



DOKUMENTACJA

DAMIAN KWIATKOWSKI

PAWEŁ PAŃCZYK

DAMIAN PUZIO

PRZEMYSŁAW STODOLNY

SPIS TREŚCI

1.	Charakterystyka produktu	3
1.1	Zastosowanie	3
1.2	Cechy wyróżniające	4
1.3	Cel produktu	4
1.4	Opis użytkowników systemu i ich funkcje	4
1.5	Struktura funkcjonalna (hierarchia funkcji)	4
1.5	Model rozwiązań (główni użytkownicy, sposób funkcjonowania, struktura sprzętowa i jej rozmieszczenie w przestrzeni)	5
2.	Harmonogram realizacji projektu.....	5
3.	Diagramy UML.....	6
3.1	Przypadki użycia	6
3.2	Diagramy sekwencji	8
4.	Testy	10
5.	Dziennik projektowy.....	12
6.	Bibliografia	13

1. CHARAKTERYSTYKA PRODUKTU

Program statystyczny RIKI TIKI STATISTIKI jest aplikacją służącą do statystycznej analizy danych, tworzenia wykresów, pracy na bazach danych, wykonywania transformacji danych. Zadaniem programu jest przetworzenie oraz filtrowanie dużej ilości danych używanych w obliczeniach statystycznych. Następnie wykonanie obliczeń statystycznych oraz prezentacja wyników w postaci graficznej.

System RIKI TIKI STATISTIKI oferuje wybrane operacje statystyczne na danych zawartych w programie Microsoft Excel. Jest skierowany dla studentów, którzy potrzebują prostą w obsłudze aplikację liczącą wybrane funkcje statystyczne. Program charakteryzuje się prostym interfejsem użytkownika. W tabeli nr 1 przedstawione są funkcjonalności programu w porównaniu z konkurencyjnymi programami statystycznymi. Są to Microsoft Excel oraz Statistica. Porównania dokonano na podstawie podręcznika dla programu Statistica¹, kompendium aplikacji Microsoft Excel² oraz własnych doświadczeń.

Tabela 1. Porównanie programu RIKI TIKI STATISTIKI z programami konkurencyjnymi

Funkcje	RIKI TIKI STATISTIKI	Statistica	Microsoft Excel
Mediana	●	●	●
Odchylenie standardowe	●	●	●
Kowariancja	●	●	●
Rozkład istotności	●	●	●
Rozkład prawdopodobieństwa	●	●	●
Łatwość obsługi	●	●	●
Cena	●	●	●
Obciążenie systemu	●	●	●

Rozwiązanie RIKI TIKI STATISTIKI dostępne jest w postaci aplikacji desktopowej. Wspierane systemy operacyjne to Windows XP x86 Sp3, Windows XP x64 Sp2, Windows Vista x86, Windows Vista x64, Windows 7 x86, Windows 7 x64, Windows 8 x32, Windows 8 x64, Windows 8.1 x86, Windows 8.1 x64, Windows 10 x86, Windows 10 x64.

1.1 ZASTOSOWANIE

Program służy do obliczeń statystycznych. W programie znajdują się funkcje tj.: rozkład prawdopodobieństwa, mediana, odchylenie standardowe, kowariancja, rozkład istotności.

¹ M. Rabiej, *Statystyka z programem Statistica*, Gliwice: Wydawnictwo Helion 2012

² J. Walkenbach, *Microsoft Excel 2016 Bible*, Indianapolis: Wydawnictwo John Wiley & Sons, Inc. 2015

1.2 CECHY WYRÓŻNIAJĄCE

- Przejrzysty i zrozumiały interfejs użytkownika
- Program nie wymaga instalacji
- Program nie zajmuje wiele miejsca na dysku

1.3 CEL PRODUKTU

Program pobiera dane z arkusza kalkulacyjnego Microsoft Office. Następnie filtruje dane według określonych kryteriów. Są one automatycznie dobierane przez aplikację po wybraniu przez użytkownika żądanej funkcji statystycznej. Program wykonuje obliczenia statystyczne: rozkład prawdopodobieństwa, medianę, odchylenie standardowe, kowariancję, rozkład istotności. Po obliczeniach aplikacja prezentuje wyniki w postaci liczbowej oraz graficznej oraz eksportuje je do pliku „.pdf”.

