РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ

**Факультет физико-математических и естественных наук**

**Кафедра прикладной информатики и теории вероятностей**

ОТЧЕТ

по лабораторной работе № 5

дисциплина: Дополнительные главы математической статистики

Студент: Назарьин Артем

Группа: НПМбд-01-19

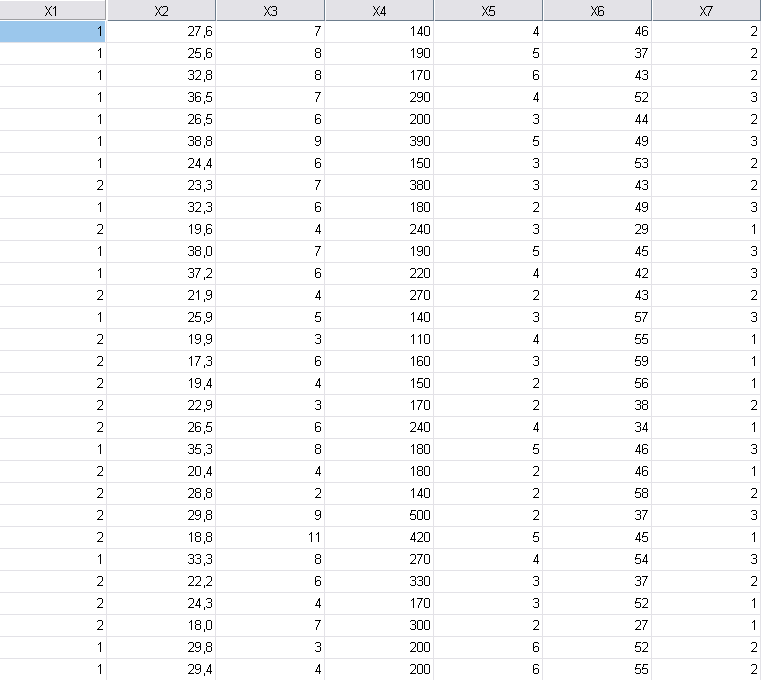
**МОСКВА**

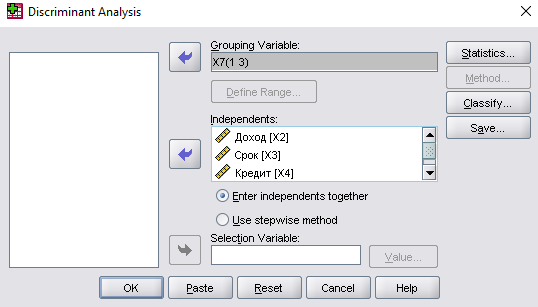
2022 г.

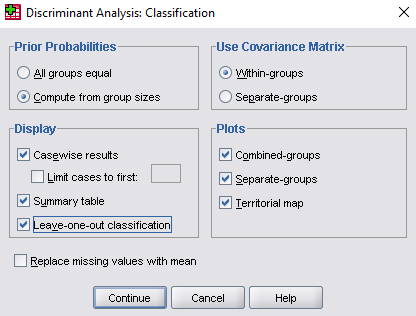
Тема: **«Дискриминантный анализ»**

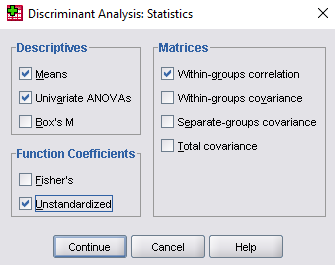
**Вариант 8**

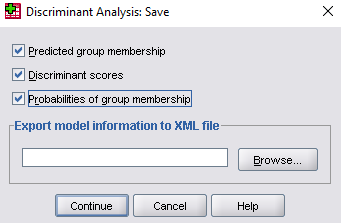
Для начала проведу анализ по обучающей выборке, затем добавлю троих клиентов, для которых нужно будет классифицировать по вероятности погашения кредита.

