



PIERWSZE KROKI W ANALIZIE
DANYCH

JAROSŁAW GÓRNIAK
JANUSZ WACHNICKI

IBM SPSS STATISTICS

Spis treści

Wprowadzenie do pracy z programem IBM SPSS Statistics

- 1 Wstępne informacje o programie IBM SPSS Statistics ■ 11**
- 2 Podstawowe konwencje**
 - 2.1 Okna ■ 13
 - 2.2 Rozwijane menu ■ 14
 - 2.3 Okna dialogowe ■ 14
 - 2.4 Pasek narzędzi ■ 15
 - 2.5 Pasek stanu ■ 15
 - 2.6 Pliki ■ 16
- 3 Automatyzacja zadań**
 - 3.1 Język poleceń ■ 17
 - 3.2 Język skryptowy ■ 19
- 4 Podstawowe operacje na plikach danych**
 - 4.1 Struktura zbiorów danych ■ 21
 - 4.2 Edytor danych ■ 25
 - 4.3 Tworzenie nowego zbioru danych ■ 27
 - 4.4 Wczytywanie i zapisywanie zbioru danych ■ 32
 - 4.5 Współpraca z relacyjnymi bazami danych i arkuszami kalkulacyjnymi z wykorzystaniem ODBC ■ 37
- 5 Zarządzanie zbiorami danych**
 - 5.1 Łączenie zbiorów danych ■ 43
 - 5.2 Agregacja danych ■ 48
 - 5.3 Sortowanie obserwacji ■ 49
 - 5.4 Analiza danych w podgrupach ■ 50
 - 5.5 Wybór obserwacji do analizy ■ 51
 - 5.6 Ważenie obserwacji ■ 52
- 6 Przekształcanie danych**
 - 6.1 Obliczanie wartości zmiennej (COMPUTE) ■ 53
 - 6.2 Transformacje warunkowe (IF) ■ 56
 - 6.3 Rekodowanie wartości zmiennych (RECODE) ■ 58
 - 6.4 Zliczanie wystąpień wartości (COUNT) ■ 62
 - 6.5 Zamiana zmiennej tekstowej na numeryczną (AUTOMATIC RECODE) ■ 64



- 6.6 Przypisywanie rang wartościom zmiennych (RANK VARIABLES) ■ 65
- 6.7 Zastępowanie brakujących obserwacji (RMV) ■ 66

7 Okno raportów

- 7.1 Obiekty tabelaryczne – tabele przestawne ■ 69
- 7.2 Obiekty tekstowe ■ 77
- 7.3 Wykresy w oknie edytora raportów ■ 78

Pierwsze kroki w analizie danych

8 Garść użytecznych pojęć

- 8.1 Statystyczna analiza danych a rzeczywistość ■ 83
- 8.2 Dwa podejścia w statystyce ■ 85
- 8.3 Co to jest zmienna? ■ 86
- 8.4 Badania korelacyjne a badania eksperymentalne ■ 87
- 8.5 Zmienne zależne i niezależne ■ 87
- 8.6 Poziom pomiaru ■ 88
- 8.7 Uwagi o sposobie zapisu sumowania ■ 90

9 Analiza jednej zmiennej: rozkład empiryczny zmiennej i miary tendencji centralnej

- 9.1 Rozkład częstości zmiennej ■ 91
- 9.2 Co to jest miara tendencji centralnej? ■ 96
- 9.3 Miara tendencji centralnej dla zmiennych nominalnych ■ 97
- 9.4 Mediana ■ 99
- 9.5 Inne kwantyle ■ 101
- 9.6 Średnia arytmetyczna ■ 103
- 9.7 Uwagi końcowe o stosowaniu miar tendencji centralnej ■ 113

10 Miary rozproszenia

- 10.1 Obszar zmienności, rozstęp ■ 115
- 10.2 Odchylenie przeciętne ■ 116
- 10.3 Wariancja i odchylenie standardowe ■ 116
- 10.4 Współczynnik zmienności i współczynnik koncentracji Giniego ■ 120
- 10.5 Rozstęp ćwiartkowy, odchylenie ćwiartkowe i inne miary rozproszenia oparte na kwantylach ■ 122
- 10.6 Uwagi końcowe ■ 123

