



Spis treści

- 1 Wprowadzenie
 - 2 Dlaczego arkusze kalkulacyjne są popularne
 - 7 Alternatywa dla arkuszy kalkulacyjnych
 - 10 Krzywa uczenia się z IBM SPSS Statistics
 - 11 Licencjonowanie oprogramowania metodą SPSS Statistics
-

Ryzyko związane z używaniem arkuszy kalkulacyjnych do analizy statystycznej

Wprowadzenie

Arkusze kalkulacyjne są szeroko wykorzystywane do analizy statystycznej. I chociaż są niezwykle użytecznymi narzędziami, to są użyteczne tylko do pewnego momentu.

Gdy są używane do zadań, których nie są przeznaczone lub do zadań na poziomie lub przekraczającym poziom ich możliwości, korzystanie z arkuszy kalkulacyjnych może być dość ryzykowne. W niniejszym dokumencie przedstawiono kilka punktów, które należy wziąć pod uwagę, jeśli używasz lub planujesz używać arkusz kalkulacyjny do analiz statystycznych. Opisuje on również inne narzędzia, które w wielu przypadkach będą bardziej odpowiednie.

Dlaczego arkusze kalkulacyjne są popularne

Łatwość użycia czyni z arkusza kalkulacyjnego atrakcyjne narzędzie do wykonywania obliczeń. Większość z nas wie (lub myśli, że wie), jak go używać. Ponadto programy do obsługi arkuszy kalkulacyjnych są standardowym zasobem komputera stacjonarnego, więc są już dostępne.

Arkusz kalkulacyjny to wspaniały wynalazek i doskonałe narzędzie do niektórych zadań. Zbyt często jednak arkusze kalkulacyjne muszą wykonywać zadania wykraczające poza ich możliwości. To jest jak w starym powiedzeniu: „Jeśli masz tylko młotek, każdy problem wygląda jak gwóźdź”. Ale niektóre problemy można lepiej rozwiązać używając śrubokrętu, kleju lub kłamy pasa.



Co więcej, opinia, że arkusz kalkulacyjny jest łatwy w użyciu, jest w pewnym stopniu iluzją. Uzyskanie odpowiedzi z arkusza kalkulacyjnego jest zawsze łatwe, ale uzyskanie prawidłowej odpowiedzi już nie zawsze jest łatwe.

Jednak decyzja o użyciu czegoś innego, na przykład nieznannej technologii lub narzędzia, nie zawsze jest łatwa. Przy rozważaniu innych możliwości rodzą się dwa pytania:

Jak przydatne jest to narzędzie? I jak trudno jest się go nauczyć?

Odpowiedź na pierwsze pytanie zależy od skali i złożoności analizy danych. Typowy arkusz kalkulacyjny ma ograniczenie liczby rekordów, które może obsłużyć, więc jeśli skala zadania jest duża, bardzo przydatne może być narzędzie inne niż arkusz kalkulacyjny.

Arkusze kalkulacyjne mogą być przydatne do analizy statystycznej, ale gdy są używane do zadań, do których nie są przeznaczone, mają pewne ograniczenia.

Złożoność....., jeśli potrzebujesz tylko powierzchownego przeglądu danych, arkusz kalkulacyjny może być odpowiednim narzędziem. Ale jeśli przypuszczasz, że w danych znajdują się cenne informacje, które nie są od razu oczywiste, lub jeśli musisz wykonać szczegółową analizę bądź znaleźć ukryte wzorce, arkusz kalkulacyjny nie zapewni potrzebnej funkcjonalności.

Kolejnym czynnikiem, który należy wziąć pod uwagę, jest wymagany stopień dokładności. Wyniki arkusza kalkulacyjnego mogą być niewiarygodne podczas pracy z dużymi zbiorami danych i/lub wykonywania skomplikowanych obliczeń. Ale arkusz kalkulacyjny może nie wystarczyć, jeśli wymagana jest bezwzględna dokładność. I dlatego należy rozważyć inne, bardziej niezawodne narzędzie.

Wreszcie, jeśli trzeba wykonać po prostu analizę ograniczonej ilości danych historycznych, wystarczy arkusz kalkulacyjny. Ale jeśli trzeba tworzyć wiarygodne prognozy lub narysować trendy, szczególnie jeśli dotyczą dużych zbiorów danych, to są znacznie lepsze narzędzia.

Później wrócimy do rozważenia odpowiedzi na drugie pytanie – jak trudno jest się nauczyć? gdy szukamy alternatywy dla arkuszy kalkulacyjnych do obliczeń statystycznych.

Zanim jednak przejdziemy dalej, warto zauważyć, że ludzie używają arkuszy kalkulacyjnych do zadań innych niż obliczenia liczbowe. Na przykład arkusze kalkulacyjne są często używane, jakby były bazami danych – do tworzenia list i zarządzania nimi. Obowiązują jednak zasady skali i złożoności. Poza pewnymi ograniczeniami, o wiele bardziej odpowiednia jest odpowiednia baza danych, z wbudowanymi zasadami strukturyzacji danych, zachowywania integralności danych, opracowywania ścieżek audytu itp.

