**РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ**

**Факультет физико-математических и естественных наук**

**Кафедра прикладной информатики и теории вероятностей**

Практикум по математической статистике

Лабораторная работа №5

**Тема: «Дискриминантный анализ»**

Вариант 10

Выполнил

Студент: Феоктистов Владислав

Группа: НПМбд-01-19б

№ c/б: 1032192939

Преподаватель: Матюшенко Сергей Иванович

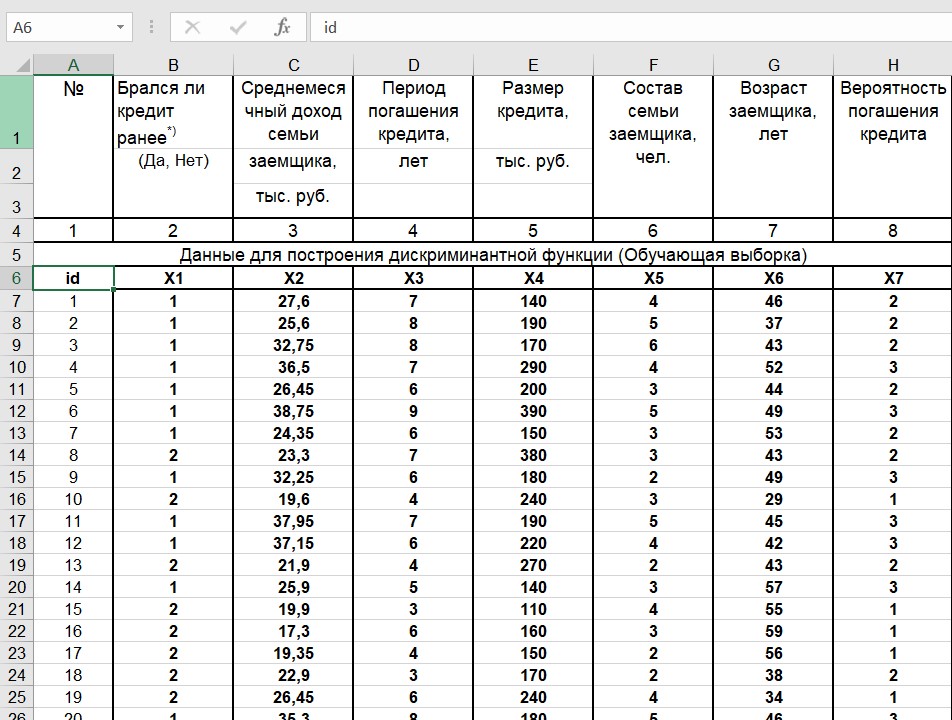
**МОСКВА**

2022 г.

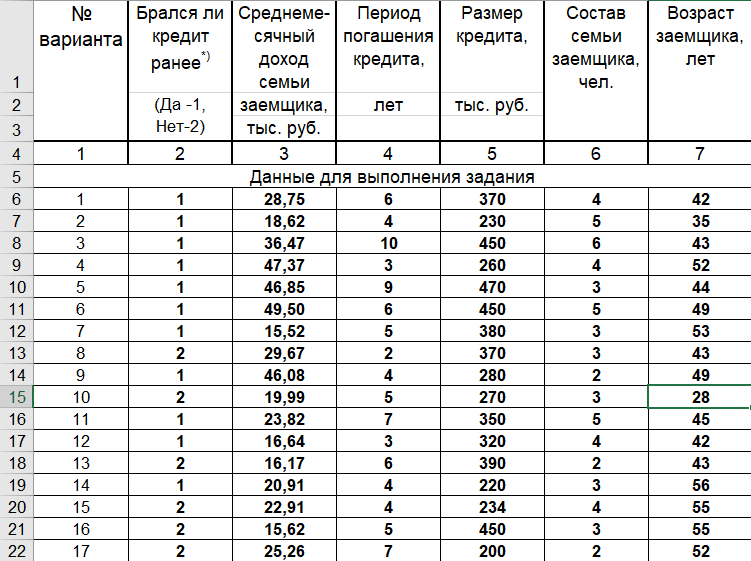
**Цель работы:** приобрести практические навыки применения дискриминантного анализа для решения конкретных задач с использованием статистического пакета SPSS.