11 Jak posługiwać się tabelami współzależności?

- 11.1 Tabela jako narzędzie analizy danych ■ 125
- 11.2 Analiza zależności z użyciem profili wierszy i kolumn ■ 126
- 11.3 Analiza zależności za pomocą reszt ■ 131
- 11.4 Testowanie statystycznej istotności związku pomiędzy zmiennymi ■ 134

12 Mierzenie siły związku pomiędzy zmiennymi w tabelach kontyngencji

- 12.1 Odzwierciedlenie zależności w strukturze tabeli ■ 139
- 12.2 Pożądane cechy mierników współzależności ■ 141
- 12.3 Model proporcjonalnej redukcji błędu (PRE) ■ 142

12.4	Przykład miernika PRE: Lambda Goodmana i Kruskala	■	143
12.5	Miary siły związku oparte na chi-kwadrat	■	146
12.6	Miary siły związku dla zmiennych porządkowych	■	149
12.7	Kilka wskazówek dotyczących używania miar siły związku	■	157
13	Testowanie hipotez o równości średnich		
13.1	Analiza opisowa różnicy pomiędzy średnimi	■	161
13.2	Testowanie istotności różnicy dwóch średnich	■	165
	Zakończenie	■	177
	Literatura	■	179
	Skorowidz	■	181



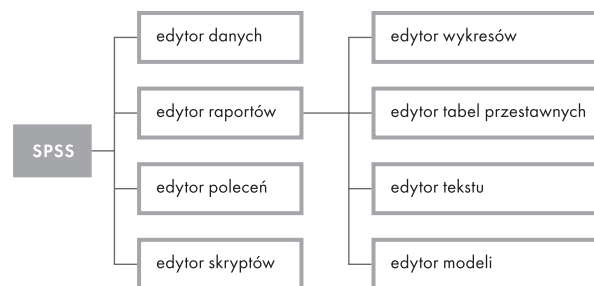
Fragment 1

Podstawowe konwencje

SPSS jest typowym programem pracującym pod kontrolą środowiska Windows, zatem po menu i oknach dialogowych możemy poruszać się korzystając z myszy lub klawiatury. Najczęściej, a w szczególności w początkowym okresie użytkowania programu, korzysta się z myszy i taki sposób obsługi programu będzie opisany poniżej.

2.1 Okna

W programie SPSS funkcjonuje osiem typów okien (rys. 2.1).



Rysunek 2.1
Typy okien
w programie SPSS

Oto ich opis:

- OKNO EDYTORA DANYCH — służy do pracy na zbiorze danych.
- OKNO EDYTORA RAPORTÓW — umieszczane są w nim wyniki i komunikaty programu w postaci obiektów tabelarycznych, graficznych i tekstowych; każdy z tych obiektów ma własne okno edycyjne. Są to:
 - OKNO EDYTORA WYKRESÓW;
 - OKNO EDYTORA TABEL PRZESTAWNYCH;
 - OKNO EDYTORA TEKSTU;
 - OKNO EDYTORA MODELI.
- OKNO EDYTORA POLECEŃ — służy do przygotowywania i uruchamiania poleceń języka programowania SPSS (mogą w nim być umieszczane także odpowiedniki poleceń wydawanych przez użytkownika za pośrednictwem interfejsu graficznego).
- OKNO EDYTORA SKRYPTÓW — pozwala na zaawansowaną edycję wyników analiz, automatyzację zadań związanych z tworzeniem raportów, dostosowywanie programu do potrzeb użytkownika, w tym także budowę własnych aplikacji na bazie pakietu SPSS.



Po uruchomieniu programu na ekranie otwiera się okno edytora danych (w ustawieniach domyślnych programu możemy także wybrać otwieranie na początku sesji okna edytora poleceń). Jednocześnie możemy pracować z wieloma oknami — edytorem raportów, edytorem poleceń i edytorem skryptów.

Pomiędzy oknami można poruszać się za pomocą poleceń z menu OKNO (lub przy użyciu paska zadań systemu Windows).

W menu tym, oprócz poleceń określających układ okien na ekranie, znajduje się także lista otwartych okien. Aktywne okno jest zaznaczone na liście parafką. Możemy przejść do danego okna, wybierając je z listy.