Dwie rzeczy do zapamiętania na temat arkuszy kalkulacyjnych

Podczas pracy z arkuszami kalkulacyjnymi trzeba pamiętać o dwóch ważnych kwestiach: arkusze kalkulacyjne mogą być skomplikowane i mogą być podatne na błędy.

Arkusze kalkulacyjne to w rzeczywistości programy komputerowe

Projektując arkusz kalkulacyjny piszesz program komputerowy. Programy arkusza kalkulacyjnego, takie jak Microsoft Excel, używają tak zwanego „nieproceduralnego języka programowania”. Chociaż możliwe jest również pisanie programów proceduralnych dla programu Excel w języku Visual Basic, codzienna praca z wpisywaniem formuł w komórki jest ćwiczeniem w programowaniu nieproceduralnym.

Tworzenie arkusza kalkulacyjnego może być złożone podobnie, jak programowanie komputerowe.

Gdy zazwyczaj myślimy o językach programowania, myślimy o BASIC, C, Java™, FORTRAN i tak dalej. Są to wszystko „języki proceduralne” i każdy z nich ma spójną metodologię opracowaną dla programów w takich językach. Jest to skutkiem tego, że przez lata stało się jasne, że ścisłe przestrzeganie tych zasad jest kluczowe dla prawidłowego działania programów. Mimo to uzyskanie skomplikowanego programu do generowania właściwych liczb może wymagać wykonania ogromnej ilości testów i usuwanie błędów.

Programowanie nieproceduralne wymaga także podejmowania wielu decyzji, jest pełne złożoności i szans na pomyłkę, jak wszystkie – poza najprostszym – programem proceduralnym.

Dzięki standardowej metodologii tworzenia oprogramowania, programy komputerowe są dwukrotnie i i trzykrotnie sprawdzane. Natomiast arkusz kalkulacyjny, choć może mieć kluczowe znaczenie dla funkcjonowania firmy, jest zwykle dziełem jednej osoby. Niemal nigdy nie jest to sprawdzane ani testowane szczegółowo i dość często wchodzi do produkcji po niewielkiej weryfikacji lub bez weryfikacji. Jednak ważne decyzje dotyczące zarządzania, prognozowanie przychodów i plany dotyczące przyszłych inwestycji opierają się na wytwarzanych danych liczbowych.

Arkusze kalkulacyjne są podatne na błędy

Przeprowadzono wiele badań dotyczących częstości występowania błędów w arkuszach kalkulacyjnych. I na tej podstawie można sądzić, że 90 procent wszystkich arkuszy zawiera co najmniej jeden błąd. Badania przeprowadzono poprzez wizualne kontrole arkuszy kalkulacyjnych o znaczeniu krytycznym, więc można założyć, że nie znaleziono wielu innych błędów. Stwierdzono również, że próby poprawienia błędów często wprowadzają nowe.

Badania wykazują, że 90 procent wszystkich arkuszy kalkulacyjnych zawiera co najmniej jeden błąd.

Przykłady kosztownych błędów arkusza kalkulacyjnego z całego świata:

- „... błąd wprowadzenia danych w wysokości 118 387 USD”¹
- „... ogromny błąd w wysokości 11 milionów dolarów doprowadził do wadliwego arkusza kalkulacyjnego”²
- „... firma generuje błąd arkusza kalkulacyjnego o wartości 30 mln USD”.³
- „... błędne dane o wartości 644 mln USD: niewykryty błąd arkusza kalkulacyjnego w procesie „ad hoc”.⁴

Rodzaje błędów arkusza kalkulacyjnego

Błędy arkusza kalkulacyjnego można podzielić na trzy główne rodzaje.

„Najbardziej przyjazny” rodzaj można nazwać błędami funkcjonalnymi. Błędy te są najłatwiejsze do znalezienia, ponieważ po prostu przerywają działanie arkusza kalkulacyjnego. Zamiast podawać błędne liczby, dają komunikaty o błędach lub nic.

Następnie są błędy zewnętrzne. Powodują, że arkusz kalkulacyjny wydaje się działać, ale liczby są nieprawidłowe. Często takie błędy są wykrywane przez kogoś, kto ma pojęcie, jakie powinny być wyniki i zwraca uwagę na ich niezgodność z oczekiwaniami.

Rodzaje błędów arkusza kalkulacyjnego to błędy funkcjonalne, błędy odbiegające od normy i błędy skradające się, które mają nasilenie od niskiego do wysokiego.

