

**mgr Alicja Antonowicz
dr Paweł Antonowicz
dr Sylwester Kania
mgr Renata Opieka
dr Jolanta Sala
dr Anita Szymańska**

**Symulator korzyści stosowania elastycznych
form zatrudnienia
– podręcznik wdrażania**

EUROSTER 

Gdańsk 2008

Wydawca

Instytut Badań nad Gospodarką Rynkową – Gdańska Akademia Bankowa
ul. Do Studzienki 63, 80-227 Gdańsk
tel.: +48 58 524 49 01, faks.: +48 58 524 49 09
www.gab.com.pl
e-mail: poczta@gab.com.pl

Redakcja i skład



www.oddk.pl

Ośrodek Doradztwa i Doskonalenia Kadr Sp. z o.o.
80-317 Gdańsk Oliwa, ul. Obrońców Westerplatte 32A

ISBN 978-83-88835-17-9

Gdańsk 2008



Publikacja jest finansowana ze środków Europejskiego Funduszu Społecznego

Spis treści

Wstęp	7
Słownik pojęć użytych w podręczniku	11
Rozdział 1.	
Znaczenie elastycznych form zatrudnienia na współczesnym rynku pracy	15
Wprowadzenie	15
1.1. Pojęcie i formy elastycznego zatrudnienia	17
1.2. Korzyści i zagrożenia stosowania elastycznych form zatrudnienia	25
Rozdział 2.	
Organizacyjne aspekty warunkujące kształt modelu elastycznych form zatrudnienia	27
2.1. Organizacja na tle zmian w otoczeniu	27
2.2. Organizacyjne uwarunkowania kształtujące możliwości stosowania elastycznych form zatrudnienia	30
2.3. Uwarunkowania z poziomu misji, strategii i celów organizacji	33
2.4. Uwarunkowanie na poziomie strategii personalnej i zarządzania zasobami ludzkimi	36
2.5. Uwarunkowania na poziomie ścieżki kariery	39
2.6. Uwarunkowania na poziomie stanowiska pracy	41
Rozdział 3.	
Metodologia budowy, weryfikacji i testowania modelu elastycznych form zatrudnienia	43
3.1. Cykl życia modelu elastycznych form zatrudnienia	43
3.2. Metodologia budowy, weryfikacji i testowania modelu (CYKL 1)	45
3.2.1. Sformułowanie celów modelu	47
3.2.2. Utworzenie zespołu ekspertów budujących model	48
3.2.3. Analiza rynku pracy i organizacji, która będzie wykorzystywać model	49
3.2.4. Analiza możliwości stosowania elastycznych form zatrudnienia	50
3.2.5. Wybór i grupowanie zawodów poddanych modelowaniu	50
3.2.6. Wybór elastycznych form zatrudnienia	51
3.2.7. Wybór aspektów – kryteriów oceny	51
3.2.8. Sformułowanie wariantów sytuacyjnych preferencji decydenta (w postaci pytań)	51
3.2.9. Instalowanie i wdrożenie edytora modelu decyzyjnego	53
3.2.10. Tworzenie drzewa decyzyjnego	54
3.2.11. Opracowanie treści modelu (formularzy i załączników)	55
3.2.12. Testowanie poprawności logicznej i edycyjnej	56

Spis treści

- 3.3. Wskazania dotyczące drugiego okresu modelowania (instalacja i wdrożenie, stosowanie i rozwój). 58

Rozdział 4.

Wybrane aspekty modelowania drzewa decyzyjnego w zakresie wyboru

form zatrudnienia	61
Wprowadzenie.	61
4.1. Finansowe kryteria oceny decyzji dopuszczalnych	63
4.1.1. Umowa o pracę	64
4.1.2. Samozatrudnienie – działalność gospodarcza	70
4.1.3. Umowa zlecenie.	75
4.1.4. Umowa na zastępstwo usprawiedliwionej nieobecności pracownika	79
4.1.5. Praca tymczasowa – umowa za pośrednictwem agencji pracy tymczasowej.	79
4.2. Pozafinansowe kryteria oceny decyzji dopuszczalnych	81
4.2.1. Godziny pracy i dyspozycyjność pracownika	81
4.2.2. Urlopy pracownicze	83
4.2.3. Możliwość podejmowania działalności konkurencyjnej	86
4.2.4. Prawa do świadczeń z tytułu okresowej niezdolności do pracy	86
4.2.5. Profilaktyczne badania lekarskie	87
4.2.6. Szkolenia z zakresu bezpieczeństwa i higieny pracy	88
4.2.7. Podnoszenie kwalifikacji zawodowych	90
4.2.8. Okresy wypowiedzenia umowy	91

Rozdział 5.

Modelowanie drzewa decyzyjnego przy wykorzystaniu Edytora DMML

– interfejs Eksperta	95
5.1. Opcje i narzędzia wykorzystywane w pakiecie modeli decyzyjnych	97
5.2. Projektowanie założeń do drzewa decyzyjnego	104
5.3. Przeniesienie założeń projektu do drzewa decyzyjnego w Edytorze DMML.	106
5.4. Opis funkcji „Decyzja dopuszczalna”	109
5.5. Sposoby dodawania, usuwania bądź zmiany kolejności wybranych elementów drzewa decyzyjnego	111
5.6. Opis funkcji „Kryterium oceny decyzji”	111
5.7. Opis funkcji „Zmienna lub Parametr”	115
5.8. Tworzenie algorytmów finansowych na poziomie Wyników opisowych przy wykorzystaniu Zmiennych lub Parametrów	121
5.9. Opis funkcji Własność (dana).	125
5.10. Opis funkcji Wybór możliwości	129
5.11. Opis funkcji Wstaw model decyzyjny	135
5.12. Opis dodatkowych funkcji Edytora DMML	139
5.13. Instalacja Edytora DMML (pakietu modeli decyzyjnych) na serwerze www.	140
5.14. Wdrożenie Serwisu Symulatora SKEFZ	141

Rozdział 6.