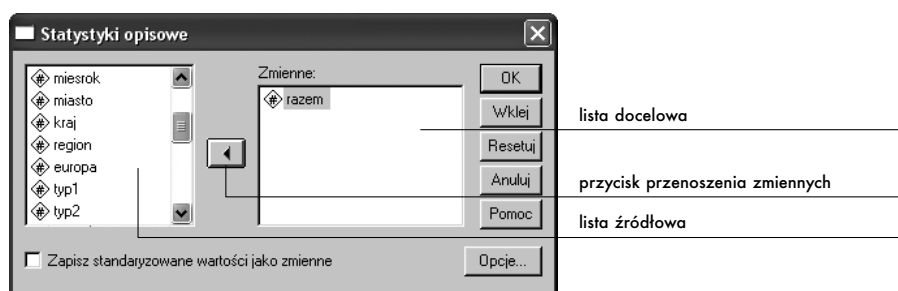
2.2 Rozwijane menu

W każdym z okien programu dostępne jest rozwijane menu. Znajdują się w nim polecenia i opcje programu. Niektóre polecenia są dostępne tylko w oknach określonego typu i związane są ze specyfiką zadań danego okna programu. Rozwijane menu może być modyfikowane przez użytkownika po wybraniu z menu NARZĘDZIA opcji EDYTOR MENU.

Do rozwijanego menu przechodzimy, klikając na nie myszą lub naciskając klawisz lewy ALT. Po menu poruszamy się, wybierając poszczególne opcje za pomocą myszy bądź za pomocą klawiszy kursorów. Do wybranego polecenia możemy też szybko przejść, naciskając klawisz litery podkreślonej w jego nazwie.

2.3 Okna dialogowe

Po wybraniu polecenia z menu na ekranie zwykle pojawia się okno dialogowe.



Rysunek 2.2
Przykład okna dialogowego

Standardowo okna dialogowe zawierają trzy podstawowe elementy (rys. 2.2):

- listę dostępnych zmiennych (LISTA ŹRÓDŁOWA);
- listę zmiennych wybranych do analizy (LISTA DOCELOWA);
- zestaw przycisków poleceń (np. OK, WKLEJ, ANULUJ itp.).

Zmienne pomiędzy listami przenosimy, zaznaczając je myszą i wciskając przycisk ze strzałką w odpowiednią stronę. Przyciski przenoszące zmienne znajdują się pomiędzy LISTĄ ŹRÓDŁOWĄ i LISTĄ DOCELOWĄ.

Typowe przyciski poleceń to:

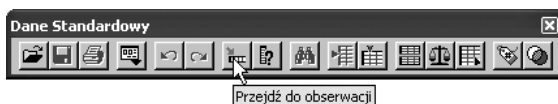
- OK, ZASTOSUJ lub DALEJ — zaakceptowanie ustawionych w oknie dialogowym parametrów;
- WKLEJ — umieszczenie w oknie edytora poleceń zestawu komend odpowiadających ustawieniom w oknie dialogowym;
- RESETUJ — powrót do ustawień początkowych;
- ANULUJ — porzucenie zmian dokonanych w oknie dialogowym (identyczny wynik daje naciśnięcie klawisza ESC);
- POMOC — przywołanie pomocy kontekstowej.

W większości okien znajdują się także przyciski wywołujące kolejne okna dialogowe, które zwykle służą do ustawiania dodatkowych opcji i parametrów, np. określenie funkcji statystycznych, jakie będą wyliczane. Niektóre parametry określamy, zaznaczając je przez kliknięcie myszą, inne podajemy, wpisując je w odpowiadające im pole tekstowe (np. rząd kwantyla, który ma zostać wyznaczony) lub wybieramy z rozwijanych list.

2.4 Pasek narzędzi

W górnej części poszczególnych okien programu SPSS znajdują się paski narzędzi (rys. 2.3). Umożliwiają one szybki dostęp do wybranych opcji programu. Opcjom tym przypisane są przyciski (ikony). Jeżeli opcja jest niedostępna, to odpowiadający jej przycisk jest wyświetlany w szarym kolorze.

Rysunek 2.3
Pasek narzędzi



Paski narzędzi mogą być modyfikowane przez użytkownika po wybraniu z menu WIDOK opcji PASKI NARZĘDZI.

