Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет Информационных технологий и управления

Кафедра Интеллектуальных информационных технологий

**ОТЧЁТ**

по дисциплине «Статистические основы индуктивного вывода»

ИПР

Тема: Построение бинарного классификатора средствами MS Excel

Выполнил: Самута Д. В.

гр. 221703

Проверил: Ефремов А. А.

Минск 2024

**Оценка выручки ресторана:**

Известна следующая информация:

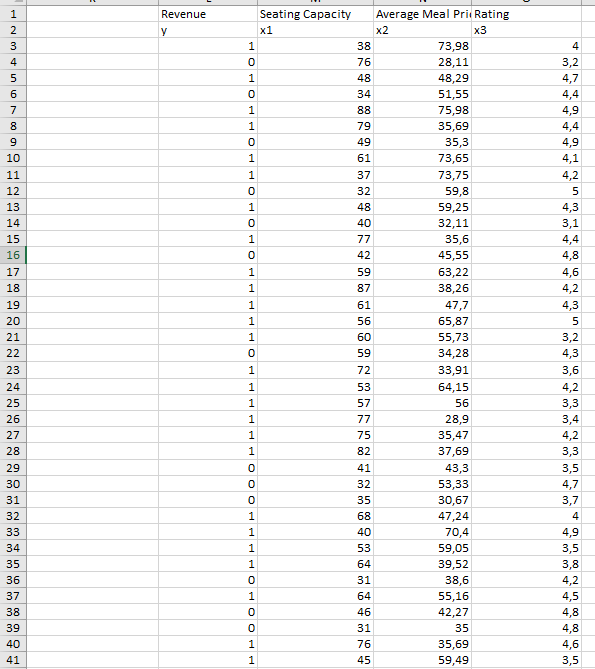
* Выручка ресторана больше 500000(да – 1, нет - 0)
* Кол-во мест в ресторане (коэфф. x1).
* Средняя цена блюда (коэфф. x2)
* Рейтинг (коэфф. x3).

Коэффициенты x1, x2, x3 потенциального ресторана равны 52, 65,1 и 4,5, соответственно.

Требуется:

1. построить линейную регрессионную модель для оценки выручки ресторана, с помощью которой определить, больше ли она 500000 или нет;
2. построить регрессионную дискриминантную модель, найти граничное значение и отнести ресторан с заданными параметрами в одну из категорий.

Информация приведена в следующей таблице

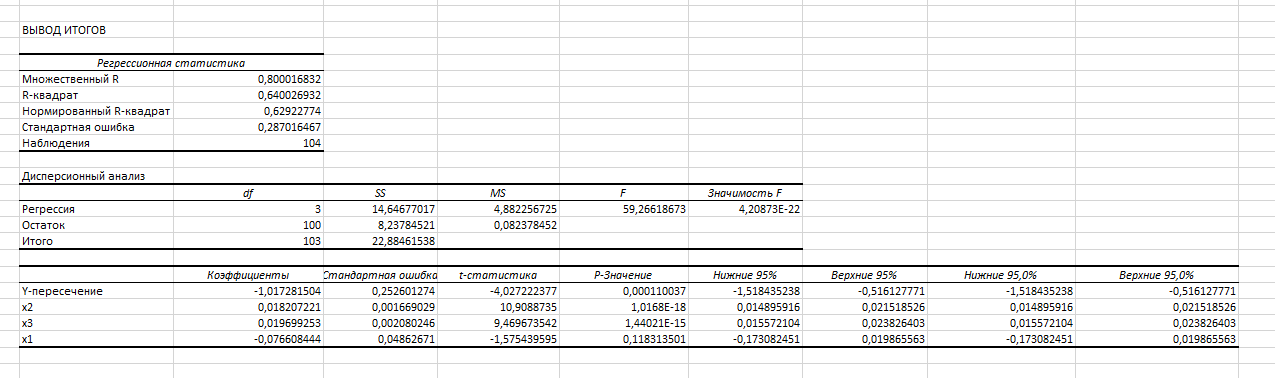


**Пункт 1** Введем таблицу с данными в Excel. Линейная регрессионная модель для оценки выручки в данном случае имеет вид: z = B0 + B1\*x1 + B2\*x2 + B3 \* x3 + e.

Для оценки коэффициентов Bk, будем использовать модуль «Анализ данных», который вызывается из «Сервиса» в главном меню. В «Анализе данных» найдем инструмент «Регрессия» и вызовем его. В появившемся окне укажем входные интервалы y и x.

Входной интервал Y – это массив ячеек (в таблице исходных данных), содержащих значения объясняемой переменной . Входной интервал x – это массив ячеек, содержащих значения объясняющих переменных x1, x2, x3.

Результат выполнения операции над данными:



Вероятность выручки ресторана, больше 500000 вычисляется по формуле: -1,017 + 0,018x1 + 0,02x2 -0,077x3

Посчитаем вероятность для исходных данных 52, 65,1 и 4,5. Результат: 0,87.

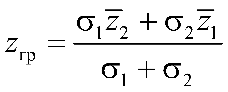
**Пункт 2** В качестве регрессионной дискриминантной модели можно взять модель из п.1.

Для каждого наблюдения вычисляются прогнозные значения показателя по формуле:

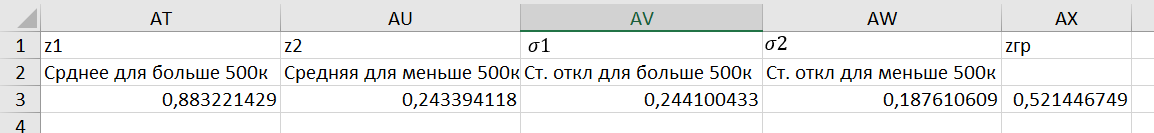
z = B0 + B1\*x1i + B2\*x2i + B3 \* x3i, где i = 1..N.

,    .

Затем с помощью функций СРЗНАЧ и СТАНДОТКЛОН  нужно найти средние значения и стандартные отклонения и для наблюдений с выручкой больше 500к (1-й массив) и для наблюдений с выручкой меньше 500к. (Для этого предварительно следует упорядочить таблицу соответствующим образом.)



Граничное значение вычисляется по формуле:  .



Поскольку z2<z1, то вероятность выручки больше 500000 потенциального ресторана оценивается как высокая, если z > z гр и как низкая, если z < z гр.

**Оценка надёжности классификатора через расчёт процента ошибок.**

N = 104 – кол-во наблюдений

FP = 3 – неверно предсказанные значения для класса 1.

FN = 2 – неверно предсказанные значения для класса 0.

Процент ошибок = (2 + 3) / 104 \* 100% = 4,8%.