Najpoważniejsze są błędy, które można nazwać błędami skradającymi się. Powodują one nieprawidłowe wyniki, ale nikt nie wierzy, że są one nieprawidłowe. Przechodzą przez kontrolę i są akceptowane jako prawdziwe. Błędy skradające się pojawiają się albo dlatego, że nikt nie wie, jaki powinien być poprawny wynik (co często ma miejsce w przypadku obliczeń statystycznych), albo liczby są jedynie nieznacznie różne od oczekiwań i wydają się uzasadnione.

Istnieje wiele opowieści o kłopotliwych konsekwencjach błędów w arkuszu kalkulacyjnym. Jedna dotyczy Nevady City w Kalifornii, w którym w styczniu 2006 roku odkryto, że ma deficyt budżetowy w wysokości pięciu milionów dolarów. Arkusz kalkulacyjny budżetu był tym samym, którego użyto wcześniej, ale przy wprowadzaniu danych na nowy rok formuła została przypadkowo zastąpiona. Na szczęście był to błąd izolowany i szybko zauważony przez radnych miejskich. Jednak dyrektor finansowy musiał go poprawiać przez cały dzień (i znalazł w tym czasie szereg innych błędów).

Inna opowieść, pochodząca z 2003 roku, dotyczy uniwersytetu, który odkrył błędy w średnich ocenach niektórych studentów: liczby po prostu nie miały sensu. Po ręcznym wykonaniu obliczeń, egzaminatorzy poprawili oceny i odkryli, że błąd w równaniach arkusza kalkulacyjnego spowodowało wycinanie i wklejanie, które nie uwzględniały różnicy między bezwzględnym i względnym adresem komórek. I chociaż arkusz kalkulacyjny sprawdził członek personelu wyższego szczebla, dokładnie obejrzał tylko wiersz, który okazał się być jedynym poprawnym.

Przyczyny błędów arkusza kalkulacyjnego

Użytkownicy arkusza kalkulacyjnego powinni wiedzieć, jakie czynniki powodują błędy. Niestety, istnieje zbyt wiele przyczyn, które należy tutaj wymienić, ale najważniejsze z nich to:

- **Błędy w logice:** Może to być coś prostego, na przykład wywołanie niewłaściwej funkcji, odjęcie zamiast dodawania lub pominięcie nawiasu przy tworzeniu formuły. Tego rodzaju błędy mogą być również spowodowane przez domniemaną relację komórek w arkuszu kalkulacyjnym
- **Nieprawidłowo skopiowane formuły:** Wpisywanie równania podczas czytania z innego miejsca często prowadzi do błędów, podobnie jak wycinanie i wklejanie. Kopiowanie istniejących równań do nowych miejsc zwykle zmienia dane komórki, co sprawia, że dokładne sprawdzanie wyników jest ważne
- **Przypadkowo zastąpione formuły:** Komórka zawierająca równanie wygląda jak liczba widziana przez wszystkich użytkowników na pierwszy rzut oka jako wynik. Przypadkowe wstawienie liczby do komórki zawierającej już formułę spowoduje zastąpienie równania i przekształcenie zawartości komórki w stałą. Jeśli inne formuły polegają na wynikach z tej komórki, błąd może zostać znacząco spotęgowany
- **Niewłaściwe użycie wbudowanych funkcji:** Można użyć niewłaściwej funkcji, na przykład używając AVERAGEA, która ocenia tekst i fałszywe wpisy na zero, zamiast AVERAGE, która je ignoruje. Tego rodzaju błąd jest niestety bardzo łatwo popełnić

- **Pomijane czynniki:** Bardzo łatwo jest po prostu coś zostawić. Może to być równanie, dane lub oba czynniki. Błędy tego rodzaju występują dość często po dodaniu nowych danych do już wypełnionego arkusza kalkulacyjnego. Możliwe są też sytuacje, gdy nie wszystkie dane zostały wprowadzone lub niektóre z nowych komórek nie zostały uwzględnione we wszystkich odpowiednich równaniach
- **Błędne dane wejściowe:** Jeśli masz szczęście, błędy wprowadzania danych będą prowadzić do błędów odbiegających od normy. Lecz nie zawsze tak jest. Na przykład, jeśli zamiast 3.5 zostanie wprowadzona wartość 3,5, arkusz kalkulacyjny zakłada, że jest to ciąg, a nie liczba. W rezultacie wartość zero jest używana w dowolnych formułach odwołujących się do tej komórki. Innym błędem byłoby wpisanie 3/5, która staje się datą ,ogromną liczbą w dowolnym obliczeniu.

Istnieje wiele innych możliwości. Na przykład, jeśli posortujesz kolumnę zawierającą zarówno liczby, jak i równania, posortujesz równania w uzupełnieniu do liczb, co może spowodować błędy obliczeniowe.

Błędy arkusza kalkulacyjnego powstają z wielu powodów, w tym błędów logicznych, nieprawidłowych formuł, przypadkowego zastąpienia formuł i niewłaściwego użycia wbudowanych funkcji.