Statystyczna analiza danych – aspekt analityka 143

Wprowadzenie	143
6.1. Funkcjonalność Edytora DMML w zakresie archiwizacji i prezentacji danych	144
6.2. Metody statystyczne w analizie danych gromadzonych przez Edytora DMML.	148

Spis treści

6.3. Statystyka opisowa	149
6.3.1. Miary tendencji centralnej (średnia, dominanta, kwartyle, mediana, decyle)	150
6.3.2. Miary zróżnicowania (odchylenie standardowe, odchylenie przeciętne, rozstęp, odchylenie ćwiartkowe, współczynnik zmienności)	155
6.3.3. Miary asymetrii (wskaźnik skośności, współczynnik skośności, współczynnik asymetrii)	158
6.3.4. Miary koncentracji (kurtoza)	160
Zakończenie	163
Literatura	165
Akty prawne	166

dr Paweł Antonowicz, mgr Alicja Antonowicz
Uniwersytet Gdański

Rozdział 6. Statystyczna analiza danych – aspekt analityka

Wprowadzenie

Przedstawione w poprzednich rozdziałach możliwości i zastosowania, jakie daje ekspertowi (administratorowi) oraz użytkownikowi (beneficjentowi) Edytor DMML, skupione były do tej pory w obszarze tworzenia samego pakietu modeli decyzyjnych. Warto podkreślić, iż pomimo że stworzenie tego programu było ściśle związane ze wspomaganiem procesów decyzyjnych w zakresie wyboru elastycznych form zatrudnienia, możliwości zastosowania Edytora DMML są znacznie szersze. Wykraczają one bowiem poza obszar samego rynku pracy, gdyż opracowany program może być generalnie wykorzystany wszędzie tam, gdzie możliwe bądź konieczne będzie wsparcie dowolnego procesu decyzyjnego. Warunkiem implementacji tego programu jest jednak opracowanie odpowiedniego ciągu przyczynowo-skutkowego zdarzeń, a także odnalezienie właściwych powiązań logicznych pomiędzy elementami wprowadzanymi do drzewa decyzyjnego.

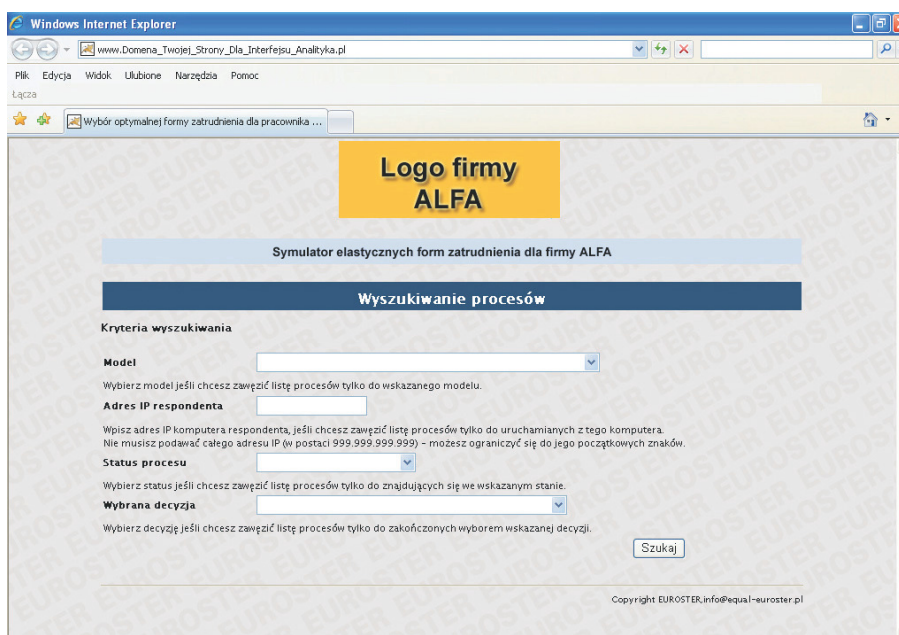
Celem niniejszego rozdziału jest omówienie ostatniego obszaru wykorzystania Edytora DMML, który wynika z archiwizowania przez program w bazie danych wszelkich rekordów, będących wynikiem wykorzystywania go przez użytkowników zewnętrznych. Analiza informacji gromadzonych w bazie danych może być przeprowadzana na różne sposoby. W zależności od konstrukcji drzewa decyzyjnego model analityk będzie posługiwał się innymi metodami i technikami badawczymi. Edytor DMML posiada interfejs analityka umożliwiający śledzenie udzielanych przez użytkowników odpowiedzi, a także ich eksport do arkusza kalkulacyjnego w formie pliku Excel. Podstawowa funkcjonalność programu w tym zakresie została omówiona w kolejnym punkcie niniejszego rozdziału. Natomiast w dalszej części rozdziału zostały przedstawione statystyczne metody analizy struktury zjawisk. Wiedza ta może być wykorzystana przez analityka w celu przeprowadzenia szerszych badań nad

preferencjami użytkowników z zakresu przeprowadzanych przez Edytora DMML symulacji.

6.1. Funkcjonalność Edytora DMML w zakresie archiwizacji i prezentacji danych

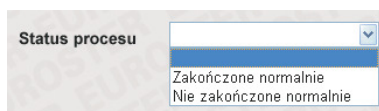
Wszystkie udzielane przez użytkowników zewnętrznych odpowiedzi, a także dokonywane przez nich wybory zapisywane są w bazie danych Edytora DMML. Zgromadzone tam informacje analityk może importować do arkusza kalkulacyjnego i poddawać dalszej analizie, bowiem taka postać „surowych” danych nie będzie stanowiła dla ich odbiorcy wystarczającej wartości poznawczej. Warto jednak zauważyć, iż już sam Edytor DMML poprzez interfejs analityka pozwala stworzyć ogólną charakterystykę udzielanych przez użytkowników odpowiedzi. Opcja ta została wprowadzona do programu, aby ułatwić ekspertowi podsumowywanie statystyk udzielanych odpowiedzi bez konieczności wykorzystywania specjalistycznej wiedzy statystycznej. Oczywiście przeprowadzenie szczegółowych badań nad strukturą udzielanych odpowiedzi (preferencjami użytkowników) będzie wymagało posłużenia się wybranymi metodami statystycznej analizy danych. Podstawową analizę danych należy jednak rozpocząć od przyjrzenia się liczbie udzielanych odpowiedzi. W tym celu należy otworzyć okno interfejsu analityka (rys. 50) i wybrać dowolną konfigurację dostępnych w nim funkcji.

