczby prób bootstrapowych do wygenerowania.
- Określenie dodatkowych kryteriów iteracji, w tym ustawienie wartości krytycznej dla sprawdzania pochodnych oraz zdefiniowanie kryterium zbieżności dla korelacji między resztami a pochodnymi.

Dodatkowe kryteria dla komendy CNLR (regresja nieliniowa z ograniczeniami) umożliwiają:

- Określenie maksymalnej liczby iteracji drugorzędnych dozwolonych w każdej głównej iteracji.
- Ustawienie wartości krytycznej dla sprawdzania pochodnych.
- Ustawienie limitu kroków.
- Określenie tolerancji w celu ustalenia, czy wartości początkowe mieszczą się w określonych granicach.

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

Estymacja wag

W standardowych modelach regresji liniowej zakłada się, że wariancja jest stała w badanej populacji. Jeśli tak nie jest (na przykład, gdy obserwacje z wysokimi wartościami jakiegoś atrybutu wykazują większą zmienność niż obserwacje, z niskimi wartościami tego atrybutu), regresja liniowa z wykorzystaniem zwykłych najmniejszych kwadratów (OLS) nie zapewnia już optymalnych oszacowań modelu. Jeśli różnice

w zmienności można przewidzieć na podstawie innej zmiennej, procedura szacowania wag może obliczyć współczynniki modelu regresji liniowej przy użyciu ważonych najmniejszych kwadratów (WNK), tak aby bardziej precyzyjne obserwacje (tj. te z mniejszą zmiennością) miały większą wagę przy określaniu współczynników regresji. Procedura szacowania wag testuje zakres przekształceń wagowych i wskazuje, które z nich będzie najlepiej dopasowane do danych.

Przykład. Jakie jest wpływ inflacji i bezrobocia na zmiany cen akcji? Ponieważ akcje o wysokiej wartości często wykazują większą zmienność niż te o niskiej wartości, metoda zwykłych najmniejszych kwadratów nie będzie generować optymalnych oszacowań. Szacowanie wag umożliwia uwzględnienie wpływu cen akcji na zmienność cen podczas obliczania modelu liniowego.

Statystyki. Wartości logarytmu wiarygodności dla każdej potęgi testowanej zmiennej źródłowej, wielorakie R , R kwadrat, skorygowane R kwadrat, tabela ANOVA dla modelu WLS, niestandardyzowane i standardyzowane oszacowania parametrów oraz logarytm wiarygodności dla modelu WLS.

Wymagania dotyczące danych przy szacowaniu wag

Dane. Zmienne zależne i niezależne powinny być zmiennymi ilościowymi. Zmienne jakościowe, takie jak wyznanie, dyscyplina studiów lub miejsce zamieszkania, muszą być zakodowane w zmiennych binarnych (sztucznych) lub innych zmiennych kontrastowych. Zmienna ważąca powinna być ilościowa i powinna być powiązana z zmiennością zmiennej zależnej.

Założenia. Dla każdej wartości zmiennej niezależnej rozkład zmiennej zależnej musi być normalny. Relacja między zmienną zależną a każdą zmienną niezależną powinna być liniowa, a wszystkie obserwacje powinny być niezależne. Wariancja zmiennej zależnej może być różna dla różnych poziomów zmiennych niezależnych, ale różnice muszą być przewidywalne na podstawie zmiennej ważącej.

Procedury pokrewne. Do monitorowania danych można również użyć procedury eksploracji. Dzięki procedurze eksploracji możliwe jest przeprowadzenie testów na normalność i jednorodność wariancji oraz uzyskanie graficznego przedstawienia danych. Jeśli zmienna zależna wydaje się mieć równe wariancje na różnych poziomach zmiennych niezależnych, można użyć procedury regresji liniowej. Jeśli dane użytkownika wydają się naruszać jakieś założenie (takie jak założenie o normalności rozkładu), należy spróbować je przekształcić. Jeśli dane nie są powiązane liniowo a przekształcenie nie pomaga, należy użyć modelu alternatywnego w procedurze estymacji krzywej. Jeśli zmienna zależna jest dychotomiczna (na przykład określa, czy dana sprzedaż została zrealizowana, lub czy dany produkt jest wadliwy), należy użyć procedury regresji logistycznej. Jeśli zmienna zależna jest oceniana (np. czas przeżycia po operacji), należy użyć tabel trwania życia, procedury Kaplana-Meiera lub regresji Coxa, dostępnych w modułach Custom Tables i Advanced Statistics. Jeśli dane nie są niezależne (np. obserwacje dotyczą tych samych osób w różnych warunkach), należy użyć procedury Powtarzane pomiary, dostępnej w modułach Custom Tables i Advanced Statistics.

Uzyskiwanie szacowania wag

1. Z menu wybierz:

Analiza > Regresja > Szacowanie wagi ...

2. Wybierz jedną zmienną zależną.

3. Wybierz co najmniej jedną zmienną niezależną.

4. Wybierz zmienną, która jest źródłem heteroskedastyczności, jako zmienną ważącą.

Zmienna ważąca

Dane są ważone na podstawie odwrotności tej zmiennej podniesionej do potęgi. Dla każdego określonego zakresu wartości wykładnika wyliczane jest równanie regresji i wskazywany jest wykładnik maksymalizujący funkcję logarytmu wiarygodności.

Wykładnik potęgi od

Ta opcja jest używana w połączeniu z zmienną wagi do obliczania wag. Do każdej wartości z zakresu wykładnika dopasowane zostanie po jednym równaniu regresji. Wartości wpisane w polu testu Wykładnik potęgi od i za pośrednictwem pola tekstowego muszą należeć do zakresu od -6,5

do 7,5 włącznie. Zakres wartości wykładnika od małego do dużego z przyrostami określonymi przez podaną wartość. Całkowita liczba wartości w zakresie wykładników nie może przekraczać 150.

Szacowanie wag: Opcje

Można określić opcje szacowania wag:

Zapisz najlepszą wagę jako nową zmienną. Dodaje zmienną ważącą do aktywnego pliku. Ta zmienna ma nazwę *WGT_n*, gdzie *n* jest numerem wybranym w celu nadania zmiennej unikalnej nazwy.

Pokaż statystyki ANOVA i oszacowania. Umożliwia określenie sposobu wyświetlania statystyk w wynikach. Dostępne są następujące alternatywy: Dla najlepszego wykładnika i Dla każdego wykładnika potęgi.

Dodatkowe właściwości komendy WNK

Język składni komend umożliwia również:

- Podanie jednej wartości mocy.
- Określenie listy wartości mocy lub kombinacji zakresu wartości z listą wartości mocy.

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

Procedura Podwójna MNK

W standardowych modelach regresji liniowej zakłada się, że błędy w zmiennej zależnej są nieskorelowane ze zmiennymi niezależnymi. Jeśli tak nie jest (na przykład, gdy relacje między zmiennymi są dwukierunkowe), regresja liniowa z wykorzystaniem zwykłych najmniejszych kwadratów (OLS) nie zapewnia już optymalnych oszacowań modelu. Regresja podwójną metodą najmniejszych kwadratów wykorzystuje zmienne instrumentalne, które nie są skorelowane ze składnikami błędów, do obliczenia wartości szacunkowych problematycznych predyktorów (pierwszy etap), a następnie wykorzystuje te obliczone wartości do oszacowania modelu regresji liniowej zmiennej zależnej (drugi etap). Ponieważ obliczone wartości są oparte na zmiennych, które nie są skorelowane z błędami, wyniki modelu podwójnego są optymalne.

Przykład. Czy popyt na towar jest związany z jego ceną i dochodami konsumentów? Trudność w tym modelu polega na tym, że cena i popyt mają wpływ na siebie nawzajem. Oznacza to, że cena może wpływać na popyt, a popyt może wpływać na cenę. Model regresji podwójną metodą najmniejszych kwadratów może wykorzystać dochody konsumentów i opóźnione ceny do obliczenia zastępczej wartości ceny, która będzie nieskorelowana z błędami w pomiarze popytu. Ta wartość zastępcza jest podstawiana zamiast ceny w pierwotnie określonym modelu, który jest następnie szacowany.

Statystyki. Dla każdego modelu: standaryzowane i niestandaryzowane współczynniki regresji, wieloraki R , R^2 , skorygowane R^2 , błąd standardowy estymacji, tabela analizy wariancji, wartości przewidywane i reszty. Ponadto 95% przedziały ufności dla każdego współczynnika regresji oraz macierze korelacji i kowariancji oszacowań parametrów.

Wymagania dotyczące danych dla regresji podwójną metodą najmniejszych kwadratów

Dane. Zmienne zależne i niezależne powinny być zmiennymi ilościowymi. Zmienne jakościowe, takie jak wyznanie, dyscyplina studiów lub miejsce zamieszkania, muszą być zakodowane w zmiennych binarnych (sztucznych) lub innych zmiennych kontrastowych. *Endogeniczne* zmienne objaśniające powinny być zmiennymi ilościowymi (nie jakościowymi).

Założenia. Dla każdej wartości zmiennej niezależnej rozkład zmiennej zależnej musi być normalny. Wariancja rozkładu zmiennej zależnej powinna być stała dla wszystkich wartości zmiennej niezależnej. Zależność między zmienną zależną a każdą zmienną niezależną powinna być liniowa.

Procedury pokrewne. Jeśli wiadomo, że żaden z predyktorów nie jest skorelowany z błędami w zmiennej zależnej, można użyć procedury Regresja liniowa. Jeśli dane wydają się naruszać jakieś założenie (takie

jak założenie o normalności rozkładu lub stałej wariancji), należy spróbować je przekształcić. Jeśli dane nie są powiązane liniowo a przekształcenie nie pomaga, należy użyć modelu alternatywnego w procedurze estymacji krzywej. Jeśli zmienna zależna jest dychotomiczna, np. określa, czy dana sprzedaż została zrealizowana, czy nie, należy użyć procedury regresji logistycznej. Jeśli dane nie są niezależne, np. obserwacje dotyczą tych samych osób w różnych warunkach, należy użyć procedury Powtarzane pomiary.

Uzyskiwanie regresji podwójną metodą najmniejszych kwadratów

1. Wybierz z menu następującą opcję:

Analizuj > Regresja > 2-Faza najmniejszych kwadratów ...

2. Wybierz jedną zmienną zależną.

3. Wybierz jedną lub więcej zmiennych objaśniających (predyktorów).

4. Wybierz co najmniej jedną zmienną instrumentalną.

- *Instrumentalne*. Są to zmienne używane do obliczenia przewidywanych wartości dla zmiennych endogennych w pierwszym etapie analizy najmniejszych kwadratów w dwóch etapach. Te same zmienne mogą być wyświetlane na listach Zmienne objaśniające i Zmienne instrumentalne. Liczba zmiennych instrumentalnych musi być co najmniej równa liczbie zmiennych wyjaśniających. Jeśli wszystkie wyświetlane zmienne wyjaśniające i instrumentalne są takie same, wyniki są takie same, jak wyniki z procedury regresji liniowej.

Zmienne objaśniające nieokreślone jako instrumentalne uznaje się za endogeniczne. Zwykle wszystkie zmienne egzogeniczne na liście Zmienne objaśniające są również określane jako zmienne instrumentalne.

Regresja podwójną metodą najmniejszych kwadratów: Opcje

Dostępne są następujące opcje analizy:

Zapisz zmienne wynikowe. Umożliwia dodanie nowych zmiennych do aktywnego pliku. Dostępne opcje to Przewidywane i Reszty.

Pokaż kowariancję parametrów. Umożliwia przedstawienie macierzy kowariancji oszacowań parametrów.