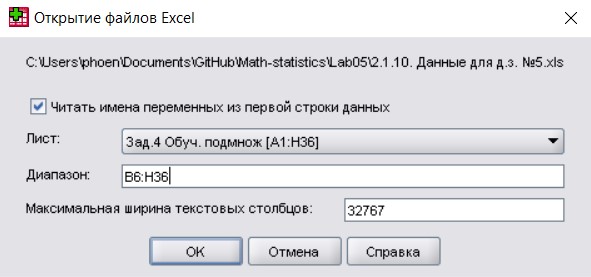
**Ход работы:**

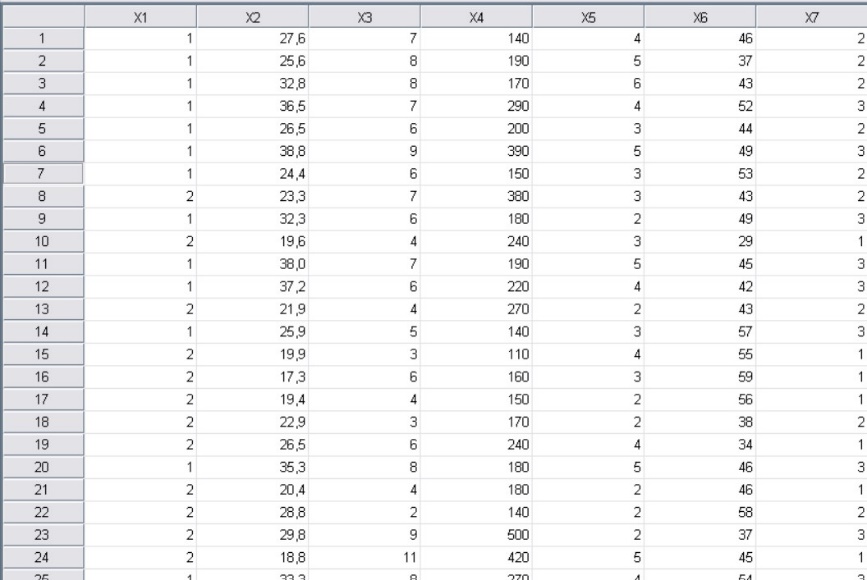
1. Изучил теоретические основы дискриминантного анализа, используя материалы учебного пособия.
2. Разобрал пример использования SPSS для реализации дискриминантного анализа.
3. Запустим программу SPSS и введем исходные данные по обучающей выборке, а также добавим данные, вероятности которых нужно классифицировать.

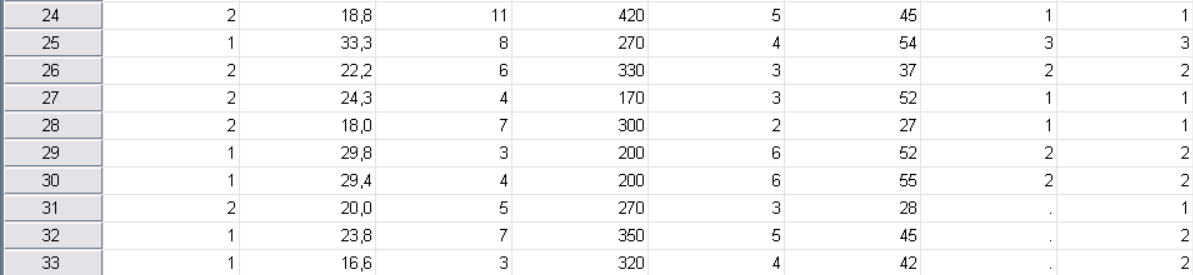


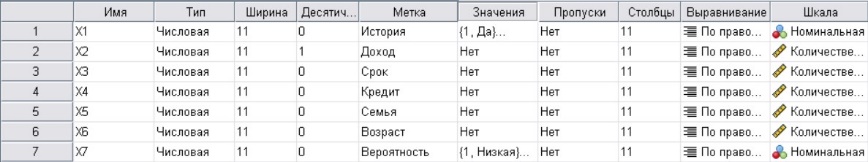
Отсюда брались только 10, 11, 12 варианты