2.5 Pasek stanu

Pasek stanu znajduje się w dolnej części głównych okien programu. Umieszczane są w nim informacje:

- o stanie programu (np. SPSS PROCESOR GOTOWY);
- o zaawansowaniu obecnie wykonywanego polecenia (np. numer aktualnie przetwarzanej obserwacji);
- czy analiza dotyczy tylko podgrupy obserwacji (FILTR WŁ.);
- o przypisaniu wag obserwacjom (WAGA WŁ.);
- o rozdzieleniu zbioru danych na podzbiory, w oparciu o wartości którejs ze zmiennych (PODZBIORY WŁ.).



2.6 Pliki

Z pakietem SPSS związanych jest osiem typów plików:

- zbiory danych w formacie SPSS (*.sav);
- pliki raportów (*.spv);
- pliki z programami w języku poleceń SPSS (*.sps);
- pliki z programami w języku skryptowym (*.wvd, *.sbs);
- zbiory danych w formacie SPSS for DOS (*.sys);
- zbiory danych w formacie wykorzystywanym do przenoszenia danych z SPSS for Windows do SPSS na innych platformach systemowych, czyli tzw. pliki transportowe (*.por);
- pliki z przygotowanymi wzorami wykresów w formacie SPSS (*.sgt);
- pliki z szablonami tabel wykorzystywanych w oknie edytora raportów SPSS (*.stt);

Pliki typu *.sps i *.wvd to zbiory tekstowe, z którymi możemy pracować także za pomocą edytorów tekstu.



Fragment 2

Testowanie hipotez o równości średnich

Do tej pory omawialiśmy zastosowanie średniej arytmetycznej do syntetycznego opisu jednej zmiennej w próbie lub populacji generalnej. Często jednak interesuje nas porównanie średnich z analizowanej zmiennej w różnych grupach lub porównanie średnich różnych zmiennych w jednej grupie badanych, pod warunkiem, oczywiście, że zmienne mierzone są za pomocą takiej samej skali.

W badaniach, zwłaszcza eksperymentalnych, często zadajemy pytania w rodzaju: czy przeciętna wartość pewnej zmiennej jest taka sama, czy też różna w grupie eksperymentalnej i w grupie kontrolnej; albo: czy podanie leku spowodowało istotną zmianę w zakresie symptomu X w badanej grupie pacjentów.

W badaniach sondażowych interesuje nas to, czy badani z dwóch segmentów różnią się istotnie pod względem średniej wieku lub czy płeć, albo inna zmienna nominalna, różnicuje istotnie średnie wartości uzyskane na skali do pomiaru określonej postawy.

Czasami pytania badawcze mają bardziej skomplikowany charakter i dotyczą porównania średnich w grupach zdefiniowanych przez kombinacje dwóch lub więcej cech. Chcemy na to pytanie odpowiedzieć na podstawie pomiaru wykonanego na próbie losowej pobranej z populacji będącej przedmiotem zainteresowania.

SPSS stawia nam do dyspozycji świetne narzędzia statystyczne wspomagające uzyskanie odpowiedzi na tak postawione pytania. Nie możemy tym narzędziom poświęcić tutaj należytej uwagi, ze względu na założone ramy książki. Omówimy tylko techniki podstawowe, koncentrując się na praktycznych aspektach ich użycia w analizie danych.

13.1 Analiza opisowa różnicy pomiędzy średnimi

Jeżeli celem naszej analizy jest przede wszystkim porównanie średnich na poziomie opisowym, możemy użyć znanej już procedury **EKSPLORACJA**. Tym razem skorzystamy z możliwości prowadzenia analizy w podgrupach wyznaczonych przez wartości zmiennej grupującej.

