

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ   
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

ФАКУЛЬТЕТ БІОМЕДИЧНОЇ ІНЖЕНЕРІЇ

КАФЕДРА БІОМЕДИЧНОЇ КІБЕРНЕТИКИ

**Практична робота №4**

## з дисципліни «Вступ до інтелектуального аналізу даних»

# на тему: «Дискретний аналіз»

**Виконав:**

студент гр. БС-03

Затуловський Г.А.

**Перевірив:**

доц. Павлов В.А.

Зараховано від \_\_\_.\_\_\_.\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис викладача)

Київ-2022

Варіант 7

***Частина 1.*** Розробка лінійних класифікаторів прогнозу ускладнень після операції на серці. Відпрацювання застосування параметрів алгоритму для визначення моделей з кращими показниками моделі на навчальній вибірці даних

***Завдання***

*Для одержання класифікаторів (варіація параметрів) керуватися вказівками викладача на занятті та методичним посібником – додаток 4.*

Для вибраного критерію якості класифікації (один з п’яти вибраних вами варіантів) виконати:

1.Для одержаних даних (додаток 4) одержати класифікатор для 4 (3) класів ускладнень операції на серці для значень F за замовчуванням.

2.Знижувати F*вкл* до тих пір поки будуть покращуватися результати класифікації

1. *Виділити окремо навчальну та тестові вібірки однорідно по дисперсії у класах у співвідношенні 70/30-75/25-80/20 (один з варіантів)*
2. *На навчальній послідовності одержувати класифікатори, порівнюючи якість класифікації (чутливість у класах) на тестовій вибірці*
3. Навести найкращі моделі класифікації за найкращим значенням чутливості у класах на тесті.

*.* ***Частина 2.*** Розробка нелінійних класифікаторів прогнозу ускладнень після операції на серці. Аналіз моделей різного рівня складності.

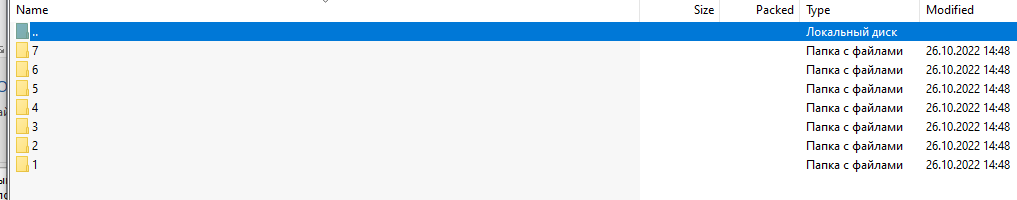
***Завдання.***

Розширити поле узагальнених змінних за рахунок функціонального перетворення первинних змінних: додати до первинних «х» - обернені 1/х, та ***прямі та обернені*** множення ***первинних х*** (та хто захоче кореня 2,3) ***другого порядку****. Для множини узагальнених змінних провести дослідження як в частині 1. Лабораторної роботи*