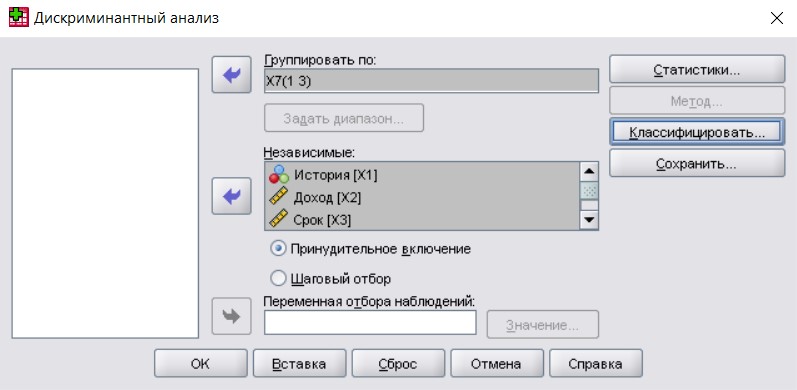


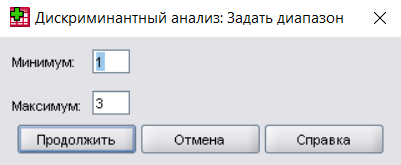


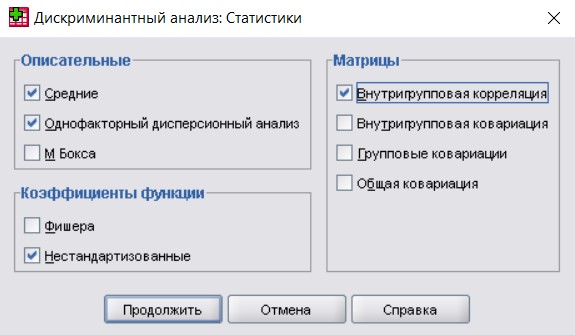


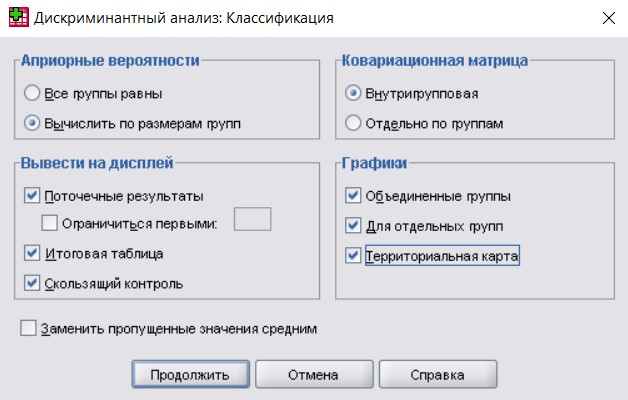


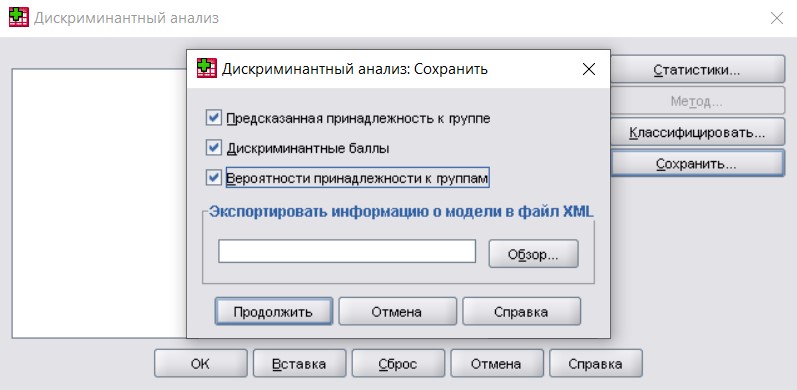
Далее проделаем дискриминантный анализ.