Dodatkowe właściwości komendy 2SLS

Język komend pozwala również na oszacowanie wielu równań jednocześnie. Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Schematy kodowania zmiennych jakościowych

W wielu procedurach można zażądać automatycznego zastąpienia niezależnej zmiennej jakościowej zbiorem kontrastów, które będą następnie wprowadzane do równania lub usuwane z równania jako jeden blok. Można określić sposób kodowania kontrastów, zwykle w opcji komendy CONTRAST. W niniejszym dodatku omówiono działanie różnych typów kontrastów żądanych w opcji komendy CONTRAST oraz przedstawiono odpowiednie przykłady.

Odchylenie

Odchylenie od średniej ogólnej. Macierzowa postać tych kontrastów jest następująca:

```

mean ( 1/k 1/k ... 1/k 1/k)
dF(1) (1-1/k -1/k ... -1/k -1/k)
dF(2) ( -1/k 1-1/k ... -1/k -1/k)
.
.
dF(k-1) ( -1/k -1/k ... 1-1/k -1/k)

```

gdzie k jest liczbą kategorii zmiennej niezależnej, a ostatnia kategoria jest domyślnie pominięta. Na przykład kontrasty odchylenia dla zmiennej niezależnej z trzema kategoriami są następujące:

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3  2/3 -1/3)
```

Aby pominąć kategorię inną niż ostatnia, podaj numer kategorii do pominięcia w nawiasach za słowem kluczowym DEVIATION. Na przykład poniższa opcja komendy wyznacza odchylenia z pierwszej i trzeciej kategorii, a pomija drugą:

```
/CONTRAST (FACTOR)=DEVIATION (2)
```

Założmy, że zmienna *factor* (czynnik) ma trzy kategorie: Oto wynikowa macierz kontrastów:

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3 -1/3  2/3)
```

Proste

Kontrasty proste. Porównuje każdy poziom czynnika z ostatnim. Ogólna postać macierzy jest następująca:

```
mean (1/k  1/k  ...  1/k  1/k)
df(1) (  1   0  ...   0  -1)
df(2) (  0   1  ...   0  -1)
.
.
df(k-1) (  0   0  ...   1  -1)
```

gdzie k jest liczbą kategorii zmiennej niezależnej. Na przykład proste kontrasty dla zmiennej niezależnej z czterema kategoriami są następujące:

```
(1/4  1/4  1/4  1/4)
(  1   0   0  -1)
(  0   1   0  -1)
(  0   0   1  -1)
```

Aby jako kategorii odniesienia użyć kategorii innej niż ostatnia, podaj numer kolejny kategorii odniesienia w nawiasach za słowem kluczowym SIMPLE. Numer kolejny nie musi być równy z wartością powiązaną z kategorią. Na przykład poniższa opcja komendy CONTRAST wyznacza macierz kontrastów z pominięciem drugiej kategorii:

```
/CONTRAST (FACTOR) = SIMPLE (2)
```

Założmy, że zmienna *factor* (czynnik) ma cztery kategorie: Oto wynikowa macierz kontrastów:

```
(1/4  1/4  1/4  1/4)
(  1  -1   0   0)
(  0  -1   1   0)
(  0  -1   0   1)
```

Helmerta

Kontrasty Helmerta. Porównuje kategorie zmiennej niezależnej ze średnią z następnych kategorii. Ogólna postać macierzy jest następująca:

```
mean (1/k  1/k  ...  1/k  1/k  1/k)
df(1) (  1 -1/(k-1)  ... -1/(k-1) -1/(k-1) -1/(k-1))
df(2) (  0   1  ... -1/(k-2) -1/(k-2) -1/(k-2))
.
.
df(k-2) (  0   0  ...   1  -1/2  -1/2)
df(k-1) (  0   0  ...   0   1  -1)
```

gdzie k jest liczbą kategorii zmiennej niezależnej. Na przykład zmienna niezależna z czterema kategoriami ma macierz kontrastów Helmerta o następującej postaci:

```
(1/4  1/4  1/4  1/4)
(  1 -1/3 -1/3 -1/3)
```

```
( 0 1 -1/2 -1/2)
( 0 0 1 -1)
```

Różnica

Kontrasty różnicy lub odwrotne kontrasty Helmerta. Porównuje kategorie zmiennej niezależnej ze średnią z poprzednich kategorii zmiennej. Ogólna postać macierzy jest następująca:

```
mean ( 1/k 1/k 1/k ... 1/k)
df(1) ( -1 1 0 ... 0)
df(2) ( -1/2 -1/2 1 ... 0)
.
.
df(k-1) (-1/(k-1) -1/(k-1) -1/(k-1) ... 1)
```

gdzie k jest liczbą kategorii zmiennej niezależnej. Na przykład kontrasty różnicy dla zmiennej niezależnej z czterema kategoriami są następujące:

```
( 1/4 1/4 1/4 1/4)
( -1 1 0 0)
(-1/2 -1/2 1 0)
(-1/3 -1/3 -1/3 1)
```

Wielomianowe

Ortogonalne kontrasty wielomianowe. Pierwszy stopień swobody zawiera efekt liniowy obejmujący wszystkie kategorie; drugi stopień swobody zawiera efekt kwadratowy; trzeci stopień swobody – sześcienny; i tak dalej w przypadku efektów wyższego rzędu.

Można określić odstępy między poziomami terapii mierzone przez daną zmienną jakościową. Równe odstępy, które obowiązują domyślnie w przypadku pominięcia metryki, można określić jako kolejne liczby całkowite od 1 do k , gdzie k jest liczbą kategorii. Jeśli zmienna *drug* (lek), ma trzy kategorie, to opcja komendy

```
/CONTRAST (DRUG)=POLYNOMIAL
```

jest równoważna

```
/CONTRAST (DRUG)=POLYNOMIAL(1,2,3)
```

Jednak nie zawsze potrzebne są równe odstępy. Załóżmy na przykład, że zmienna *drug* reprezentuje różne dawki leku podawane trzem grupom. Jeśli dawka podawana drugiej grupie jest dwa razy większa od dawki podawanej pierwszej grupie, a dawka podawana trzeciej grupie jest trzy razy większa od dawki podawanej pierwszej grupie, to odstępy między kategoriami terapii są równe, a odpowiednia metryka w tej sytuacji składa się z kolejnych liczb całkowitych:

```
/CONTRAST (DRUG)=POLYNOMIAL(1,2,3)
```

Jeśli jednak dawka podawana drugiej grupie jest cztery razy większa od dawki podawanej pierwszej grupie, a dawka podawana trzeciej grupie jest siedem razy większa od dawki podawanej pierwszej grupie, to odpowiednia będzie metryka:

```
/CONTRAST (DRUG)=POLYNOMIAL(1,4,7)
```

W obu przypadkach pierwszy stopień swobody zmiennej *drug* zawiera efekt liniowy poziomów dawki, a drugi stopień swobody zawiera efekt kwadratowy.

Kontrasty wielomianowe są szczególnie użyteczne przy testowaniu trendów i badaniu charakteru powierzchni odpowiedzi. Kontrastów wielomianowych można też używać do nieliniowego dopasowywania krzywych, np. regresji krzywoliniowej.

Powtórzone

Porównuje sąsiednie poziomy zmiennej niezależnej. Ogólna postać macierzy jest następująca:

```
mean (1/k 1/k 1/k ... 1/k 1/k)
df(1) ( 1 -1 0 ... 0 0)
df(2) ( 0 1 -1 ... 0 0)
      .
      .
df(k-1) ( 0 0 0 ... 1 -1)
```

gdzie k jest liczbą kategorii zmiennej niezależnej. Na przykład kontrasty powtórzone dla zmiennej niezależnej z czterema kategoriami są następujące:

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 1 -1 0)
( 0 0 1 -1)
```

Kontrasty te są przydatne w analizie profili i wszędzie tam, gdzie potrzebne są oceny różnic.

Specjalne

Kontrast zdefiniowany przez użytkownika. Umożliwia wprowadzanie kontrastów specjalnych w postaci macierzy kwadratowej z tyloma wierszami i kolumnami, ile kategorii ma dana zmienna niezależna. W przypadku analizy MANOVA i LOGLINEAR pierwszy wprowadzony wiersz jest zawsze efektem średnim lub stałym i odzwierciedla zbiór wag wpływających na sposób uśredniania pozostałych zmiennych niezależnych (jeśli są wybrane) według danej zmiennej. Zasadniczo ten kontrast jest wektorem jedynek.

Pozostałe wiersze macierzy zawierają kontrasty specjalne określające porównania między kategoriami zmiennej. Zwykle najbardziej użyteczne są kontrasty ortogonalne. Kontrasty ortogonalne są statystycznie niezależne i nienadmiarowe. Kontrasty są ortogonalne, jeśli:

- Dla każdego wiersza suma współczynników kontrastu wynosi 0.
- Suma iloczynów odpowiednich współczynników ze wszystkich par rozłącznych wierszy również wynosi 0.

Założmy na przykład, że terapia ma cztery poziomy i chcemy porównać ze sobą nawzajem te różne poziomy terapii. Oto odpowiedni kontrast specjalny:

```
(1 1 1 1) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0 2 -1 -1) compare 2nd with 3rd and 4th
(0 0 1 -1) compare 3rd with 4th
```

który definiuje się za pomocą następującej opcji komendy CONTRAST w przypadku analizy MANOVA, LOGISTIC REGRESSION i COXREG:

```
/CONTRAST(TREATMNT)=SPECIAL( 1 1 1 1
                             3 -1 -1 -1
                             0 2 -1 -1
                             0 0 1 -1 )
```

W przypadku analizy LOGLINEAR należy zdefiniować:

```
/CONTRAST(TREATMNT)=BASIS SPECIAL( 1 1 1 1
                                    3 -1 -1 -1
                                    0 2 -1 -1
                                    0 0 1 -1 )
```

Każdy wiersz z wyjątkiem wiersza średnich sumuje się do 0. Ilorazy każdej pary rozłącznych wierszy sumują się do 0, a także:

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

Kontrasty specjalne nie muszą być ortogonalne. Jednak nie mogą być liniowymi kombinacjami siebie nawzajem. W takim przypadku procedura poinformuje o zależności liniowej i zakończy działanie. Kontrasty Helmerta, różnic i wielomianowe są kontrastami ortogonalnymi.

Wskaźnik

Kodowanie zmiennych wskaźnikowych. Nazywane także kodowaniem sztucznym, jest niedostępne w przypadku analiz LOGLINEAR i MANOVA. Liczba kodowanych nowych zmiennych wynosi $k-1$. Obserwacje w kategorii odniesienia mają kod 0 dla wszystkich $k-1$ zmiennych. Obserwacja w i -tej kategorii ma kod 0 dla wszystkich zmiennych wskaźnikowych z wyjątkiem i -tej, która ma kod 1.

Regresja grzbietowa jądra

Regresja grzbietowa jądra jest procedurą rozszerzenia, która wykorzystuje klasę Python `sklearn.kernel_ridge.KernelRidge` do oszacowania modeli regresji grzbietowej jądra. Modele regresji grzbietowej jądra są nieparametrycznymi modelami regresji, które są zdolne do modelowania liniowych i nieliniowych relacji między zmiennymi predyktora i wynikami. Wyniki mogą być bardzo wrażliwe na wybór hiperparametrów modelu. Regresja grzbietowa jądra ułatwia wybór wartości hiperparametrów poprzez k -krotną walidację krzyżową w określonych siatkach wartości przy użyciu klasy `sklearn.model_selection.GridSearchCV`.

Przykład

Statystyka

addytywne CHI2, CHI2, cosinus, laplasjan, liniowy, wielomianowy, RBF, sigmoid, alfa, gamma, Coef0, stopień, walidacja krzyżowa, obserwacje a przewidywania, reszty a przewidywania, podwójne współczynniki wagowe, współczynniki wagowe przestrzeni jądra.

Zagadnienia dotyczące danych

Dane

- Możliwe jest określenie dowolnej lub wszystkich ośmiu różnych funkcji jądra.
- Wybrana funkcja jądra określa, które hiperparametry są aktywne.
- Hiperparametry obejmują alfa regularyzację grzbietu, które są wspólne dla wszystkich jąder, plus trzy inne hiperparametry dla każdej konkretnej funkcji jądra.
- Jeśli określono wiele podkomend jądra lub określono więcej niż jedną wartość dla dowolnego parametru, wykonywane jest wyszukiwanie w siatce z walidacją krzyżową w celu oceny modeli, a następnie wybierany jest najlepiej dopasowany model oparty na wstrzymanych danych.
- Rozszerzenie akceptuje zmienne dzielone z procedury Split File (podział pliku) i wagi z wykorzystaniem procedury Weight Cases (ważenie obserwacji).
- Podczas uwzględniania wag są one używane do tworzenia dopasowanych wartości we wszystkich analizach. Ze względu na ograniczenia w metodzie oceny w klasie `sklearn.model_selection.GridSearchCV` oceny walidacji krzyżowej, które są używane do wyboru modelu, nie są ważone.

Założenia

Uzyskiwanie regresji grzbietowej jądra

1. Wybierz z menu następującą opcję:

Analizuj > Regresja > Grzbiet jądra...

2. Wybierz zmienną **zależną**.
3. Wybierz przynajmniej jedną zmienną **niezależną**.
4. Domyślne ustawienie **Pojedynczy model** jest używane tylko wtedy, gdy określono tylko jedną wartość dla każdego parametru funkcji jądra. Jeśli wybrano ustawienie **Pojedynczy model**, nie można określić dodatkowych funkcji i wag **jądra**, które są w pełni stosowane w całej analizie, wartościowaniu i ocenianiu wyników. Można również użyć przycisków strzałek w górę i w dół, aby zmienić kolejność funkcji jądra.

Opcjonalnie wybierz opcję **Wybór modelu** z listy **Tryb**.

Po wybraniu opcji **Wybór modelu** z listy **Tryb** można dodać wiele funkcji jądra do listy **Jądro**.

- a. Kliknij ikonę dodawania (+), aby dołączyć dodatkowe funkcje jądra.
- b. Aby wybrać funkcję jądra, należy kliknąć pustą komórkę w kolumnie **Jądro**.
- c. Dwukrotnie kliknij dowolną komórkę wiersza funkcji jądra, aby określić wartości parametrów funkcji jądra dla odpowiedniej kolumny (**Alfa**, **Gamma**, **Coef0**, **Stopień**). Więcej informacji na ten temat zawiera sekcja "Parametry jądra" na stronie 30. Poniżej wymieniono domyślne parametry dostrajania funkcji jądra.

Additive_CHI2

ALPHA=1 GAMMA=1

CHI2

ALPHA=1 GAMMA=1

Cosinus

ALPHA=1

Laplasjan

ALPHA=1 GAMMA=1/p

Liniowe

Domyślna funkcja jądra. ALPHA=1

Wielomianowe

ALPHA=1 GAMMA=1/p COEF0=1 DEGREE=3

RBF

ALPHA=1 GAMMA=1/p

Sigmoid

ALPHA=1 GAMMA=1/p COEF0=1

Uwaga: Jeśli określono więcej niż jedną wartość dla dowolnego parametru funkcji jądra, przeprowadzane jest wyszukiwanie w siatce z walidacją krzyżową w celu oceny modeli, a następnie wybierany jest najlepszy model dopasowania oparty na wstrzymanych danych.

5. Opcjonalnie można kliknąć opcję **Opcje**, aby określić krotność walidacji krzyżowej, opcje wyświetlania, ustawienia wykresu i elementy do zapisania. Więcej informacji na ten temat zawiera sekcja "[Regresja grzbietowa jądra: Opcje](#)" na stronie 30.
6. Kliknij **OK**.

Parametry jądra

Okno dialogowe **Parametry jądra** udostępnia opcje służące do określania wartości parametrów funkcji pojedynczego jądra oraz do określania, że wybór modelu jest wykonywany przy użyciu funkcji wyszukiwania siatki w kombinacjach jądra i określonych wartości parametrów siatki.

Określ pojedyncze parametry

Włącz to ustawienie, aby określić wartości dla wybranego parametru funkcji jądra.

- Wprowadź wartość i kliknij przycisk **Dodaj**, aby dołączyć wartość do parametru funkcji jądra.
- Wybierz wartość parametru i kliknij przycisk **Zmień**, aby zaktualizować wartość.
- Wybierz wartość parametru i kliknij przycisk **Usuń**, aby usunąć wartość.

Określ parametry siatki

Włącz to ustawienie, aby określić, że wybór modelu jest wykonywany przy użyciu wyszukiwania siatki w kombinacjach jądra i określonych wartości parametrów siatki.

Regresja grzbietowa jądra: Opcje

W oknie dialogowym **Wykresy** dostępne są opcje służące do określania krotności walidacji krzyżowej, opcji wyświetlania, ustawień wykresu oraz elementów do zapisania.

Wartość krotności walidacji krzyżowej

Liczba podziałów lub krotności w walidacji krzyżowej z wyszukiwaniem siatki w celu wyboru modelu. Wprowadź liczbę całkowitą większą od 1. Wartością domyślną jest 5. Ustawienie jest dostępne tylko wtedy, gdy **Wybór modelu** jest wybierany jako **Tryb** w podstawowym oknie dialogowym **Regresja grzbietowej jądra**.

Wyświetl

Udostępnia opcje służące do określania, które dane wyjściowe mają być wyświetlane w przypadku, gdy walidacja krzyżowa jest aktywna.

Najlepiej

Ustawienie domyślne wyświetla tylko podstawowe wyniki dla wybranego najlepszego modelu.

Porównaj

Wyświetla podstawowe wyniki dla wszystkich ocenionych modeli.

Porównywanie modeli i krotności

Wyświetla pełne wyniki dla każdego podziału lub krotności dla każdego ocenianego modelu.

Wykres

Udostępnia opcje określające wykresy wartości obserwowanych lub rezydualnych w porównaniu z wartościami przewidywanymi.

Obserwowane a przewidywane

Wyświetla wykres rozrzutu obserwowanych i przewidywanych wartości dla określonego lub najlepszego modelu.

Reszty a przewidywane

Wyświetla wykres rozrzutu reszt w porównaniu z wartościami przewidywanymi dla określonego lub najlepszego modelu.

Zapisz

Tabela zawiera opcje umożliwiające określenie zmiennych, które mają zostać zapisane w aktywnym zbiorze danych.

Wartości przewidywane

Zapisuje przewidywane wartości z określonego lub najlepszego modelu w aktywnym zbiorze danych. Można dołączyć opcjonalną nazwę zmiennej.

Reszty

Umożliwia zapisanie reszt z określonych lub najlepszych predykcji modelu do aktywnego zbioru danych. Można dołączyć opcjonalną nazwę zmiennej.

Podwójne współczynniki

Umożliwia zapisanie podwójnych współczynników lub współczynników wagowych przestrzeni jądra z określonego modelu do aktywnego zbioru danych. Można dołączyć opcjonalną nazwę zmiennej. Ustawienie jest dostępne tylko wtedy, gdy określono opcję **Wybór modelu** dla opcji **Tryb** w podstawowym oknie dialogowym **Regresja grzbietowa jądra**.

Modele czasu przyspieszonego błędu parametrycznego

Analiza modelu AFT (parametrycznego przyspieszonego czasu niepowodzenia) wywołuje procedurę parametrycznego modelu przeżycia z niepowtarzającymi się danymi czasu życia. Parametryczne modele przeżycia zakładają, że czas przeżycia jest zgodny ze znanym rozkładem, a ta analiza pasuje do przyspieszonych modeli czasu awarii z ich modelowym wpływem proporcjonalnym w odniesieniu do czasu przeżycia.

Uzyskiwanie analizy modelowania parametrycznego przyspieszonego czasu awarii

1. Z menu wybierz:

Analiza > Przeżycie > Modele AFT (Parametric Przyspiesza czasu awarii)

2. Wybierz zmienną źródłową.

Czas

Przeżycie

Pojedyncza zmienna numeryczna oznaczająca czas trwania czasu przeżycia.

Początek/koniec

Zmienne numeryczne oznaczające **Czas rozpoczęcia** i **Czas zakończenia**.

Status

Pojedyncza opcjonalna zmienna łańcuchowa lub numeryczna, która określa jeden z następujących ustawień statusu:

Niepowodzenie/zdarzenie

Odwzorowuje rekord na kategorię usterki/zdarzenia. Wartością domyślną zmiennej statusu łańcucha jest F.

Obciążenie prawostronne

Odwzorowuje rekord na odpowiednią kategorię cenzurowania. Wartością domyślną zmiennej statusu łańcucha jest R.

Obciążenie lewostronne

Odwzorowuje rekord do lewej kategorii cenzurowania. Wartością domyślną zmiennej statusu łańcucha jest L.

Obciążenie odstępu

Odwzorowuje rekord do kategorii cenzurowania interwału. Tylko dla opcji **Start/End**. Wartością domyślną dla zmiennej statusu łańcucha jest I.

Przetwarzanie nieodwzorowanych wartości

Określa kategorię, do której ma zostać odwzorowana nieodwzorowana rekordy. Aby usunąć rekordy, których odwzorowanie nie powiodło się, wybierz opcję **Wyklucz je z analizy**.

W przypadku opcji **Survival** domyślny status dla wszystkich obserwacji to **Niepowodzenie/Zdarzenie**.

W przypadku opcji **Start/End** domyślny status to **Censoring przedziału czasu**. Kliknij przycisk **Zdefiniuj zdarzenie**, aby zdefiniować zdarzenie dla zmiennej statusu.