I w końcu pojawia się kwestia niezawodności oprogramowania arkusza kalkulacyjnego. Programiści arkuszy kalkulacyjnych ciągle wydają poprawki i poprawki do swoich programów. Jednak w 2008 roku Gregg Keizer⁵ zgłosił, że poprawka, która naprawiła błąd w programie Excel, spowodowała nowe błędy w obliczeniach: w rzeczywistości spowodowała więcej błędów niż wcześniej. (Oprogramowanie zostało wydane pięć lat wcześniej, ale albo problem nie został wykryty, albo jeśli został wykryty, nie podjęto żadnej próby jego rozwiązania). To nie tylko kwestia poprawek. Kilka badań pokazuje, że nawet poprawnie zakodowane arkusze kalkulacyjne nie są wystarczająco dokładne dla skomplikowanych procedur matematycznych lub dużych zestawów danych.

Ponieważ jest on tak szeroko wykorzystywany, nawet w nauczaniu statystyki, wiele artykułów szczegółowo opisało błędy w procedurach statystycznych w Excelu i więcej niż kilka stron internetowych podkreśla jego niedociągnięcia w kontekście zaawansowanej analityki. (Zobacz cytaty na końcu tego dokumentu). Podsumowując ich wyniki, „poważna analiza programu Excel wskazuje, że nie jest on odpowiedni dla dużych ani złożonych zestawów danych, ponieważ może to wymagać od użytkownika w pewnym stopniu naruszenia dokładności wyników”.

Inne problemy związane z używaniem arkuszy kalkulacyjnych

Użytkownicy arkusza kalkulacyjnego mogą również napotkać ograniczenia przy pracy ze specjalnymi typami danych, prognozowanie i zarządzanie danymi.

Uwzględnianie specjalnych typów danych

Kilka typów danych, wspólnych dla wielu rodzajów badań, wymaga specjalnego uwzględniania.

Podczas zapisywania wyników badania z użyciem arkuszy kalkulacyjnych może być szczególnie trudne dokładne przedstawienie brakujących lub kategoryalnych danych.

Jednym z częstych problemów jest postępowanie z brakującymi wartościami danych. Podczas pracy z arkuszami kalkulacyjnymi musisz zachować ostrożność, gdy masz do czynienia z brakującymi wartościami. Jeśli chcesz przypisać wartość zerową do takich danych, spowoduje to zniekształcenie, na przykład, średniej z zakresu wartości. Jeśli wprowadzisz ciąg w komórce, aby wskazać brakującą wartość, niektóre równania zignorują ciąg, a inne ocenią go jako zero. Ponieważ w niektórych przypadkach zero jest prawidłową wartością, potrzebny jest inny sposób wskazania brakującej wartości. Jednak wtedy nie tylko musisz uważać, aby konsekwentnie używać oznaczenia brakującej wartości przy wprowadzaniu danych, ale musisz również dokładnie dokumentować swoje podejście, aby późniejsze modyfikacje arkusza kalkulacyjnego nie unieważniały używanych konwencji danych. Należy zauważyć, że żadne z powyższych podejść nie zapewnia przyjętej metody wiarygodnego obliczania brakujących wartości.

Inna szczególna sytuacja pojawia się w przypadku danych kategoryjnych (często spotykanych w wynikach ankiet). Załóżmy na przykład, że cztery wartości 1, 2, 3 i 4 są przypisane do reprezentowania odpowiedzi „Tak”, „Nie”, „Nie wiem” i „Odmawiam odpowiedzi” w pytaniu ankiety. Jeśli używasz arkusza kalkulacyjnego do przechowywania tego typu danych, musisz udokumentować te wartości i ich znaczenie, aby zapewnić prawidłowe wprowadzenie takich danych (prawidłowa wartość przypisana do każdej odpowiedzi) i że są one przetwarzane w sposób znaczący. W przeciwnym razie znaczenie zostanie utracone, gdy tylko osoba, która opracowała arkusz kalkulacyjny, będzie nieobecna.

Projekcja przyszłości

Zazwyczaj arkusze kalkulacyjne są używane do wyodrębniania informacji i relacji dotyczących przeszłych zdarzeń. Coraz częściej jednak organizacje chcą wiedzieć, co może się wydarzyć w przyszłości. Najnowsze wersje programu Excel mają specjalne funkcje, na przykład PROGNOZA, TREND i WZROST do przewidywania nowych wartości na podstawie istniejących danych; mają także szereg programów wtyczkowych. Chodzi jednak o niezawodność i dokładność tych możliwości, które w żadnym wypadku nie dają żadnego z testów, które poważni matematycy użyliby do sprawdzenia poprawności wyników.