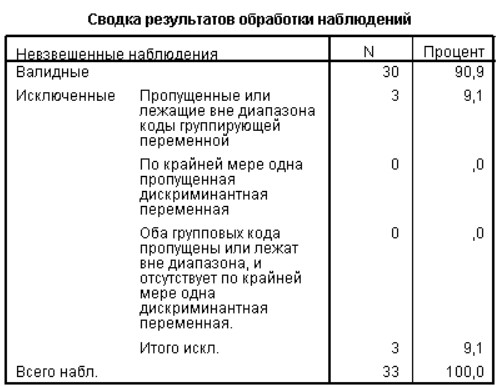




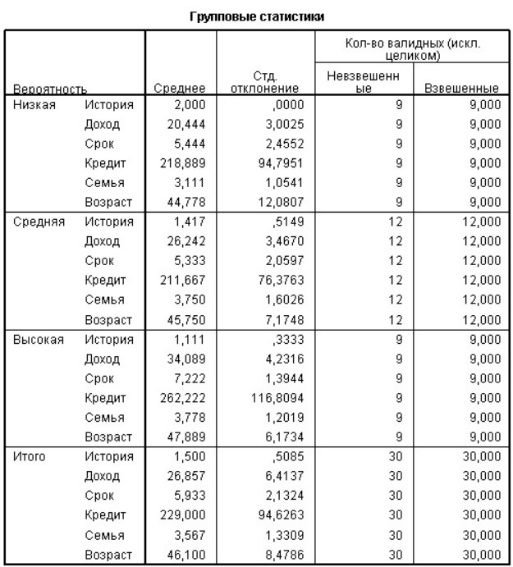


После нажатия на кнопку «ОК» получаем следующий вывод:

Общее количество наблюдений составило 33 единицы, в том числе 30 действительных и 3 подлежащих дискриминации:



По средним значениям групп заметно, что в 3й группе самый большой размер кредита (262,222), самый маленький – во 2ой (211,667); также в 3ей группе самый большой размер дохода (34,089), самый маленький – в 1й (20,444); и кроме того, 3я группа чаще всего брала кредит ранее (1,111 наиболее близкое к 1).



Из данных таблицы «Критерий равенства групповых средних» следует, что переменные «Кредит» («Размер кредита»), «Семья» («Состав семьи заемщика»), «Возраст» («Возраст заемщика») незначимо различаются по группам, поскольку для них уровень значимости Знч. > 0.05, поэтому классификацию заемщиков целесообразно проводить по первым двум переменным: «История» («Брался ли кредит») и «Доход» («Среднемесячный доход семье заемщика»).



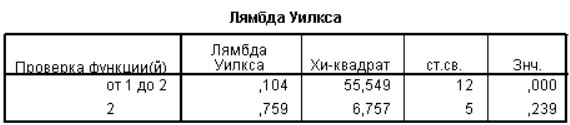
Анализ матрицы коэффициентов в таблице «Объединенные внутригрупповые матрицы» свидетельствует об отсутствии мультиколлинеарности, поэтому коэффициенты корреляции малы.

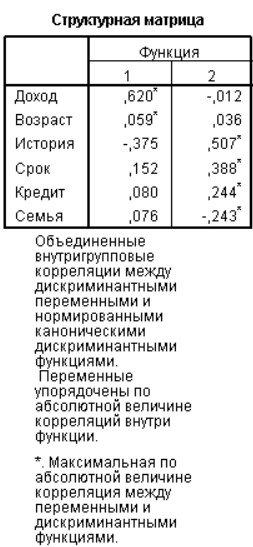


Данные таблицы «Собственные значения» показывают, что первая функция учитывает 95,2% дисперсии, а корреляция между исходными данными и данными, полученными по модели, высокая и составляет 0,929. Для второй функции эти значения намного меньше.



Оценка значимости дискриминантных функций проводится по коэффициенту Уилкса (). Из данных таблицы «Лямбда Уилкса» видно, что для первой функции значимость Знч. < 0,001, следовательно, она позволяет значимо и надежно дискриминировать наблюдения. В то же время значимость второй функции составляет лишь 0,239. Поэтому в дальнейшем для классификации целесообразно использовать только первую дискриминантную функцию.



