7.Przetwarzanie danych statystycznych SPSS

Do tej pory zdobyłeś już podstawową wiedzę na temat statystyki, manipulacji danymi, tworzenia symulacji, modelowania i analizy w ramach logistycznych łańcuchów dostaw, a także prostych metod regresji liniowej. Podczas gdy statystyka oferuje różnorodny zakres

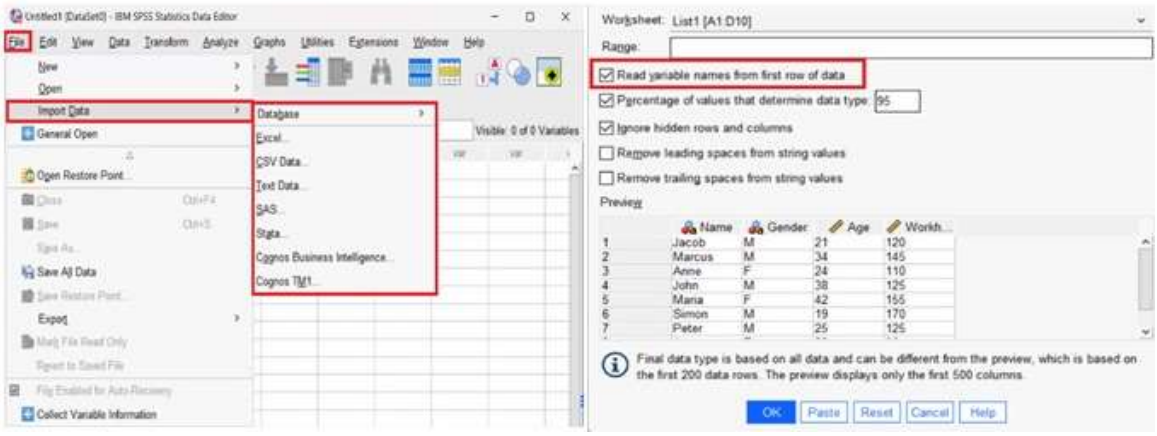
modeli i technik w celu zwiększenia wysiłków optymalizacyjnych, przeprowadzania analiz i identyfikowania potencjalnych ulepszeń, być może zauważyłeś, że wraz ze wzrostem złożoności analizowanych danych i obliczeń, tradycyjne podejścia mogą stać się coraz bardziej skomplikowane i trudne do obliczenia. Wraz ze wzrostem złożoności danych i obliczeń, konwencjonalne metody mogą stać się przestarzałe, a w niektórych przypadkach mogą zagrozić wiarygodności wyników. Aby temu zaradzić, statystyka wykorzystuje różne programy, które automatyzują analizę i interpretację zebranych danych, zapewniając jednocześnie wiele modeli i funkcji zapewniających wiarygodne wyniki. Jednym z takich programów jest IBM SPSS, który będzie kluczowym narzędziem w tym rozdziale. W tym rozdziale przedstawione zostanie zwięzłe wprowadzenie do podstawowego wykorzystania oprogramowania SPSS, badając jego funkcje i praktyczne zastosowania. Po wstępnym wprowadzeniu nastąpi praktyczne zastosowanie programu poprzez cztery podstawowe testy do obliczania wyników: test T, korelacje, Chi - kwadrat i ANOVA. Aby ułatwić naukę, przedstawione zostaną proste problemy i ich rozwiązania, które pomogą w zapoznaniu się z tymi testami.

7.1 Podstawy programu IBM SPSS

Być może masz już pewne doświadczenie z oprogramowaniem SPSS. Jeśli jednak nie, to pozwól nam zaoferować krótkie wprowadzenie. SPSS, podobnie jak jego bardziej znany odpowiednik, Excel, ułatwia manipulowanie danymi, ich analizę i wizualizację. Niemniej jednak, w przeciwieństwie do Excela, który czasami może być pracochłonny i skomplikowany w programowaniu funkcji, SPSS oferuje przyjazny dla użytkownika interfejs do analizy statystycznej (IBM, 2021). Oferuje różnorodne funkcje i metodologie do efektywnej obsługi dostarczonych danych. Chociaż oprogramowanie SPSS wyróżnia się szerokimi możliwościami

analizy statystycznej, poruszanie się po danych i konfigurowanie ustawień początkowych do analizy może czasami stanowić wyzwanie (IBM, 2021). W związku z tym zagłębimy się w podstawy importu danych i przygotowania danych do późniejszych testów statystycznych.

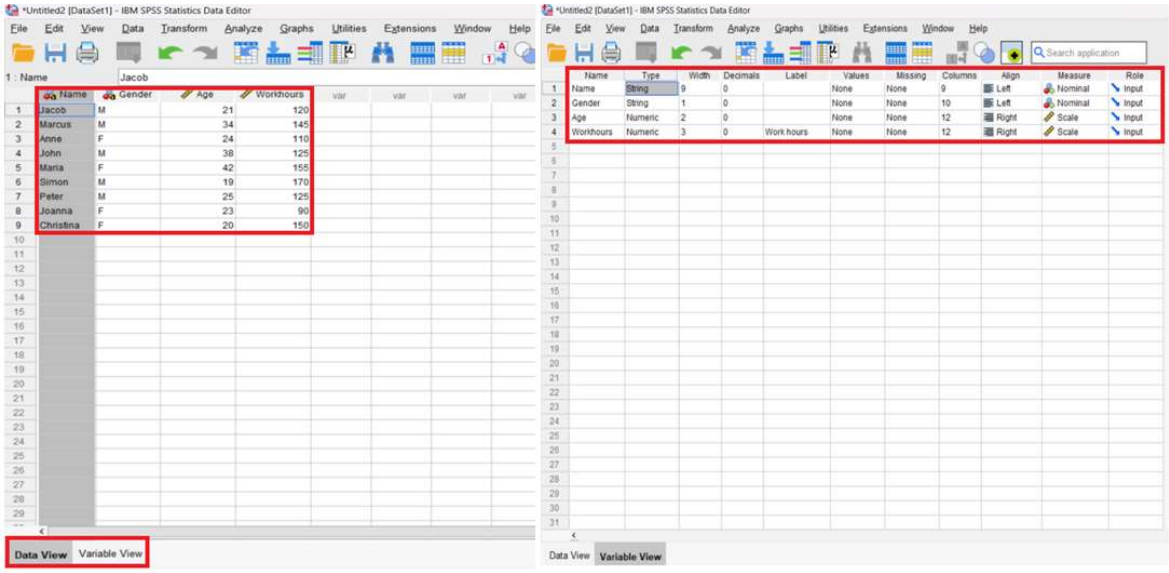
Biorąc pod uwagę powszechne wykorzystanie programu Excel do obsługi danych numerycznych, dane mogą być uzyskane lub przygotowane w arkuszu kalkulacyjnym Excel. Na szczęście SPSS może importować dane z różnych formatów plików do swoich arkuszy kalkulacyjnych. Po przygotowaniu gotowych arkuszy kalkulacyjnych Excel, otwórz oprogramowanie SPSS. Na ekranie początkowym przejdź do zakładki „File" (*Plik*) i wybierz „Import Data" (*Importuj dane*). W kolejnym oknie możesz wybrać format danych, który chcesz zaimportować (patrz rysunek 7.1). Po wykonaniu tego kroku należy zlokalizować przygotowany plik, wybrać go i przejść do następnego okna. Okno to wyświetli monit o skonfigurowanie dodatkowych ustawień. Jeśli nazwy kolumn zostały już uwzględnione w pierwszym wierszu danych, wybierz opcję „Read variable names from first row of data" (*pl. Odczytaj nazwy zmiennych z pierwszego wiersza danych*) (patrz rysunek 7.1), a następnie kliknij przycisk „Finish" (*Zakończ*), aby dane pojawiły się w arkuszu kalkulacyjnym (IBM, 2021).



Rysunek 7.1 Ustawienia importowania danych w programie SPSS.