****

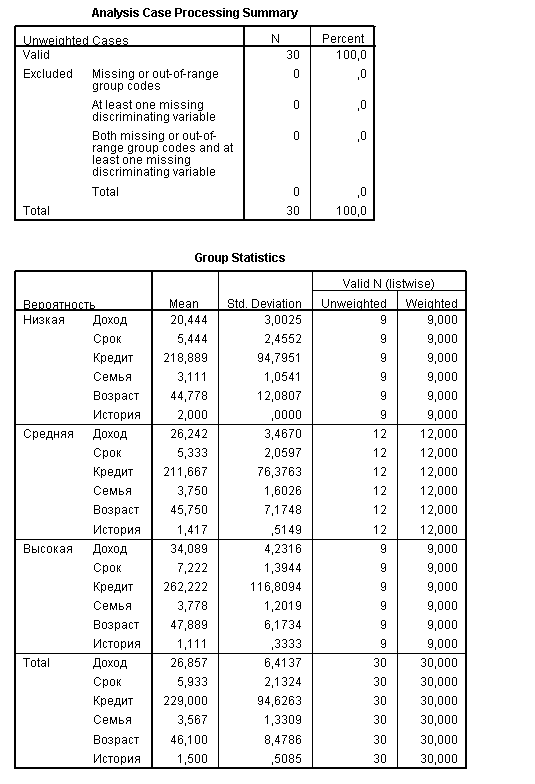






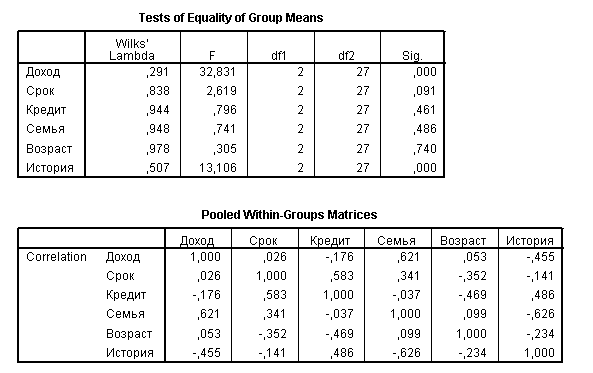


**Результат**:

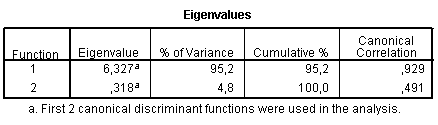


Переменные «Кредит», «Семья», «Возраст», «Срок» незначимо различаются по группам, поскольку для них уровень значимости (sig > 0.05). Следовательно, классификацию заемщиков целесообразно проводить по переменным «История» и «Доход».

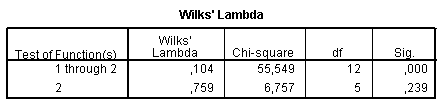
Анализируя матрицу коэффициентов, можно сделать вывод об отсутствии мультиколлинеарности – нет ни одного коэффициента корреляции > 0.8.

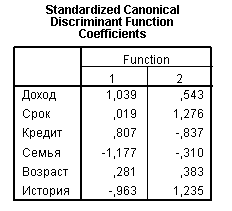
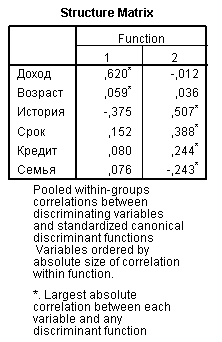


Функция 1 учитывает 95,2% дисперсии, а корреляция между исходными данными и данными, полученными по модели, высокая и составляет 0,929. У второй функции эти значения меньше.



По таблице «Wilks’ Lambda» оценивается уровень значимости дискриминантных функций. Определяем, что в дальнейшем для классификации целесообразно использовать только первую функцию, т.к. ее уровень значимости (Sig.) < 0,001. У второй же функции он составляет 0,239.