Zarządzanie danymi

Nacisk na poziomie komórek powoduje, że arkusze kalkulacyjne stawiają przed użytkownikiem szereg wyzwań w zarządzaniu danymi. Konceptualnie prosta zmiana, taka jak modyfikacja czasu rozpoczęcia, dodanie nowych składowych lub zmiana formuły, może wymagać dziesiątek, a nawet setek innych zmian.

Chociaż niektóre aplikacje arkusza kalkulacyjnego zawierają funkcje do przewidywania przyszłych trendów i wyników, metody te są często niewiarygodne i niedokładne.

Nawet jedna prosta modyfikacja może wymagać wstawiania lub usuwania komórek/wierszy/kolumn, edytowania lub kopiowania formuł w całym zakresie komórek lub ponownego konfigurowania całego arkusza kalkulacyjnego. Operacje te są nie tylko czasochłonne, ale mogą prowadzić do większej liczby błędów.

Niemal niezmiennie nowe dane muszą być dodawane do gotowego arkusza kalkulacyjnego. Ale w jaki sposób należy uwzględnić nowe liczby? Jednym ze sposobów jest ustawienie arkusza kalkulacyjnego na rozszerzanie równań w celu uwzględnienia wszystkich nowych danych: problem polega na tym, że niektóre równania mogą zostać nierozważnie rozszerzone, aby uwzględnić dane, które nie powinny się w nich znajdować. Z drugiej strony, jeśli arkusz nie zostanie ustawiony automatycznie, niektóre dane, które powinny zostać uwzględnione, prawdopodobnie zostaną pominięte. Tak czy inaczej, jest mało prawdopodobne, żeby arkusz kalkulacyjny dawał prawidłowe wyniki, chyba że zmiany zostaną starannie sprawdzone.

Alternatywa dla arkuszy kalkulacyjnych

Jak dotąd w tym dokumencie omówiono sytuacje, w których arkusze kalkulacyjne mogą okazać się niewystarczające lub co najmniej kłopotliwe dla analizy statystycznej. Nie znaczy to, że są bezwartościowe. Jeśli zadaniem jest przeprowadzenie prostych testów na niewielkiej liczbie zmiennych, to arkusz kalkulacyjny jest równie dobrym narzędziem, jak każde inne.

Z tego wynika, że arkusz kalkulacyjny jest, jak stwierdzono wcześniej, oprogramowaniem ogólnego przeznaczenia. Z wtyczkami lub bez, zakres narzędzi analitycznych jest ograniczony, a algorytmy programu arkusza kalkulacyjnego nie są tak rygorystycznie zaprojektowane ani przetestowane, jak programy specjalnie zaprojektowane do analizy statystycznej.

Oprogramowanie IBM SPSS Statistics oferuje organizacjom możliwość przeprowadzania rzetelnej, dogłębnej analizy informacji statystycznych bez konieczności programowania.

Podobnie jak stolarz może użyć piły ręcznej do cięcia kilkunastu kawałków drewna, ale sięgnąłby do specjalistycznych narzędzi do produkcji szafek i elektronarzędzi do obróbki wystarczającej ilości drewna do budowy, każdy, kto chciałby wykonać solidną i dogłębnią analizę, powinien użyć narzędzia specjalnie do tego przeznaczonego. Jednym z nich jest IBM® SPSS Statistics.

SPSS Statistics jest stale rozwijany i testowany od 1968 roku. W tym okresie wiele programów statystycznych zostało wbudowanych w oprogramowanie, a algorytmy wykonujące te równania, zostały przetestowane zarówno przez programistów, jak i użytkowników w środowisku akademickim, w laboratoriach i praktycznie we wszystkich rodzajach działalności. W rezultacie użytkownicy mogą mieć pewność, że oprogramowanie zostało dokładnie przetestowane, a jego wyniki uznane za wiarygodne.

Nie wykonując żadnego programowania, użytkownicy mogą wykonywać bardzo szeroki zakres analiz statystycznych. Ponadto wraz z postępem zrozumienia przez użytkowników, mogą oni natychmiast zastosować bardziej zaawansowane metody, które są już w oprogramowaniu.

Oczywiście SPSS Statistics jest zoptymalizowany do obsługi obliczeń statystycznych w sposób, w jaki arkusz kalkulacyjny nigdy nie mógł być. W rzeczywistości oprogramowanie jest zoptymalizowane pod kątem pracy statystycznej na każdym etapie, od wprowadzania danych po tworzenie raportów dla decydentów.

Wprowadzanie danych metodą SPSS Statistics

W przypadku SPSS Statistics proces wprowadzania danych rozpoczyna się od definicji rodzajów danych, które zostaną użyte. Są one dość szczegółowe. Na przykład, każdy rodzaj danych ma zarówno długą, jak i krótką nazwę. (Najlepiej pasująca nazwa to taka, która jest najlepsza do wpisywania do tabel i wykresów). Ponadto można podać rodzaj danych, jakie można wprowadzić; prostym przykładem mogą być liczby lub tekst. W tym momencie następuje pierwszy poziom sprawdzania błędów. Dane muszą być zgodne z charakterystyką określonego rodzaju, gdyż nie będą akceptowane. Rodzaju danych ani żadnej innej charakterystyki układu nie można zmienić przypadkowo. Nie można również zmieniać relacji między danymi. Wprowadzanie danych to po prostu wprowadzanie danych: nie ma w ogóle nic wspólnego z programowaniem.