Teraz, gdy mamy już dane w naszym arkuszu kalkulacyjnym, zauważysz wyraźną różnicę w ich prezentacji w porównaniu do Excela. SPSS kategoryzuje dane na dwa podstawowe typy, każdy z dwoma dodatkowymi podtypami. Jak pokazano na rysunku 7.2, dane można sklasyfikować jako numeryczne lub kategoryczne. Dane numeryczne składają się z liczb i można je

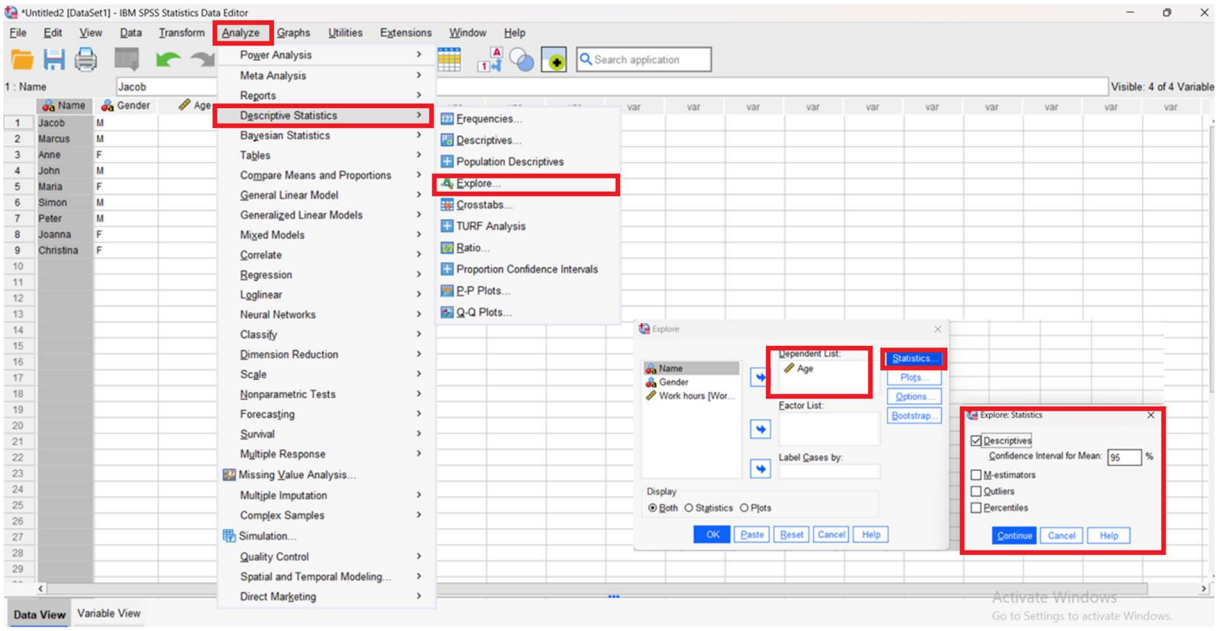
sklasyfikować jako dyskretne (ze skończonymi opcjami) lub ciągłe (oferujące nieskończone opcje). Z drugiej strony, dane kategoryczne obejmują słowa i mogą być dalej rozróżniane jako porządkowe (posiadające hierarchię) lub nominalne (bez hierarchii). W zależności od charakteru danych może być konieczne skonfigurowanie zmiennych w celu dostosowania ich do pożądanej analizy. W większości przypadków SPSS automatycznie odpowiednio skategoryzuje zmienne. Załóżmy, że chcesz dokonać dalszej manipulacji typami danych. W takim przypadku można uzyskać dostęp do opcji „View" (*Widok*) i w sekcji „Variable view" (*Widok zmiennej*) dostosować informacje o zmiennej, takie jak Name (*Nazwa*), Type (*Typ*), Width (*Szerokość*), Measure (*Miara*) i inne (IBM, 2021).



Rysunek 7.2 Okna widoku danych i zmiennych.

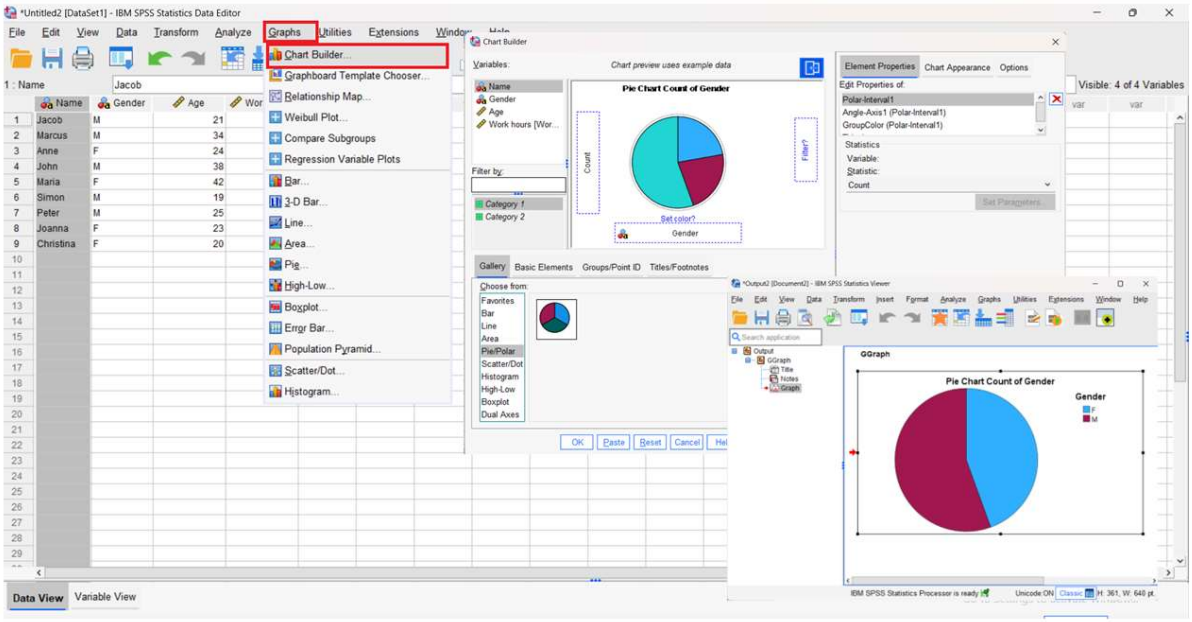
Po prawidłowym skonfigurowaniu danych można je eksplorować w SPSS. SPSS umożliwia użytkownikom wykonywanie podstawowych analiz statystycznych bez polegania na predefiniowanych funkcjach. Na ekranie początkowym (patrz rysunek 7.3), przejdź do „Analyze" (*pl. Analizuj*), a potem „Descriptive Statistics" (*pl. Statystyki opisowe*), a następnie wybierz „Explore" (*Eksploruj*). W sekcji „Explore" (*Eksploruj*) można znaleźć różne opcje w zależności od charakterystyki dostarczonych danych. W tym trybie SPSS dostarczy informacje o danych w postaci „statystyk opisowych". Chociaż jest to cenne dla wstępnej analizy danych, oferuje jedynie podstawowe spostrzeżenia i nie zagłębia się w bardziej szczegółową analizę statystyczną, która zostanie omówiona w kolejnych

rozdziałach. Zanim przejdziemy dalej, zbadamy również inną funkcję w SPSS - wizualizację wykresów (IBM, 2021).



Rysunek 7.3 Ustawienia statystyk opisowych

SPSS oferuje szereg opcji wizualizacji danych, w tym histogramy, wykresy pudełkowe, wykresy słupkowe, wykresy punktowe, wykresy liniowe, wykresy kołowe i inne. Do tego momentu powinieneś mieć podstawową wiedzę na temat tego, co reprezentuje każdy typ wykresu i jak interpretować przedstawione na nich wyniki. Dlatego też skupimy się na sposobie tworzenia tych wykresów w oprogramowaniu SPSS. Aby utworzyć wykresy, wybierz zakładkę „Graphs" (*Wykresy*) na ekranie początkowym, a następnie "Chart Builder" (*Kreator wykresów*). W nowym oknie możesz wybrać typ wykresu, który chcesz utworzyć i wybrać zmienne, które mają zostać uwzględnione. Po wybraniu opcji "Finish" (*Zakończ*) pojawi się nowe okno z wynikami zwizualizowanymi w wybranym formacie wykresu. W tym nowym oknie można aktywnie wchodzić w interakcje z wykresem, umożliwiając modyfikowanie kolorów i czcionek zmiennych, eksplorowanie rozkładów zmiennych na wykresie i nie tylko (IBM, 2021).