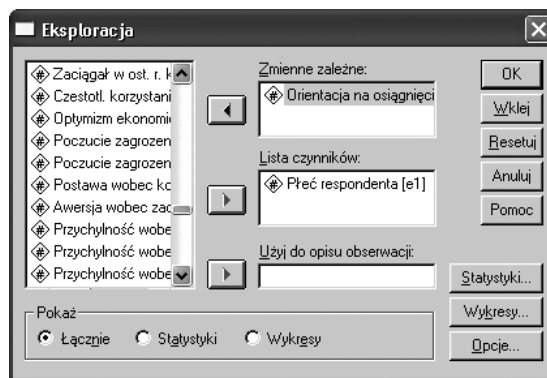
Przeanalizujemy różnicę pomiędzy mężczyznami i kobietami w zakresie orientacji na osiągnięcia. Cecha ta została zmierzona za pomocą skali Likerta. Wartości skali wahają się od 1 (niska orientacja na osiągnięcia) do 5 (wysoka



orientacja na osiągnięcia). Badanie wykonane zostało na losowej próbie dorosłych mieszkańców Polski w 1995 r.

Wartości dla każdego respondenta uzyskane zostały przez zsumowanie punktów za odpowiedzi na siedem pytań, po uprzednim sprawdzeniu jednowymiarowości i rzetelności skali (procedura weryfikacji skal wykracza poza ramy tej publikacji). W celu uzyskania wartości mieszczących się w granicach 1–5, takich jakie miały odpowiedzi na pytania, które zostały użyte do skonstruowania skali, wynik każdej osoby podzielono przez 7 (wartość uzyskana na skali jest więc równa średniej wartości odpowiedzi udzielonych na siedem pytań tworzących skalę).

W celu uzyskania statystyk opisowych dla grup klikamy: ANALIZA ▷ OPIS STATYSTYCZNY ▷ EKSPLOKACJA i do okienka ZMIENNE zależne wkładamy zmienną, której średnie wartości chcemy porównać w interesujących nas grupach. Grupy te wyznaczają kody zmiennej umieszczonej w okienku LISTA CZYNNIKÓW (rys. 13.1).



Rysunek 13.1
Definiowanie analizy
w podgrupach
w procedurze Eksploracja

Możemy użyć również polecenia w języku SPSS:

```
EXAMINE  
VARIABLES=przeds BY e1.
```

Tabela z wynikami analizy EKSPLOKACJA (rys. 13.2) została nieco zmodyfikowana w stosunku do pierwotnego kształtu przez przedstawienie poszczególnych wymiarów za pomocą PANELU PRZESTAWIANIA. Z porównania średnich widać, że mężczyźni charakteryzują się, przeciętnie biorąc, wyższą motywacją do osiągnięć niż kobiety.

Statystyki opisowe

Orientacja na osiągnięcia

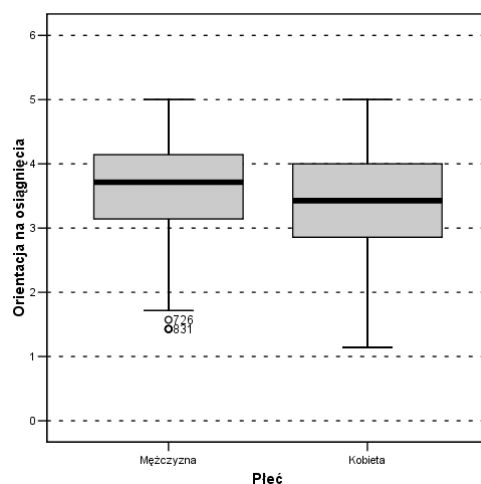
	Statystyka		Błąd standardowy		
	Płeć respondenta		Płeć respondenta		
	Mężczyzna	Kobieta	Mężczyzna	Kobieta	
Średnia	3,6201	3,3840	,03230	,03420	
95% przedział ufności dla średniej					
	Dolna granica	3,5566	3,3168		
	Górna granica	3,6836	3,4512		
5% średnia obciąża	3,6481	3,4048			
Mediana	3,7143	3,4286			
Wariancja	,521	,585			
Odchylenie standardowe	,72148	,76484			
Minimum	1,43	1,14			
Maksimum	5,00	5,00			
Rozstęp	3,57	3,86			
Rozstęp ćwiartkowy	1,00	1,14			
Skośność	-,604	-,335	,109	,109	
Kurtozja	,177	-,253	,218	,218	

Rysunek 13.2
Statystyki opisowe dla grup
uzyskane w procedurze
Eksploracja

Różnica jest nieduża, co świadczy o tym, że płeć nie jest czynnikiem silnie różnicującym pod względem orientacji na osiągnięcia.

