Я проанализировал 3 регрессионные модели на основе представленных в задаче исходных данных.

Type R^2

Linear Regression 0.20066938 4.9632554

Logistic Regression 0.20227031 3.5151098

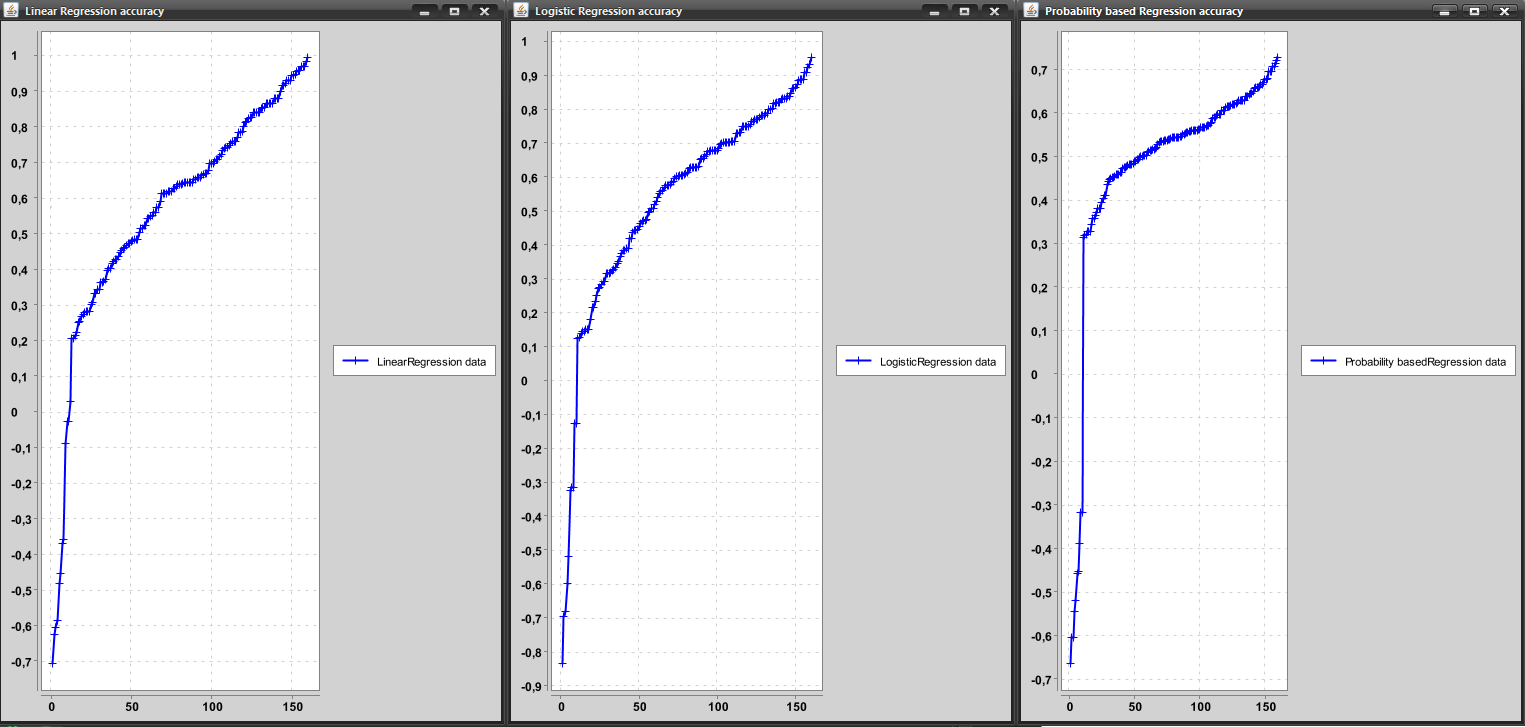
Probit Regression 0.4353688 1.2098458

Все, кроме Probit Regression, модели можно считать значимыми на уровне 5%

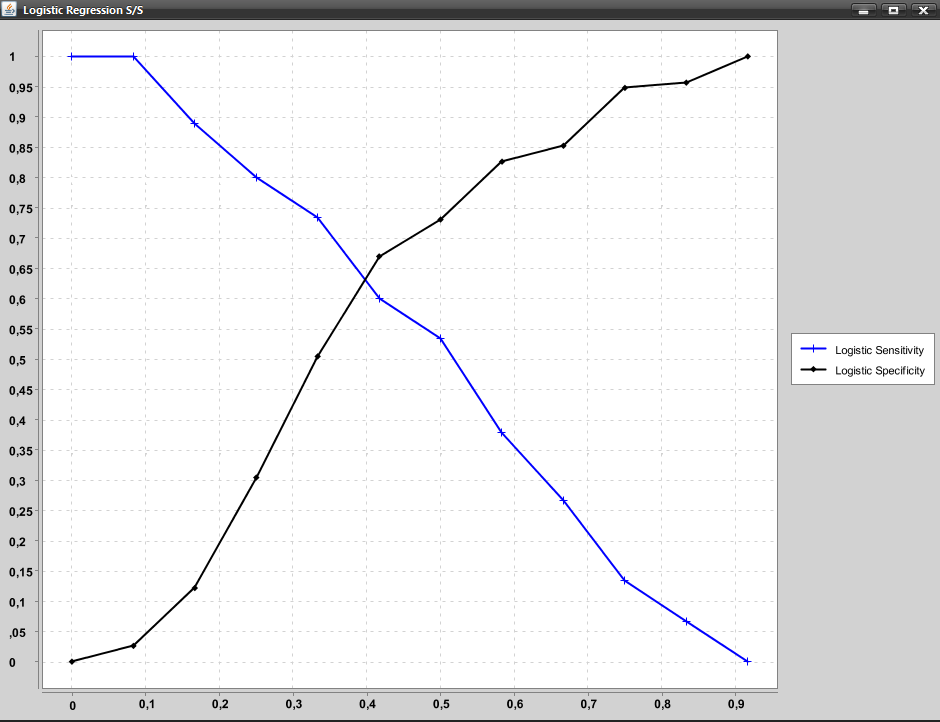
В таблице представлены вероятности для оставшихся десяти матерей

|  |  |  |
| --- | --- | --- |
| Linear Regression | Logistic Regression | Probit model |
| 0.6286446 | 0.6744652 | 0.6117933 |
| 0.39494348 | 0.31702623 | 0.39370218 |
| 0.64090306 | 0.6838879 | 0.54551095 |
| 0.5464616 | 0.3020434 | 0.39486432 |
| 0.97117877 | 0.872744 | 0.68279046 |
| 0.29241315 | 0.1642829 | 0.33482042 |
| 0.97117877 | 0.872744 | 0.68279046 |
| 0.41341922 | 0.47939393 | 0.47901928 |
| 0.37497035 | 0.40018803 | 0.45562005 |
| 0.51594955 | 0.68498075 | 0.5415515 |

Наибольшее внимание следует уделить женщинам, у которых значение вероятности близко к 0.5

На графиках представлена точность оценок каждой модели (Линейная, Логит, Пробит)

Коэффициенты чувствительности и специфичности логистической модели



Оптимальный уровень ~ 0.65