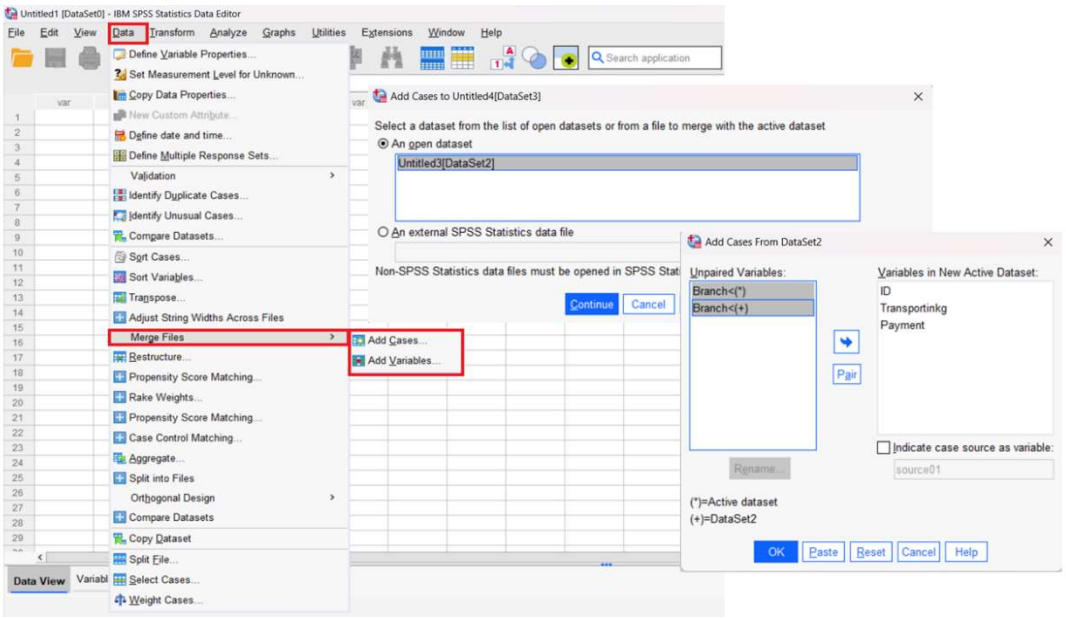
Rysunek 7.4 Ustawienia kreatora wykresów w programie SPSS

Do tego momentu zostały omówione trzy z czterech zasad „Eksploracji danych", które obejmują przeglądanie danych (eksploracja danych surowych), identyfikację danych (określanie typów danych) oraz, do pewnego stopnia, tworzenie wykresów i opisywanie danych za pomocą statystyk opisowych. Ostatnią zasadą jest „Formułowanie pytań", gdzie zadajemy sobie pytanie, co chcemy osiągnąć poprzez analizę danych i odpowiednio konfigurujemy wykresy i statystyki opisowe, aby uzyskać odpowiedzi na nasze konkretne pytania. Na przykład w naszym obecnym przykładzie pytanie może brzmieć: „Czy w analizowanej populacji przeważają kobiety?". Wykorzystując zarówno wykresy, jak i statystyki opisowe, możemy stwierdzić, że nasza populacja składa się głównie z mężczyzn. Podczas formułowania pytań należy zawsze brać pod uwagę dostępne dane izidentyfikowane zmienne (Garth, 2008). Na tym kończy się pierwsza część analizy danych SPSS, a teraz przejdziemy do przeprowadzania testów.

7.2 Zarządzanie danymi

Podczas pracy z kluczowymi danymi w oprogramowaniu SPSS istotne staje się zrozumienie technik manipulowania informacjami w poszczególnych aktywnych zbiorach danych. SPSS zapewnia funkcje ułatwiające manipulowanie istniejącymi danymi zawartymi w aktywnych

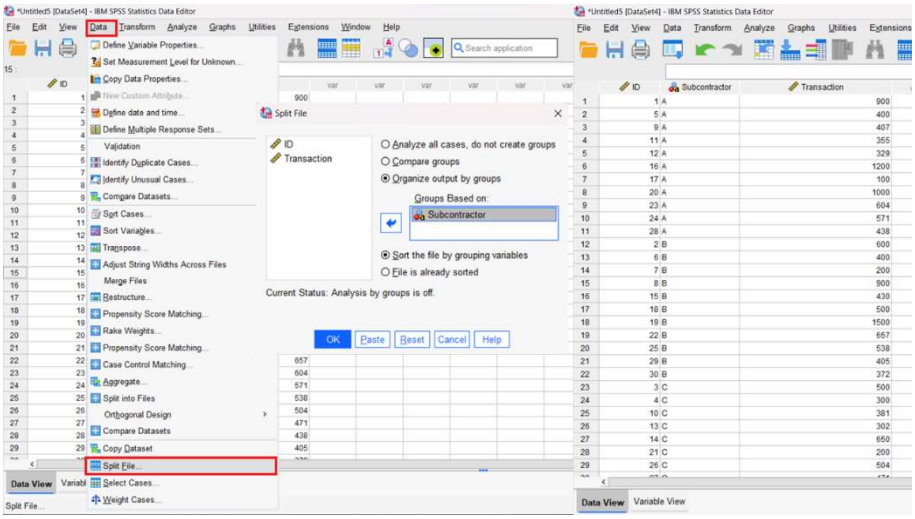
zestawach danych. Czasami można napotkać dwie bazy danych oddzielnie zaimportowane do dwóch plików, zestawów danych, ale preferowane jest ich połączenie w celu lepszej analizy. Rozważmy firmę logistyczną z dwoma oddziałami, z których każdy dostarcza dane dotyczące kosztów i przewożonych ładunków w kilogramach. Celem kierownictwa jest analiza ogólnej wydajności firmy. W programie SPSS osiągnięcie tego celu wymaga przejścia do sekcji "Data" (*pl. Dane*), wybrania opcji "Merge files" (*pl. Scal pliki*) i skorzystania z dwóch różnych opcji. Jedna z nich obejmuje wybranie "Cases" (*pl. Przypadki*) i określenie zmiennej do scalenia, usunięcie tej zmiennej podczas scalania innych. Alternatywnie, wybranie opcji "Variable" (*pl. Zmienna*) zachowuje zmienną w nowym zestawie danych. Praktyczne zastosowanie jest widoczne w naszym scenariuszu logistycznym, w którym scalanie zestawów danych upraszcza kompleksową analizę wydajności firmy.



Rysunek 7.5 Okno scalania plików

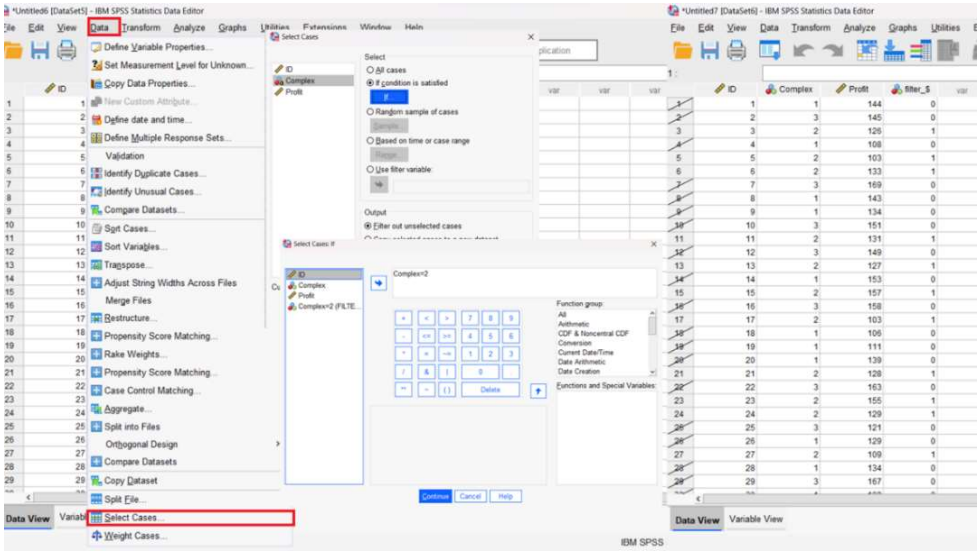
Podczas gdy funkcje scalania i dzielenia umożliwiają określone manipulacje danymi, opcja "Select cases" (*pl. Wybierz przypadki*) oferuje wyraźne korzyści. Wyobraź sobie, że masz dane dla sklepów B, C i D w jednej bazie danych, a skupiasz się wyłącznie na porównaniu sklepu A i sklepu C. Wybierając "Data" (*pl. Dane*) i "Select cases" (*pl. Wybierz przypadki*), można określić interesujące zmienne, skutecznie odfiltrowując niechciane dane. Na przykład ustawienie sklepu C, czyli wartości 2 nakazuje oprogramowaniu skoncentrowanie się wyłącznie na sklepie C, generując dane wyjściowe, które są następnie dostępne

do późniejszych analiz, takich jak statystyki opisowe, koncentrując się wyłącznie na wybranych przypadkach. Takie podejście umożliwia również analizę porównawczą tylko wartości sklepu A i sklepu C.