6. Statystyczna analiza danych – aspekt analityka



Rys. 50. Interfejs analityka – ogólna charakterystyka danych

Interfejs analityka zawiera cztery rozwijane pola wyboru, w których można (nie jest to warunek konieczny) wybrać szczegółowe kryteria, w oparciu o które program wygeneruje odpowiednio posortowaną listę odpowiedzi. Pierwszym z kryteriów wyszukiwania jest **Model**. Jeżeli analityk chce, aby została wygenerowana lista odpowiedzi dla wybranego modelu, powinien wskazać właściwe rozwinięcie, w przeciwnym wypadku program wyświetli udzielone odpowiedzi dla wszystkich modeli. Może zaistnieć również konieczność sprawdzenia udzielanych odpowiedzi w ramach wybranego numeru **IP** (*Internet Protocol address* – unikatowy numer przyporządkowany urządzeniom sieci komputerowych). Trzecie kryterium wyszukiwania to **Status procesu**. W ramach tej opcji wygenerować można listę odpowiedzi, które po rozwinięciu wyglądają następująco:



Wybór opcji „**zakończone normalnie**” spowoduje wygenerowanie listy tych odpowiedzi, które przeszły wszystkie kroki dialogu, tzn. zakończyły się podjęciem przez użytkownika decyzji (wskazania, którą decyzję dopuszczalną wy-

brał) oraz przesłaniem opinii (uwag, pytań i spostrzeżeń do eksperta (rys. 51). Opcja „nie zakończone normalnie” pozwala wylistować te odpowiedzi, które zakończyły się na pierwszym etapie dialogu (bez przekazania informacji o podjętej decyzji oraz bez wyrażenia przez użytkownika opinii).

Wyszukiwanie procesów

Kryteria wyszukiwania

Model

Wybierz model jeśli chcesz zawęzić listę procesów tylko do wskazanego modelu.

Adres IP respondenta

Wpisz adres IP komputera respondenta, jeśli chcesz zawęzić listę procesów tylko do uruchamianych z tego komputera. Nie musisz podawać całego adresu IP (w postaci 999.999.999.999) – możesz ograniczyć się do jego początkowych znaków.

Status procesu

Wybierz status jeśli chcesz zawęzić listę procesów tylko do znajdujących się we wskazanym stanie.

Wybrana decyzja

Wybierz decyzję jeśli chcesz zawęzić listę procesów tylko do zakończonych wyborem wskazanej decyzji.

Wyniki wyszukiwania

Znaleziono: 61
Strona: 1 / 4

Lp.	Model	Adres IP użytkownika	Data i godzina	Podjęta decyzja	Opinia
1	Formy zatrudnienia dla zawodu:	194.181.108.34	2007-08-28 08:57:21	Umowa o pracę na zastępstwo nieobecnego pracownika	+
2	Formy zatrudnienia dla zawodu:	195.117.155.155	2007-08-28 09:40:16	Własna działalność gospodarcza	+
	Formy zatrudnienia dla zawodu:	195.117.155.155	2007-08-28 09:47:43	Własna działalność gospodarcza	+

Logo firmy ALFA

Symulator elastycznych form zatrudnienia dla firmy ALFA

Szczegóły procesu

Podaj kwotę brutto, dla której zostanie przeprowadzona symulacja.	4500
Jaki jest Twój staż pracy?	5
Czy dopuszczasz możliwość założenia własnej działalności gospodarczej?	tak
Czy chciałbyś sam organizować sobie czas pracy?	nie

Rys. 51. Lista odpowiedzi użytkowników w interfejsie analityka dla kryteriów wyszukiwania: „zakończone normalnie”

Wskazanie kursorem wybranego wyniku wyszukiwania (poprzez naciśnięcie odpowiedniej strzałki) spowoduje otworzenie w nowym oknie szczegółów zawierających udzielone przez danego użytkownika odpowiedzi. Podobną listę można wygenerować w oparciu o wskazanie ostatniego kryterium wyszukania, jakim jest **Wybrana decyzja**. W niniejszym podręczniku został stworzony pakiet modeli decyzyjnych dla trzech opisanych w poprzednim rozdziale form zatrudnienia: *umowa o pracę*, *samozatrudnienie* oraz *umowa zlecenie*. Dlatego też wybór w tym polu właściwego kryterium wyszukiwania spowoduje, iż Edytor DMML wygeneruje listę udzielonych odpowiedzi dla wybranej formy zatrudnienia.

Jak zostało przedstawione na powyższych rysunkach, interfejs analityka daje pewien ogólny obraz preferencji użytkowników symulatora i umożliwia szczegółowe przyjrzenie się konkretnym odpowiedziom wybranych grup użytkowników. Niemniej jednak, aby móc sformułować szersze wnioski na temat użytkowników programu, analityk będzie musiał wykorzystać wiedzę z zakresu stosowania podstawowych miar statystycznych.