Mechanizmy sprawdzania poprawności danych i błędów zapewnione przez SPSS Statistics są naprawdę kompletne.

Umożliwia porównanie dwóch plików danych lub dwóch zestawów danych według metadanych całego dokumentu lub porównanie poszczególnych wartości zmiennych w celu zidentyfikowania rozbieżności między nimi. Automatyczne procedury lokalizują wartości, które wydają się być niezgodne i polega to na lokalizowaniu większości literówek. Jeśli jednak wartość znajdowała się w zakresie, ale w pewnym sensie była nietypowa w porównaniu do innych wprowadzonych liczb, SPSS Statistics zauważy i zapyta o to.

Wbudowane mechanizmy walidacji i sprawdzania błędów w SPSS Statistics przyczyniają się do zapewnienia ważności i prawidłowości danych.

Przygotowanie danych do analizy: podejście SPSS Statistics

Jak zauważono powyżej, często zdarza się, że dane dostępne do analizy są niekompletne. Na przykład, w ankiecie niektórzy mogą pominąć jakieś pytania lub nie odpowiedzieć na pytania. Jak wspomniano wcześniej, obsługa niekompletnych danych w arkuszu kalkulacyjnym stwarza liczne trudności. Dzięki oprogramowaniu SPSS Statistics naukowcy mogą badać dostępne dane i obliczać wartości brakujących elementów (proces zwany „imputacją”). W celu wykrycia brakujących wzorców danych mogą badać dane wykorzystując jeden z sześciu raportów diagnostycznych.

Mogą też szacować statystyki podsumowujące i imputować brakujące wartości, używając automatycznej procedury, która wybiera najbardziej odpowiednią metodę imputacji w oparciu o charakterystykę danych. Następnie analizę można przeprowadzić tak, jakby wszystkie dane były obecne, co w bardzo rzeczywistym i matematycznie uzasadnionym znaczeniu, faktycznie jest.

Inne etapy gromadzenia danych do analizy obejmują analizę rozkładu danych, sprawdzanie wartości

odstających i organizowanie lub binningowanie danych, tak aby algorytmy zaplanowane do użycia, takie jak Naïve Bayes lub modele logitowe, działały wydajnie. SPSS Statistics wykonuje te etapy przygotowania danych, do czego nie jest przeznaczony żaden program arkusza kalkulacyjnego.

Analiza statystyczna z użyciem SPSS Statistics

Gdy oprogramowanie SPSS Statistics przełączy się w tryb analizy i wykona czynności niezbędne do wytworzenia danych wyjściowych, dane nie zostaną zmodyfikowane: są one używane tylko jako dane wejściowe do procesu, a dane wyjściowe dostępne w różnych formatach, w tym imponujące bogactwo nakresów i wykresów, znajdują się w osobnym oknie.

Wyniki z analizy danych z wykorzystaniem oprogramowania SPSS Statistics można przekazywać w różnych formatach, w tym w imponującym bogactwie nakresów i wykresów.

Ponadto, gdy wykonywany jest dowolny rodzaj analizy, oprogramowanie automatycznie zapisuje program w postaci składni, którą można zapisać i uruchamiać wielokrotnie na różnych zestawach danych, bez potrzeby jej zmiany. (W razie potrzeby można ją jednak zmienić.)

Oprogramowanie SPSS Statistics ma także tę zaletę, że umożliwia zaawansowanym użytkownikom wdrażanie nowych procedur i funkcjonalności za pośrednictwem Programmability Extension. Ta zaawansowana funkcja pozwala użytkownikom wygodnie postępującym się językiem programowania statystycznego R, Python, .NET lub Java, na osadzanie nowych algorytmów lub funkcji bezpośrednio w produkcie. Mogą nawet stworzyć natywny graficzny interfejs użytkownika (GUI) dla nowej funkcji, którą utworzyli, aby udostępnić ją osobom innym niż programiści, które następnie mogą szybko i wydajnie wykonywać analizy samodzielnie.

Prezentowania najlepszych wyników dzięki SPSS Statistics

Korzystając z kreatora wykresów (Chart Builder) użytkownicy SPSS Statistics mogą łatwiej tworzyć najczęściej używane wykresy, takie jak matryce rozproszenia (SPLOM), histogramy i piramidy populacji. Osoby pracujące z wykresami statystycznej kontroli procesów mogą zażądać sprawdzenia reguł zarówno w pierwotnych, jak i wtórnych wykresach kontrolnych, co zapewnia większą dokładność i lepsze zrozumienie, czy proces działa normalnie.