1.4 OPIS UŻYTKOWNIKÓW SYSTEMU I ICH FUNKCJE

Użytkownik standardowy – obsługa programu w zakresie głównych funkcji tj.: importu danych, obliczeń statystycznych wraz z prezentacją wyników oraz eksportu do pliku „.pdf”.

1.5 STRUKTURA FUNKCJONALNA (HIERARCHIA FUNKCJI)

Na poniższych rysunkach przedstawiono schemat struktury funkcjonowania (hierarchie funkcji) oprogramowania RIKI TIKI STATISTIKI dla poszczególnych funkcji statystycznych.



Rys. 1. Schemat struktury funkcjonowania dla mediany



Rys. 2. Schemat struktury funkcjonowania dla rozkładu istotności



Rys. 3. Schemat struktury funkcjonowania dla odchylenia standardowego



Rys. 4. Schemat struktury funkcjonowania dla kowariancji



Rys. 5. Schemat struktury funkcjonowania dla rozkładu prawdopodobieństwa

1.5 MODEL ROZWIĄZAŃ (GŁÓWNI UŻYTKOWNICY, SPOSÓB FUNKCJONOWANIA, STRUKTURA SPRZĘTOWA I JEJ ROZMIESZCZENIE W PRZESTRZENI)

- System operacyjny Microsoft Windows
- Arkusz kalkulacyjny Microsoft Excel
- Python 2.7
- Pycharm
- Qt

2. HARMONOGRAM REALIZACJI PROJEKTU

- 05-11/03/2017 - Utworzenie wstępnej dokumentacji projektu. Podział obowiązków,
- 12-18/03/2017 - Opracowanie funkcji matematycznych dla projektu,
- 19-25/03/2017 - Opracowanie UML - przypadki użycia, diagramy sekwencji. Oddanie dokumentacji,
- 26/03/2017-01/04/2017 - Stworzenie interfejsu użytkownika aplikacji,
- 02-22/04/2017 - Implementacja funkcji matematycznych w kodzie. Import/eksport,
- 22/04/2017 - 13/05/2017 - Testy aplikacji,
- 14/05/2017 - 10/06/2017 - Oddanie gotowej aplikacji do użytku.

3. DIAGRAMY UML

3.1 PRZYPADKI UŻYCIA

Rysunek nr 6 przedstawia diagram przypadków użycia w systemie.

Import danych z excela - po wybraniu opcji importu danych wybieramy ścieżkę do pliku. Następnie importujemy dane co spowoduje wyświetlenie danych w głównym oknie.

Funkcje matematyczne - wyświetla pięć funkcji matematycznych i pozwala na wybór tej która nas interesuje.

Rozkład prawdopodobieństwa - po wybraniu funkcji wybieramy zakres danych, następnie funkcja wykonuje obliczenia przekazuje wyniki do klasy wyniki.

Rozkład istotność - po wybraniu funkcji wybieramy zakres danych, następnie funkcja wykonuje obliczenia przekazuje wyniki do klasy wyniki. Funkcja wykonuje test t-studenta dla dwóch populacji. Wynik, wraz z poziomem istotności ($\alpha=0.05$) jest porównywany z tabelą rozkładu t-studenta i na tej podstawie jest stwierdzana istotność.