Widzimy też, że średnie ucięte i mediany dwóch grup różnią się w podobnym stopniu, co średnie arytmetyczne, a to w pewnej mierze upewnia nas, że zaobserwowana różnica nie jest efektem wpływu nietypowych, skrajnych obserwacji na jedną ze średnich. Odchylenie standardowe w obu grupach jest zbliżone i stosunkowo nieduże. Przy dużej liczebności każdej z grup skutkuje to małymi błędami standardowymi średniej i w konsekwencji wąskimi przedziałami ufności — 95% przedziały ufności średnich z obu grup nie zachodzą na siebie, co świadczy o tym, że różnica między średnimi jest istotna statystycznie (na poziomie istotności 0,05), tzn. można ją przypisać różnicy między średnimi w populacji generalnej, a nie tylko zmienności losowej prób.

Dobrą ilustracją porównawczą jest wspólny wykres skrzynkowy rozkładu wartości skali orientacji na osiągnięcia w obu grupach, również tworzony przez procedurę EKSPLOACJA (rys. 13.3).



Rysunek 13.3
Wykres skrzynkowy
ilustrujący wyniki pomiaru
motywacji do osiągnięć
w grupie kobiet i mężczyzn

Pokazuje on, że wartość mediany jest wyższa wśród mężczyzn. Przy okazji stwierdzamy, że rozkłady nie są silnie skośne, co pozwala uznać średnią za niezłą miarę tendencji centralnej i jest ważne dla testu t , który omawiamy niżej. Nieliczne obserwacje mieszczą się w zakresie wewnętrznego obszaru obserwacji odstających; nie ma obserwacji ekstremalnych, które mogłyby silniej wpłynąć na średnią i odchylenie standardowe. Rozkład jest nieco skośny (ujemnie) w obu grupach (wśród mężczyzn nieco bardziej), jednak skośność ta jest niewielka. Wykres ten potwierdza informacje zawarte w statystykach opisowych: mamy do czynienia z silniejszą orientacją na osiągnięcia wśród mężczyzn niż wśród kobiet, choć różnice są nieduże.

Alternatywą dla procedury EKSPLOACJA jest analiza średnich w grupach ŚREDNIE. Aby ją wykorzystać, musimy kliknąć: ANALIZA ▷ PORÓWNYWANIE ŚREDNICH ▷ ŚREDNIE. Pojawi się wówczas okno dialogowe (rys. 13.4), podobne do znanego z procedury EKSPLOACJA.



Rysunek 13.4
Okno definiujące
procedury Średnie



I tu musimy w okienko ZMIENNE ZALEŻNE wstawić zmienną zależną, której średnie w grupach chcemy analizować, a w okienku ZMIENNE NIEZALEŻNE należy umieścić zmienną, której wartości wyznaczają podział na grupy.

Można wskazać więcej niż jedną zmienną zależną: SPSS wyprodukuje wspólną tabelę dla tych zmiennych, ułatwiającą porównanie ich średnich (rys. 13.5).

Rysunek 13.5
Tabela z wynikami
procedury Średnie

Raport			
Orientacja na osiągnięcia			
Płeć respondenta	Średnia	N	Odchylenie standardowe
Mężczyzna	3,6201	499	,72148
Kobieta	3,3840	500	,76484
Ogółem	3,5019	999	,75245

Można też użyć więcej niż jednej zmiennej niezależnej: SPSS utworzy wówczas odrębną tabelę dla każdej zmiennej niezależnej. Przycisk NASTĘPNA, odnoszący się do WARSTWY, umożliwi wstawienie kolejnych zmiennych niezależnych do następnych warstw tabeli. Umieszczenie poszczególnych zmiennych niezależnych (grupujących) w kolejnych warstwach prowadzi do zagnieżdżenia ich kategorii w obrębie kategorii zmiennych umieszczonych we wcześniejszych warstwach tabeli, a nie — jak to ma miejsce w przypadku umieszczenia kilku zmiennych w jednej warstwie — do wyprodukowania odrębnych tabel.

Korzystając z przycisku OPCJE, możemy zdefiniować szczegółowo zestaw statystyk wyświetlanych w tabeli opisującej analizowane grupy. Domyślnie wyświetlane są średnie, odchylenia standardowe i liczba obserwacji w grupie. Dodatkowo można wybrać m.in. medianę, sumę wartości, minimum i maksimum. Można ponadto zaznaczyć opcję obliczenia testu ANOVA i współczynnika eta (rys. 13.6).