System grafiki prezentacji w SPSS Statistics zapewnia kontrolę zarówno na etapie tworzenia, jak i edycji, co umożliwia pracę w warunkach produkcyjnych.

W przeciwieństwie do arkuszy kalkulacyjnych wyniki analityczne można przeglądać geograficznie, korzystając z szablonów map dostępnych w selektorze szablonów wykresów (Graphboard Template Chooser). Można tworzyć różne rodzaje wizualizacji, w tym mapy kartogramów (mapy kolorów), mapy z mini-mapami i mapy nakładki, aby pomóc użytkownikom w bardziej efektywnym planowaniu, prognozowaniu i targetowaniu. Oprogramowanie SPSS Statistics jest dostarczane z kilkoma plikami map – lub narzędzie do konwertowania map (Map Conversion Utility) można wykorzystać do konwersji istniejących plików kształtów map do użycia z selektorem szablonów wykresów.

Można przygotować dane wyjściowe z SPSS Statistics do prezentacji, tworząc tabele bezpośrednio z okien dialogowych procedury z formatowaniem warunkowym. Można również zautomatyzować typowe modyfikacje dokumentu wyjściowego.

Wykresy i tabele można przeglądać w dowolnym czasie i dowolnym miejscu na różnych urządzeniach, w tym w środowiskach Windows, Mac i Linux, iPod, iPhone i iPad, telefonach i tabletach z systemem Android (wersje 2.1 i nowsze), urządzeniach z systemem Windows 8.

Przyszłość z IBM SPSS Statistics

Arkusze kalkulacyjne są często wykorzystywane do prognozowania przyszłych zdarzeń na podstawie danych historycznych lub niepewnych danych wejściowych. Na przykład, powszechnym zastosowaniem biznesowym byłoby prognozowanie przychodów z dwóch kolejnych kwartałów na podstawie wyniku z poprzedniego roku. Chociaż możliwe jest przeprowadzenie takiego obliczenia w arkuszu kalkulacyjnym, to uwzględnienie takich czynników, jak sezonowość w firmie, analiza typu „co jeśli” lub tworzenie scenariuszy opartych na wielu innych zmiennych jest możliwe tylko przy użyciu oprogramowania matematycznie odpornego, takiego jak SPSS Statistics.

Chociaż arkusze kalkulacyjne mogą być używane do tworzenia niektórych projekcji, jednak potrzebne jest narzędzie, takie jak SPSS Statistics do uwzględniania złożonych zmiennych lub sytuacji.

Krzywa uczenia się z IBM SPSS Statistics

Na początku tego dokumentu postawiliśmy pytanie, które często przychodzi nam na myśl, gdy rozważamy użycie nowego oprogramowania: Jak trudno jest się go nauczyć?

W przypadku SPSS Statistics odpowiedź brzmi: „Wcale nie jest trudno”. Podobnie, jak w arkuszu kalkulacyjnym, ma interfejs WYSIWIG (What You See Is What You Get), więc wszystko jest dobrze widoczne, a jego funkcje są dostępne dzięki znanym układom menu i paskowi narzędzi. Funkcje statystyczne programu są pogrupowane logicznie: Po wybraniu jednej z nich w wyskakującym okienku pojawiają się odpowiednie opcje, a obliczenia są wykonywane poprzez wybranie wymaganych opcji i kliknięcie przycisku OK lub Uruchom (Run).

SPSS Statistics jest łatwym do nauczenia się narzędziem pomagającym twojej organizacji przejść poza arkusze kalkulacyjne do matematycznej analizy złożonych danych.

Ponadto oprogramowanie SPSS Statistics zawiera bardzo obszerne samouczki, bardzo szczegółowe pliki pomocy i studia przypadków, które szczegółowo opisują wykorzystanie analizy statystycznej w sytuacjach biznesowych i badawczych. Razem mogą dość szybko uczynić z nowicjusza statystycznego do roli kompetentnego analityka. Firma oferuje oczywiście szereg opcji szkoleniowych, w tym szkolenia internetowe po otrzymaniu stosownego wniosku. Ponadto ze względu na długą historię używania tego oprogramowania przez analityków we wszystkich typach środowisk, strony trzecie zapewniają dodatkowe zasoby edukacyjne, w tym internetowe fora dyskusyjne z poradami innych użytkowników, książkami instruktażowymi i filmami wideo, podręcznikami i zeszytami ćwiczeń.

Licencjonowanie oprogramowania metodą SPSS Statistics

Ponieważ różne organizacje mają różne preferencje i potrzeby w zakresie wspierania instalacji oprogramowania, SPSS Statistics oferuje wiele opcji licencjonowania, które pasują do każdej organizacji. Są to opcje licencji wieczystej, rocznej i oparte na subskrypcji. Oznacza to, że można kupić pełny dostęp do oprogramowania lub miesięczne subskrypcje bądź obie te usługi – w zależności od tego, która z nich najlepiej odpowiada potrzebom użytkowników.