**Виконання:**

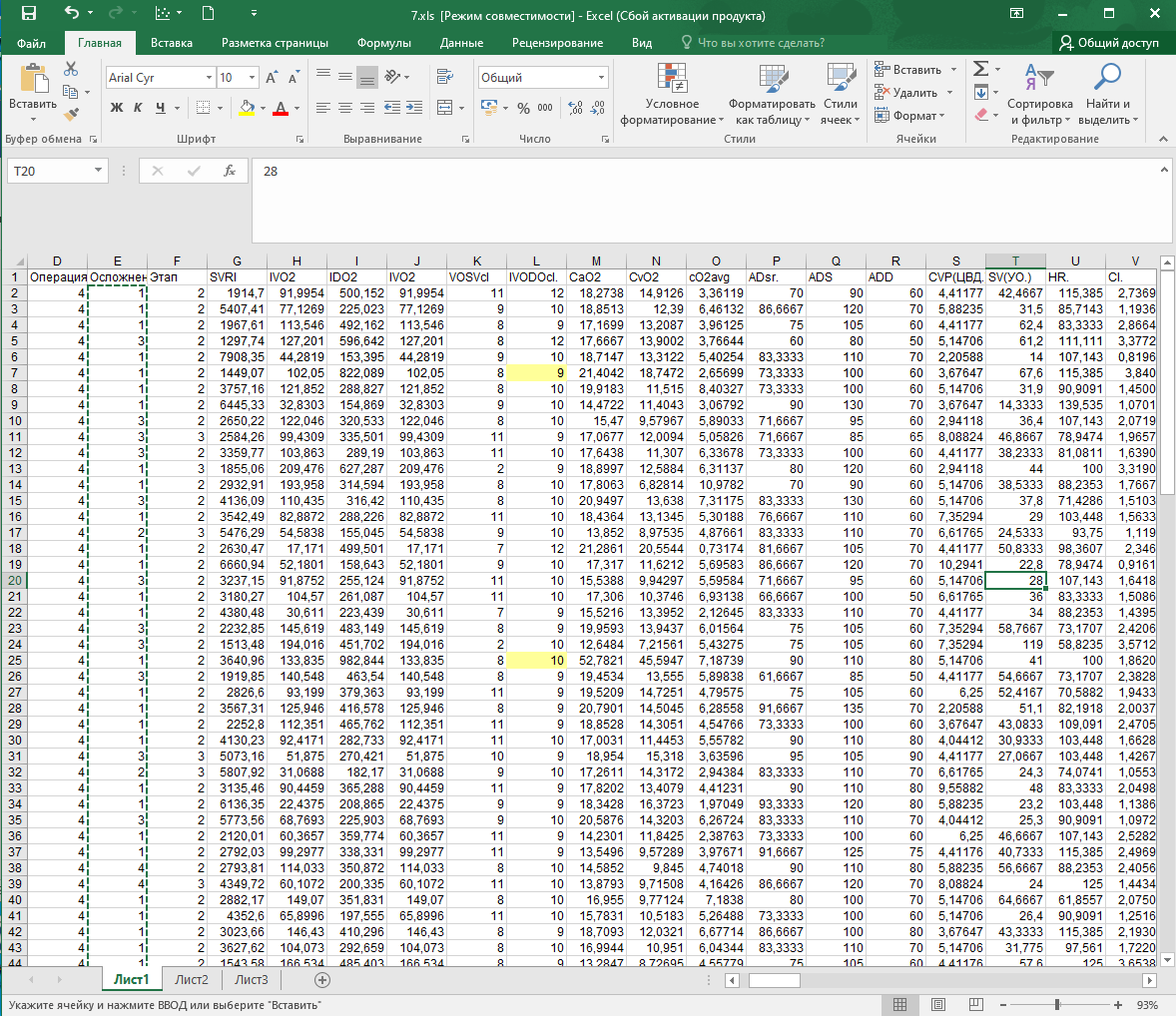
Обираємо варіант 7 у папці з варіантами та редагуємо таблицю у IBM SPSS:

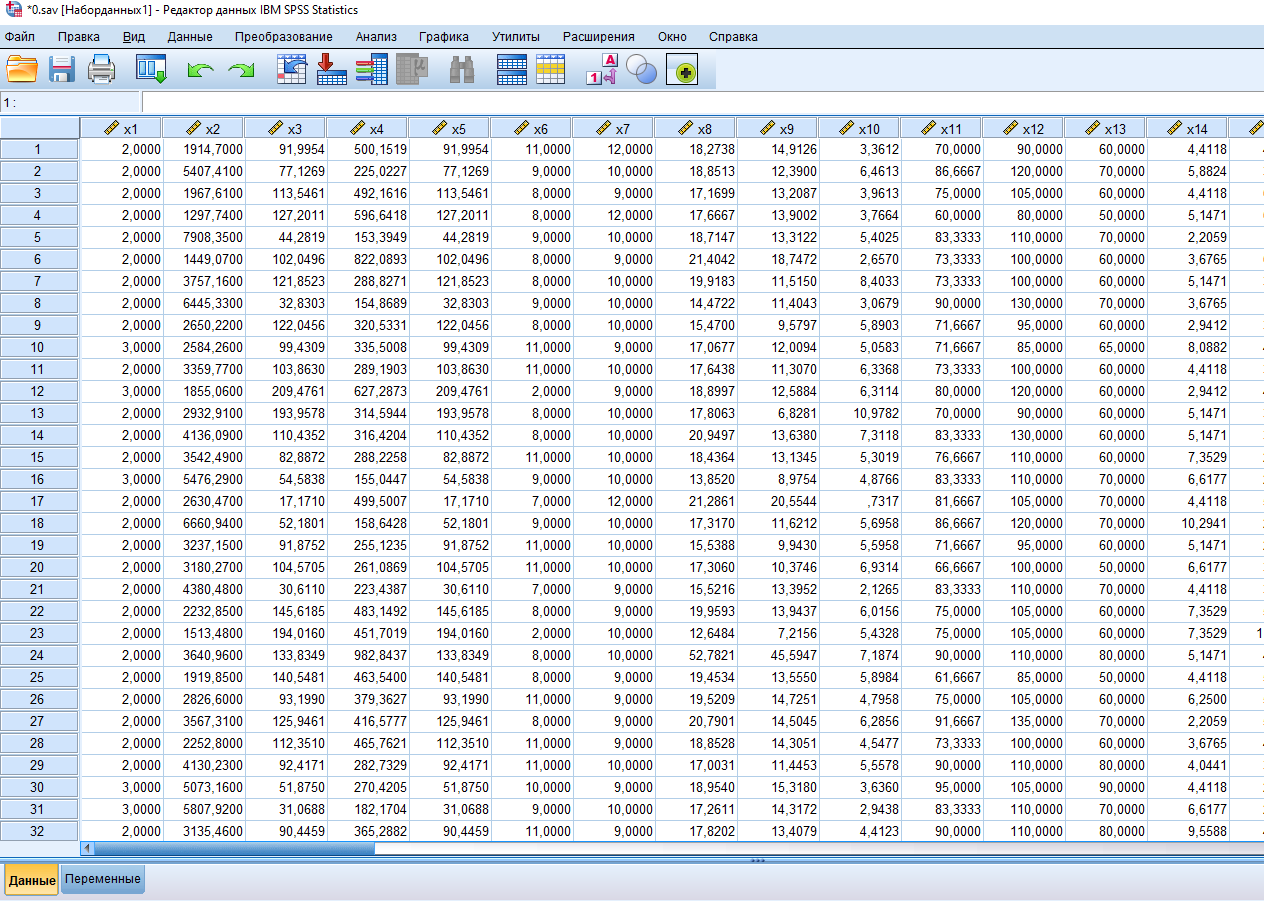
Папка під назвою “2 ПОЧТА – ВАРИАНТЫ”



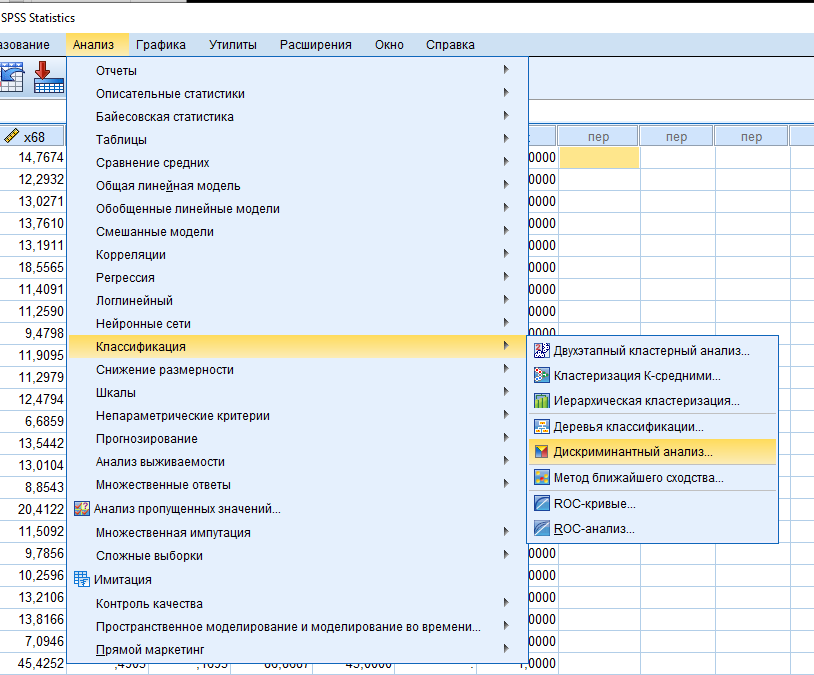
Переносимо дані з таблиці 7.xls у таблицю 0.sav

Cкорюємо всі дані почитаючи зі стовпчику Этап та до HbE02 , стовпчик під назвою Ускладнення будемо використовувати як коефіціент по котрому будемо робити вирівнювання