Warto w tym miejscu podkreślić, iż aby móc uogólniać wnioski na temat populacji w oparciu o określoną próbę, należy przede wszystkim zastosować właściwy dobór próby badawczej. Prawdopodobnie będzie to bardzo trudne w przypadku ogólnego (powszechnego) zastosowania Symulatora EFZ na terenie Polski, gdyż wykorzystywanie jego jest ograniczone wieloma czynnikami, m.in. brakiem dostępu do sieci internetowej wszystkich potencjalnie zainteresowanych użytkowników (np. osób w wieku produkcyjnym). Dlatego też warto pamiętać, iż formułowane wnioski w tym wypadku nie powinny odnosić się do uogólniania obrazu całej populacji Polaków, a jedynie do przedstawiania analiz odpowiedzi udzielonych przez użytkowników samego programu. Sytuacja ta jednak nie będzie miała miejsca w przypadku opracowania przez eksperta symulatora dla wybranego przedsiębiorstwa. Wówczas użytkownikami tego programu będą wszyscy jego pracownicy i wysoce prawdopodobne będzie, iż określona liczba osób korzystających z tego narzędzia, przy zagwarantowaniu jego powszechnej dostępności, będzie reprezentatywnie odzwierciedlała preferencje większości pracowników firmy. Te założenia są oczywiście podstawowymi i nie stanowią opisu wszystkich zasad związanych z zachowaniem reprezentatywności próby badawczej. Czytelnika szerzej zainteresowanego tą problematyką odesłać należy do literatury specjalistycznej z zakresu statystyki.

W kolejnym punkcie poddane zostały analizie podstawowe miary statystyczne, którymi może posłużyć się analityk opracowujący zebrane przez Edytor

DMML w bazie danych odpowiedzi. Funkcjonalność Edytora DMML nie została rozbudowana o te elementy analiz z uwagi na to, iż szerokie spektrum możliwości zastosowania programu de facto oznaczałoby konieczność dopasowania raportów z analizy danych indywidualnie do każdego z opracowanych symulatorów. Jest to oczywiście niemożliwe, dlatego też warto zapoznać się z teorią analizy danych, która może mieć szerokie zastosowanie w dalszych badaniach struktury udzielanych przez użytkowników programu odpowiedzi.

6.2. Metody statystyczne w analizie danych gromadzonych przez Edytor DMML

Statystyka to nauka, której przedmiotem zainteresowania są metody pozyskiwania i prezentacji, a przede wszystkim analizy danych opisujących zjawiska masowe. Możemy wyróżnić następujące metody statystyczne, które wykorzystywane są na kolejnych etapach badania statystycznego:

- ◆ analiza struktury zjawisk:
 - miary tendencji centralnej,
 - miary zróżnicowania,
 - miary asymetrii,
 - miary koncentracji.
- ◆ analiza współzależności zjawisk:
 - miary korelacji,
 - miary regresji.
- ◆ analiza dynamiki zjawisk:
 - miary tendencji rozwojowej,
 - miary wahań okresowych,
 - miary wahań przypadkowych,
 - indeksy statystyczne.

Z uwagi na charakter badań, które w większości przypadków będzie realizował analityk korzystający z bazy danych utworzonej w oparciu o działanie Edytora DMML, w niniejszym podręczniku przedstawione zostały metody analizy struktury zjawisk. Badania te stanowią pierwszy etap charakterystyki zbiorowości. Metodami opisu danych statystycznych uzyskanych podczas badania statystycznego zajmuje się dział statystyki, który określany jest **statystyką opisową**.

6.3. Statystyka opisowa

Celem stosowania metod statystyki opisowej jest podsumowanie zbioru danych i wyciągnięcie podstawowych wniosków na temat badanej zbiorowości. Wykorzystywane są w tym celu tzw. **parametry opisowe**, które w sposób syntetyczny określają właściwości badanych zbiorowości statystycznych poprzez sumaryczny opis rozkładu występującej w niej cechy, a także porównanie różnych zbiorowości pod względem rozkładu tej samej cechy. Parametry te dzieli się na **klasyczne**⁵⁷ oraz **pozycyjne**⁵⁸. Statystyka opisowa stanowi zazwyczaj pierwszy krok analizy danych, który jest wykonywany przez analityka. Wykorzystywane są w jej obszarze następujące miary opisujące badaną zbiorowość:

- ◆ miary tendencji centralnej (położenia, skupienia),
- ◆ miary zróżnicowania (zmienności, rozproszenia, rozrzutu, dyspersji),
- ◆ miary asymetrii (skośności),
- ◆ miary koncentracji (spłaszczenia, kurtozy)⁵⁹.

W grupie miar służących pomiarowi tendencji centralnej do najczęściej stosowanych miar klasycznych należy zaliczyć: **średnią arytmetyczną**, **średnią harmoniczną** oraz **średnią geometryczną**. Miary pozycyjne w tej grupie stanowią: **dominanta** (wartość najczęstsza, modalna), oraz **kwantyle** – **kwartyle** (podział zbiorowości na cztery części, kwartył II stanowi Mediana), **decyle** (podział zbiorowości na dziesięć części), oraz **centyle** (percentyle, podział zbiorowości na 100 części). Oczywiście nie wszystkie z wymienionych miar analityk powinien zastosować w swoich badaniach. Zastosowanie tych ostatnich jest uzasadnione wyłącznie w przypadku posiadania bardzo dużej liczby obserwacji.

Wśród miar zróżnicowania wyróżniamy miary klasyczne, takie jak: **wariacja**, **odchylenie standardowe**, **odchylenie przeciętne** oraz **współczynnik zmienności**. Natomiast do miar pozycyjnych w tej grupie zaliczamy **rozstęp** i **odchylenie ćwiartkowe**. Zastosowanie tych miar pozwala badaczowi określić, jakie jest zróżnicowanie wartości cechy w badanej zbiorowości statystycznej.

⁵⁷ Parametry klasyczne stanowią wypadkową wartości przyjmowanych przez wszystkie jednostki badanej zbiorowości statystycznej. Ponadto w ramach danej grupy parametrów wykluczają się wzajemnie, a przy ich obliczaniu nie jest konieczne porządkowanie jednostek.