Test jest wykonywany na podstawie wzoru:

$$T = \frac{x1 - x2}{Sx1 - x2} \quad , \text{gdzie:}$$

$x1, x2$ - średnie arytmetyczne populacji

$$Sx1 - x2 = \sqrt{\frac{(n1 - 1) \cdot s1 + (n2 - 1) \cdot s2}{n1 + n2 - 2} \cdot \left(\frac{1}{n1} + \frac{1}{n2}\right)} \quad , \text{gdzie:}$$

$n1, n2$ - liczebność populacji 1 i 2;

$s1, s2$ - wariancja pierwszej i drugiej populacji;

Mediana - po wybraniu funkcji wybieramy zakres danych, następnie funkcja wykonuje obliczenia przekazuje wyniki do klasy wyniki.

Obliczenia są wykonywane zgodnie ze wzorem:

$$Me = \frac{(n + 1)}{2} \quad , \text{gdzie:}$$

n – ilość elementów;

Kowariancja - po wybraniu funkcji wybieramy zakres danych, następnie funkcja wykonuje obliczenia przekazuje wyniki do klasy wyniki.

Program liczy kowariancję na podstawie wzoru:

$$\text{cov}(X, Y) = E(X * Y) - (E(X) * E(Y)) \text{ , gdzie:}$$

$E(X)$, $E(Y)$ - wartości oczekiwane populacji X i Y.

Odchylenie standardowe - po wybraniu funkcji wybieramy zakres danych, następnie funkcja wykonuje obliczenia przekazuje wyniki do klasy wyniki.

W pierwszej kolejności program liczy wartość oczekiwaną populacji na podstawie wzoru:

$$\sum_{i=1}^n x_i \cdot p_i \text{ , gdzie:}$$

x_i - i-ty element populacji;

p_i - prawdopodobieństwo wystąpienia elementu.

Następnie, odchylenie standardowe populacji jest liczone zgodnie ze wzorem:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}} \text{ , gdzie:}$$

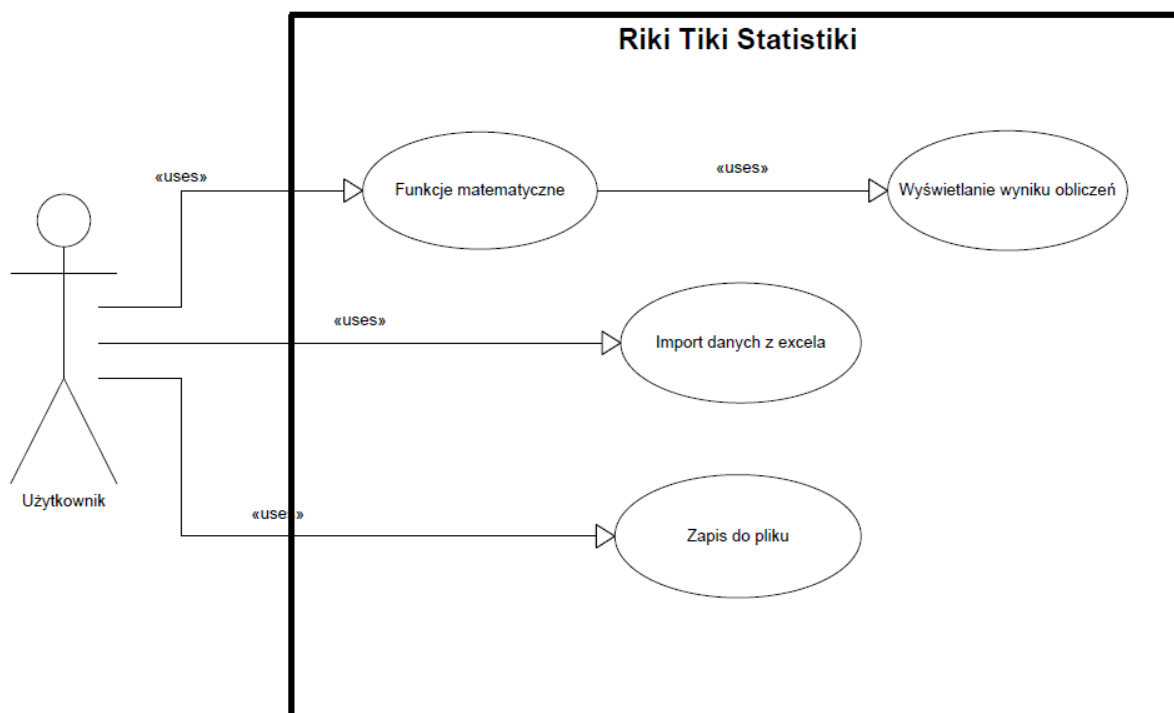
σ - odchylenie standardowe;