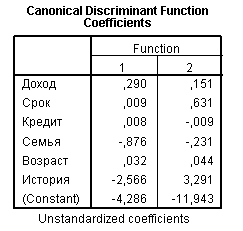
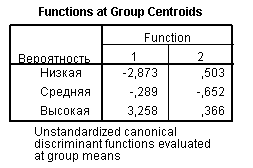
По таблице «Canonical Discriminant Function Coefficients» можно построить две дискриминантные функии:

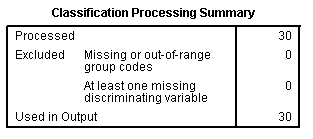
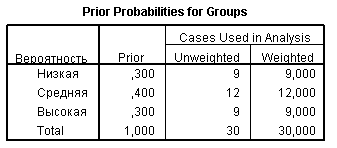
D1(X) = -4,286 – 2,566X1 + 0,290X2 + 0,09X3 + 0,08X4 – 0,876X5 + 0,032X6

D2(X) = -11,943 + 3,291X1 + 0,151X2 + 0,631X3 – 0,09X4 – 0,231X5 + 0,044X6

Однако поскольку коэффициент значимости второй функции >0,001, ее для дискриминации использовать нецелесообразно.

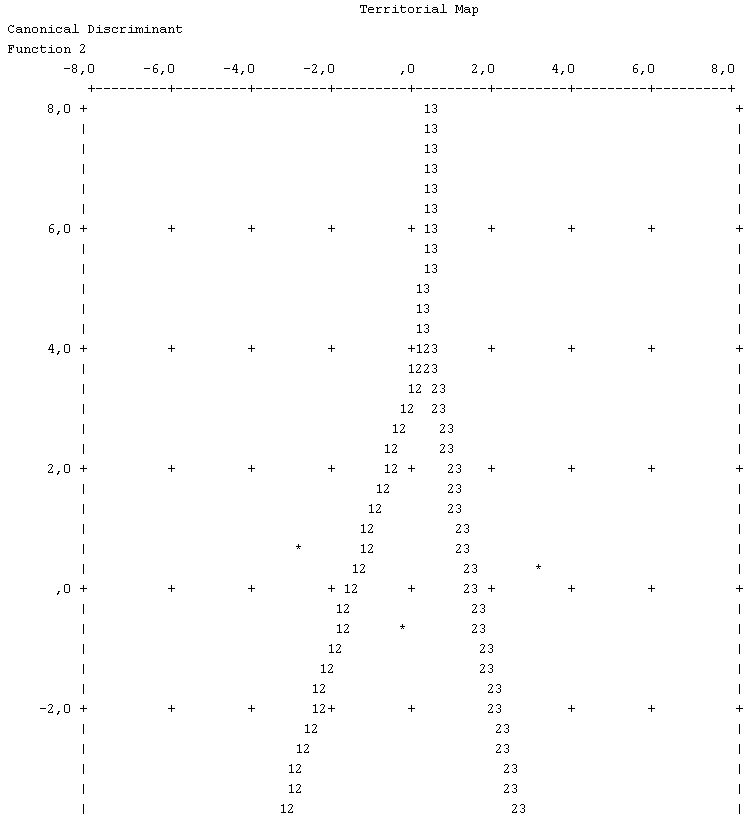
Координаты центроидов по группам приведены в таблице «Functions at Group Centroids». Они используются для нанесения центроидов на карту восприятия.