⁵⁸ Parametry pozycyjne są wyznaczane na podstawie wartości jednej lub kilku jednostek zajmujących szczególną pozycję w badanej zbiorowości statystycznej. W ramach danej grupy parametrów uzupełniają się wzajemnie, a przy ich obliczaniu konieczne jest uporządkowanie jednostek według wartości badanej cechy.

⁵⁹ W literaturze w różny sposób nazywane są poszczególne grupy miar. Dlatego też w nawiasach podano najczęściej stosowane i przyjęte w statystyce nazewnictwo.

Ostatnia grupa miar pozwala analitykowi określić, czy wartości badanej cechy są rozłożone równomiernie w stosunku do wartości średniej, czy też mają tendencję do skupiania się przy dolnej lub górnej granicy przedziału zmienności tej cechy. Wśród miar asymetrii wyróżnić należy **współczynnik skośności** oraz **współczynnik asymetrii**. Natomiast pośród najczęściej wymienianych w literaturze przedmiotu miar koncentracji rozkładu znajduje się **kurtoza** (miara spłaszczenia rozkładu).

6.3.1. Miary tendencji centralnej (średnia, dominanta, kwartyle, mediana, decyle)

Pierwszą i zarazem najczęściej stosowaną miarą położenia jest **średnia arytmetyczna**. Definiuje się ją jako iloraz sumy wartości cechy mierzalnej i liczebności populacji. Średnia jest wielkością mianowaną tak samo jak badana cecha (wyrażona jest w tej samej jednostce). Wyraża ona przeciętny poziom badanej cechy (zmiennej) w analizowanej zbiorowości. Średnią arytmetyczną powinno się stosować w przypadku posiadania informacji o wszystkich wartościach badanej cechy w jednorodnej zbiorowości. Rozkład tej cechy powinien charakteryzować się niewielkim zróżnicowaniem i niewielką asymetrią. Przy dużej dysproporcji wartości badanej cechy (powyżej ± 3 odchyłeń standardowych od wartości średniej) pojedyncze **obserwacje odstające** (tzw. *outliers value*) można usunąć w celu eliminacji zakłóceń zniekształcających obraz średniego poziomu cechy badanej zbiorowości.

Wzór 1. Średnia arytmetyczna (a) prosta i (b) ważona

$$\text{a) } \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{N} = \frac{\sum_{i=1}^N x_i}{N}$$
$$\text{b) } \bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{N} = \frac{\sum_{i=1}^k x_i n_i}{N}$$

gdzie:

N – liczebność całej zbiorowości statystycznej,

$x_i = x_1, x_2, x_3, \dots, x_n$ – wartości cechy mierzalnej poszczególnych jednostek statystycznych.

Średnia arytmetyczna przedstawiona wzorem 1 nosi nazwę średniej arytmetycznej prostej (nieważonej) i ma zastosowanie w przypadku, gdy liczba obserwacji jest niewielka. W sytuacji, gdy mamy do czynienia z licznymi zbiorowościami (poszczególne wartości cechy występują więcej niż jeden raz), to przy wyliczaniu średniej wartości można każdą z cech uwzględnić we właściwej proporcji do częstości jej występowania. Taką średnią nazywamy średnią arytmetyczną ważoną. Średnia arytmetyczna spełnia następujące reguły:

- ♦ $x_{\min} < \bar{x} < x_{\max}$ – wartość średniej arytmetycznej będzie zawsze oscylowała pomiędzy wartością minimalną cechy w zbiorowości a jej wartością maksymalną.
- ♦ $\sum_{i=1}^n (x_i - \bar{x}) = 0$, oraz $\sum_{i=1}^k (x_i - \bar{x}) n_i = 0$ – co oznacza, iż suma odchyleń poszczególnych wartości cechy od średniej arytmetycznej równa się zero.
- ♦ $\sum_{i=1}^n (x_i - \bar{x})^2 = \min$ oraz $\sum_{i=1}^k (x_i - \bar{x})^2 n_i = \min$ – suma kwadratów odchyleń poszczególnych wartości cechy od średniej arytmetycznej jest minimalna.
- ♦ $n\bar{x} = \sum_{i=1}^n x_i$ lub $n\bar{x} = \sum_{i=1}^k x_i n_i$ – suma wartości cechy jest równa iloczynowi średniej arytmetycznej i liczebności zbiorowości (oznacza to, iż znając wartość średnią oraz liczebność zbioru można odtworzyć sumę wartości cechy, co niekiedy w zastosowaniach ekonomicznych tej miary okazuje się przydatne).
- ♦ Jeżeli wszystkie wartości zmiennej powiększymy, pomniejszymy bądź pomnożymy o daną stałą, to średnia arytmetyczna będzie równa sumie, różnicy lub iloczynowi średniej arytmetycznej wyjściowych zmiennych i tej stałej.
- ♦ Średnia arytmetyczna jest wrażliwa na wartości skrajnie odstające (*outliers value*).
- ♦ Średnia arytmetyczna z próby jest dobrym przybliżeniem wartości przeciętnej.

W przypadku gdy skala pomiarowa nie jest liniowa lub gdy badamy średnie tempo zmian wielkości w czasie, należy zastosować średnią geometryczną. Stosuje się ją zatem w badaniach średniego tempa zmian zjawisk, a więc gdy zjawiska są ujmowane dynamicznie.

Wzór 2. Średnia (a) geometryczna oraz (b) harmoniczna

$$\text{a) } \bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

$$\text{b) } \bar{x}_H = \frac{\sum_{i=1}^k l_i}{\sum_{i=1}^k m_i} = \frac{\sum_{i=1}^k l_i}{\sum_{i=1}^k \frac{l_i}{x_i}}$$

gdzie:

x_i – wartość i -tego wariantu badanej cechy,

l_i – wartość i -tego wariantu licznika badanej cechy,

m_i – wartość i -tego wariantu mianownika badanej cechy.