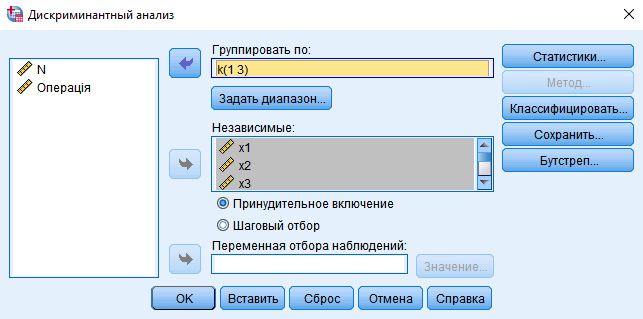
**

**

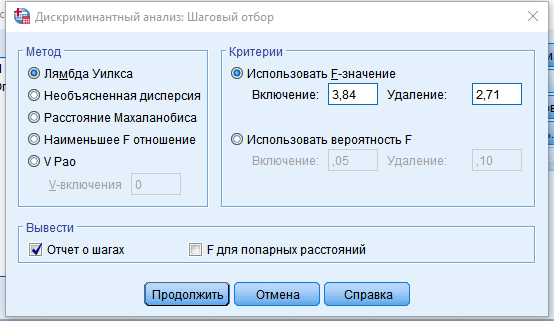
Починаємо аналізувати

**

Угрупуємо наші дані по ускадності;

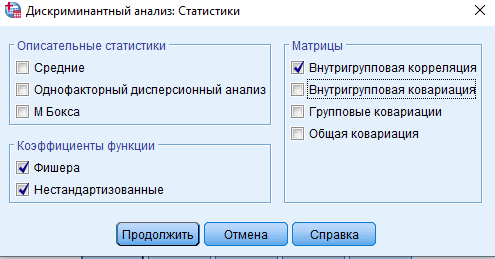
**

Залишаємо незміною шаговий отбір

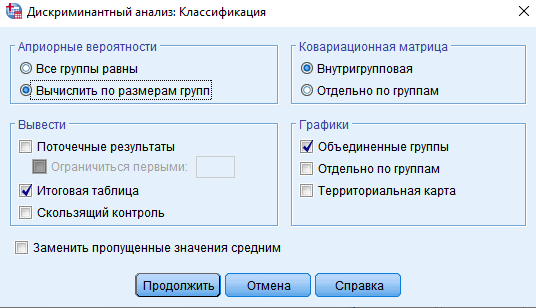


Обираємо коефіціент функції Фішера та Нестандартизовані

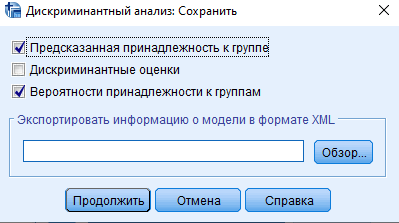
У Матриці обираємо “Внутригрупповая корреляция “



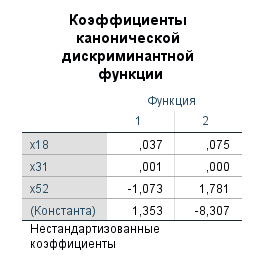
У класифікації ставимо “Вычислить по размерам групп” “Ковариационная матрица : Внутригрупповая “,“Графики Обьединенные группы“ та виводимо “Итоговою таблицю”



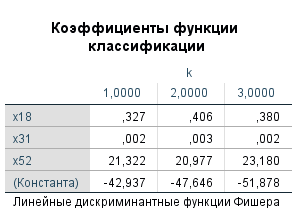
У вкладці “Сохранить“ обираємо “Предсказанная принадлежность к группе“ “Вероятности принадлежности к группам“



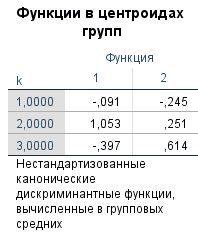
Отримаємо результати

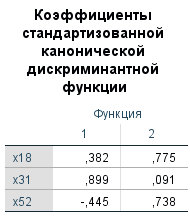
**

У таблиці Коефіцієнти канонічної дискримінантної функції показують коефіцієнти, що використовують для побудови графіку функції