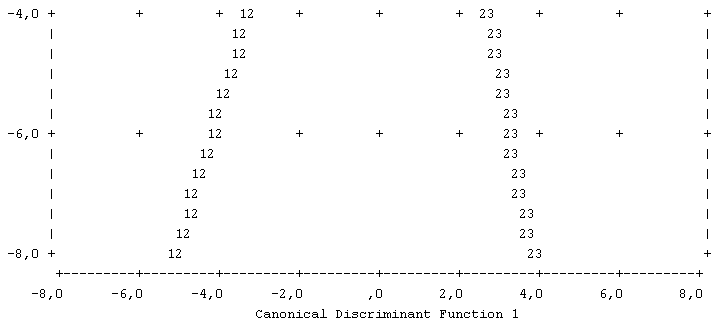
 

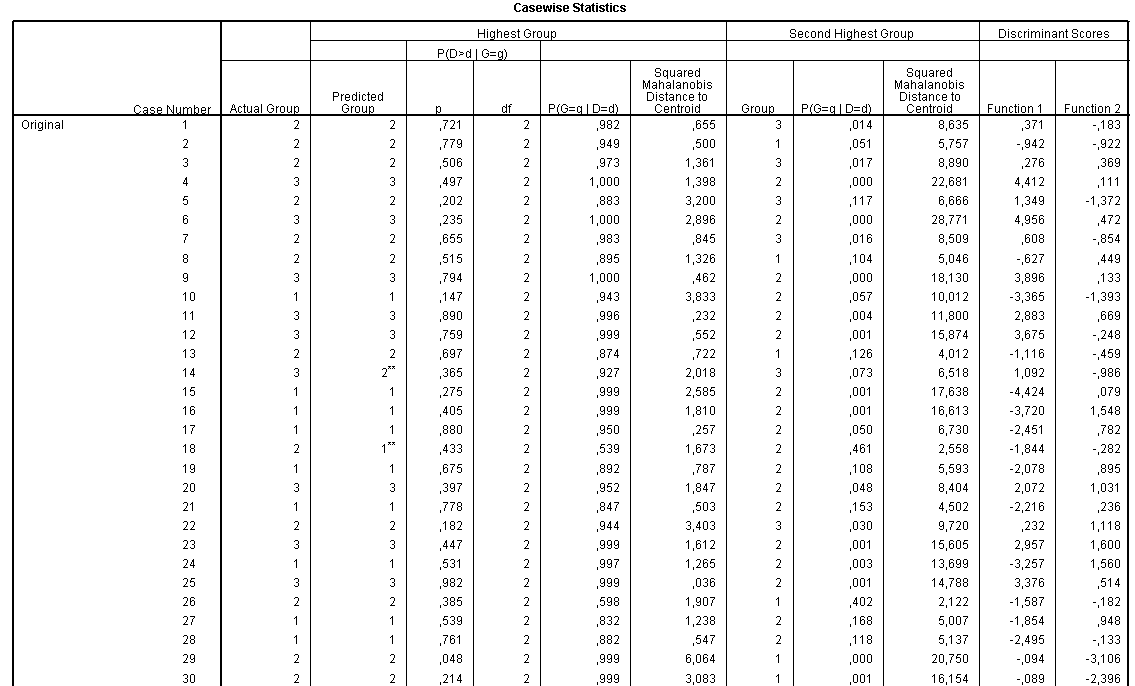
Территориальная карта (Territorial Map) визуализирует разделение наблюдений функциями. Так, первая функция делит наблюдения на 2 группы: 1,2 и 2,3; вторая функция отделяет наблюдения 2 от остальных.