Rysunek 13.6
Współczynnik eta uzyskany
w procedurze Średnie

Miara związku		
	Eta	Eta kwadrat
Orientacja na osiągnięcia *		
Płeć respondenta	,157	,025

Test ANOVA służy sprawdzaniu hipotezy, że w populacji generalnej wszystkie wartości średnich grupowych są równe. Omówimy go niżej. Współczynnik eta, po podniesieniu do kwadratu (co SPSS automatycznie robi), pozwala powiedzieć, jaka część całkowitej wariancji zmiennej zależnej wyjaśniana jest przez różnice średnich między grupami, a więc wynika z przynależności do kategorii zdefiniowanych przez zmienną niezależną lub kombinacje wartości zmiennych niezależnych.

Innymi słowy, eta i eta-kwadrat pozwala ustalić, jaka jest siła związku pomiędzy zmienną ilościową a zmienną nominalną. W naszym przykładzie siła tego związku jest bardzo słaba: $\eta = 0,157$, natomiast $\eta^2 = 0,025$.

13.2 Testowanie istotności różnicy dwóch średnich

13.2.1 Test t równości dwóch średnich w próbach niezależnych

Wyżej wskazywaliśmy już, w oparciu o analizę przedziałów ufności, że średnie wartości uzyskane na skali orientacji na osiągnięcia różnią się w grupie kobiet i mężczyzn w niezbyt dużym stopniu, choć stwierdziliśmy, że jest to różnica istotna statystycznie, tzn. odzwierciedlająca różnicę w populacji, a nie tylko losową zmienność prób. Na ogół jednak hipotezy dotyczące równości średnich w populacji na podstawie wyników z próby sprawdzamy za pomocą testów statystycznych, jak np. test t Studenta i test ANOVA. Odrzucenie hipotezy o równości średnich prowadzi do stwierdzenia istnienia statystycznie istotnego związku pomiędzy analizowaną zmienną jakościową i ilościową.

Hipotezę o średnich w dwóch próbach weryfikujemy w oparciu o test t Studenta. W przypadku **prób niezależnych** obserwacje są losowane w sposób niezależny, a więc dobór osób do jednej próby nie wpływa na dobór osób do drugiej próby. Jeśli próba została dobrana losowo, a następnie podzielona na dwie grupy (lub więcej) za pomocą zmiennej klasyfikującej, to podpróby uzyskane w wyniku takiego podziału są niezależnymi próbami losowymi z odpowiadających im subpopulacji. Takimi próbami niezależnymi są w naszym przykładzie mężczyźni i kobiety, których poddaliśmy badaniu orientacji na osiągnięcia.

W przypadku **prób zależnych** istnieje naturalny związek pomiędzy każdą z obserwacji w jednej próbie a obserwacją z drugiej próby. Najczęściej z próbami zależnymi mamy do czynienia w sytuacji, gdy każda próba składa się z tych samych obiektów, natomiast pomiar w każdej z prób następuje w innym czasie, np. próbę pacjentów poddaje się pomiarowi przed podaniem i po podaniu leku. Próbami zależnymi są też np. próby matek i ich dzieci, które spotykamy w badaniach marketingowych w przypadku badania zgodności ocen cech jakiegoś produktu ferowanych przez matki i ich dzieci, a więc gdy osoby dobierane były do próby parami.

Hipoteza zerowa testu t — brzmiąca przewrotnie w stosunku do naszego zainteresowania istotnością różnicy pomiędzy średnimi — jest następująca: w populacji generalnej średnie w obu grupach są jednakowe. Hipoteza alternatywna stwierdza, że te średnie nie są jednakowe. SPSS wylicza wartość sprawdzianu testu na podstawie wzoru, w którego liczniku mamy różnicę między wartością średnich, w mianowniku — błąd standardowy tej różnicy. Im większa wartość różnicy pomiędzy średnimi, tym większa wartość statystyki t , przy danej wielkości błędu standardowego. Podobnie, im większy błąd standardowy, tym mniejsza wartość statystyki t , przy danej wielkości różnicy między średnimi.