Średnią harmoniczną stosuje się natomiast wówczas, gdy wartości cechy są podane w przeliczeniu na stałą jednostkę innej zmiennej, czyli w postaci wskaźników natężenia. Powinna być ona stosowana wówczas, gdy wartości badanej cechy wyrażają stosunek między dwoma zjawiskami powiązаныmi ze sobą w logiczny sposób. Z wymienionych powyżej miar określających średni poziom badanej cechy w analizowanej zbiorowości najczęściej stosowana jest średnia arytmetyczna. Niemniej jednak w zależności od rodzaju udzielanych przez użytkowników programu odpowiedzi analityk może w badaniach wykorzystać również dwie pozostałe miary tendencji centralnej.

Do grupy pozycyjnych miar tendencji centralnej zaliczyć należy **dominantę** (modalną, modę, wartość najczęstszą), czyli miarę charakteryzującą wartość występującą w rozkładzie empirycznym najczęściej, tzn. taką, której odpowiada najwyższa liczebność (częstość). Warto pamiętać, iż dominanty nie oblicza się, gdy w szeregu występuje więcej niż jedno maksimum, przedziały szeregu mają różną rozpiętość, bądź w sytuacji gdy dominanta znajduje się w przedziale skrajnym, który jest otwarty i nie można go sztucznie zamknąć. W zależności od realizowania badania na określonym szeregu, wartość dominanty oblicza się inaczej:

- ♦ dla szeregów szczegółowych oraz rozdzielczych punktowych – jest wartością cechy, która występuje najczęściej (posiada największą liczebność, częstość),

- ♦ dla szeregów rozdzielczych z przedziałami klasowymi – można ją wyznaczyć jako przedział, a jej przybliżoną wartość wyznacza się graficznie w oparciu o histogram.

Wzór 3. Wartość najczęstsza (dominanta)

$$D(x) = x_{om} + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} \cdot h_m$$

gdzie:

m – numer przedziału (klasy), w którym występuje modalna,

x_{om} – dolna granica przedziału, w którym występuje modalna,

n_m – liczebność przedziału modalnej, tzn. klasy o numerze m ,

n_{m-1} ; n_{m+1} – liczebność klasy poprzedzającej i następniej,

h_m – rozpiętość przedziału klasowego, w którym występuje modalna.

Ważnych informacji na temat badanej zbiorowości mogą dostarczyć również wskaźniki określone jako tzw. **kwantyle**⁶⁰. W zależności od podziału zbiorowości na mniejsze części wyróżniamy ich następujące odmiany: **kwartyle**, **decyle** i **percentyle**. W analizie danych, uzyskanych na podstawie korzystania przez użytkowników z Edytora DMML, analityk będzie zazwyczaj posługiwał się poziomem kwartyli (w tym medianą, czyli wartością środkową), a także niekiedy decylami. Wartość środkowa dzieli badaną zbiorowość na dwie równe części. Interpretacja tej miary mówi, iż mniejszą bądź równą wartości mediany przyjmuje połowa badanych jednostek. Druga połowa przyjmuje natomiast wartości większe bądź równe medianie.

Wzór 4. Mediana dla skokowego rozkładu empirycznego

$$M(x) = \begin{cases} \frac{x_{n+1}}{2}, & \text{gdy } n \text{ jest nieparzyste} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n+2}{2}}}{2}, & \text{gdy } n \text{ jest parzyste} \end{cases}$$

⁶⁰ Kwantyle definiuje się jako wartości cechy badanej zbiorowości, przedstawionej w postaci szeregu statystycznego, które dzielą zbiorowość na określone części pod względem liczby jednostek.

Wzór 5. Mediana dla rozdzielczego rozkładu empirycznego

$$Me = x_{0m} + h_m \cdot \frac{N_{Me} - \sum_{i=1}^{m-1} n_i}{n_m}$$

gdzie:

m – numer przedziału (klasy), w której występuje mediana,

x_{0m} – dolna granica przedziału, w którym występuje mediana,

n_m – liczebność przedziału mediany – klasy o numerze m ,

$\sum_{i=1}^{m-1} n_i$ – liczebność skumulowana wyrażona sumą liczebności przedziałów poprzedzających przedział mediany,

h_m – rozpiętość przedziału klasowego z medianą,

N_{Me} – pozycja mediany wyznaczona za pomocą wzoru: $N_{Me} = \frac{n}{2}$.

Należy pamiętać, iż aby dokonywać podziału zbiorowości na kwartyli (w tym $Q_2 = Me$), analityk musi wcześniej uporządkować obserwacje według rosnących wartości cechy. Każdy z kwartyli dzieli zbiorowość na określone dwie części:

- ♦ Q_1 (kwartyl pierwszy) – 25% znajduje się z lewej, natomiast 75% zbiorowości z prawej strony kwartyla, $N_{Q_I} = \frac{1}{4}n$
- ♦ Q_2 (kwartyl drugi) – 50% znajduje się z lewej oraz 50% zbiorowości z prawej strony kwartyla (wartość środkowa, mediana), $N_{Q_{II}} = \frac{2}{4}n = \frac{1}{2}n$
- ♦ Q_3 (kwartyl trzeci) – 75% z lewej i 25% zbiorowości z prawej strony kwartyla, $N_{Q_{III}} = \frac{3}{4}n$.

Różnica pomiędzy trzecim i pierwszym kwartylem nosi nazwę rozstępu ćwiartkowego lub **rozstępu kwartylnego**. Zróznicowanie cechy jest tym większe, im większa jest wartość rozstępu ćwiartkowego.

6.3.2. Miary zróżnicowania (odchylenie standardowe, odchylenie przeciętne, rozstęp, odchylenie ćwiartkowe, współczynnik zmienności)

Miary zmienności pozwalają określić, jakie jest zróżnicowanie wartości cechy w badanej zbiorowości statystycznej. Dają zatem odpowiedź na dwa pytania, w konsekwencji czego dzielą się na dwie grupy:

- ♦ **miary absolutne** – odpowiadają na pytanie, o ile średnio różnią się wartości cechy przyjmowane przez poszczególne jednostki zbiorowości od swej przeciętnej,
- ♦ **miary stosunkowe** – odpowiadają na pytanie, jak duże są to różnice w stosunku do wartości przeciętnej.