Формально по данным таблицы «Коэффициенты канонической дискриминантной функции» можно построить две дискриминантные функции:

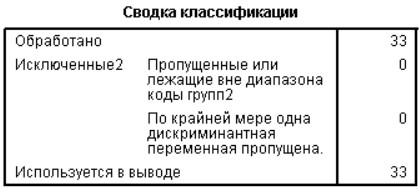
;



Однако поскольку значимость второй функции более 0,001, ее для дискриминации использовать нецелесообразно.

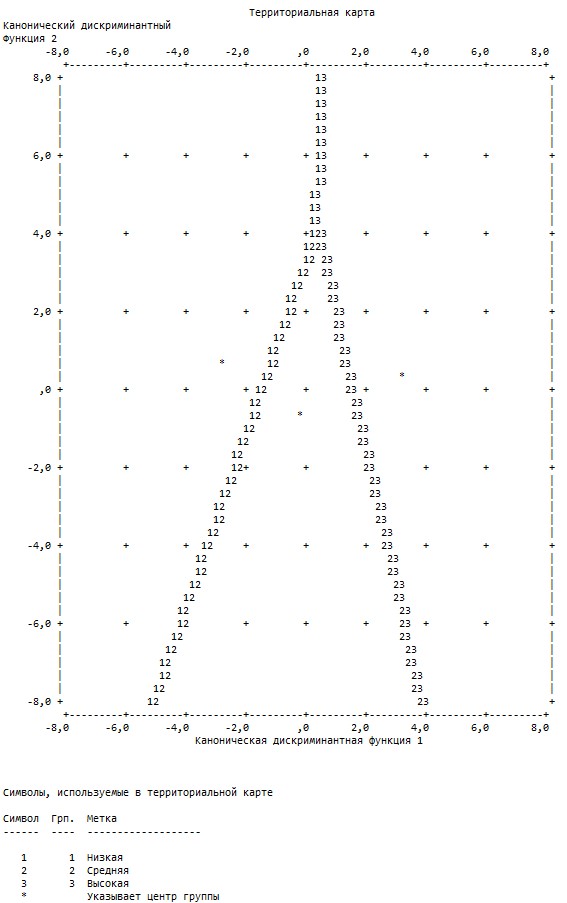
Координаты центроидов по группам приведены в таблице «Функции в центроидах групп». Они используются для нанесения центроидов на карту восприятия.