n - liczba obserwacji/elementów populacji;

x_i - i-ty element;

μ - wartość oczekiwana.

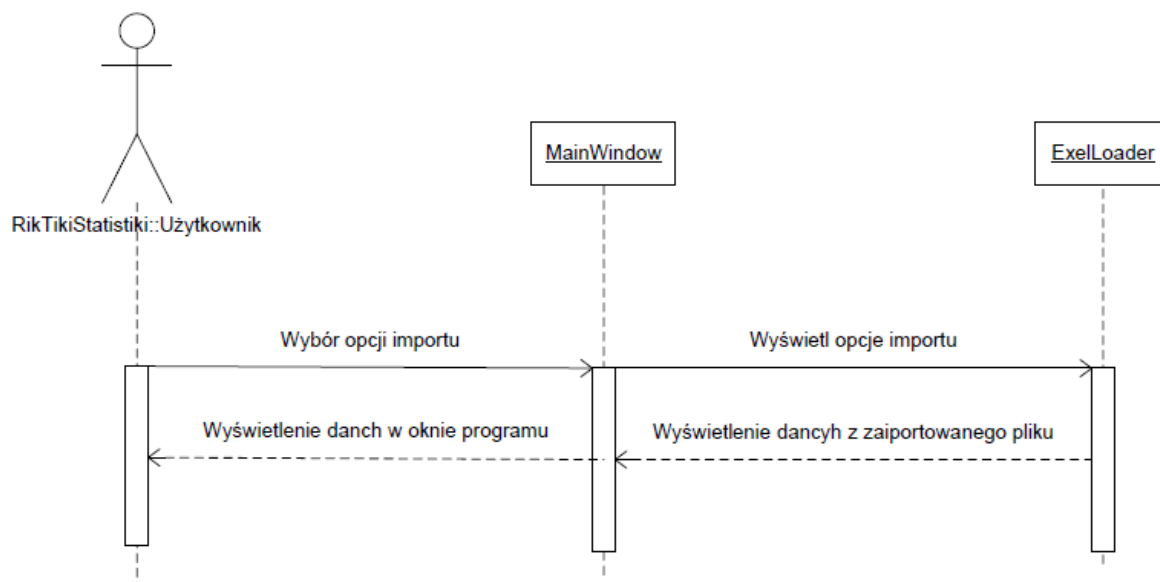
Zapis do pliku - wybranie opcji zapisu spowoduje zapis do pliku .pdf wyników obliczeń.