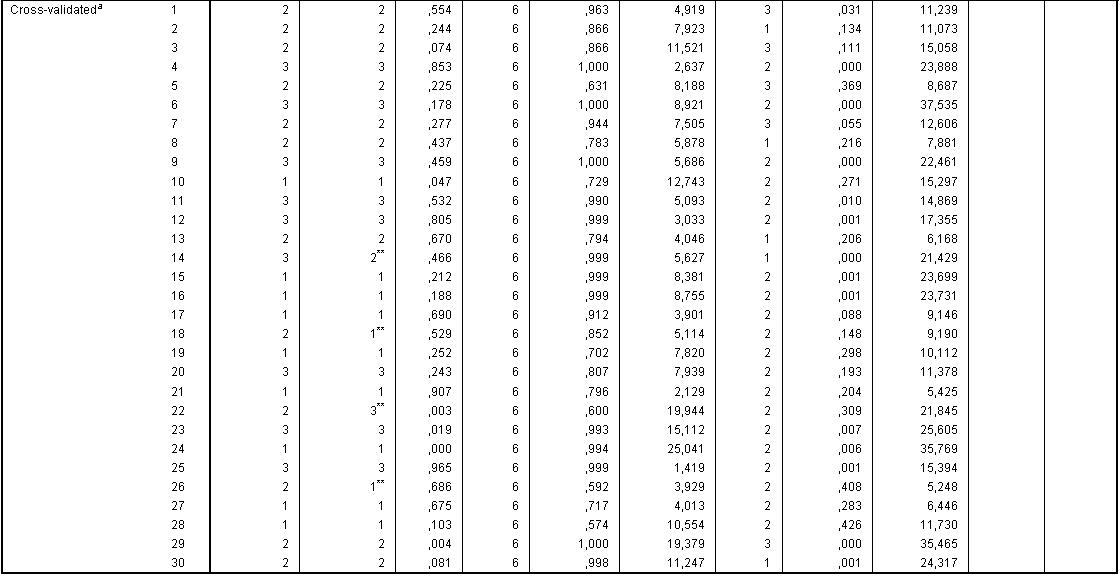
Поле графика разделено дискриминантными функциями на три части: в левой части находятся преимущественно наблюдения третьей группы с высокой вероятностью своевременного погашения кредита; в правой части – первой группы с малой вероятностью и в нижней части – второй группы со средней вероятностью.



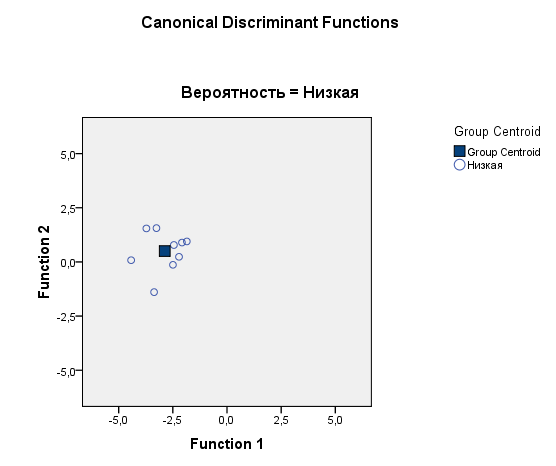
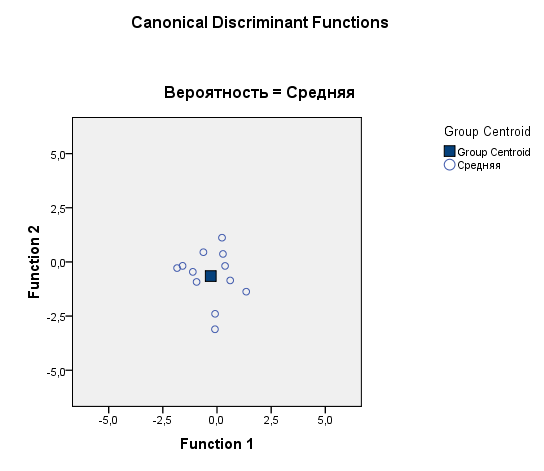


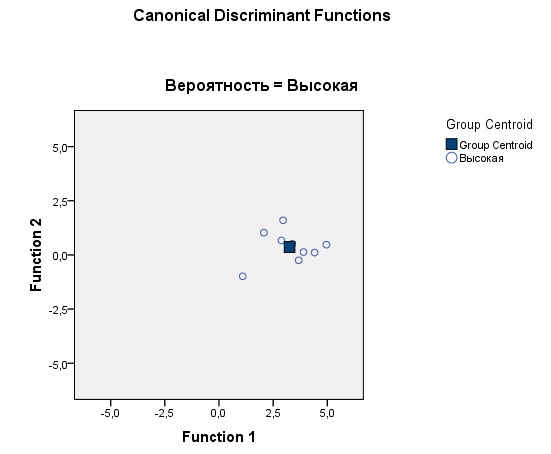
В таблице «Поточечные статистики» (Casewise statistics) размещена информация о фактических (Actual Group) и предсказанных (Predicted Group) группах для каждого заемщика и соответствующие дискриминантные баллы (Discriminant Scores), полученные при подстановке значений переменных в уравнения дискриминантных функций D1(X), D2(X).