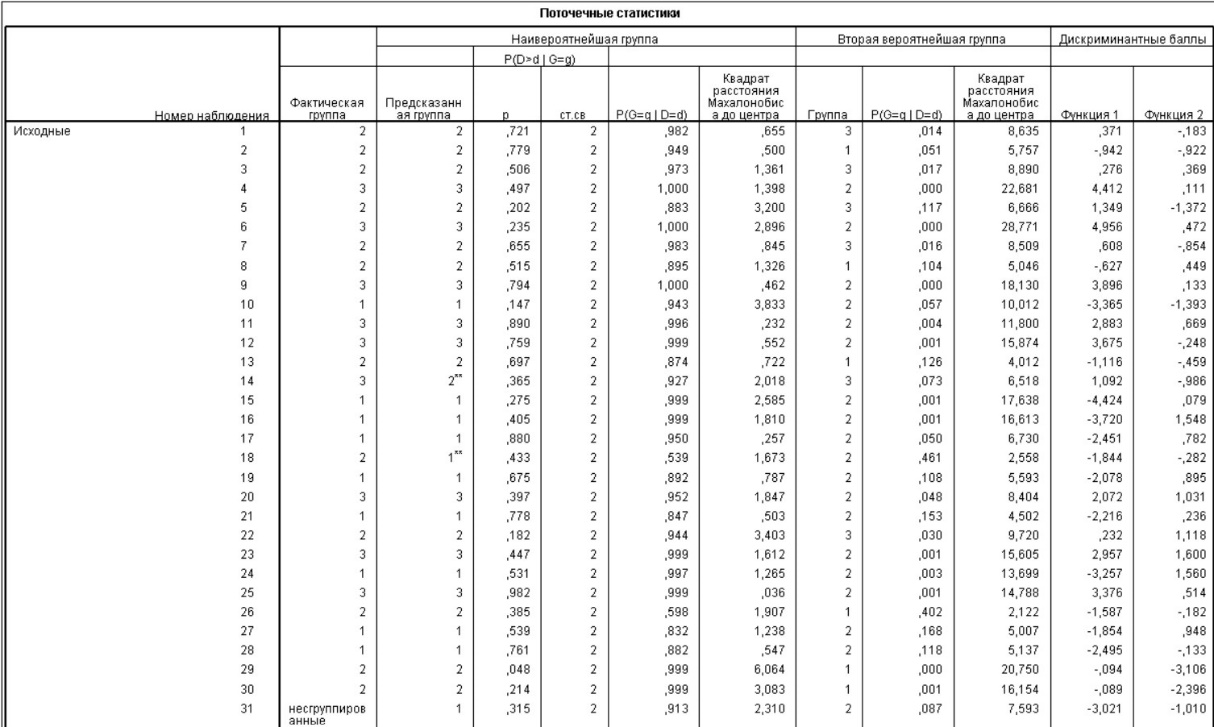
 

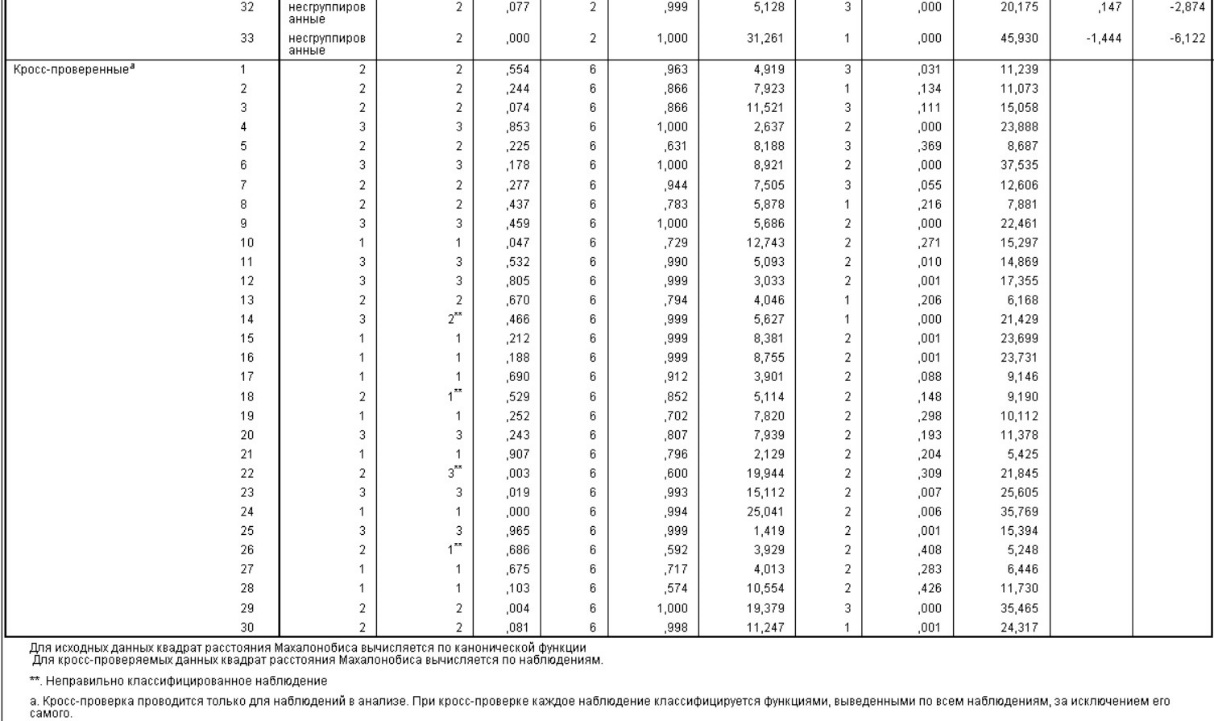
Карта восприятий визуализирует разделение наблюдений функциями. Так, первая функция делит наблюдения на две группы: 1, 2 и 2, 3, вторая функция отделяет наблюдения 2 от всех остальных.