Zakończenie

Pisząc ten dokument, autor odkrył kilka rzeczy wartych poruszenia w tym momencie: Po pierwsze, arkusze kalkulacyjne są używane znacznie szerzej niż nam się powszechnie wydaje, często bez szukania innych rozwiązań. Po drugie, wskaźnik błędów w użyciu arkusza kalkulacyjnego może być wyższy niż dopuszczalny poziom błędów w innych formach wykonywania obliczeń. Po trzecie, arkusze kalkulacyjne są przeznaczone do wielu różnych rodzajów problemów i niektóre z nich mogą nie być w pełni dostosowane do możliwości tych programów.

Każdy zestaw danych jest niepowtarzalny, podobnie jak sposób korzystania z arkusza kalkulacyjnego.

Aby dowiedzieć się, czy arkusz kalkulacyjny jest wystarczający dla danych potrzeb lub czy można skorzystać z wyspecjalizowanego narzędzia, takiego jak SPSS Statistics, najlepiej jest sprawdzić, jak każdy program działa z konkretnymi danymi, wykonując zadania analityczne, które są zwykle potrzebne.

Łatwo jest przetestować SPSS Statistics. Można się skontaktować z firmą lub pobrać darmową wersję próbną oprogramowania na stronie ibm.biz/GetStats.

Jeśli szczególne dane znajdują się już w arkuszu kalkulacyjnym, SPSS Statistics może je łatwo zaimportować. A gdy dane znajdują się już w tym oprogramowaniu, można ocenić dostępne rodzaje analiz i sprawdzić, czy w danym przypadku lub w niektórych sytuacjach istnieją zalety korzystania z narzędzia zaprojektowanego do analizy statystycznej, zamiast programu arkusza kalkulacyjnego ogólnego zastosowania.

Więcej informacji.

Odwiądź stronę ibm.com/analytics/spss, aby uzyskać więcej informacji na temat oferty oprogramowania IBM SPSS.



IBM Polska
ul. 1 sierpnia
02-134 Warszawa

IBM, logo IBM, **ibm.com** i SPSS są znakami towarowymi International Business Machines Corp., zastrzeżonymi w jurysdykcjach wielu krajów. Nazwy innych produktów i usług mogą być znakami towarowymi IBM lub innych podmiotów. Aktualna lista znaków handlowych IBM jest dostępna w sieci WWW, na stronie [“Copyright and trademark information”](#) na stronie ibm.com/legal/copytrade.shtml

Microsoft i Windows jest znakiem towarowym Microsoft Corporation w Stanach Zjednoczonych i (lub) w innych krajach.

Linux jest zarejestrowanym znakiem towarowym Linusa Torvaldsa w Stanach Zjednoczonych i (lub) w innych krajach.

Java i wszystkie znaki towarowe oraz logo zawierające słowo Java są znakami towarowymi firmy Oracle i/lub jej spółek zależnych.

Niniejszy dokument jest aktualny w dniu początkowej publikacji i może zostać zmieniony przez IBM w dowolnym momencie. Nie wszystkie oferty są dostępne w każdym kraju, w którym firma IBM prowadzi działalność.

Cytowane przykłady klientów są prezentowane wyłącznie w celach poglądowych. Rzeczywiste wyniki mogą różnić się w zależności od konkretnej konfiguracji i środowiska. Odpowiedzialność za ocenę i weryfikację działania dowolnych produktów lub programów wraz z produktami i programami IBM ponowi użytkownik. INFORMACJE W NINIEJSZYM DOKUMENCIE SĄ UDOSTĘPNIANE W STANIE, W JAKIM SIĘ ZNAJDUJĄ („AS IS”), BEZ UDZIELANIA JAKICHKOLWIEK GWARANCJI, WYRAŻNYCH ANI DOMNIEMANYCH, W TYM GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU ANI GWARANCJI LUB WARUNKÓW NIENARUSZALNOŚCI. Produkty IBM są objęte gwarancją zgodnie z warunkami umowy, na mocy której są oferowane.

© Copyright IBM Corporation 2018

¹ <http://archive.columbiatribune.com/2006/feb/20060222news009.asp>

² http://articles.marketwatch.com/2005-11-09/news/30780581_1_eastman-kodak-robot-brust-kodak-spokesman-gerard-meuchner

³ <http://www.abc.net.au/news/newsitems/200506/s1394937.htm>

⁴ <http://www.gao.gov/atext/d04754t.txt>.

⁵ Gregg Keizer, “Microsoft fixes Excel math mistake,” Computerworld (March 2008).



Nadaje się do recyklingu