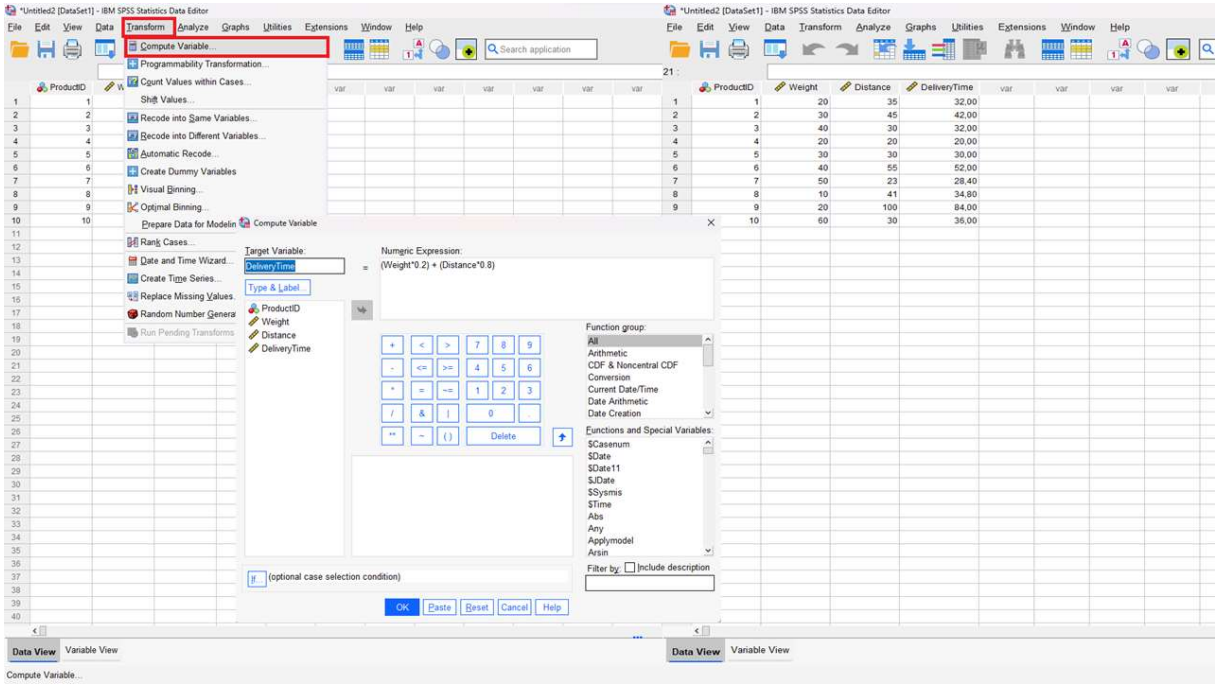
Rysunek 7.6 Okno podziału pliku

Podczas gdy funkcje scalania i dzielenia umożliwiają pewną manipulację danymi, istnieje również opcja "Select cases" (*pl. Wybierz przypadki*). Wyobraźmy sobie, że wiemy na pewno, że sklep A ma średnio 120 € zysku i chcemy porównać to ze sklepem C. Niestety w naszej bazie danych mamy dane dla sklepów B, C i D w jednej bazie danych, a analiza obejmowałaby dane ze wszystkich trzech sklepów. Klikając "Data" (*pl. Dane*) i "Select cases" (*pl. Wybierz przypadki*) można wybrać przypadki o interesującej nas wartość zmiennej, na której chcemy się skupić. Dlatego ustawiliśmy, że zmienna sklep ma wynosić 2, a następnie utworzyliśmy funkcję dla oprogramowania, aby skupić się tylko na sklepie C, który ma kod 2. Dane wyjściowe można następnie wykorzystać do późniejszej analizy, wybierając tę nową kolumnę (np. statystyki opisowe).



Rysunek 7.7 Wybór procedury postępowania

Czasami zbiory danych mogą już zawierać zmienne, ale istnieje potrzeba wprowadzenia nowych zmiennych w oparciu o istniejące. Weźmy na przykład menedżera firmy logistycznej, który posiada dane dotyczące wagi i przebytej odległości dla różnych produktów, ale potrzebuje danych o czasie dostawy w celu optymalizacji tras. W SPSS osiągnięcie tego wymaga kliknięcia "Transform" (*Przekształć*), a następnie "Compute Variables" (*Oblicz zmienne*). Nowa zmienna, DeliveryTime (*CzasDostawy*), jest tworzona w nowym oknie poprzez ustawienie wyrażeń numerycznych. W tym przypadku przypisanie skali 0,8 do odległości i 0,2 do wagi skutkuje nową zmienną reprezentującą czas dostawy, co jest kluczowym dodatkiem do zbioru danych. Istnieje również elastyczność obliczania dodatkowych zmiennych, stworzonych na potrzeby testów statystycznych.



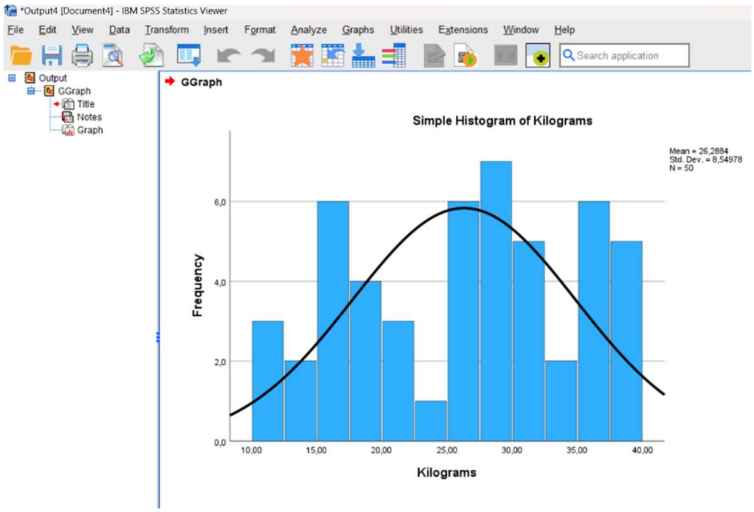
Rysunek 7.8 Procedura obliczania zmiennych

Na tym kończy się mały przegląd funkcji zarządzania danymi, które obejmuje SPSS, a które mogą być przydatne podczas kolejnych testów modeli omówionych w tym rozdziale. Będziemy kontynuować etapy wymagane przed przeprowadzeniem testu statystycznego w oprogramowaniu SPSS.

7.3 Przygotowanie testu

Przed przystąpieniem do testów statystycznych konieczne jest przestrzeganie standardowego procesu analizy danych, który obejmuje eksplorację danych (jak opisano w rozdziałach 7.1 i 7.2), analizę danych i interpretację wyników (Garth, 2008; George & Mallery, 2022). W tym rozdziale skupiamy się na analizie danych przy użyciu oprogramowania SPSS. Ponieważ hipotezy zostały już omówione w poprzednich rozdziałach, skupimy się przede wszystkim na przeprowadzeniu testów normalności w SPSS. Istnieją trzy metody oceny normalności: histogram, wykres QQ i test normalności. Wskazane jest zastosowanie co najmniej dwóch, jeśli nie wszystkich trzech z tych opcji, ponieważ każda z nich dostarcza odrębnych informacji (Ghasemi & Zadesiasl, 2012). Aby utworzyć histogram, przejdź do "Graphs" (*Wykresy*), a następnie "Chart Builder" (*Kreator wykresów*). W nowym oknie wybierz „Histogram”. Jeśli masz wiele zmiennych, musisz powtórzyć ten proces dla każdej z nich, aby uzyskać wyniki.

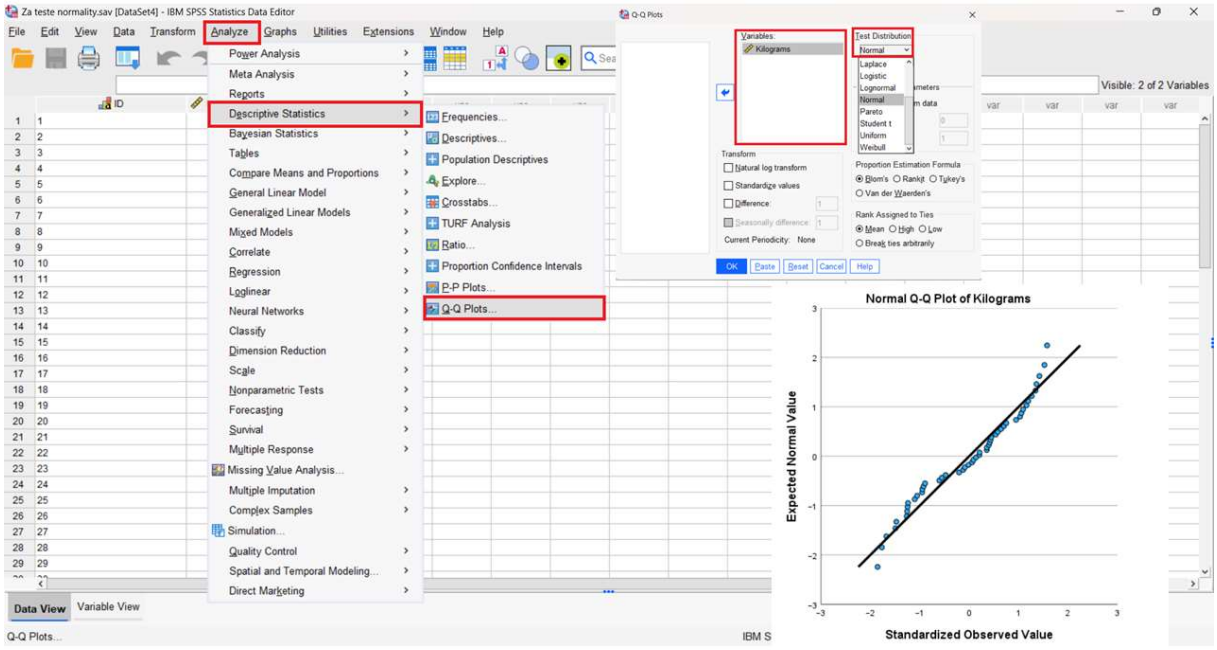
Histogram potwierdza test rozkładu normalnego, jeśli słupki reprezentujące wartości zmiennych przypominają krzywą dzwonową. Jeśli słupki przechylają się bardziej w lewą lub prawą stronę, może to wskazywać na rozkład wykładniczy. Na przykład została wygenerowana baza danych zawierająca 100 rekordów, obserwacji, z których każdy miał zmienną reprezentującą wagę w kilogramach. Postępując zgodnie z instrukcjami, utworzony został histogram, jak pokazano na rysunku 7.9. Jak widać na rysunku, słupki są rozmieszczone na wykresie i chociaż mogą nie odzwierciedlać idealnie krzywej, niemniej jednak sugerują rozkład normalny i pozytywny wynik testu (George & Mallery, 2022; Goeman & Solari, 2021).