Поле графика разделено дискриминантными функциями на три области: в левой части находятся преимущественно наблюдения первой группы с низкой вероятностью своевременного погашения кредита; в правой части – третьей группы с высокой вероятностью и в нижней части – второй группы со средней вероятностью.



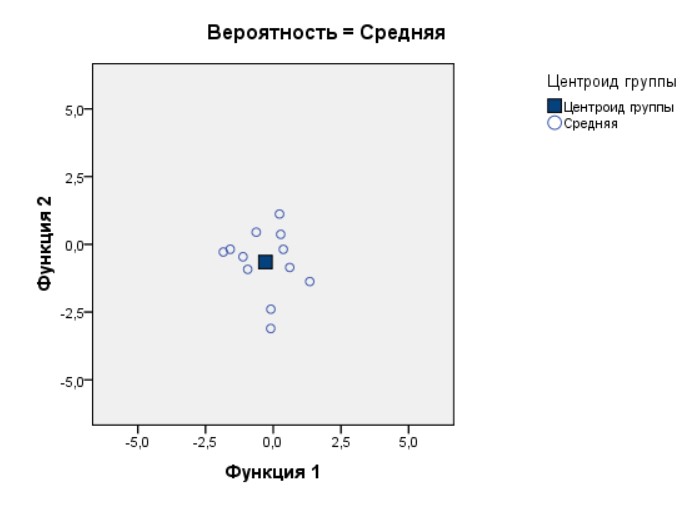
В таблице «Поточечные статистики» размещена информация о фактических (*Actual Group*) и предсказанных (*Predicted Group*) группах для каждого заемщика и соответствующие дискриминантные баллы (*Discriminant Scores*), полученные при подстановке значений переменных в уравнениях дискриминантных функций и .

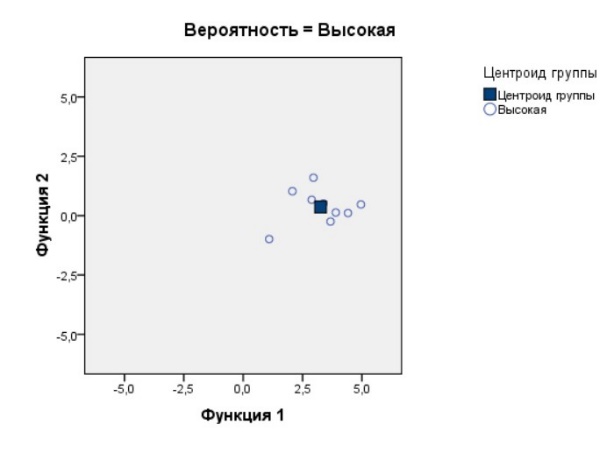
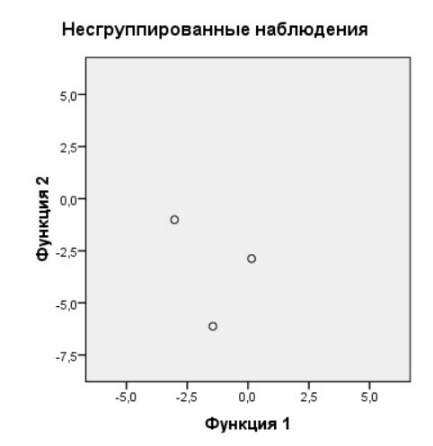