Współzmiennie

Jedna lub kilka opcjonalnych zmiennych numerycznych, które mają być traktowane jako współzmiennie. Należy zauważyć, że zmienna nie może być określona jednocześnie przez zmienną **Współzmiennie** i **Współczynnik (y) stały**.

Czynniki stałe

Jedna lub kilka opcjonalnych zmiennych, które mają być traktowane jako czynniki. Zmienna nie może być określona jednocześnie przez **Współczynnik (y) stały** i **Współzmiennie**.

Obcinanie lewostronne

Pojedyncza opcjonalna zmienna numeryczna dla obcinania po lewej stronie tylko dla opcji **Survival**.

Modelowe modele czasu niepowodzenia w trybie przyspieszonym: Kryteria

Kryteria

Opcjonalny panel do określania kryteriów ogólnych.

Przedział ufności

Opcjonalny procent w celu określenia poziomu dla przedziałów ufności parametrów regresji. Musi to być pojedyncza wartość liczbowa z zakresu od 0 do 100. Domyślną wartością jest 95.

Braki danych

Opcja do sterowania sposobem traktowania braków danych przez użytkownika:

Wyklucz systemowe i zdefiniowane przez użytkownika braki danych

Traktuje braki danych użytkownika jako poprawne wartości. Jest to ustawienie domyślne.

Zdefiniowane braki danych są traktowane jako ważne wartości

Ignoruje braki danych zdefiniowane przez użytkownika i traktuje je jako poprawne wartości.

Traktowanie statusu

Tylko dla opcji **Start/End** . Opcja do sterowania sposobem obsługi rekordów z niepoprawnymi polami statusu:

Odrzuć rekord powodujący konflikt

Usuwa rekordy powodujące konflikt. Jest to ustawienie domyślne.

Uzyskaj informacje o czasie zgodnie ze statusem

Pobiera informacje o czasie w zależności od statusu.

Wylicz status zgodnie z informacjami o czasie

Zmienia status zgodnie z informacjami o czasie.

Modelowe modele czasu niepowodzenia w trybie przyspieszonym: Model

model

Opcjonalny panel, w którym można określić opcje modelu i ustawienia.

Rozkład czasu przeżycia

Opcja określania rozkładu czasu przeżycia.

Weibulla

Określa rozkład Weibulla. Jest to ustawienie domyślne.

Wykładniczy

Określa rozkład wykładniczy.

Log-normalny

Określa rozkład logarytmiczno-normalny.

Log-logistyczny

Określa rozkład logarytmiczno-logistyczny.

Ustawienia współzmiennych

Określ zmienne współzmiennie.

Ustawienia czynników

Określ zmienne czynnika.

Początkowa wartość przechwyty

Opcja określania początkowej wartości okresu przechwycenia. Jeśli zostanie podana, musi to być pojedyncza wartość liczbowa i nie może być równa 0.

Początkowa wartość parametru skali

Opcja sterująca ustawieniem parametru skali.

Błąd standardowy odpowiedniej regresji OLS

Używa standardowego błędu odpowiadającego regresji zwykłej najmniejszych kwadratów jako wartości początkowej.

Odwraca błąd standardowy odpowiedniej regresji OLS

Korzysta z odwrotności błędu standardowego.

Wartość podana przez użytkownika

Jeśli podano pojedynczą wartość liczbową, wartość ta jest używana jako wartość początkowa. Jeśli zostanie podana, musi być większa od 0.

Modelowe modele czasu niepowodzenia w trybie przyspieszonym: szacowanie

Oszacowanie

Opcjonalny panel, w którym można określić ustawienia w celu sterowania szacowaniem przyspieszonych modeli czasu awarii oraz opcjonalnego procesu wyboru składników.

Alternatywna metoda kierunkowska lub mnożniki (ADMM)

Uproszczone

Stosuje metodę szybkiego naprzemiennego kierunku dla mnożników (ADMM). Jest to ustawienie domyślne.

Tradycyjny

Stosuje tradycyjny algorytm ADMM.

Zastosuj regularyzację L-1

Prowadzi proces do sterowania wyborem opcji. Pole **Parametr Penalty** określa parametr kary, który steruje procesem regularyzacji. Musi to być pojedyncza wartość większa od 0. Ustawieniem domyślnym jest 0,001.

Modelowanie kryteriów zbieżności

Zbieżność parametru:

Określa kryteria zbieżności dla parametru. Musi to być pojedyncza wartość liczbowa należąca do $[0, 1)$. Ustawieniem domyślnym jest 0,000001. W polu **Typ** można wybrać opcję **ABSOLUTE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub **RELATIVE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalne pole **Wartość** określa słowo kluczowe.

Zbieżność funkcji celu

Określa kryteria zbieżności dla funkcji celu strategicznego. Musi to być pojedyncza wartość liczbowa należąca do $[0, 1)$. Ustawieniem domyślnym jest 0, co nie ma zastosowania do kryteriów zbieżności. W polu **Typ** można wybrać opcję **ABSOLUTE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub **RELATIVE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalne pole **Wartość** określa słowo kluczowe.

Zbieżność Hessego

Określa kryteria zbieżności dla macierzy Hessego. Musi to być pojedyncza wartość liczbowa należąca do $[0, 1)$. Ustawieniem domyślnym jest 0, co nie ma zastosowania do kryteriów zbieżności. W polu **Typ** można wybrać opcję **ABSOLUTE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub **RELATIVE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalne pole **Wartość** określa słowo kluczowe.

Kryteria zbieżności resztowej

Opcja sterowania procesem optymalizacji.

Reszty pierwotne i dualne

Stosuje zarówno kryterium zbieżności pierwotne, jak i podwójne. Jest to ustawienie domyślne.

Tylko reszty pierwotne

Stosuje pierwotne kryterium zbieżności rezydualnej.

Tylko reszty dualne

Stosuje podwójne kryterium zbieżności rezydualnej.

Metoda losowania

Opcjonalny parametr do określenia metody estymacji.

Automatycznie

Automatycznie wybiera metodę w oparciu o przykładowy zestaw danych. Jest to ustawienie domyślne. Pole **Próg liczby predyktorów** określa próg liczby predyktorów i musi być pojedynczą liczbą całkowitą większą niż 1. Wartością domyślną jest 1000.

Newton-Raphson

Stosuje metodę Newtona-Raphsona.

L-BFGS

Stosuje algorytm BFGS z ograniczonymi pamięciami. Pole **Aktualizacja** określa liczbę poprzednich aktualizacji obsługiwanych przez algorytm BFGS z ograniczonymi pamięciami i musi być pojedynczą liczbą całkowitą większą lub równą 1. Wartość domyślna to 5.

Iteracja

Maksymalna liczba iteracji

Określa maksymalną liczbę iteracji. Musi to być pojedyncza liczba całkowita należąca do [1, 100]. Ustawieniem domyślnym jest 20.

Maksymalna liczba kroków połowienia

Określa maksymalną liczbę kroków połowienia. Musi to być pojedyncza liczba całkowita należąca do [1, 20]. Ustawienie domyślne to 5.

Maksymalna liczba wyszukiwań linii

Określa maksymalną liczbę wyszukiwań linii. Musi to być pojedyncza liczba całkowita należąca do [1, 100]. Ustawieniem domyślnym jest 20.

Bezwzględna zbieżność dla procesu iteracji

Określa zbieżność bezwzględną dla procesu iteracji zewnętrznej. Musi to być pojedyncza wartość liczbowa należąca do (0, 1). Wartością domyślną jest 0.0001.

Zbieżność względna dla procesu iteracji

Określa względną zbieżność dla procesu iteracji zewnętrznej. Musi to być pojedyncza wartość liczbowa należąca do (0, 1). Wartością domyślną jest 0,01.

Modelowe modele czasu niepowodzenia w trybie przyspieszonym: Drukowanie

Drukuj

Opcjonalny panel sterujący danymi wyjściowymi tabeli.

Szczegóły kodowania czynników

Wybranie tej opcji powoduje wyświetlenie i wydrukowanie szczegółów kodowania dla czynników. Proces jest ignorowany, jeśli nie istnieją żadne czynniki.

Wartości początkowe przypisane do parametrów regresji

Po wybraniu tej opcji zostaną wyświetlone wartości początkowe używane w procesie oszacowania.

Historia iteracji modelu

Jeśli ta opcja jest wybrana, wyświetla historię iteracji analizy przeżycia. W polu **Liczba kroków** określ liczbę kroków od 1 do 99999999. Ustawieniem domyślnym jest 1.

Wyniki wyboru zawierające

Steruje wyświetlaniem szczegółów wyboru opcji.

Zarówno wybrane, jak i niewybrane zmienne

Wyświetl w tabeli zarówno wybrane, jak i niewybrane zmienne.

Tylko wybrane zmienne

Wyświetl tylko wybrane zmienne.

Tylko niewybrane zmienne

Wyświetl tylko niewybrane zmienne. Pole **Maksymalna liczba zmiennych do wyświetlenia** określa maksymalną liczbę zmiennych wydrukowanych w tabeli. Ustawienie domyślne to 30.

Modelowe modele czasu niepowodzenia w trybie przyspieszonym: predykcja

Predykcja

Opcjonalny panel, który umożliwia ocenę i zapisanie przewidywanych statystyk w aktywnym zestawie danych.

Wartości czasu do oceny

Wartości czasu zdefiniowane przez zmienne zależne

Wyniki **predykcje** są oparte na zmiennej czasu określonej dla modelu przeżycia parametrycznego.

Przedziały regularne

Umożliwia ocenę **predykcji** w oparciu o przyszłe wartości czasu. Pole **Przedział czasu** określa przedział czasu i musi być pojedynczą wartością liczbową większą niż 0. Pole **Liczba okresów** określa liczbę przedziałów czasu i musi to być pojedyncza liczba całkowita z zakresu od 2 do 100.

Przedział czasu

Wyniki **predykcji** wyznaczone są na podstawie czasu trwania w celu zdefiniowania przyszłych wartości czasu. Musi to być pojedyncza zmienna numeryczna.

Predykcje

Przeżycie

Zapisuje wyniki i zapisuje przewidywane statystyki przeżycia w aktywnym zestawie danych. Domyślną nazwą zmiennej niestandardowej (lub nazwą użytkownika root) jest PredSurvival.

Hazard

Zapisuje wyniki i zapisuje przewidywane zagrożenia dla aktywnego zestawu danych. Domyślną nazwą zmiennej niestandardowej (lub nazwą użytkownika root) jest PredHazard.

Skumulowane zagrożenie

Zapisuje wyniki i zapisuje przewidywane skumulowane zagrożenia dla aktywnego zestawu danych. Domyślną nazwą zmiennej niestandardowej (lub nazwą użytkownika root) jest PredCumHazard.

Przeżycie warunkowe

Zapisuje wyniki i zapisuje przewidywane warunkowe statystyki przeżycia do aktywnego zestawu danych. Domyślna nazwa zmiennej niestandardowej (lub nazwa użytkownika root) to PredConditionalSurvival. Jeśli parametr PASTTIME nie zostanie określony, proces zostanie zignorowany. Wartość **Czas przeżycia przeszłości** jest wymagana i określa minione wartości czasu dla oceniania. Musi to być pojedyncza zmienna numeryczna.

Modelowe modele czasu awarii z akceleracją parametryczną: wykres

Wykres

Wykres reszt Coxa-Snella