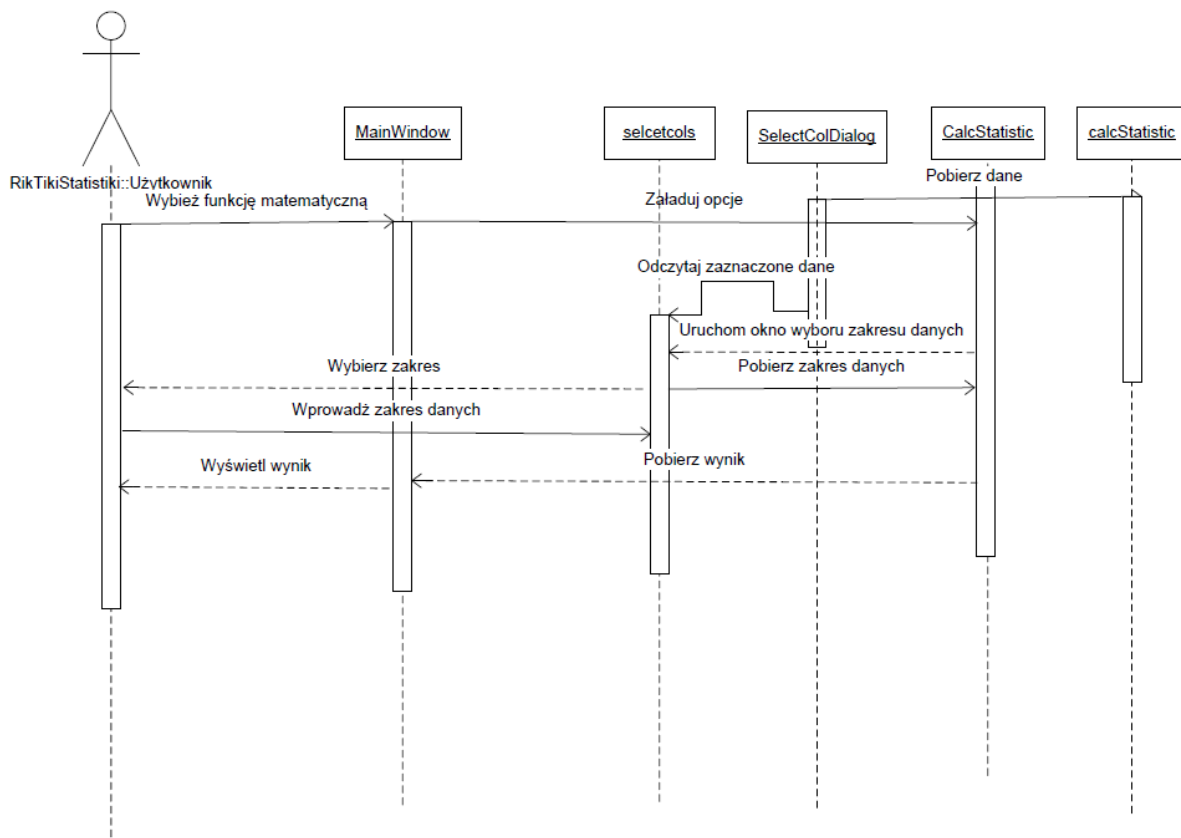
Rys. 6. Przypadki użycia

3.2 DIAGRAMY SEKWENCJI

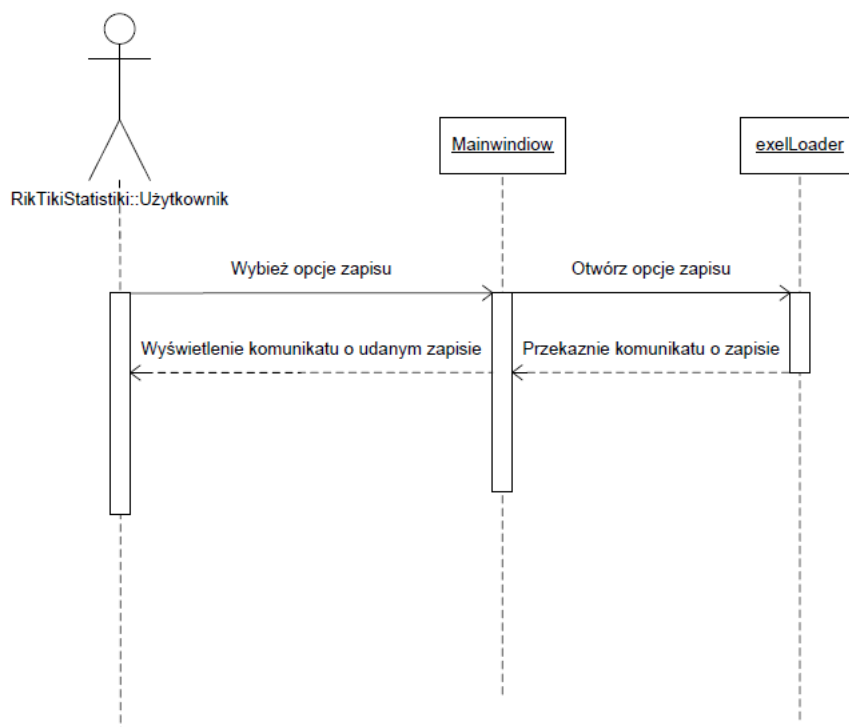
Poniższe diagramy sekwencji przedstawiają przepływ komunikatów pomiędzy danymi obiektami systemu.