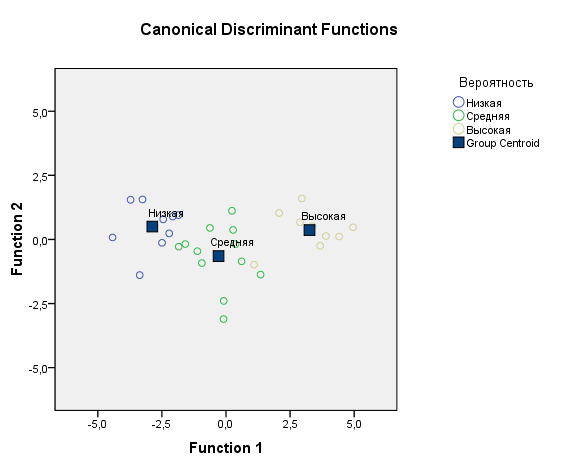


На данных графиках представлено расположение заемщиков каждой из трех групп на плоскости двух дискриминантных функций D1(X), D2(X). По ним можно проводить анализ вероятностей погашения кредита внутри каждой группы, судить о характере распределения заемщиков и оценивать степень их удаленности от соответствующего центроида.

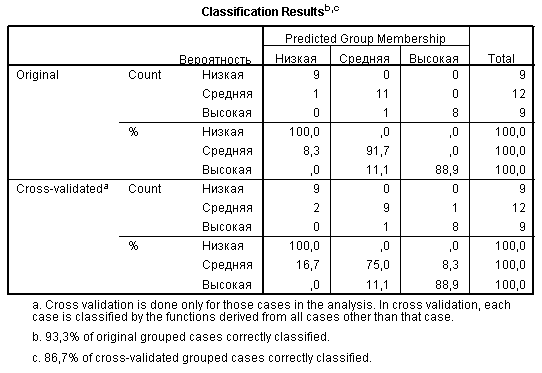
 



На следующем рисунке в той же системе координат приведен объединенный график распределения всех групп заемщиков вместе со своими центроидами; его можно использовать для проведения визуального анализа характера взаимного расположения групп заемщиков банка с разными вероятностями погашения кредита. В левой части графика расположены заемщики с высокой вероятностью погашения кредита, в правой – с низкой, в средней части – со средней вероятностью. Поскольку по результатам расчета вторая дискриминантная функция D2(X) оказалась незначима, то различия координат центроидов по этой оси незначительны. Этот факт подтверждается картой восприятия (выше).