Do miar zmienności absolutnych zaliczamy: **odchylenie standardowe**, **odchylenie przeciętne**, **rozstęp** oraz **odchylenie ćwiartkowe**. Do stosunkowych miar zmienności zaliczamy natomiast: **współczynnik zmienności** oraz **pozytywny współczynnik zmienności**.

Odchylenie standardowe stanowi klasyczną, absolutną miarę zmienności, która określa, o ile średnio różnią się poszczególne wartości cechy od średniej arytmetycznej. Jest ono obliczane jako pierwiastek z wariancji (wzór 6).

Wzór 6. Odchylenie standardowe

$$S = \sqrt{S^2}$$

gdzie:

S – odchylenie standardowe,

S^2 – wariancja.

Wariancję definiuje się natomiast jako średnią arytmetyczną kwadratów odchyleń wartości cechy od średniej arytmetycznej zbiorowości. Jest ona pomocniczą miarą zmienności, której się nie interpretuje (wzór 7).

Wzór 7. Wariancja

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{N}$$

gdzie:

S^2 – wariancja,

x_i – wartość i -tego wariantu badanej cechy,

\bar{x} – średnia arytmetyczna,

N – liczebność zbiorowości statystycznej.

Odchylenie przeciętne jest średnią arytmetyczną bezwzględnych odchyleń wartości cechy od jej średniej arytmetycznej. Interpretuje się je podobnie jak odchylenie standardowe.

Współczynnik zmienności stanowi klasyczną, względną miarę zmienności, która pozwala określić siłę zmienności i porównać zmienność różnych cech i różnych zbiorowości. Stanowi on iloraz odchylenia standardowego (lub przeciętnego) i średniej arytmetycznej. Im wyższa jest wartość współczynnika zmienności, tym zmienność zbiorowości jest większa.

Wzór 8. Współczynnik zmienności

$$V(s) = \frac{S}{\bar{x}} \text{ lub } V(s) = \frac{S}{\bar{x}} \cdot 100$$

gdzie:

$V(S)$ – współczynnik zmienności,
 S – odchylenie standardowe,
 \bar{x} – średnia arytmetyczna.

Dokonując interpretacji uzyskanej wartości współczynnika zmienności, warto pamiętać o następujących zasadach:

- ♦ $V(S) = 0$ lub $V(S) = 0\%$ – brak zmienności,
- ♦ $(0 - 0,34)$ lub $(0 - 34\%)$ – zmienność mała,
- ♦ $\langle 0,34 - 0,67 \rangle$ lub $\langle 34 - 67\% \rangle$ – zmienność średnia,
- ♦ $\langle 0,67 - 1 \rangle$ lub $\langle 67 - 100\% \rangle$ – zmienność duża.

Odchylenie ćwiartkowe to pozycyjna, absolutna miara zmienności, określająca średnie odchylenie kwartyła pierwszego i trzeciego od mediany. Definiuje się je jako połowę różnicy pomiędzy trzecim i pierwszym kwartyłem. Odchylenie ćwiartkowe mierzy zróżnicowanie w tzw. **zawężonym obszarze zmienności** – odrzucane są jednostki o wartościach badanej cechy poniżej pierwszego kwartyła (25%) oraz powyżej trzeciego kwartyła (25%).

Wzór 9. Odchylenie ćwiartkowe

$$Q = \frac{Q_3 - Q_1}{2}$$

gdzie:

Q – odchylenie ćwiartkowe,
 Q_1 – kwartył 1,
 Q_3 – kwartył 3.

Pozycyjny współczynnik zmienności jest miarą względną, która określa siłę zmienności w dwóch środkowych ćwiartkach zbiorowości oraz pozwala porównać zmienność różnych cech i różnych zbiorowości.

Wzór 10. Pozycyjny współczynnik zmienności

$$V(Q) = \frac{Q}{Me} \quad \text{lub} \quad V(Q) = \frac{Q}{Me} \cdot 100$$

gdzie:

$V(S)$ – pozycyjny współczynnik zmienności,

Q – odchylenie ćwiartkowe,

Me – mediana.

Typowy obszar zmienności wyznacza przedział najbardziej typowych wartości w całej zbiorowości (typowy obszar zmienności klasyczny) lub w dwóch środkowych ćwiartkach zbiorowości (typowy obszar zmienności pozycyjny).

Wzór 11. Typowy obszar zmienności (a) klasyczny i (b) pozycyjny

$$\text{a) } \bar{x} - S(x) < x_{\text{typ}} < \bar{x} + S(x)$$

$$\text{b) } Me(x) - Q(x) < x_{\text{typ}} < Me(x) + Q(x)$$

gdzie:

x_{typ} – typowy obszar zmienności,

$S(x)$ – odchylenie standardowe,

\bar{x} – średnia arytmetyczna,

$Me(x)$ – mediana,

$Q(x)$ – odchylenie ćwiartkowe.

Jeżeli rozkład jest symetryczny, to 68% (ok. 2/3) jednostek przyjmuje wartości z tego przedziału. Pozycyjny przedział zmienności jest węższy od przedziału dla miar klasycznych.

Rozstęp definiuje się jako różnicę pomiędzy największą i najmniejszą wartością cechy. Jest miarą bardzo prostą, posiadającą jednocześnie małą wartość poznawczą, gdyż nie daje informacji o zróżnicowaniu poszczególnych wartości cechy w zbiorowości (wzór 12).

Wzór 12. Rozstęp

$$R = x_{\max} - x_{\min}$$

gdzie:

R – rozstęp,

X_{\max} – najwyższa wartość cechy,

X_{\min} – najniższa wartość cechy.