Wybierz opcję **Wyświetl wykres**, aby utworzyć wykres resztkowy Cox-Snella. W polu **Number of binning cut points** (Liczba punktów podziału na kategoryzację) podaj liczbę z zakresu od 1 do 10000. Ustawieniem domyślnym jest 100.

Wykresy funkcji

Opcja sterująca wykresami funkcji.

Typ

Przeżycie

Tworzy wykres dla funkcji przeżycia.

Hazard

Tworzy wykres dla funkcji hazardu.

Gęstość

Tworzy wykres dla funkcji gęstości.

Liczba punktów do wyświetlenia

Określa liczbę punktów funkcji z zakresu od 1 do 200. Ustawieniem domyślnym jest 100.

Wartości współzmiennych do wykresu

Opcjonalnie, aby określić wartości podane przez użytkownika i przypisać je do predyktorów. Domyślnie wyznaczone wykresy zostaną utworzone przy **Średniej** każdej współzmiennnej, a także częstotliwość kategorii dla każdego czynnika. Jeśli zostanie określona, wyznaczone wykresy zostaną utworzone na podstawie ustawienia wzorca. W obecności zduplikowanych zmiennych, jeden określony jako pierwszy zostałby rozpoznany, a reszta zostanie zignorowana. Poprawna zmienna musi być zawarta w efekcie modelu. W przypadku współzmiennnej wartość podana przez użytkownika musi być wartością liczbową. Pominięcie zmiennej w efekcie wskazuje, że

częstotliwość kategorii i **Średnia** są domyślnie używane dla współczynnika i współzmiennej. Jeśli do zmiennej zostanie przypisana niepoprawna wartość, żądany wzorzec nie zostanie wykreślony.

Wartości czynników do wykresu

Opcjonalnie, aby określić wartości podane przez użytkownika i przypisać je do predyktorów. W obecności zduplikowanych zmiennych, jeden określony jako pierwszy zostałby rozpoznany, a reszta zostanie zignorowana. Poprawna zmienna musi być zawarta w efekcie modelu. Pominięcie zmiennej w efekcie wskazuje, że częstotliwość kategorii i średnia będą używane odpowiednio dla współczynnika i współzmiennej. Jeśli do zmiennej zostanie przypisana niepoprawna wartość, żądany wzorzec nie zostanie wykreślony.

Oddzielne wiersze dla

Opcja określania zmiennej kategoryjnej, za pomocą której rysowane będą wykresy liniowe.

Maksymalna liczba wierszy na wykresie

Określa maksymalną liczbę wierszy na wykresie, jeśli określono **Oddzielne linie dla**. Ustawienie domyślne to 10.

Modele czasu przyspieszonego niepowodzenia parametrycznego: Eksport

Eksportuj

Wybierz opcję **Eksportuj informacje o modelu do pliku XML**, aby zapisać informacje o modelu i parametrze w pliku PMML w celu uzyskania oceny. Należy określić katalog i nazwę pliku PMML, który ma zostać zapisany.

Czas przeżycia AFT-definiowanie zdarzeń dla zmiennych statusu

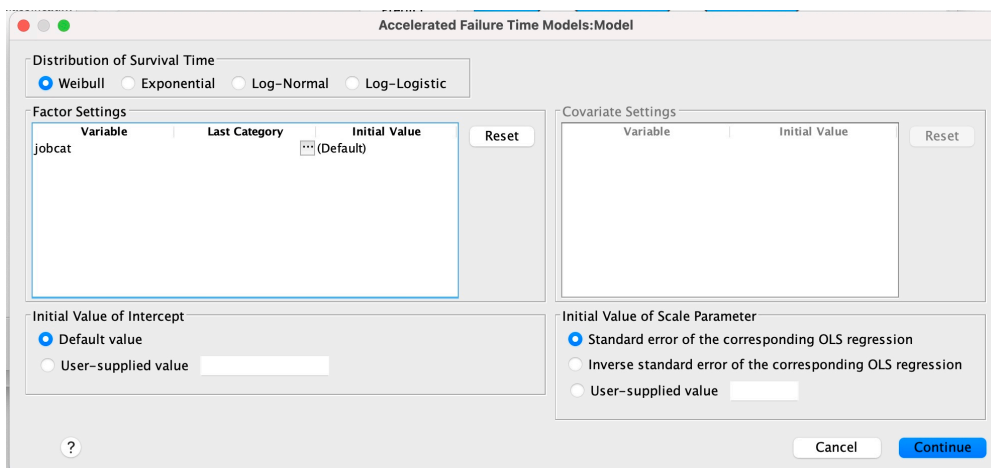
Wystąpienia wybranej wartości lub wartości dla zmiennej statusu wskazują, że wystąpiło zdarzenie terminalu dla tych obserwacji. Wszystkie pozostałe obserwacje uznaje się za cenzurowane. Wprowadź jedną wartość lub zakres wartości identyfikujący interesujące zdarzenie.

Modelowanie parametrycznych przyspieszonych modeli czasu: Wybierz kategorię

Ustawienie Wybierz kategorię udostępnia opcje umożliwiające wybór wartości oznaczających kategorię, która ma być modelowana jako linia bazowa dla porównania.

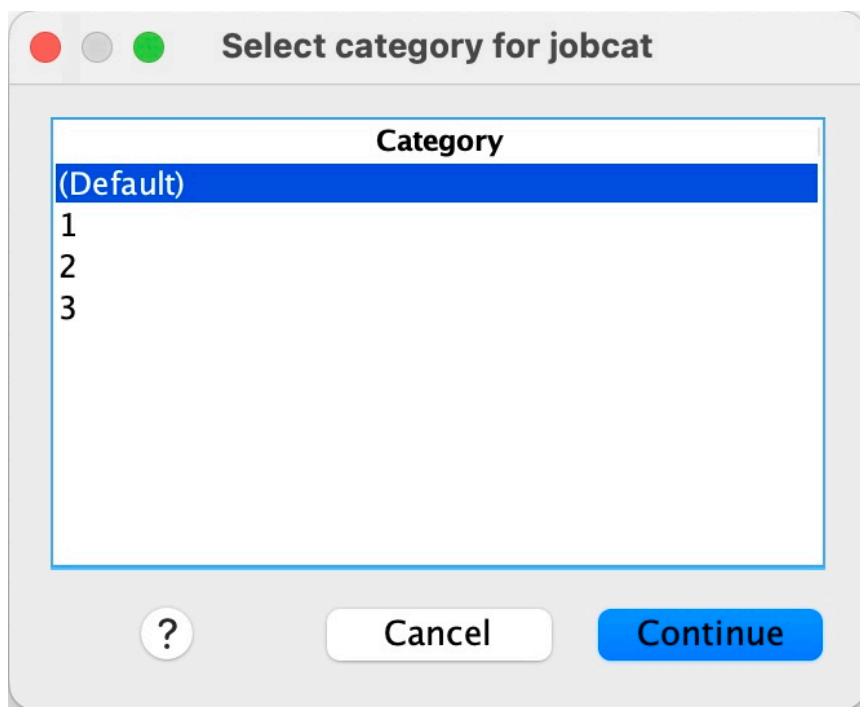
Wybieranie kategorii

Kliknij opcję 'Ostatnia kategoria', aby otworzyć okno dialogowe 'Wybierz kategorię'.



Rysunek 1. Przyspieszone modele czasu życia-okno dialogowe-Kategoria

Aby wyznaczyć kategorię jako linię bazową, wybierz wartość z okna dialogowego 'Wybierz kategorię'.



Rysunek 2. Przyspieszone modele czasu życia-okno dialogowe-Wyberz kategorię

Kliknij przycisk Kontynuuj.

Modele parametrycznych współużytkowanych modeli Frailty

Analiza Parametric Shared Frailty Models Survival uruchamia procedurę parametrycznego modelu przeżycia z rekurencyjnym wejściem danych w czasie życia. Parametryczne modele przeżycia zakładają, że czas przeżycia jest zgodny ze znanym rozkładem, a ta analiza zawiera frailty termin w parametrycznym modelu przeżycia. Jest on traktowany jako komponent losowy w celu uwzględnienia efektu nieobserwowanego z powodu zmienności poziomu indywidualnego lub grupy.

Uzyskiwanie analizy parametrycznej współużytkowanych modeli Frailty

1. Z menu wybierz opcję:

Analiza > Survival > Parametryczne współużytkowane modele Frailty

2. Wybierz zmienną źródłową.

Czas

Przeżycie

Czas przeżycia jest reprezentowany przez jedną zmienną w celu oznaczenia czasu zakończenia. Czas rozpoczęcia zostanie ustawiony na 0.

Początek/koniec

Zmienne liczbowe, które oznaczają **Czas rozpoczęcia** i **Czas zakończenia**.

Obiekt

Wymagane do uruchomienia procedury. Określa pojedynczą zmienną dla identyfikatora podmiotu.

Przedział

Określa jedną i zmienną numeryczną dla numeru interwału, który jest używany do identyfikowania różnych rekordów powtarzających się, które mają ten sam identyfikator podmiotu.

Status

Pojedyncza opcjonalna zmienna łańcuchowa lub numeryczna, która określa jeden z następujących ustawień statusu:

Niepowodzenie/zdarzenie

Odwzorowuje rekord na kategorię usterki/zdarzenia. Wartością domyślną zmiennej statusu łańcucha jest F.

Obcięcie prawostronne

Odwzorowuje rekord na odpowiednią kategorię cenzurowania. Wartością domyślną zmiennej statusu łańcucha jest R.

Przetwarzanie nieodwzorowanych wartości

Określa kategorię, do której ma zostać odwzorowana nieodwzorowana rekordy. Aby usunąć rekordy, których odwzorowanie nie powiodło się, wybierz opcję **Wyklucz je z analizy**.

Kliknij przycisk **Zdefiniuj zdarzenie**, aby zdefiniować zdarzenie dla zmiennej statusu.

Współzmiennie

Jedna lub kilka opcjonalnych zmiennych numerycznych, które mają być traktowane jako współzmiennie. Należy zauważyć, że zmienna nie może być określona jednocześnie przez zmienną **Współzmiennie** i **Współczynnik (y) stały**.

Czynniki stałe

Jedna lub kilka opcjonalnych zmiennych, które mają być traktowane jako czynniki. Zmienna nie może być określona jednocześnie przez **Współczynnik (y) stały** i **Współzmiennie**.

Modele parametryczne współużytkowane frailty: kryteria

Kryteria

Opcjonalny panel do określania kryteriów ogólnych.

Przedział ufności

Opcjonalny procent w celu określenia poziomu dla przedziałów ufności parametrów regresji. Musi to być pojedyncza wartość liczbowa z zakresu od 0 do 100. Domyślną wartością jest 95.

Poziom istotności

Opcja umożliwia określenie poziomu istotności testu ilorazu wiarygodności dla komponentu frailty. Musi to być pojedyncza wartość liczbowa z zakresu od 0 do 1. Ustawienie domyślne to 0.05.

Braki danych

Opcja do sterowania sposobem traktowania braków danych przez użytkownika:

Wyklucz systemowe i zdefiniowane przez użytkownika braki danych

Traktuje braki danych użytkownika jako poprawne wartości. Jest to ustawienie domyślne.

Zdefiniowane braki danych są traktowane jako ważne wartości

Ignoruje braki danych zdefiniowane przez użytkownika i traktuje je jako poprawne wartości.

Traktowanie odstępu czasu

Opcja do sterowania sposobem obsługi rekordów, których interwał jest w konflikcie z czasem rozpoczęcia i zakończenia. Działanie to staje się skuteczne, jeśli istnieją dwie zmienne czasu ze zmienną przedziału czasu określoną w głównym dialogu.

Odrzuć rekordy z konfliktem

Usuwa wszystkie rekordy szeregowy tematu, jeśli wartość przedziału czasu jest w konflikcie z czasem rozpoczęcia i zakończenia. Jest to ustawienie domyślne.

Wykryj wartości przedziałów w oparciu o czas rozpoczęcia i zakończenia

Wykrywa wartość przedziału czasu od czasu rozpoczęcia i zakończenia.

Modele parametryczne współużytkowane Frailty: Model

Model

Opcjonalny panel, w którym można określić opcje modelu i ustawienia.

Rozkład czasu przeżycia

Opcja określania rozkładu czasu przeżycia.

Weibulla

Określa rozkład Weibulla. Jest to ustawienie domyślne.

Wykładniczy

Określa rozkład wykładniczy.

Log-normalny

Określa rozkład normalny dziennika.

Log-logistyczny

Określa rozkład log-logistyczny.

Ustawienia współzmiennych

Określ zmienne współzmiennne.

Ustawienia czynników

Określ zmienne czynnika.

Początkowa wartość przechwyty

Opcja określania początkowej wartości okresu przechwycenia. Jeśli zostanie podana, musi to być pojedyncza wartość liczbowa i nie może być równa 0.

Początkowa wartość parametru skali

Opcja sterująca ustawieniem parametru skali.

Błąd standardowy odpowiedniej regresji OLS

Używa standardowego błędu odpowiadającego regresji zwykłej najmniejszych kwadratów jako wartości początkowej.

Odwraca błąd standardowy odpowiedniej regresji OLS

Korzysta z odwrotności błędu standardowego.

Wartość podana przez użytkownika

Jeśli podano pojedynczą wartość liczbową, wartość ta jest używana jako wartość początkowa. Jeśli zostanie podana, musi być większa od 0.

Komponent słabości

Opcjonalny parametr, który służy do określania **dystrybucji** komponentu frailty.

Rozkład gamma

Określa dystrybucję Gamma. Jest to ustawienie domyślne.

Odwrotność rozkładu gaussowskiego

Określa rozkład odwrotności-Gaussa.

Początkowa wartość wariancji

Określa początkową wartość wariancji komponentu frailty. Musi to być pojedyncza wartość liczbowa większa od 0. Wartość domyślna to 1.0 dla dystrybucji Gamma i 0.1 dla dystrybucji odwrotnej-Gaussa.

Modele współużytkowanych ramek parametrycznych: szacowanie

Oszacowanie

Opcjonalny panel, który umożliwia określenie ustawień sterujących szacowaniem współużytkowanych modeli ramek oraz opcjonalnym procesem wyboru predyktantów.

Alternatywna metoda kierunkowa lub mnożniki (ADMM)

Uproszczone

Stosuje metodę szybkiego naprzemiennego kierunku dla mnożników (ADMM). Jest to ustawienie domyślne.

Tradycyjny

Stosuje tradycyjny algorytm ADMM.

Zastosuj regularyzację L-1

Prowadzi proces do sterowania wyborem opcji. Pole **Parametr Penalty** określa parametr kary, który steruje procesem regularyzacji. Musi to być pojedyncza wartość większa od 0. Ustawienie domyślne to 0.001.

Modelowanie kryteriów zbieżności

Zbieżność parametru:

Określa kryteria zbieżności dla parametru. Musi to być pojedyncza wartość liczbowa należąca do $[0, 1)$. Ustawieniem domyślnym jest 0,000001. W polu **Typ** można wybrać opcję **ABSOLUTE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub **RELATIVE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalna wartość **Wartość** określa próg liczbowy dla typu zbieżności.