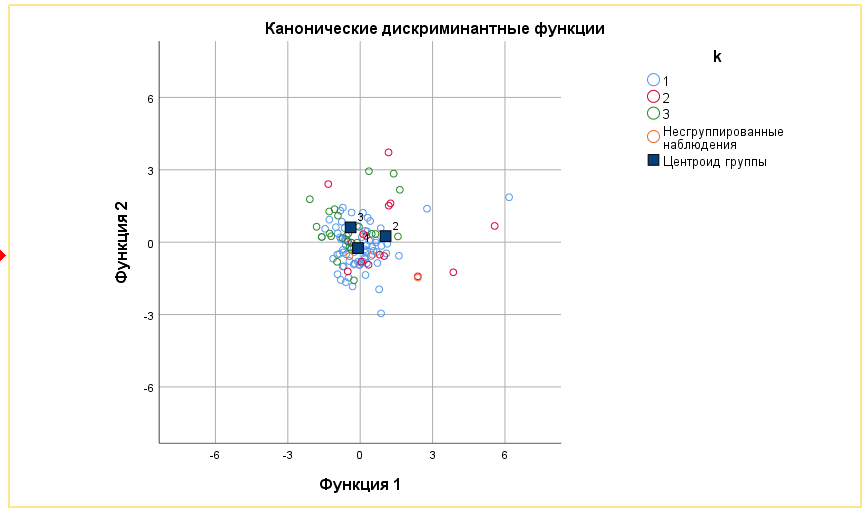
**

У таблиці Коефіцієнти функції класифікації показують коефіцієнти, що використовують для побудови графіку функції Фішера

**

**

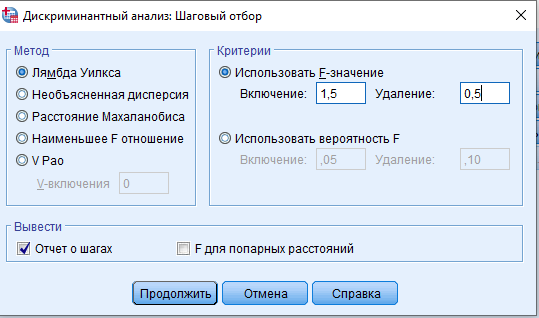
У таблиці “Коефіцієнти стандартизованої канонічної дискримінантної функції” можна побачити що змінні х18,х31,х52 можна побачити ціні зміни

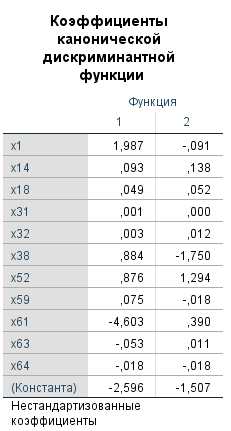
**

**

Загальна точність 69.6% нас це не задовольняє, тому ми опустимо наші порогі границі поки наша точність не буде досягати 80-100%, також класифікування другого та третього класу відбувається некоректно

Спробуємо зменшити поріг значення критарій (Включение – 1.5 ; Удаление – 0.5)