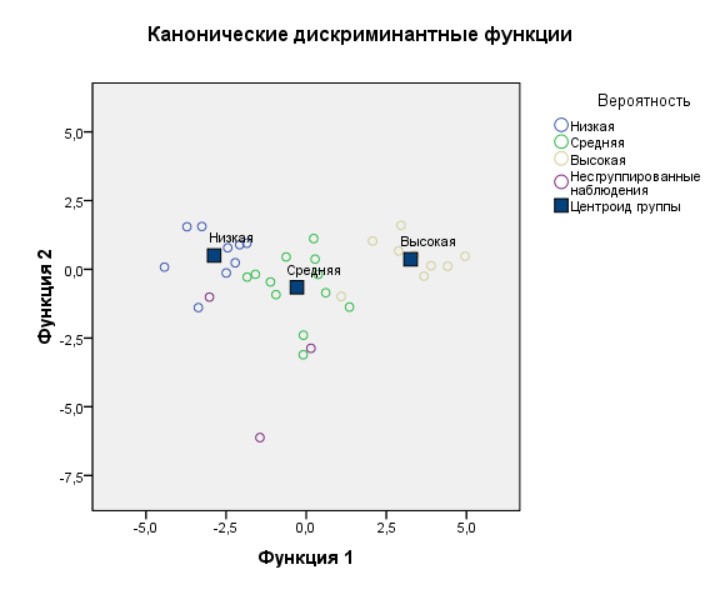


На рисунках с 1го по 3ий ниже отражено расположение заемщиков каждой из трех групп на плоскости двух дискриминантных функций и . По этим графикам можно проводить детальный анализ вероятностей погашения кредита внутри каждой группы, судить о характере распределения заемщиков и оценивать степень их удаленности от соответствующего центроида.

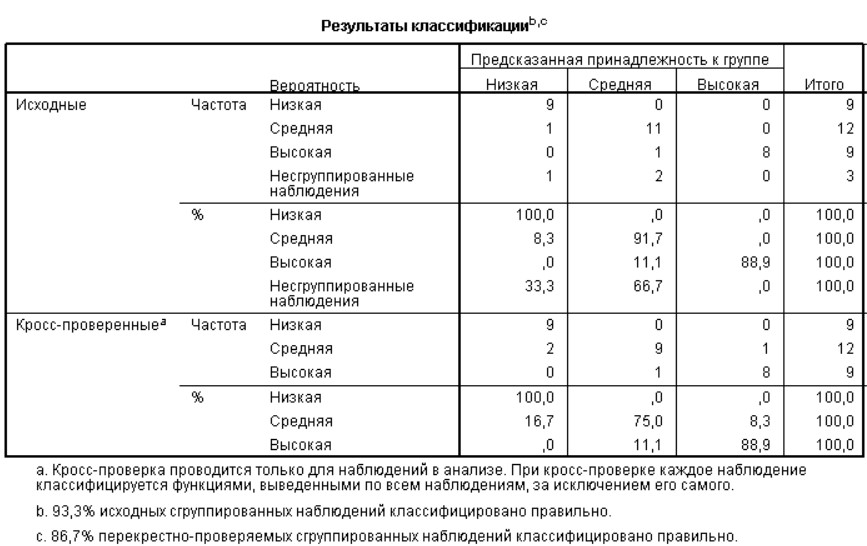
Кроме того, на 4ом рисунке в той же системе координат приведен объединенный график распределения всех групп заемщиков вместе со своими центроидами; его можно использовать для проведения сравнительного визуального анализа характера взаимного расположения групп заемщиков банка с разными вероятностями погашения кредита. В левой части графика расположены заемщики с низкой вероятностью погашения кредита, в правой – с высокой, а в средней части – со средней вероятностью. Поскольку по результатам расчета вторая дискриминантная функция оказалась незначима, то различия координат центроидов по этой оси незначительны. Этот факт подтверждается картой восприятия, которая была расположена выше.



Данные таблицы «Результаты классификации» свидетельствуют о том, что для 93,3% исходных и перекрестно-проверяемых 86,7% сгруппированных наблюдений классификация проведена корректно, высокая точность достигнута в каждой из групп, но в первой она максимальная – 100%, а в третьей несколько ниже – 88,9%, ниже всего во второй – 75%, но все еще относительно хорошая.



**Вывод:** приобрёл практические навыки применения дискриминантного анализа для решения конкретных задач с использованием статистического пакета SPSS.