Zbieżność funkcji celu

Określa kryteria zbieżności dla funkcji celu strategicznego. Musi to być pojedyncza wartość liczbowa, która należy do $[0, 1)$. Ustawieniem domyślnym jest 0, co nie ma zastosowania do kryteriów zbieżności. W polu **Typ** można wybrać opcję **ABSOLUTE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub **RELATIVE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalna wartość **Wartość** określa próg liczbowy dla typu zbieżności.

Zbieżność Hessego

Określa kryteria zbieżności dla macierzy Hessego. Musi to być pojedyncza wartość liczbowa, która należy do $[0, 1)$. Ustawieniem domyślnym jest 0, co nie ma zastosowania do kryteriów zbieżności. W polu **Typ** można wybrać opcję **ABSOLUTE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub **RELATIVE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalna wartość **Wartość** określa próg liczbowy dla typu zbieżności.

Kryteria zbieżności resztowej

Opcja sterowania procesem optymalizacji.

Reszty pierwotne i dualne

Stosuje zarówno kryterium zbieżności pierwotne, jak i podwójne. To ustawienie jest domyślnie używane.

Tylko reszty pierwotne

Stosuje pierwotne kryterium zbieżności rezydualnej.

Tylko reszty dualne

Stosuje podwójne kryterium zbieżności rezydualnej.

Metoda

Opcjonalny parametr do określenia metody estymacji.

Automatycznie

Automatycznie wybiera metodę w oparciu o przykładowy zestaw danych. Ta metoda jest wybrana domyślnie. Pole **Próg liczby predyktorów** określa próg liczby predyktorów i musi być pojedynczą liczbą całkowitą większą niż 1. Wartość domyślna: 1000.

Newton-Raphson

Stosuje metodę Newtona-Raphsona.

L-BFGS

Stosuje algorytm BFGS z ograniczonymi pamięciami. Pole **Aktualizacja** określa liczbę poprzednich aktualizacji, które są obsługiwane przez algorytm BFGS z ograniczoną pamięcią, i musi być pojedynczą liczbą całkowitą większą lub równą 1. Wartość domyślna to 5.

Iteracja

Maksymalna liczba iteracji

Określa maksymalną liczbę iteracji. Musi to być pojedyncza liczba całkowita, która należy do [1, 300]. Ustawieniem domyślnym jest 20.

Maksymalna liczba kroków połowienia

Określa maksymalną liczbę kroków połowienia. Musi to być pojedyncza liczba całkowita, która należy do [1, 200]. Ustawienie domyślne to 5.

Maksymalna liczba wyszukiwań linii

Określa maksymalną liczbę wyszukiwań linii. Musi to być pojedyncza liczba całkowita, która należy do [1, 300]. Ustawieniem domyślnym jest 20.

Bezwzględna zbieżność dla procesu iteracji

Określa zbieżność bezwzględną dla procesu iteracji zewnętrznej. Musi to być pojedyncza wartość liczbowo, która należy do (0, 1). Wartością domyślną jest 0.0001.

Zbieżność względna dla procesu iteracji

Określa względną zbieżność dla procesu iteracji zewnętrznej. Musi to być pojedyncza wartość liczbowo, która należy do (0, 1). Wartością domyślną jest 0.01.

Modele parametryczne współużytkowane szandarowe: drukowanie

Drukuj

Opcjonalny panel sterujący danymi wyjściowymi tabeli.

Szczegóły kodowania czynników

Wybranie tej opcji powoduje wyświetlenie i wydrukowanie szczegółów kodowania dla czynników. Proces jest ignorowany, jeśli nie istnieją żadne czynniki.

Wartości początkowe, które są przypisane do parametrów regresji

Po wybraniu tej opcji zostaną wyświetlone wartości początkowe, które są używane w procesie oszacowania.

Historia iteracji modelu

Jeśli ta opcja jest wybrana, wyświetla historię iteracji analizy przeżycia. W polu **Liczba kroków** określ liczbę kroków od 1 do 99999999. Ustawieniem domyślnym jest 1.

Modele sparametryzowane współużytkowane parametryczne: Predict

Predykcja

Opcjonalny panel, który umożliwia ocenę i zapisanie przewidywanych statystyk w aktywnym zestawie danych.

Wartości czasu do oceny

Wartości czasu zdefiniowane przez zmienne zależne

Wyniki **predykcje** są oparte na zmiennej czasu określonej dla modelu przeżycia parametrycznego.

Przedziały regularne

Umożliwia ocenę **predykcji** w oparciu o przyszłe wartości czasu. Pole **Przedział czasu** określa przedział czasu i musi być pojedynczą wartością liczbową większą niż 0. Pole **Liczba okresów** określa liczbę przedziałów czasu i musi to być pojedyncza liczba całkowita z zakresu od 2 do 100.

Przedział czasu

Wyniki **predykcji** wyznaczone są na podstawie czasu trwania w celu zdefiniowania przyszłych wartości czasu. Musi to być pojedyncza zmienna numeryczna.

Predykcje

Przeżycie

Zapisuje wyniki i zapisuje przewidywane statystyki przeżycia w aktywnym zestawie danych. Domyślną nazwą zmiennej niestandardowej (lub nazwą użytkownika root) jest `PredSurvival`.

Hazard

Zapisuje wyniki i zapisuje przewidywane zagrożenia dla aktywnego zestawu danych. Domyślną nazwą zmiennej niestandardowej (lub nazwą użytkownika root) jest `PredHazard`.

Skumulowane zagrożenie

Zapisuje wyniki i zapisuje przewidywane skumulowane zagrożenia dla aktywnego zestawu danych. Domyślną nazwą zmiennej niestandardowej (lub nazwą użytkownika root) jest `PredCumHazard`.

Przeżycie warunkowe

Zapisuje wyniki i zapisuje przewidywane warunkowe statystyki przeżycia do aktywnego zestawu danych. Domyślną nazwą zmiennej niestandardowej (lub nazwą użytkownika root) jest `PredConditionalSurvival`. Jeśli parametr `PASTTIME` nie został określony, proces jest ignorowany. Wartość **Czas przeżycia przeszłości** jest wymagana i określa minione wartości czasu dla oceniania. Musi to być pojedyncza zmienna numeryczna.

Przeżycie bezwarunkowe

Zapisuje wyniki i zapisuje prognozowane bezwarunkowe statystyki przeżycia do aktywnego zestawu danych. Słowo kluczowe jest domyślnie pomijane. Po określeniu tej opcji może nastąpić opcjonalna nazwa zmiennej podana przez użytkownika (lub nazwa użytkownika root) podana w nawiasach. Nazwa domyślna to `PredUnCondSurvival`.

Hazard bezwarunkowy

Wyniki i zapisuje prognozowane bezwarunkowe statystyki zagrożeń dla aktywnego zestawu danych. Słowo kluczowe jest domyślnie pomijane. Po określeniu tej opcji może nastąpić opcjonalna nazwa zmiennej podana przez użytkownika (lub nazwa użytkownika root) podana w nawiasach. Nazwa domyślna to `PredUnCondHazard`.

Bezwarunkowe ryzyko skum

Wyniki i zapisuje przewidywane bezwarunkowe skumulowane statystyki hazardu do aktywnego zestawu danych. Słowo kluczowe jest domyślnie pomijane. Jeśli zostanie podana, może to być następna opcjonalna nazwa zmiennej podana przez użytkownika (lub nazwa główna określona w nawiasach). Nazwa domyślna to `PredUnCondCumHazard`.

Modele parametryczne współużytkowane frailty: Plot

Wykres

Wykresy funkcji

Opcja sterująca wykresami funkcji.

Typ

Przeżycie

Tworzy wykres dla bezwarunkowych funkcji przeżycia.

Hazard

Tworzy wykres dla niewarunkowych funkcji hazardu.

Gęstość

Tworzy wykres dla funkcji gęstości.

Liczba punktów do wyświetlenia

Określa liczbę punktów funkcji z zakresu od 1 do 200. Ustawieniem domyślnym jest 100.

Wartości współzmiennych do wykresu

Opcja określania wartości dostarczonych przez użytkownika i przypisywania ich do predyktorów. Domyślnie wyznaczone wykresy zostaną utworzone na podstawie średniej dla każdej współzmiennych. Jeśli zostanie określona, wyznaczone wykresy zostaną utworzone na podstawie ustawienia wzorca. W obecności zduplikowanych zmiennych, jeden określony jako pierwszy

zostały rozpoznany, a reszta zostanie zignorowana. Poprawna zmienna musi być zawarta w efekcie modelu. W przypadku współzmiennych wartość podana przez użytkownika musi być wartością liczbową. Pominięcie zmiennej w efekcie wskazuje, że wartość średnia będzie używana domyślnie dla współzmiennych. Jeśli do zmiennej zostanie przypisana niepoprawna wartość, żądany wzorzec nie zostanie wykreślony.

Wartości czynników do wykresu

Opcja określania wartości dostarczonych przez użytkownika i przypisywania ich do predyktorów. Domyślnie wyznaczone wykresy zostaną utworzone z częstotliwością kategorii dla każdego efektu. Jeśli zostanie określona, wyznaczone wykresy zostaną utworzone na podstawie ustawienia wzorca. W obecności zduplikowanych zmiennych, jeden określony jako pierwszy zostanie rozpoznany, a reszta zostanie zignorowana. Poprawna zmienna musi być zawarta w efekcie modelu. Pominięcie zmiennej w efekcie wskazuje, że częstotliwość kategorii będzie używana domyślnie dla czynnika. Jeśli do zmiennej zostanie przypisana niepoprawna wartość, żądany wzorzec nie zostanie wykreślony.

Oddzielne wiersze dla

Opcja określania zmiennej kategoryjnej, za pomocą której rysowane będą wykresy liniowe.

Maksymalna liczba wierszy na wykresie

Określa maksymalną liczbę wierszy na wykresie, jeśli określono **Oddzielne linie dla** . Wartością domyślną jest 10.

Modele sparametryzowanych współużytkowanych modeli: Eksport

Eksportuj

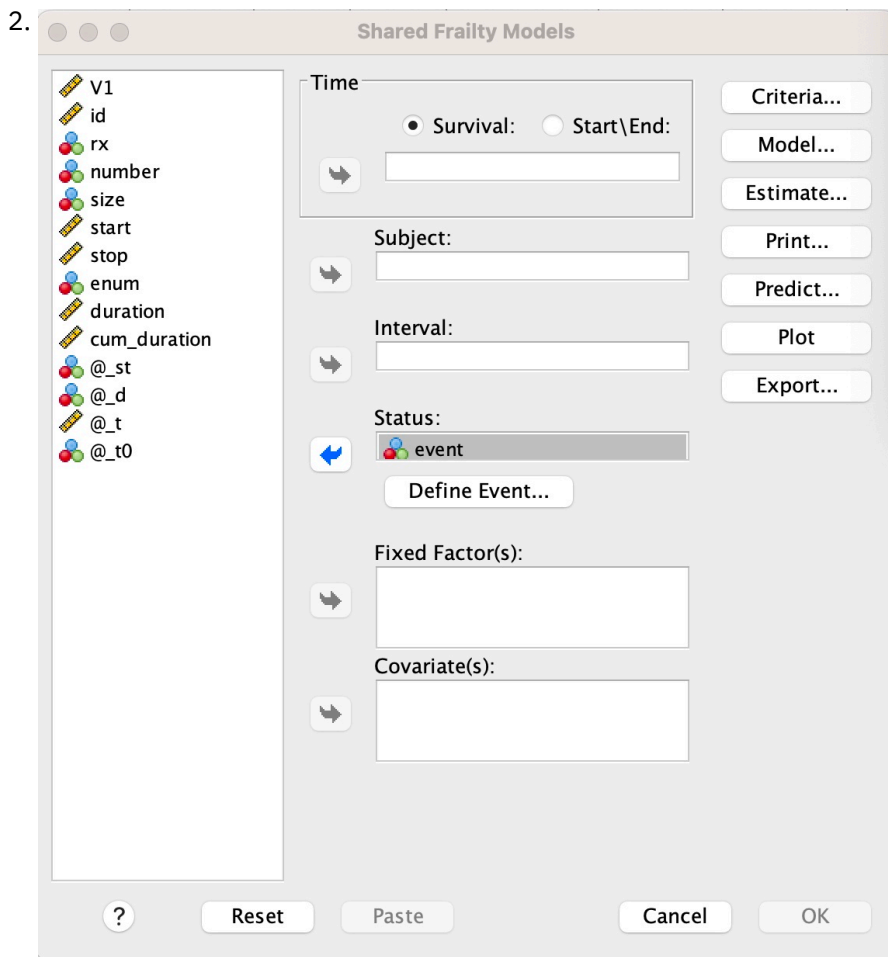
Wybierz opcję **Eksportuj informacje o modelu do pliku XML** , aby zapisać informacje o modelu i parametrze w pliku PMML w celu uzyskania oceny. Należy określić katalog i nazwę pliku PMML, który ma zostać zapisany.

Modele parametryczne współużytkowane Frailty: Definiuj zdarzenia

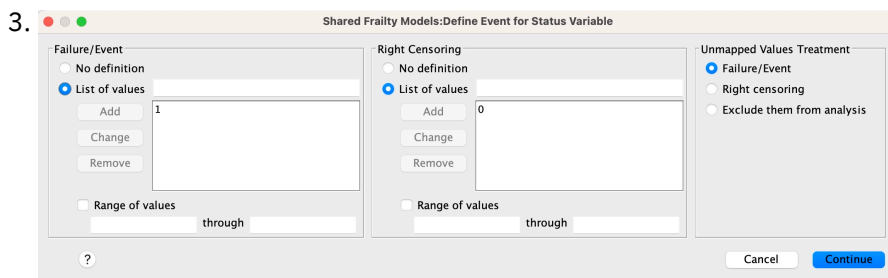
Opcja definiowania statusu. Jeśli zmienna statusu zostanie pominięta, niepowodzenie lub zdarzenie stanie się domyślnym statusem dla wszystkich obserwacji.

1. Z menu wybierz,

Analiza > Survival > Parametryczne współużytkowane modele Frailty ...



Rysunek 3. Modele współużytkowanych modeli-okno dialogowe-Status



Rysunek 4. Współużytkowane modele Frailty-okno dialogowe-Status-Definiowanie zdarzenia

Modele parametryczne współużytkowane frailty-przykłady

Przykład 1

SURVREG RECURRENT Y WITH x1 BY x2

/MODEL SUBJECT = ID FRAILTY=GAMMA DISTRIBUTION=WEIBULL.

W współzmiennnej x1 i współzmiennnej x2 dopasowany jest parametryczny model przetrwania współdzielony.

Czas przeżycia jest reprezentowany przez pojedynczą zmienną y.

Obiekty są identyfikowane za pomocą identyfikatora zmiennej.

Przyjmuje się, że czas przeżycia jest zgodny z rozkładem Weibulla.

Zakłada się, że wariancja frailty jest zgodna z rozkładem Gamma.

Wszystkie poprawne rekordy są używane w analizie przeżycia.

Przykład 2