5.3.3. Miary asymetrii (wskaźnik skośności, współczynnik skośności, współczynnik asymetrii)

Miary asymetrii charakteryzują rodzaj i stopień odstępstwa od symetrii rozkładu badanej cechy. Podobnie jak w przypadku wcześniej opisanych parametrów miary asymetrii można podzielić na klasyczne i pozycyjne:

- ♦ miary klasyczne: **współczynnik skośności** (A_s lub A_d), **współczynnik asymetrii** (A),
- ♦ miary pozycyjne: **współczynnik skośności** (A_Q).

Najprostszą do obliczenia miarą asymetrii jest **wskaźnik skośności**. Dla miar klasycznych stanowi on różnicę pomiędzy średnią arytmetyczną i modalną (wzór 13a). Dla miar pozycyjnych natomiast odległość kwartyli od mediany (wzór 13b).

Wzór 13. Wskaźnik skośności (a) klasyczny i (b) pozycyjny

$$\text{a) } W_s = \bar{x} - M_o$$

$$\text{b) } W_Q = (Q_{III} - M_e) - (M_e - Q_I) = Q_I + Q_{III} - 2M_e$$

gdzie:

W_s – wskaźnik skośności klasyczny,

\bar{x} – średnia arytmetyczna,

M_o – modalna,

W_Q – wskaźnik skośności pozycyjny,

Q_I – kwartyl 1,

Q_{III} – kwartyl 3,

M_e – mediana.

Jeżeli rozkład badanej cechy jest symetryczny, to średnia jest równa modalnej, a wskaźnik skośności jest równy zero.

$$W_s = \bar{x} - M_o = 0$$

Natomiast jeżeli rozkład badanej cechy nie jest symetryczny, to mamy do czynienia z asymetrią rozkładu. Można wyróżnić dwa rodzaje asymetrii: lewo- i prawostronną.

Dla miar klasycznych:

- ♦ asymetria lewostronna będzie miała miejsce, gdy:
 $W_s = \bar{x} - M_o < 0$
- ♦ asymetria prawostronna będzie występowała, gdy:
 $W_s = \bar{x} - M_o > 0$

Dla miar pozycyjnych natomiast:

- ♦ asymetria lewostronna będzie miała miejsce, gdy:
 $W_Q = (Q_{III} - M_e) - (M_e - Q_I) < 0$
- ♦ asymetria prawostronna będzie występowała, gdy:
 $W_Q = (Q_{III} - M_e) - (M_e - Q_I) > 0$

Dla porównania kierunku i siły asymetrii w dwóch lub więcej zbiorowościach stosowany jest **współczynnik skośności**, który może przyjąć formę klasyczną bądź pozycyjną (wzór 14). Współczynnik skośności w przeciwieństwie do wskaźnika skośności jest liczbą niemianowaną. Im większą wartość przyjmuje współczynnik, tym asymetria jest silniejsza.

Wzór 14. Współczynnik skośności (a) klasyczny i (b) pozycyjny

$$\text{a) } A_s = \frac{\bar{x} - M_o}{S}$$

$$\text{b) } A_Q = \frac{Q_I + Q_{III} - 2M_e}{2Q}$$

gdzie:

A_s – współczynnik skośności klasyczny,

\bar{x} – średnia arytmetyczna,

M_o – modalna,

S – odchylenie standardowe,

A_Q – współczynnik skośności pozycyjny,

Q_I – kwartył 1,

Q_{III} – kwartył 3,

M_e – mediana,

Q – odchylenie ćwiartkowe.

Do klasycznych miar asymetrii zaliczany jest również **współczynnik asymetrii** (wzór 15). Przyjmuje on wartość zero dla rozkładu symetrycznego, wartości ujemne dla rozkładów o lewostronnej asymetrii oraz wartości dodatnie dla rozkładów o prawostronnej asymetrii.

Wzór 15. Współczynnik asymetrii

$$A = \frac{m_3}{S^3}$$

gdzie:

A_s – współczynnik skośności klasyczny,

S – odchylenie standardowe,

m_3 – liczony jest według następującego wzoru: $m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$

5.3.4. Miary koncentracji (kurtoza)

Do oceny stopnia koncentracji stosuje się dwie metody:

- **metodę numeryczną**, polegającą na wyznaczeniu odpowiednich wskaźników liczbowych: **współczynnik skupienia** (kurtoza), **współczynnik koncentracji Lorenza**,
- **metodę graficzną**, polegającą na wykreśleniu i analizie **krzywej koncentracji Lorenza**.

Nadrzędną klasyczną miarą koncentracji jest współczynnik skupienia – kurtoza, określający stopień skupienia poszczególnych obserwacji wokół średniej. (wzór 16).

Wzór 16. Kurtoza

$$K = \frac{m_4}{S^4}$$

gdzie:

K – kurtoza,

S – odchylenie standardowe,

m_4 – liczony jest według następującego wzoru: $m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4$

Im większa jest wartość kurtozy, tym większa jest również koncentracja. W praktyce otrzymaną z obliczeń kurtozę porównuje się z kurtozą rozkładu normalnego. Zatem, jeżeli:

- ◆ $K > 3$, rozkład badanej cechy jest wyższy i smuklejszy od rozkładu normalnego,
- ◆ $K < 3$, rozkład badanej cechy jest niższy i bardziej rozłożysty.

Przedstawione w niniejszym rozdziale podręcznika podstawowe miary statystyczne mogą zostać wykorzystane przez analityka do pogłębienia badań nad znajdującymi się w bazie danych Edytora DMML odpowiedziami użytkowników, w celu poznania struktury ich preferencji. Rozdział niniejszy objął swoim zakresem najczęściej stosowane miary statystyczne, którymi można posłużyć się w analizie wygenerowanych przez Edytor DMML danych. Osoby zainteresowane przeprowadzeniem badania w głębszym zakresie odsyłamy do literatury specjalistycznej z obszaru statystyki.