Rys. 7. Import danych bezpośrednio z pliku aplikacji Microsoft Excel. Kolejność wysyłania komunikatów pomiędzy obiektami w celu wykonania importu danych z pliku do aplikacji.



Rys. 8. Funkcje matematyczne. Kolejność wysyłania komunikatów pomiędzy obiektami w celu wykonania wybranej funkcji matematycznej.



Rys. 9. Zapis do pliku „.pdf”. Kolejność wysyłania komunikatów pomiędzy obiektami w celu zapisu wyników funkcji matematycznych do plików „.pdf”.

Diagram klas (rys. 10) - diagram przedstawia klasy użyte w programie oraz funkcje jakie znajdują się w konkretnej klasie jak i zależność pomiędzy klasami.

Do diagramu zaliczają się klasy:

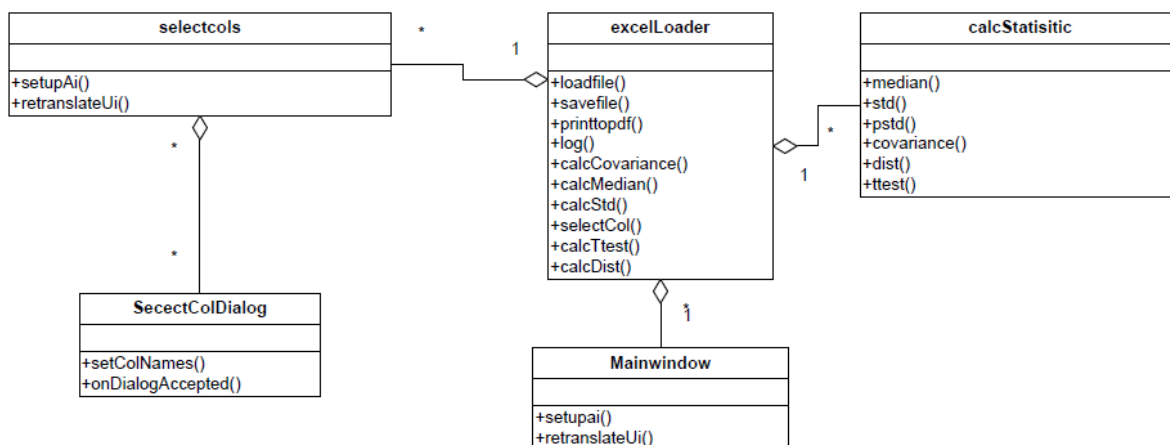
Selectcols - klasa odpowiedzialna za wyświetlenie interfejsu do wyboru zakresu danych.

SecectColDialog - klasa odpowiedzialna za wybór zakresu interesujących nas danych.

Mainwindow - klasa odpowiedzialna za interfejs graficzny programu.

excelLoader - klasa odpowiedzialna za interakcje z interfejsem programu znajdują się w niej funkcje odpowiedzialne za odczytanie pliku zapis oraz przechowuje odczytane dane.

calcStatisitic - klasa odpowiedzialna za funkcje matematyczne. Znajdują się tutaj wszystkie funkcje matematyczne wraz z ich opisem matematycznym.