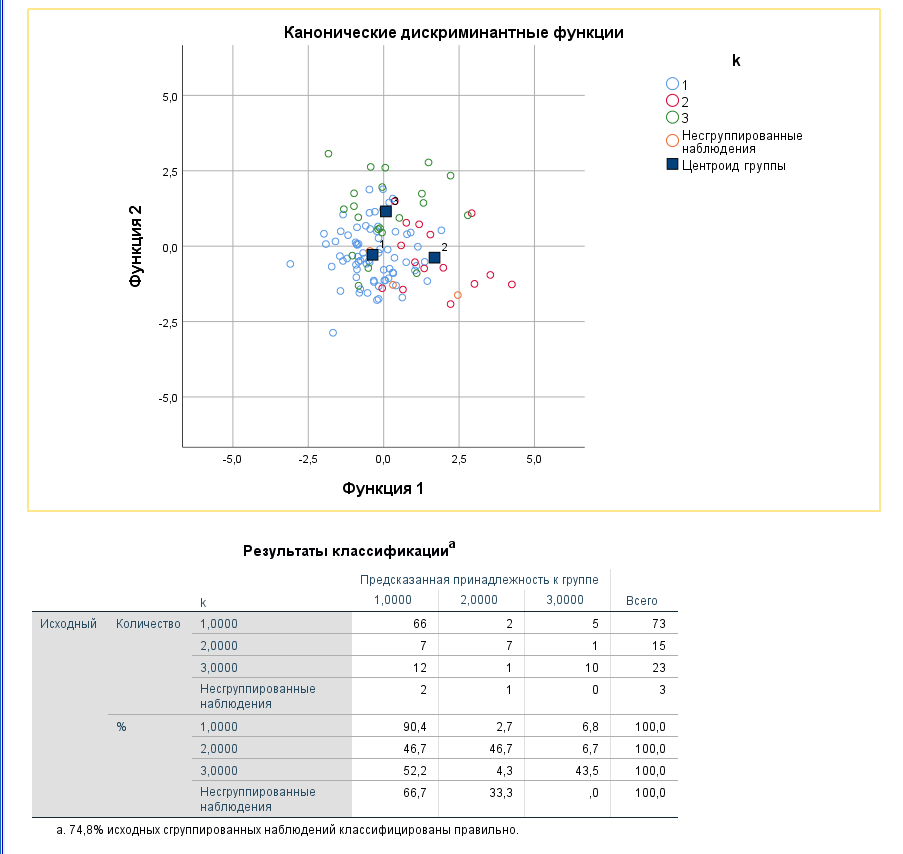




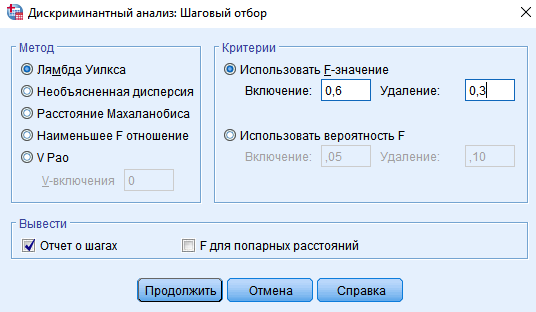
У таблиці “Коефіцієнти стандартизованої канонічної дискримінантної функції” можна побачити що збільшилось кісльуість змінних x1,x14,х18,х31,x32,x38,х52,x59.x61.x63,x64 можна побачити ціні зміни



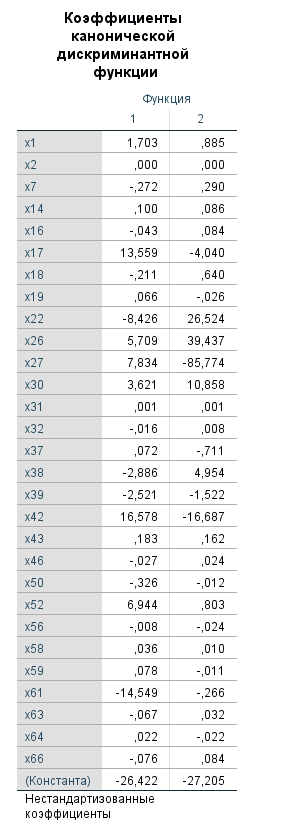
У таблиці Коефіцієнти функції класифікації показують коефіцієнти, що використовують для побудови графіку функції Фішера



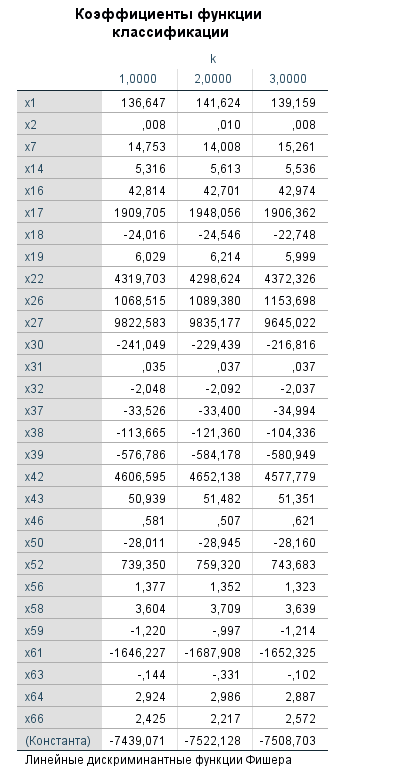
Після повторного дискретного аналізу загальна точність збільшилася з 69.4 % до 74.8 %, але цей результат нас не влаштовує (74.8<80), відповідно у другому та третему класі класифікація відбувається у більше полови випадків класифікує не коректно. Тому продовжимо зменшувати поріг границі (Включение – 0.6 ; Удаление – 0.3)