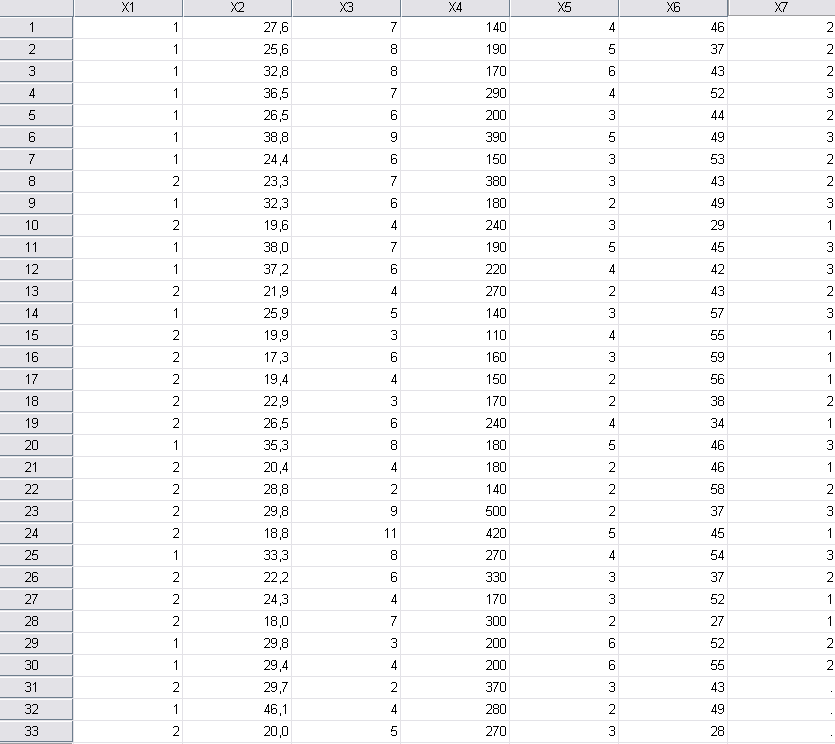


Данные таблицы «Classification Results» свидетельствуют о том, что для 93,3% исходных наблюдений и 86,7% сгруппированных кросс-валидационных (перекрестно-проверяемых) наблюдений классификация проведена корректно, высокая точность достигнута в каждой из групп. В первой группе точность максимальная (100%), во второй и третьей – поменьше (75% и 88,9% соответственно).

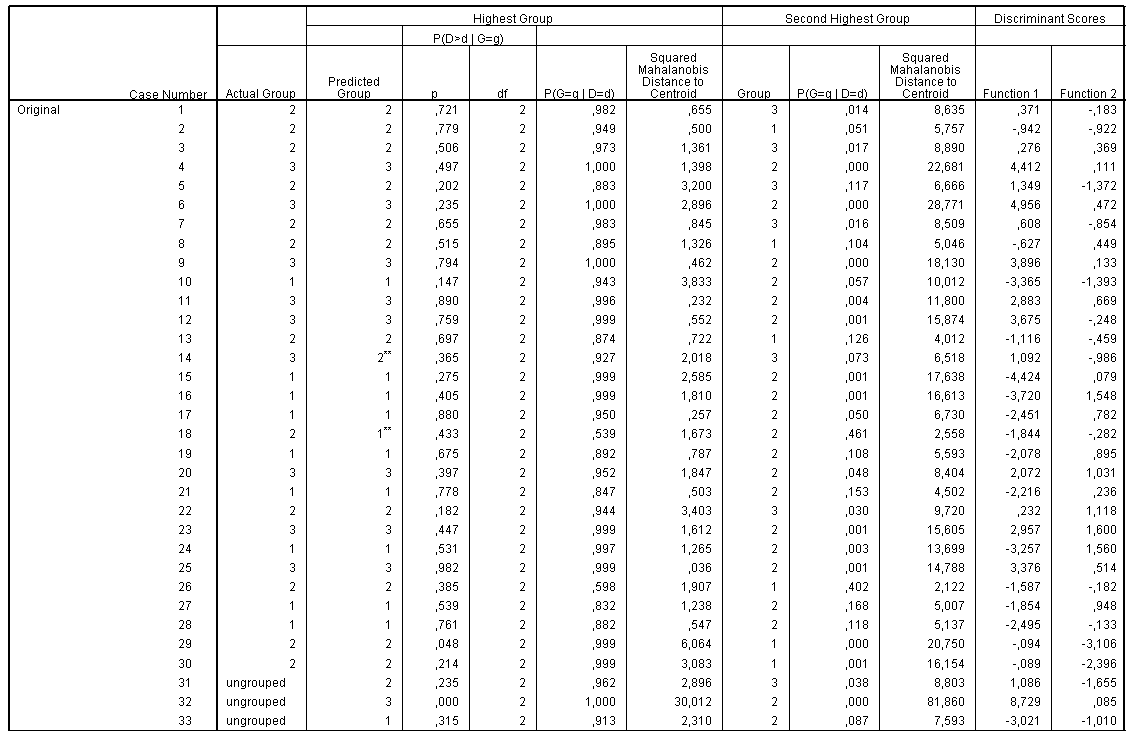


Добавлю в исходную таблицу троих клиентов: №8, 9 и 10.