```
SURVREG RECURRENT Y WITH x1 BY x2
```

```
/MODEL SUBJECT = id FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_NORMAL INTERVAL=Z.
```

W współzmiennnej x1 i współzmiennnej x2 dopasowany jest parametryczny model przetrwania współdzielony.

Czas przeżycia jest reprezentowany przez dwie zmienne y1 i y2 , oznaczające czas rozpoczęcia i zakończenia.

Obiekty są identyfikowane za pomocą identyfikatora zmiennej.

Przyjmuje się, że czas przeżycia jest zgodny z dzienną dystrybucją dziennika.

Zakłada się, że wariancja frailty jest zgodna z odwrotną dystrybucją Gaussa.

Interwały czasowe są definiowane przez zmienną z. Dla każdego podmiotu procedura korzysta tylko z niekonfliktowanych rekordów i wyklucza z analizy wszystkie rekordy po pierwszym statusie niepowodzenia.

Przykład 3

```
SURVREG RECURRENT y1 y2 Z x1 PRZEZ x2(1)
```

```
/MODEL SUBJECT = id FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_LOGISTIC
```

```
/STATUS VARIABLE=zdarzenie FAILURE=1 RIGHT=0.
```

W współzmiennnej x1 i współzmiennnej x2 dopasowany jest parametryczny model przetrwania współdzielony. Przeżycie

czas jest reprezentowany przez dwie zmienne y1 i y2 , oznaczające odpowiednio czas rozpoczęcia i zakończenia. W przypadku czynnika x2 kategoria "1" jest wyznaczana jako linia bazowa, która ma być modelowana.

Obiekty są identyfikowane za pomocą identyfikatora zmiennej.

Przyjmuje się, że czas przeżycia jest zgodny z rozkładem log-logistycznym.

Zakłada się, że wariancja frailty jest zgodna z odwrotną dystrybucją Gaussa.

Zdarzenie zmiennej jest określane w celu zdefiniowania statusu z 1 i 0 oznaczaniem usterki i odpowiednio cenzurowaniem.

Przykład 4

```
SURVREG RECURRENT Y WITH x1 BY x2
```

```
/MODEL SUBJECT = id
```

```
/STATUS VARIABLE=zdarzenie FAILURE=1 RIGHT=0
```

```
/PRZEWIDYWANIE NIEPOZORNEGO PRZEŻYCIA NIEZAGROŻENIECZEGO NIEPOZORNEGO ZAGROŻENIA
```

```
/FUNCTIONPLOT SURVIVAL HAZARD DENSITY PLOTBY (x2).
```

W współzmiennnej x1 i współzmiennnej x2 dopasowany jest parametryczny model przetrwania współdzielony. Czas przeżycia jest reprezentowany przez pojedynczą zmienną y.

Obiekty są identyfikowane za pomocą identyfikatora zmiennej.

Niewarunkowe lub oparte na populacjach przeżycia, zagrożenie i skumulowane zagrożenie są oceniane i zapisywane w aktywnym zestawie danych.

Wykreślone są bezwarunkowe lub oparte na populacjach przeżycie i krzywe zagrożenia oddzielone kategoriami w x2.

Przykład 5

```
SURVREG RECURRENT Y WITH x1 BY x2
/MODEL SUBJECT = ID FRAILITY=GAMMA DISTRIBUTION=WEIBULL
/STATUS VARIABLE=zdarzenie FAILURE=1 RIGHT=0
/ESTIMATION HCONVERGE=1e-12(RELATIVE) PCONVERGE=0 FCONVERGE=0SELECTFEATURES=TRUE
PENALTY=0.01.
```

W współzmiennnej x1 i współzmiennnej x2 dopasowany jest parametryczny model przetrwania współdzielony. Czas przeżycia jest reprezentowany przez pojedynczą zmienną y.

Obiekty są identyfikowane za pomocą identyfikatora zmiennej.

Przyjmuje się, że czas przeżycia jest zgodny z rozkładem Weibulla.

Zakłada się, że wariancja frailty jest zgodna z rozkładem Gamma.

Kryteria zbieżności są oparte na macierzy Hessego. Używa on 1e-12 jako względnej zbieżności.

Model obejmuje okres karny w celu kontroli procesu regulatyzacji. Parametr kary jest ustawiony na 0.01.

Przykład 6

```
SURVREG RECURRENT Y WITH x1 BY x2
/MODEL SUBJECT = id
/STATUS VARIABLE=infect FAILURE=1 RIGHT=0
/ESTIMATION MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20.
```

W współzmiennnej x1 i współzmiennnej x2 dopasowany jest parametryczny model przetrwania współdzielony. Czas przeżycia jest reprezentowany przez pojedynczą zmienną y.

Obiekty są identyfikowane za pomocą identyfikatora zmiennej.

Procedura określa maksymalną liczbę operacji wyszukiwania w wierszu na 100, maksymalną liczbę iteracji równą 50 i maksymalną liczbę kroków połowienia, które mają być zmniejszone o 20.

Parametryczne współużytkowane modele Frailty-studium przypadku dla danych rekurencyjnych

Parametryczne współużytkowane modele Frailty-studium przypadku dla danych rekurencyjnych

Nazwa przypadku użycia-efekt po stronie leczenia.

Podmioty-Public health investigator and practitioner.

Warunki wstępne-Wyczyszczony zestaw danych dostępny na podstawie czasu przeżycia, statusu efektu ubocznego i predyktorów, które mają zostać dostosowane.

Opis-Patrick, publiczny badacz zdrowia, bada próbkę danych, która obejmuje 20 uczestników. Uczestnicy rekrutują się w badaniu dotyczącym łagodnego efektu ubocznego, który może być spowodowany nowym leczeniem. Projektant leczenia twierdzi, że nie byłoby różnic między mężczyznami i kobietami, co do skutków ubocznych. Patrick chciałby ocenić taką hipotezę. Zmienne, które są zawarte w próbce danych, są wymienione w następujący sposób:

- patID: numer identyfikatora w celu zidentyfikowania unikalnego uczestnika.

- endTime: czas przeżycia (w dniach) działania niepożądanego po leczeniu, który jest mierzony od początku leczenia do działania niepożądanego zgłaszanego lub cenzurowania w ciągu 60 dni.

- sideEffect: status efektu ubocznego, status = 0, jeśli cenzurowany i status = 1, jeśli zgłaszany jest efekt łagodny.

-wiek: wiek uczestnika w okresie badawczym.

-żeńskie: samica = 0, jeśli samce i żeńskie = 1, jeśli samica.

Możliwe jest zastosowanie wielu zabiegów, co skutkuje wielokrotnymi rekordami czasów powtórzenia, które są mierzone dla określonego uczestnika. Czas rozpoczęcia jest zawsze równy 0 dla każdego rekordu, który został pominięty w próbkę danych. Patrick jest zainteresowany wizualizacją funkcji przetrwania i hazardu, aby narysować porównanie między mężczyzną a samica, kontrolując ich wiek i słabość. Jest świadomy, że te zabiegi, które są podawane do tego samego uczestnika, są bardziej skorelowane. Przyjmując, że czas przeżycia jest zgodny z rozkładem Weibulla, Patrick postanawia zbudować parametryczny model przetrwania współdzielony w SPSS Statistics, aby uwzględnić zależność leczenia od tego samego uczestnika.

Składnia-