Rys. 10. Diagram klas

4. TESTY

Testy przypadków użycia:

1. Import danych z pliku „*.xls”;
2. Obliczanie odchylenia standardowego z próby i populacji;
3. Obliczanie mediany z wybranego zakresu;
4. Obliczanie kowariancji wybranych zmiennych;
5. Obliczanie korelacji;
6. Obliczanie rozkładu istotności (test t-studenta);
7. Eksport danych do „*.pdf”;
8. Testy eksploracyjne interfejsu użytkownika;

Przypadek 1: Import danych.

Scenariusz: Próba importu danych z różnych wersji pliku .xls (2003, 2007, 2016). Próba importu danych z innych plików np. „*.pdf”.

Wynik: Program importuje dane z różnych wersji pliku „*.xls”.
Program nie wczytuje danych z innych plików.

Przypadek 2:	Obliczanie odchylenia standardowego z próby i populacji.
Scenariusz:	Import danych z pliku „*.xls” → Wybranie zakresu danych → Próba obliczenia odchylenia standardowego → Porównanie wyników z programem konkurencyjnym (Microsoft Excel).
Wynik:	Program importuje dane z pliku „*.xls”. Po wyborze zakresu i opcji “Odchylenie standardowe”, program pyta czy ma liczyć z próby czy z populacji. Uzyskane wyniki są prezentowane w oknie dialogowym i są zgodne z wynikami uzyskanymi w Microsoft Excel, do dziewiątego miejsca po przecinku.
Przypadek 3:	Obliczanie mediany z wybranego zakresu.
Scenariusz:	Import danych z pliku „*.xls” → Wybranie zakresu danych → Próba obliczenia mediany z wybranego zakresu → Porównanie wyników z programem konkurencyjnym (Microsoft Excel).
Wynik:	Program importuje dane z wybranego pliku „*.xls”. Po wyborze zakresu i opcji “Mediana”, program liczy medianę z wybranego zakresu i prezentuje wynik w oknie dialogowym. Wyniki są zgodne z wynikami uzyskanymi w Microsoft Excel.
Przypadek 4:	Obliczanie Kowariancji z wybranego zakresu.
Scenariusz:	Import danych z pliku „*.xls” → Wybranie zakresu danych → Próba obliczenia kowariancji → Porównanie wyników z programem konkurencyjnym (Microsoft Excel).
Wynik:	Program importuje dane z wybranego pliku .xls. Po kliknięciu w opcję “Kowariancja”, użytkownik proszony jest o wybranie zmiennych. Następnie program wylicza średnią arytmetyczną dla pierwszej zmiennej, średnią arytmetyczną dla drugiej zmiennej i kowariancję między obiema zmiennymi. Wyniki są zgodne z wynikami uzyskanymi w Microsoft Excel.
Przypadek 5:	Obliczanie korelacji.
Scenariusz:	Import danych z wybranego pliku „*.xls” → Wybranie zakresu danych → Próba obliczenia korelacji → Porównanie wyników z programem konkurencji (Microsoft Excel).
Wynik:	~~~~~!!!Na razie nie działa!!!~~~~~
Przypadek 6:	Obliczanie rozkładu istotności. Wykonanie testu t-studenta.
Scenariusz:	Import danych z wybranego pliku „*.xls” → Wybranie zakresu danych → Próba wykonania testu → Porównanie wyników z programem konkurencji (Microsoft Excel).

Wynik: Program prawidłowo wykonuje import danych z pliku .xls. Program wykonuje test t studenta dla dwóch wybranych populacji. Wyniki są zgodne z wynikami uzyskanymi w Excelu."

Przypadek 7: Export wyników do pliku „*.pdf”.

Scenariusz: Import danych z wybranego pliku „*.xls” → Wybranie zakresu danych → Próba obliczenia odchylenia standardowego, mediany, kowariancji, korelacji i testu t-studenta → Próba eksportu uzyskanych wyników do pliku „*.pdf”.

Wynik: ~~~~~!!!Na razie nie działa!!!~~~~~

Przypadek 8: Testy eksploracyjne interfejsu użytkownika.