Rysunek 7.9 Histogram wyników testu normalności

Inną opcją przeprowadzania testów normalności jest wykres Q-Q, który można zainicjować, klikając "Analyze" (*pl. Analizuj*), następnie "Descriptive Statistics" (*pl. Statystyki opisowe*), a następnie wybierając "Q-Q Plots" (*pl. Wykresy Q-Q*). Zaletą tego podejścia jest to, że pozwala ono na ocenę wielu zmiennych jednocześnie (Williamson, b.d). Test uznaje się za udany, gdy punkty na wykresie skupiają się blisko linii prostej, reprezentującej rozkład normalny. Jeśli punkty tworzą „ogony", oznacza to niepowodzenie testu normalności (Andersen & Dennison, 2018). Korzystając z tej samej bazy danych z testu wykresu histogramu, przeprowadzony został test Q-Q Plot. Na rysunku 7.10 poniżej można zauważyć, że większość punktów skupienia dla naszej zmiennej pokrywa się z linią prostą, co wskazuje na normalny rozkład naszych danych. Chociaż już na tym etapie można było stwierdzić,

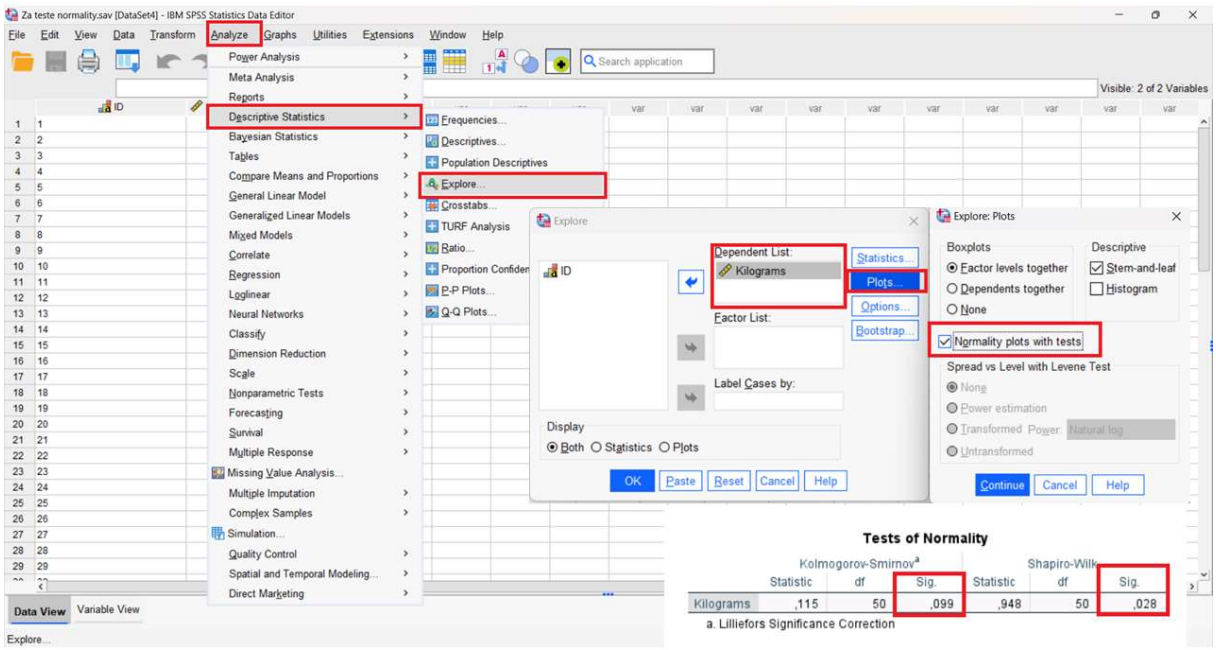
że test normalności jest pozytywny, zdecydowaliśmy się poszukać potwierdzenia we wszystkich trzech testach.



Rysunek 7.10 Ustawienia i wyniki testu normalności wykresu Q-Q.

Ostatnią opcją weryfikowania normalności danych jest tak zwany Test Normalności, uważany za test statystyczny. Zazwyczaj wykorzystuje on test Kołmogorowa-Smirnowa, ale w przypadku małych rozmiarów, do 20 obserwacji, próby należy zastosować test Shapiro-Wilka (Goeaman & Solari, 2021). W programie SPSS można wykonać ten test, klikając "Analyze" (*Analizuj*), następnie "Descriptive Statistics" (*Statystyki opisowe*) a później "Explore" (*Eksploruj*). Należy ustawić zmienne, które mają być sprawdzone w polu "Dependent List" (*Lista zależna*). Następnie w sekcji "Plots" (*pl. Wykresy*) trzeba wybrać "Normality Plots with Tests" (*Wykresy normalności z testami).* Wyjątkowo tylko w tym teście wynik testu hipotezy zerowej (H0) uznaje się za pozytywny (potwierdzający), jeśli kolumna „Sig" (p-value) w wynikach jest większa niż 0,05, co wskazuje na rozkład normalny. Vice versa, jeśli *p-value* jest mniejsze niż 0,05. Sugeruje to rozkład inny niż normalny, a wynik testu uznaje się za negatywny dla hipotezy zerowej. Test ten został przeprowadzony ponownie, korzystając z tej samej bazy danych, co w poprzednich testach. Na podstawie wyników można stwierdzić, że zgodnie ze standardem Kołmogorowa-Smirnowa test jest pozytywny, ponieważ wartość p jest wyższa niż 0,05.

Jednak w przypadku testu Shapiro-Wilka wartość *p* jest niższa, co wskazuje na negatywny wynik testu. Te różne wyniki występują, ponieważ oba podejścia mają różne ustawienia czułości i mocy w wykrywaniu odchyleń (Ghasemi & Zahediasl, 2012). Ponieważ zostały przeprowadzone już zarówno testy wykresów Q-Q, jak i histogramów, test normalności można ogólnie uznać za pozytywny. Po potwierdzeniu testów normalności można przeprowadzić główne testy, takie jak Test dla Jednej Próby.



Rysunek 7.11 Ustawienia i wyniki testu normalności

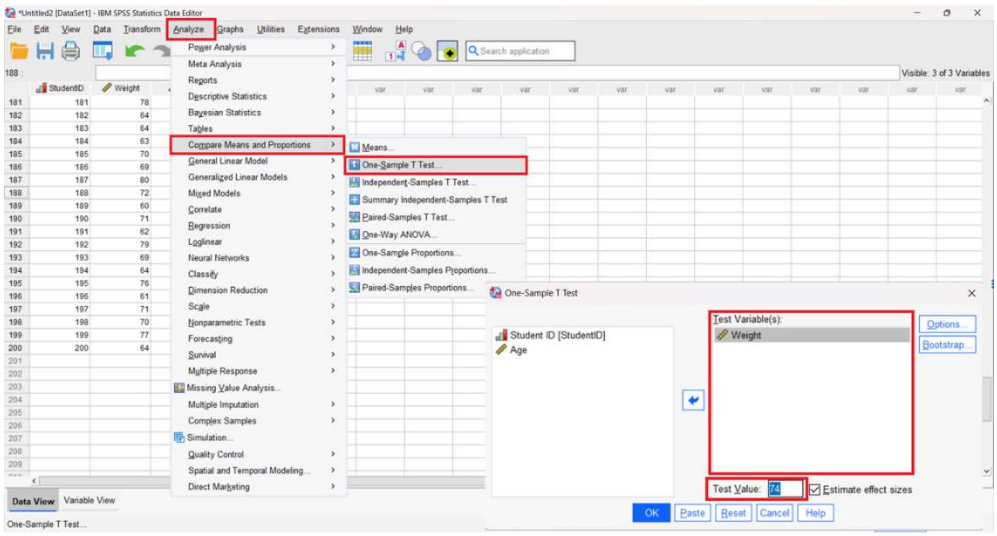
7.4 Test T dla jednej próby

W poprzednich rozdziałach omówiono już teorię stojącą za testem T dla jednej próby; skupimy się zatem przede wszystkim na przeprowadzeniu testu za pomocą oprogramowania SPSS. Na potrzeby naszego testu T dla jednej próby przygotowana została baza danych z próbą 200 danych wejściowych, która obejmuje 1 zmienną kategorialną (Student ID) (*Identyfikator studenta*) i 2 zmienne numeryczne (Weight and Age) (*Waga i Wiek*) (Kim, 2015). Postępując zgodnie ze wskazówkami z poprzednich podrozdziałów, przeprowadzamy następujące kroki:

- Zbadaj dane, a mianowicie nasze **zmienne** i **statystyki opisowe** oraz postaw **pytanie hipotezę**,

- Sprawdź **normalność**, ponieważ powinien wystarczyć tylko jeden histogram zmiennej i wykres Q-Q,
- Postaw hipotezę, w której dla **Wartości zerowej** zmienna nie różni się od określonej wartości i **Alternatywnej**, w której się różni,
- Przeprowadź **test t-Studenta**,
- Zinterpretuj wyniki, koncentrując się na **odrzuceniu** lub **nieodrzuceniu wartości zerowej**, odpowiedz na pytanie i napisz raport z naszego testu.

W naszym przypadku zdecydowaliśmy, że pytanie powinno brzmieć: „Czy średnia waga studentów przekracza 74 kilogramy?". Po zadaniu pytania ustalamy naszą hipotezę, która brzmi: „Zerowa = nie ma różnicy" i „Alternatywna = istnieje różnica". Przeprowadziliśmy histogram i wykresy Q-Q, aby sprawdzić testy normalności, a po ich zakończeniu przeprowadziliśmy test T. Aby uruchomić test T, klikamy "Analyze" (*Analizuj*), następnie "Compare Means" (*Porównaj średnie*) i "One-Sample T-test" (*test T dla jednej próby*). W polu zmiennej testowej umieszczamy „weight" (waga), ustawiamy wartość testowaną na 74 i rozpoczynamy test (patrz rysunek 7.12).



Rysunek 7.12 Ustawienia testu T dla jednej próby

Po potwierdzeniu testu pojawi się kolejne okno z wynikami naszej analizy (patrz rysunek 7.13). Okno to zawiera kilka informacji dotyczących naszej analizy. W tym przypadku obie wartości *p-value* są niższe niż 0,05, co wskazuje na istotność testu, a więc wyniki

negatywny dla H0. Dodatkowo sprawdzamy wartości *t* i *df*, które w naszym przypadku wynoszą odpowiednio -9,806 i 199. Na podstawie tych wyników możemy stwierdzić, że nasza hipoteza zerowa została odrzucona. W związku z tym pełny raport wyników jest następujący: „Średnia waga studenta jest znacząco niższa niż wartość 74 kg. Stwierdzamy poziom średniej wagi niższy niż w hipotezie (był 74 kg), gdyż zauważmy, że średnia = 69,63, a przede wszystkim wartość testu *t*, tzw. sprawdzian, okazała się ujemna: 1-próbkowy test t, *t* = -9,806, *df* = 199, wartość *p*< 0,001)".

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Weight	200	69,63	6,303	,446

One-Sample Test

Test Value = 74

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Weight	-9,806	199	<,001	<,001	-4,370	-5,25	-3,49

One-Sample Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Weight	Cohen's d	6,303	-,693	-,847	-,538
	Hedges' correction	6,326	-,691	-,844	-,536

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

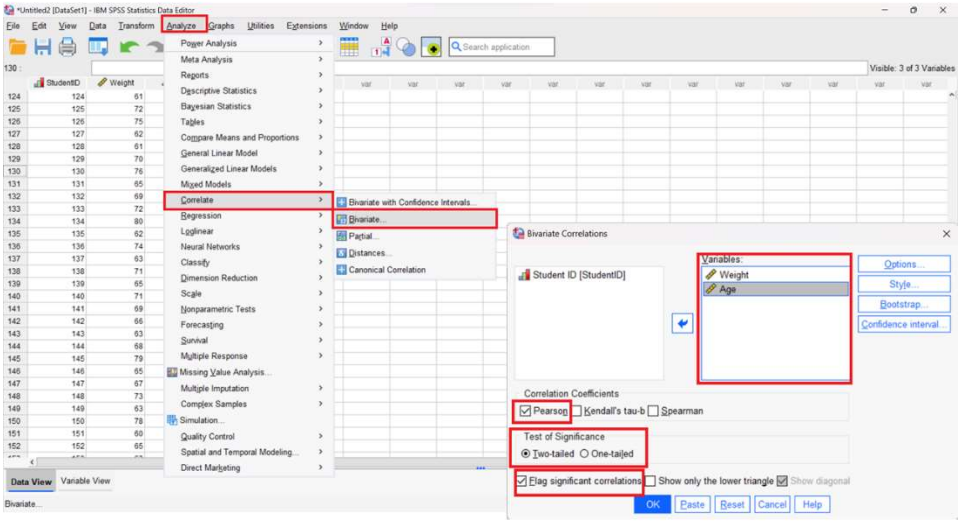
Rysunek 7.13 Wyniki testu T dla jednej próby.

7.5 Korelacja

Przejdźmy teraz do drugiego testu, którym jest test korelacji. Przeprowadzimy go przy użyciu tej samej bazy danych, co w przykładzie testu t dla jednej próby. Podobnie jak w przypadku testu t dla jednej próby, będziemy postępować zgodnie z procedurą z kilkoma modyfikacjami. Podczas przeprowadzania korelacji między dwiema zmiennymi ważne jest, aby określić, która z nich jest zmienną zależną, a która zmienną niezależną (Janse i in., 2021; Mishra i in., 2019). Wyboru tego można dokonać na podstawie pytania badawczego.

W naszym przypadku chcemy zbadać „Czy istnieje korelacja między wiekiem ucznia a jego wagą?". Zgodnie z tym pytaniem uważamy wagę za zmienną zależną, a wiek za zmienną niezależną, ponieważ chcemy zbadać, czy zmiany wieku są związane ze zmianami wagi. Definiujemy hipotezę zerową i alternatywną (patrz rysunek 7.13 i 7.14), a następnie uruchamiamy test, klikając "Analyze"

(*Analizuj*), a następnie "Correlate" (*Koreluj*) i "Bivariate" (*Współzależność prosta*). Obie zmienne powinny zostać umieszczone w polu "Variable" (*Zmienna*). Upewnij się, że „Pearson", „Two-Tailed" (*Dwustopniowy*) i „Flag Significant" (*Oznacz znaczące*) są wybrane lub ustawione (patrz rysunek 7.14). W tym przypadku wybraliśmy „Pearson", ponieważ nasze dane wskazywały na rozkład normalny i mogły być analizowane przy użyciu metod parametrycznych. Jeśli rozkład normalny nie został potwierdzony, należy zastosować metody nieparametryczne (w tym przypadku należy wybrać Spearmana zamiast Pearsona) (George & Mallery, 2022; McClure, 2005).



Rysunek 7.14 Ustawienia testu korelacji

Ponownie otrzymujemy wyniki w nowym oknie (patrz rysunek 7.15). Na podstawie wyników możemy zauważyć, że nasza korelacja Pearsona wynosi -0,038, a nasze *p-value* wynosi 0,596. W analizie korelacji im wartość korelacji jest bliższa zeru, tym słabsza jest korelacja między zmiennymi. W naszym przypadku korelacja jest bardzo bliska zeru, co wskazuje na brak istotnej korelacji między dwiema zmiennymi (McClure, 2005). Ponadto duże *p-value* (0,596) sugeruje, że nie ma istotnych dowodów na to, że istnieje znacząca korelacja między dwiema wybranymi zmiennymi (Williamson, b.d.). W rezultacie nasza hipoteza zerowa nie została odrzucona. Na tej podstawie możemy stwierdzić, że „Nie było korelacji między wiekiem a wagą uczniów".