****

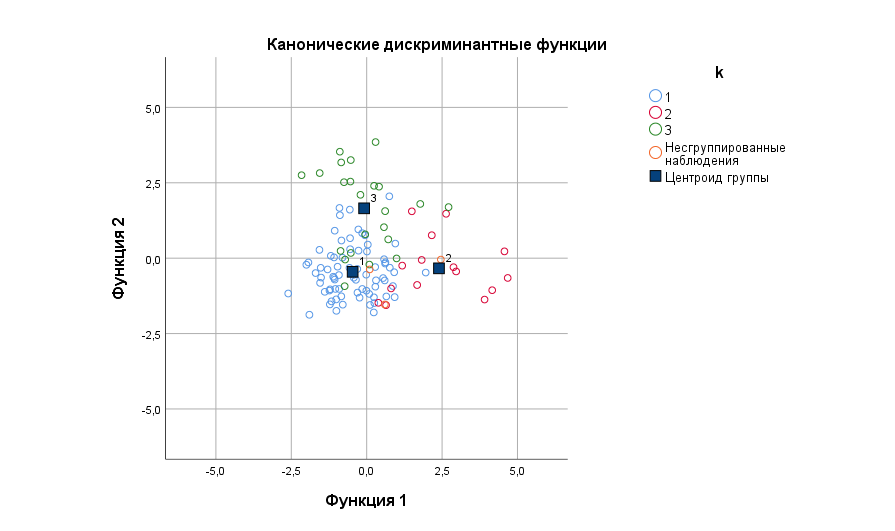
****

****

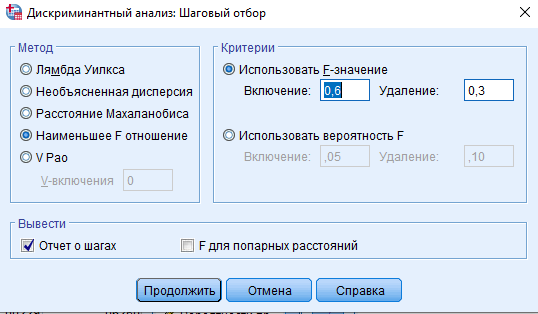
У таблиці “Коефіцієнти стандартизованої канонічної дискримінантної функції” можна побачити що кількість зміних збільшилася з 11 до 30 зміних які є з найбільшими зміною

****

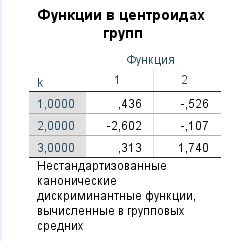
У таблиці Коефіцієнти функції класифікації показують коефіцієнти, що використовують для побудови графіку функції Фішера

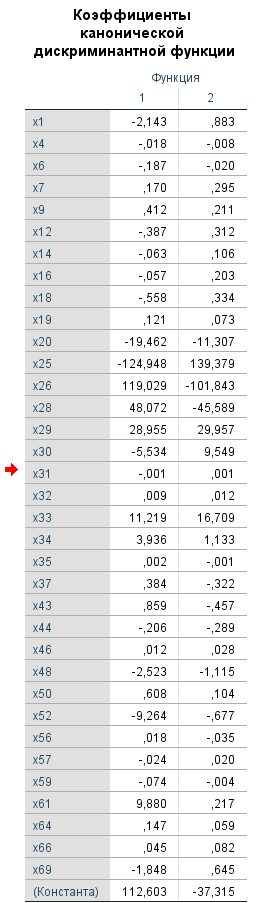
****

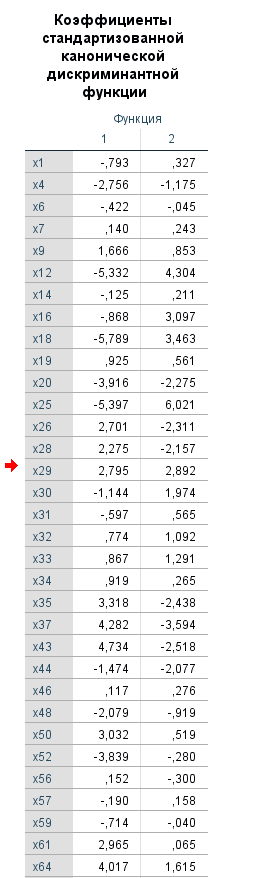
****

Після зниження порогу границі, отримали результат в 82.9%. Цей результат нас влаштовує. Класифікування другого та третього класу вже є більше 60% відсотків, такий результат нас теж влаштовує. Спробуємо змінити метод з Лямбда Уипкса на Наименьшее F отношение