5. DZIENNIK PROJEKTOWY

04 marzec 2017:

Wybór tematu projektu. Wstępny podział obowiązków:
Damian Puzio - programowanie;
Damian Kwiatkowski - dokumentacja;
Paweł Pańczyk - UML;
Przemysław Stodolny - opis matematyczny, testy;

06-12 marzec 2017:

Założone cele: Wstępna dokumentacja projektu, charakterystyka produktu.
Oficjalny podział obowiązków.
Wykonano: Wstępna dokumentacja projektu. Określenie funkcji programu

13-25 marzec 2017:

Założone cele: Wykonanie przypadków użycia, oraz opis matematyczny;
Wykonano: Określono przypadki użycia. Opis matematyczny, wraz z wzorami niezbędnymi do wykonania funkcji programu.

26 marzec - 01 kwiecień 2017:

Założone cele: Wstępny projekt interfejsu użytkownika. Diagramy sekwencji.
Struktura funkcjonalna.
Wykonano: Określono interfejs użytkownika, oraz narzędzia potrzebne do jego wykonania. Wykonano diagramy sekwencji, oraz określono struktury funkcjonalne.

03-30 kwiecień 2017:

Założone cele: Napisanie programu w pierwszej wersji. Napisanie pierwszych funkcji realizowanych przez program. Diagram klas. Testy programu.
Wykonano: Interfejs użytkownika, oraz pierwsze funkcje programu: Mediana, odchylenie standardowe, kowariancja, import danych. Wykonane zostały testy powyższych funkcji.

08-13 maj 2017:

Założone cele: Uzupełnienie diagramów uml. Poprawa dokumentacji. Dopisanie kolejnych funkcji programu, tj. Eksport do pdf. Testy regresyjne, oraz testy nowych Funkcji.

Wykonano: Dopisano eksport do pdf. Poprawiono dokumentację. Uzupełniono diagramy uml. Testy regresyjne zakończone pomyślnie. Eksport do pdf nie działa jak należy.

15-20 maj 2017:

Założone cele: Dopisanie brakujących funkcji. Poprawa funkcji eksportu. Uzupełnienie uml o nowe klasy. Poprawienie dokumentacji. Testy.

Wykonano: Uml uzupełniony. Poprawiono dokumentację.

20 maj - 10 czerwca 2017:

Założone cele: Dopisanie brakujących funkcji programu, tj. Korelacji, testu t-studenta i rozkładu prawdopodobieństwa. Stworzenie prezentacji produktu.

Wykonano: Dopisano funkcję wykonującą test t-studenta. Prezentacja została ukończona w ok. 90%.

6. BIBLIOGRAFIA

Literatura przedmiotu:

1. M. Rabiej. 2012. *Statystyka z programem Statistica*. Gliwice: Wydawnictwo Helion. ISBN 978-83-246-4110-9.
2. J. Walkenbach. 2015. *Microsoft Excel 2016 Bible*. Przeł. Piotr Cieślak. Indianapolis: Wydawnictwo John Wiley & Sons, Inc. 2015. ISBN 978-83-283-2439-8.
3. Strona internetowa opisująca moduł „XLRD” języka Python – służący do ekstrakcji informacji z arkusza kalkulacyjnego Microsoft Excel www.lexicon.net/sjmachin/xlrd.html (dostęp 20.05.2017).
4. Strona internetowa opisująca bibliotekę „PySide.QtGui” platformy programistycznej „Qt” – służącej do tworzenia interfejsu graficznego programu <https://srinikom.github.io/pyside-docs/PySide/QtGui/> (dostęp 20.05.2017).
5. Strona internetowa z funkcjami matematycznymi (statystycznymi) dla programu Python <https://docs.python.org/3/library/statistics.html> (dostęp 20.05.2017).
6. Strona internetowa na temat platformy programistycznej „Qt” <http://doc.qt.io/>.