Correlations

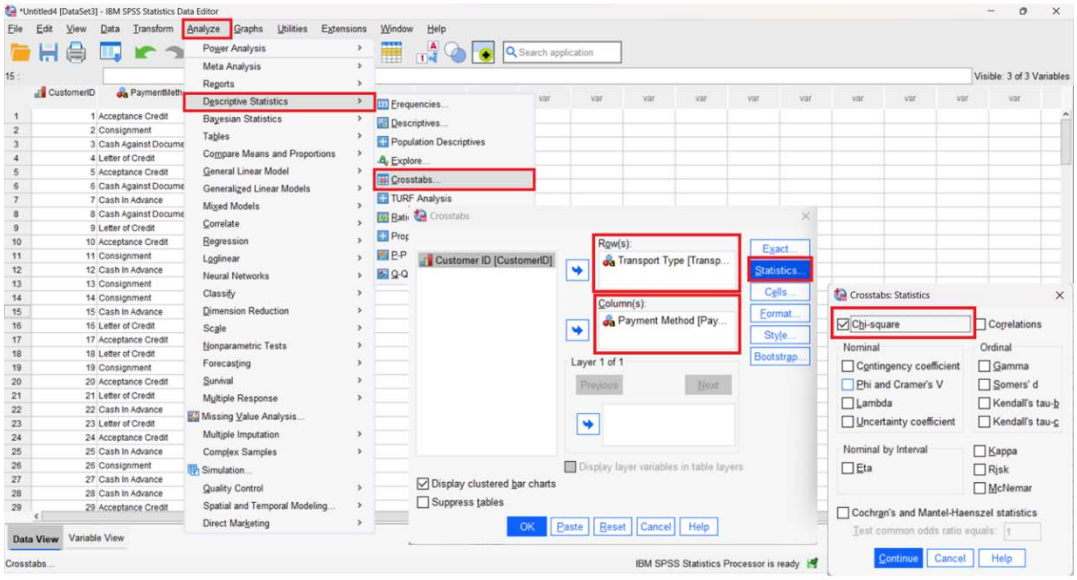
Correlations

		Weight	Age
Weight	Pearson Correlation	1	-,038
	Sig. (2-tailed)		,596
	N	200	200
Age	Pearson Correlation	-,038	1
	Sig. (2-tailed)	,596	
	N	200	200

Rysunek 7.15 Wyniki testu korelacji

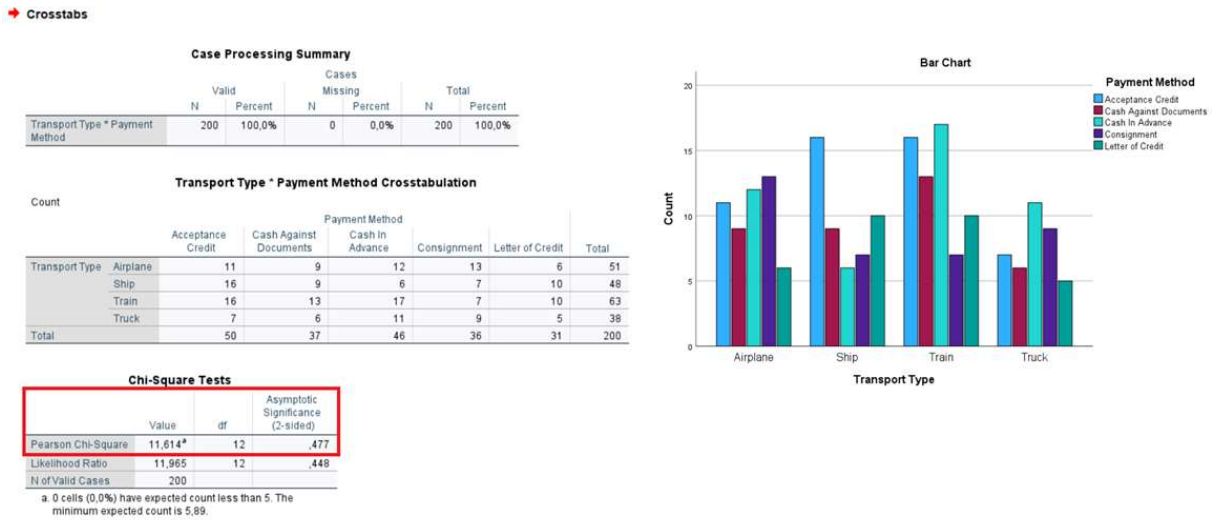
7.6 Chi-Kwadrat

Trzecim testem, który przeprowadzimy w oprogramowaniu SPSS, jest test Chi-kwadrat. W przeciwieństwie do poprzednich dwóch testów, test Chi-kwadrat porównuje dwie zmienne kategorialne, a nie zmienne liczbowe (Turhan, 2020). Podobnie jak w sekcjach 7.4 i 7.5, zaczynamy od zbadania danych i sformułowania pytania badawczego. W naszym przykładzie mamy firmę logistyczną z 200 klientami i dysponujemy danymi na temat rodzaju płatności i rodzaju transportu wybranego przez każdego klienta. Pytanie, na które chcemy odpowiedzieć, brzmi: „Czy różne rodzaje płatności wykazują różne preferencje dotyczące rodzajów transportu?”. Ponieważ mamy do czynienia tylko ze zmiennymi kategorycznymi, nie ma potrzeby przeprowadzania testu normalności. Ustalamy hipotezę zerową (H0: preferencje dotyczące rodzajów transportu są takie same wśród wszystkich rodzajów płatności) i hipotezę alternatywną. Aby przeprowadzić analizę Chi-kwadrat, zgodnie z H0 nazywaną też testem niezależności, kliknij "Analyze" (*Analizuj*), następnie "Descriptive Statistics" (*Statystyki opisowe*) i wybierz „Crosstabs” (*Tabele krzyżowe*). Kluczowe jest umieszczenie zmiennych opartych na pytaniu badawczym w polu kolumny lub wiersza (patrz rysunek 7.16) (Garth, 2008).



Rysunek 7.16 Ustawienia testu Chi-kwadrat

Po zakończeniu analizy w nowym oknie wyświetlone zostaną wyniki (patrz rysunek 7.17). W tym oknie można zauważyć, że wartość Chi-kwadrat Pearsona wynosi 11,614, wartość *df* wynosi 12, a *p-value* (istotność empiryczna) wynosi 0,477. Na podstawie tych wyników możemy stwierdzić, że nie ma istotnego związku między dwiema zmiennymi, a hipoteza zerowa nie została odrzucona. W związku z tym raport jest następujący: „Nie wykryto znaczącej preferencji między różnymi rodzajami płatności dla różnych rodzajów transportu (2- stopniowy test Chi-kwadrat, chi-sq = 11,614, *df* = 12, *p-value* = 0,477)."



Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Transport Type * Payment Method	200	100,0%	0	0,0%	200	100,0%

Transport Type * Payment Method Crosstabulation

Count

		Payment Method					
		Acceptance Credit	Cash Against Documents	Cash In Advance	Consignment	Letter of Credit	Total
Transport Type	Airplane	11	9	12	13	6	51
	Ship	16	9	6	7	10	48
	Train	16	13	17	7	10	63
	Truck	7	6	11	9	5	38
Total		50	37	46	36	31	200

Chi-Square Tests

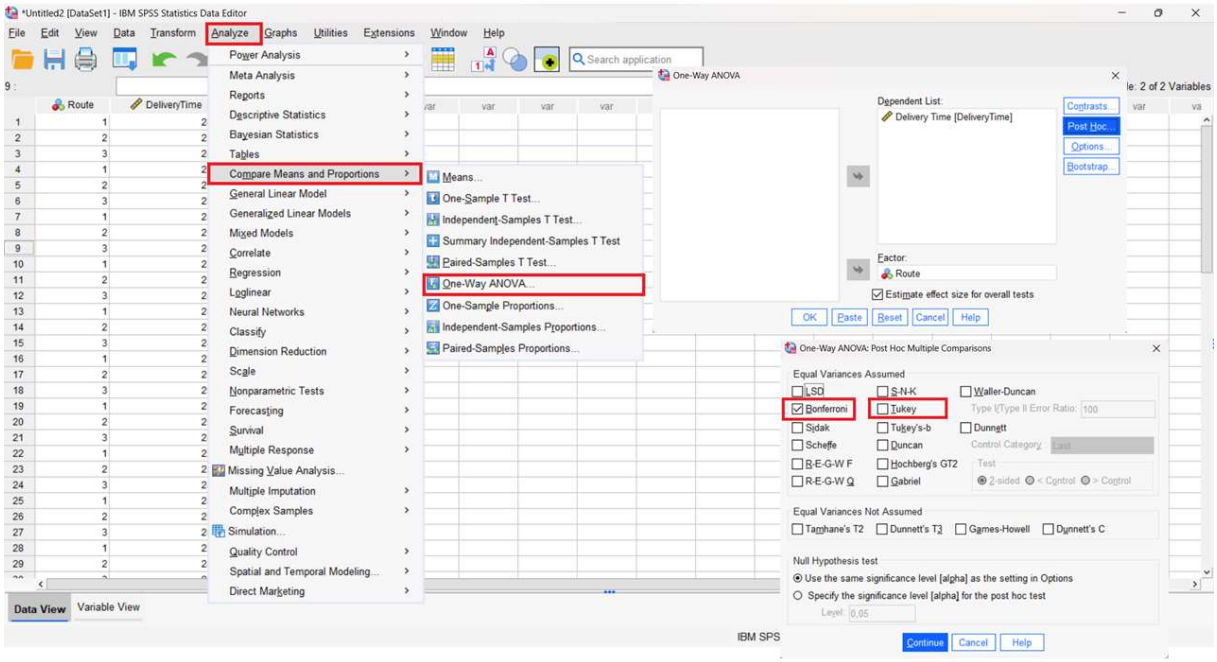
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	11,614[a]	12	,477
Likelihood Ratio	11,965	12	,448
N of Valid Cases	200		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,89.