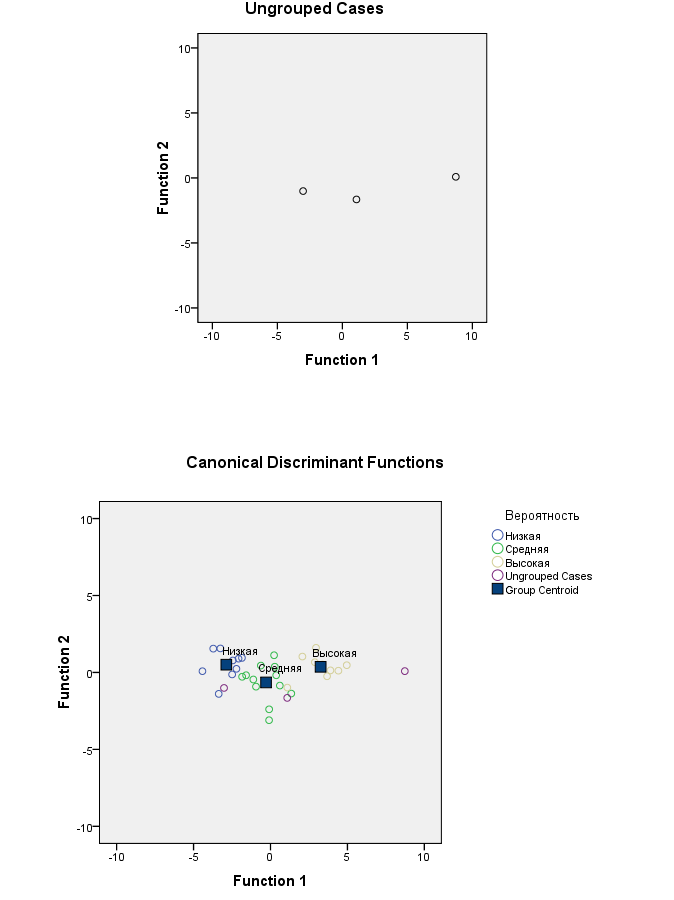
В таблице они расположены под номерами 31, 32, 33.



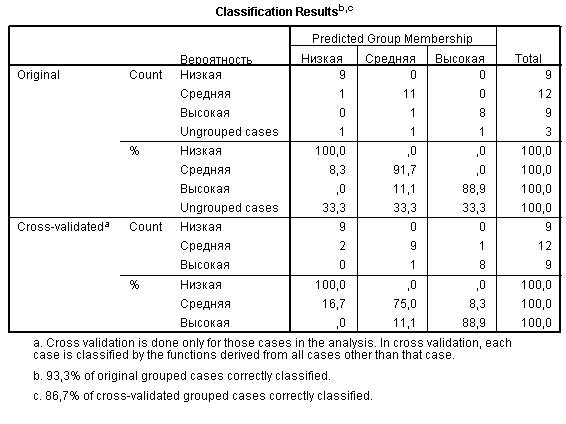
По результатам анализа и прогноза клиент 31 попал в группу 2 (средняя вероятность погашения кредита), клиент 32 попал в группу 3 (высокая вероятность погашения кредита) и клиент 33 попал в группу 1 (низкая вероятность погашения кредита).



Для этих клиентов был выведен график «Ungrouped Cases» (несгруппированные случаи), они также были отображены на графике распределения всех групп заемщиков.



По результатам классификации один клиент был отнесен к группе низкой вероятности погашения кредита, один к средней и один к высокой.



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