Отримаємо результати

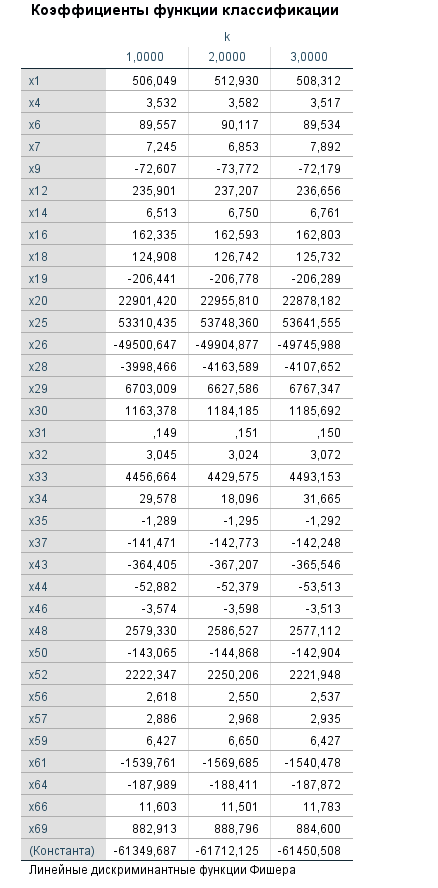
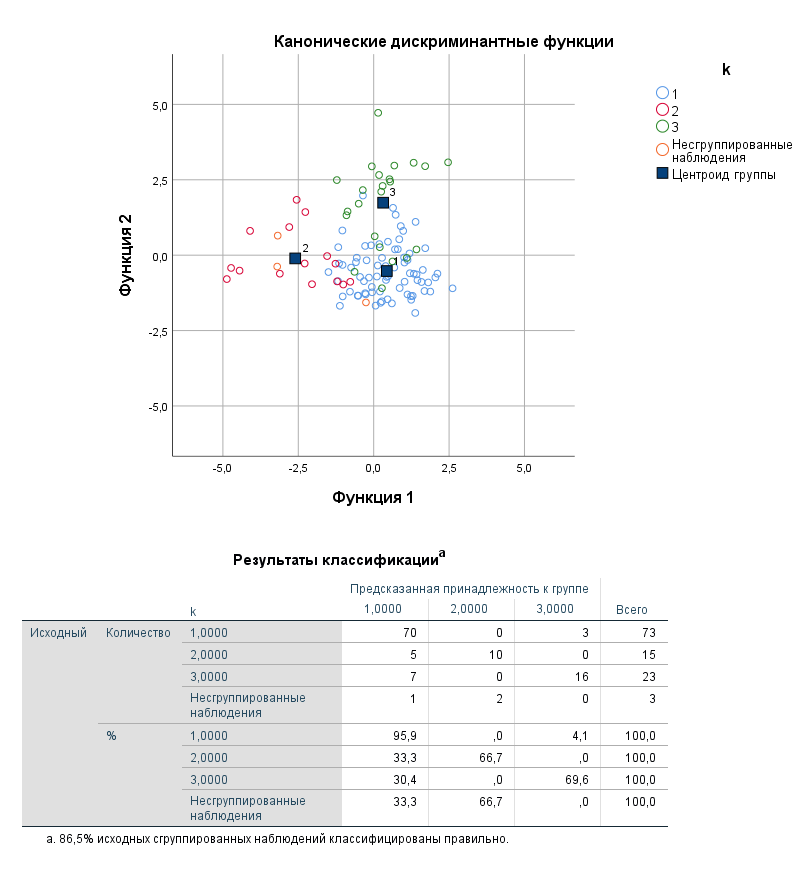








У таблиці “Коефіцієнти стандартизованої канонічної дискримінантної функції” можна побачити що кількість зміних збільшилася з 30 до 35 які є з найбільшими зміною

Після змінення методу “Лямбда Уипкса“ на “Наименьшее F отношение“ , результат збільшився з 82.9% на 86.5% . Класифікування класів теж збільшився (1 – з 93.2% до 95.9%; 3 – 60.9% до 69.6%)

**Висновок:** Під час роботи ми розробили лінійні класифікатори прогнозу ускладнень після операції на серці. Відпрацювали застосування параметрів алгоритму для визначення моделей з кращими показниками моделі на навчальній вибірці даних та отримали результати класифікації, різних класів паралельно знижуючи поріг