Rysunek 7.17 Wyniki testu Chi-kwadrat

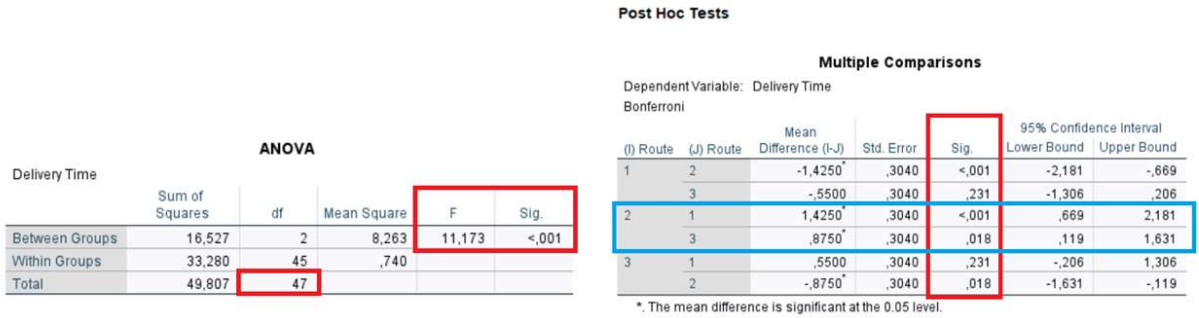
7.7 ANOVA

Ostatnim testem, który zostanie omówiony, jest test ANOVA, w szczególności wariant testu oparty na prostym modelu znanym jako jednoczynnikowa ANOVA, który obejmuje zmienną niezależną kategorialną (to jest czynnik potencjalnie oddziałujący na wartość zmiennej zależnej) i zmienną zależną liczbową (Goeman & Solari, 2021). Podobnie jak w przypadku testu T, będziemy postępować zgodnie z tą samą procedurą: zbadać dane, sformułować pytanie badawcze, przeprowadzić test i postawić hipotezy. Rozważmy studium przypadku dyspozytora transportu pracującego dla firmy logistycznej. Dyspozytor ściśle współpracuje z firmą partnerską i regularnie planuje trzy różne trasy dla ciężarówek w celu dostarczenia towarów. Ze względu na politykę „Just-in-time" kładącą nacisk na szybsze dostawy, pojawia się pytanie: „Czy wybór trasy dostawy wpływa na czas dostawy dla firmy?". Aby przeprowadzić test ANOVA w SPSS, przejdź do "Analyze" (*Analizuj*), następnie „Compare Means..." (*Porównaj średnie…*, a następnie „One-way ANOVA" (*Jednoczynnikowa ANOVA*). Umieść zmienną zależną w polu "Dependent List" (*Lista zależnych*), a zmienną czynnikową w polu "Factor" (*Czynnik)* (patrz rysunek 7.18). W celu dokładnej analizy uwzględnione zostało również ustawienie „Post Hoc", czyli kontrastów (różnic) wartości czasowych dla dróg. Ważne jest, aby pamiętać, że analiza Post Hoc powinna być przeprowadzana tylko wtedy, gdy wynik testu ANOVA jest negatywny dla H0. Stosując analizę Post Hoc, można zidentyfikować najbardziej optymalny wybór kategorii czynnika (w naszym przypadku trasę). Najbardziej niezawodnymi metodami analizy post hoc są korekta Bonferroniego lub metoda HSD Tukeya (Goeman i Solari, 2021).



Rysunek 7.18 Ustawienia ANOVA

Wyniki naszej analizy wskazują, że nasza wartość statystyki *F* wynosi 11,173 (wyższe wartości wskazują na większe różnice między grupami), a *p-value* <0,001, co oznacza, że nasza hipoteza zerowa została odrzucona (patrz rysunek 7.19). Ponieważ istnieje znacząca różnica między trzema trasami (<0,001), test post hoc jest również ważny w naszym przypadku (George & Mallery, 2022). Po przeprowadzeniu testu kontrastów Bonferroniego widzimy, że parę najmniejszych wartości *p-value* odnotowano w przypadku trasy 2 (patrz rysunek 7.19) w porównaniu z obiema pozostałymi trasami. W raporcie możemy stwierdzić, że „wystąpiła znacząca różnica w wyborze trasy dostawy w korelacji z czasem dostawy (1-kierunkowa ANOVA, *F*=11,173, *df* = 47, *p* = <0,001). Trasa 2 miała najdłuższy czas dostawy".



ANOVA

Delivery Time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16,527	2	8,263	11,173	<,001
Within Groups	33,280	45	,740		
Total	49,807	47			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Delivery Time

Bonferroni

(I) Route	(J) Route	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-1,4250*	,3040	<,001	-2,181	-,669
	3	-,5500	,3040	,231	-1,306	,206
2	1	1,4250*	,3040	<,001	,669	2,181
	3	,8750*	,3040	,018	,119	1,631
3	1	,5500	,3040	,231	-,206	1,306
	2	-,8750*	,3040	,018	-1,631	-,119

*. The mean difference is significant at the 0.05 level.

Rysunek 7.19 Wstępne wyniki ANOVA i wyniki testu post-hoc

Kończymy ten rozdział książki ze świadomością, że omówiliśmy w nim niektóre z bardziej powszechnych testów. Istnieją jeszcze inne testy, takie jak ANOVA z powtarzanymi pomiarami, testy wiarygodności i testy wrażliwości, które również można modelować i analizować za pomocą oprogramowania SPSS. Te dodatkowe testy zapewniają szerszy zakres narzędzi do analizy danych i wyciągania znaczących wniosków na temat różnych badań i praktycznych zastosowań.

BIBLIOGRAFIA DO ROZDZIAŁU 7.

- Andersen, A.J. & Dennison, J.R. (2018). An Introduction to Quantile-Quantile Plots for the Experimental Physicist. Journal Articles, 51.
- Garth, A. (2008). Analysing data using SPSS [Dostępne: https://students.shu.ac.uk/lits/it/documents/pdf/analysing_data_using_spss.pdf, dostęp October 26, 2023]
- George, D. & Mallery, P. (2022). IBM SPSS Statistics 27 Stet by Step: A Simple Guide and Reference, 17TH edition, Abingdon: Routledge
- Ghasemi, A. & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. International Journal of Endocrinology and Metabolism, 10(2), pp. 486-489.
- Goeman, J.J. & Solari, A. (2021). Comparing Three Groups. The American Statistician, 76(2), pp. 168-176
- IBM (2021). IBM SPSS Statistics 28 Brief [Dostępne: https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Statistics_Brief_Guide.pdf, dostęp October 26, 2023]

- Janse, R.J., Hoekstra, T., Jager, K.J., Zoccali, C., Tripepi, G., Dekker, F.W. & van Diepen, M. (2021). Conducting correlation analysis: important limitations and pitfalls, 14(11), pp. 2332-2337.
- Kim, T.K. (2015). T test as a parametric statistic. Korean Journal of Anesthesiology, 68(6), pp. 540-546
- Landau, S. & Everitt, B.S. (2004). A Handbook of Statistical Analyses using SPSS, 1st edition, London: Chapman & Hall/CRC
- McClure, P. (2005). Correlation Statistics Review of the Basics and Some Common Pitfalls. Journal of Hand Therapy, 18(3), pp. 378-380
- Mishra, P., Singh, U., Pandey, C.M., Mishra, P. & Pandey, G. (2019). Application of Student's t-test, Analysis of Variance, and Covariance. Annals of Cardiac Anesthesia, 22(4), pp. 407-411
- Turhan, N.S. (2020). Karl Pearson's chi-square tests. Educational Research and Reviews, 15(9), pp. 575-580
- Williamson, M. (b.d.). Data Analysis using SPSS [Dostępne: https://med.und.edu/research/daccota/_files/pdfs/berdc_resource_pdfs/data_analysis_using_spss.pdf, dostęp October 26, 2023]