```
DATA LIST FREE
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .
BEGIN DATA .
1 45 0 38.00 0
2 26 1 20.00 1
3 58 0 53.00 0
4 31 1 37.00 1
4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILTY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .
```

Synopsis:

Składnia określona przez Patryka wyznacza endTime jako zmienną czasu zależnego. Procedura automatycznie zakłada, że dla każdego rekordu godzina rozpoczęcia wynosi 0. Zmienne wiekowe i żeńskie są modelowane odpowiednio jako współzmienna i czynnik. Zakłada się, że czasy przeżycia nawrotu są zgodne z rozkładem Weibulla. Przyjmuje się, że termin frailty jest zgodny z rozkładem Gamma, a jego komponent wariancji jest modelowany. W odniesieniu do danych wyjściowych tabela Podsumowanie

modelu zawiera informacje o procedurze i modelu. Tabela Podsumowanie przetwarzania przypadków zawiera obszerny wykaz usterki/cenzurowania, a także te przypadki, które zostały wykluczone z analizy.

W próbie danych Patryka wszystkie rekordy są poprawne i uwzględnione w analizie. Porównując prawdopodobieństwo logarytmu wiarygodności z odpowiednim modelem bez komponentu frailty, model współdzielony nie osiągnie znaczącego poziomu (wartość $p = 0.168$). Patrick zastanawia się, czy konieczne jest uwzględnienie w modelu terminu dzielnicy-frailty'ego. Szacowany współczynnik przyspieszenia dla uczestnika męskiego wynosi 1.017, który jest uzyskiwany przez obliczanie wykładnika oszacowanego współczynnika regresji 0.017 z [kobieta = 0.0]. Powiązany z nim przedział ufności 95% to (.688, 1.504). Wyniki te sugerują, że osobnik płci męskiej ma niemal ten sam współczynnik przyspieszenia, co kobieta indywidualna w tym samym wieku i zapachu. Na poziomie populacji, Patrick wykreśla bezwarunkową przeżycie i krzywe zagrożenia oddzielnie dla samców i samic, które są oceniane w średniej wieku (37.45 lat).

Patrick potwierdza, że dla każdej stałej wartości czasu przeżycia, mężczyzna i samica średnio mają mieć takie samo prawdopodobieństwo przeżycia. Co ciekawe, pomimo modalnego kształtu, który jest widoczny na bezwarunkowym wykresie hazardu, Patrick odkrywa, że w ciągu 60 dni zagrożenie dla populacji jest w rzeczywistości coraz większe. To zachowanie może sugerować istnienie efektu frailty. Aby dokładniej zbadać efekt uboczny spowodowany przez zabiegi, Patrick może kontynuować model bez frailty komponent i porównać zachowanie samców i samic. Ponadto, może rozważyć następujące działania z udziałem uczestników przez okres dłuższy niż 60 dni w celu zebrania większej ilości danych.

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

Analiza Probit (PROBIT)
definiuj zakres [11](#)
dodatkowe właściwości komendy [12](#)
granice ufności [11](#)
iteracje [11](#)
kryteria [11](#)
proporcja reakcji naturalnej [11](#)
statystyki [11](#)
test równoległości [11](#)
względna moc medialna [11](#)
analiza przeżycia
w regresji grzbietowej jądra [29](#)
analiza regresji logistycznej [2](#)

B

binarna regresja logistyczna [1](#), [2](#)

C

Chi-kwadrat Pearsona
do estymacji wartości skalowania rozproszenia [9](#)
dobroć dopasowania [8](#)

D

D Cooka
w regresji logistycznej [4](#)
DfBeta
w regresji logistycznej [4](#)
dobroć dopasowania
w wielomianowej regresji logistycznej [8](#)

E

eliminacja wsteczna
w regresji logistycznej [3](#)
Estymacja wag
dodatkowe właściwości komendy [24](#)
logarytm wiarygodności [22](#)
przykład [22](#)
statystyki [22](#)

F

funkcja odchylenia
do estymacji wartości skalowania rozproszenia [9](#)

G

granice ufności
w analizie Probit [11](#)
Grzbiet jądra
alfa [29](#)

Grzbiet jądra (*kontynuacja*)

coef0 [29](#)
gamma [29](#)
pojedynczy model [29](#)
stopień [29](#)
wybór modelu [29](#)

I

iloraz wiarygodności
do estymacji wartości skalowania rozproszenia [9](#)
dobroć dopasowania [8](#)
iteracje
w analizie Probit [11](#)
w regresji logistycznej [5](#)
w wielomianowej regresji logistycznej [8](#)

K

kategoria odniesienia
w wielomianowej regresji logistycznej [7](#)
klasyfikacja
w wielomianowej regresji logistycznej [6](#)
komórki z zerową liczbą obserwacji
w wielomianowej regresji logistycznej [8](#)
kontrasty
w regresji logistycznej [4](#)
kroki połowienia
w wielomianowej regresji logistycznej [8](#)
kryterium zbieżności
w wielomianowej regresji logistycznej [8](#)

L

logarytm wiarygodności
w szacowaniu wag [22](#)
w wielomianowej regresji logistycznej [8](#)

M

macierz korelacji
w wielomianowej regresji logistycznej [8](#)
macierz kowariancji
w wielomianowej regresji logistycznej [8](#)
model Gaussa
w regresji nieliniowej [20](#)
model gęstości
w regresji nieliniowej [20](#)
model Gompertza
w regresji nieliniowej [20](#)
model Johnsona-Schumachera
w regresji nieliniowej [20](#)
model Michelisa Mentena
w regresji nieliniowej [20](#)
model Morgan-Mercer-Florin

- model Morgan-Mercer-Florin (*kontynuacja*)
w regresji nieliniowej [20](#)
- model Pearl-Reed
w regresji nieliniowej [20](#)
- model Richardsa
w regresji nieliniowej [20](#)
- model stosunków kwadratów
w regresji nieliniowej [20](#)
- model stosunku sześciątów
w regresji nieliniowej [20](#)
- model Verhulst
w regresji nieliniowej [20](#)
- model Von Bertalanffy
w regresji nieliniowej [20](#)
- model Weibulla
w regresji nieliniowej [20](#)
- model zależności plonów od gęstości
w regresji nieliniowej [20](#)
- model zmodyfikowany logarytmicznie
w regresji nieliniowej [20](#)
- Modele czasu przyspieszonego błędu parametrycznego
 - analiza [31](#)
 - Dokonaj predykcji [35](#)
 - drukowanie [35](#)
 - eksport [37](#)
 - kryteria [32](#)
 - model [33](#)
 - Oszacowanie [33](#)
 - wykres [36](#)
- modele efektów głównych
w wielomianowej regresji logistycznej [7](#)
- modele nieliniowe
w regresji nieliniowej [20](#)
- Modele parametryczne Frailty
zmienne statusu przeżycia [44](#)
- Modele parametrycznych współużytkowanych modeli
 - Frailty
 - analiza [38](#)
 - Dokonaj predykcji [42](#)
 - drukowanie [42](#)
 - Eksportowanie [44](#)
 - kryteria [39](#)
 - model [40](#)
 - Oszacowanie [40](#)
 - wykres [43](#)
- modele użytkownika
w wielomianowej regresji logistycznej [7](#)

O

- ograniczenia dotyczące parametrów
w regresji nieliniowej [21](#)
- ograniczona regresja
w regresji nieliniowej [21](#)
- osobliwość
w wielomianowej regresji logistycznej [8](#)
- oszacowania parametrów
w wielomianowej regresji logistycznej [8](#)

P

- pełne modele czynnikowe
w wielomianowej regresji logistycznej [7](#)

- prawo Metcherlicha przyrostów mniej niż proporcjonalnych
w regresji nieliniowej [20](#)
- Procedura Podwójna MNK
 - przykład [24](#)
 - statystyki [24](#)
 - zmienne instrumentalne [24](#)
- przebieg iteracji
w wielomianowej regresji logistycznej [8](#)
- przedziały ufności
w wielomianowej regresji logistycznej [8](#)
- Przeżycie AFT
 - Okno dialogowe przeżycia-zmienne kategorii [37](#)
- przyrost
jako korekta dla komórek z zerową liczbą obserwacji [8](#)

R

- R-kwadrat Coxa i Snella
w wielomianowej regresji logistycznej [8](#)
- R-kwadrat McFaddena
w wielomianowej regresji logistycznej [8](#)
- R-kwadrat Nagelkerke'a
w wielomianowej regresji logistycznej [8](#)
- regresja asymptotyczna
w regresji nieliniowej [20](#)
- Regresja grzbietowa jądra
 - krotność walidacji krzyżowej [30](#)
 - parametry [30](#)
 - parametry siatki [30](#)
 - wykresy [30](#)
 - wyświetlanie [30](#)
 - zapisywanie [30](#)
- Regresja kwantylowa
 - eksport [17](#)
 - kryteria [13](#)
 - model [14](#)
 - przykład [12](#)
 - statystyki [12](#)
 - wyświetl [16](#)
 - zapisywanie [17](#)
- regresja liniowa
 - Procedura Podwójna MNK [24](#)
- Regresja liniowa
 - szacowanie wag [22](#)
- regresja logistyczna [2](#)
- Regresja logistyczna
 - binarna [1](#)
 - definiowanie reguły filtrowania [3](#)
 - dodatkowe właściwości komendy [5](#)
 - iteracje [5](#)
 - kontrasty [4](#)
 - metody wyboru zmiennych [3](#)
 - miary wpływu [4](#)
 - opcje wyświetlania [5](#)
 - prawdopodobieństwo w metodzie krokowej [5](#)
 - przykład [2](#)
 - punkt podziału w klasyfikacji [5](#)
 - reszty [4](#)
 - stała [5](#)
 - Statystyka dobroci dopasowania Hosmera-Lemeshowa [5](#)
 - statystyki [2](#)
 - statystyki i wykresy [5](#)

Regresja logistyczna (*kontynuacja*)
ustawianie reguły 3
wartości przewidywane 4
współczynniki 2
współzmiennie jakościowe 4
współzmiennie łańcuchowe 4
zapisywanie zmiennych wynikowych 4

Regresja logistyczna SPSS 2

Regresja nieliniowa
algorytm Levenberga-Marquardta 21
dodatkowe właściwości komendy 22
funkcja straty 20
interpretacja wyników 22
logika warunkowa 19
metody estymacji 21
model posegmentowany 19
ograniczenia dotyczące parametrów 21
parametry 19
pochodne 21
przykład 18
reszty 21
sekwencyjne programowanie kwadratowe 21
statystyki 18
szacowanie z samouzgodnieniem 21
typowe modele nieliniowe 20
wartości początkowe 19
wartości przewidywane 21
zapisywanie zmiennych wynikowych 21

Regresja podwójną metodą najmniejszych kwadratów
dodatkowe właściwości komendy 25
kowariancja parametrów 25
zapisywanie zmiennych wynikowych 25

Regresja probit
przykład 10
statystyki 10

S

selekcja krokowa
w regresji logistycznej 3
w wielomianowej regresji logistycznej 7
selekcja postępująca
w regresji logistycznej 3
separacja
w wielomianowej regresji logistycznej 8
stała
w regresji liniowej 5
Statystyka dobroci dopasowania Hosmera-Lemeshowa
w regresji logistycznej 5
Szacowanie wag
pokaż statystyki ANOVA i oszacowania 24
przebieg iteracji 24
zapisywanie najlepszych wag jako nowych zmiennych 24

T

tabele klasyfikacji
w wielomianowej regresji logistycznej 8
tabele prawdopodobieństw komórek
w wielomianowej regresji logistycznej 8
Tabele trwania życia
zmiennie statusu przeżycia 37
test równoległości

test równoległości (*kontynuacja*)
w analizie Probit 11

W

wartości wpływu
w regresji logistycznej 4
wartość skalowania rozproszenia
w wielomianowej regresji logistycznej 9
Wielomianowa regresja logistyczna
dodatkowe właściwości komendy 10
eksportowanie informacji o modelu 9
kategoria odniesienia 7
kryteria 8
modele 7
statystyki 8
zapisywanie 9
współzmiennie
w regresji logistycznej 4
współzmiennie jakościowe 4
współzmiennie łańcuchowe
w regresji logistycznej 4
wyraz wolny
uwzględnienie lub wykluczenie 7
względna moc medialna
w analizie Probit 11

