# Тестовое задание

**Цель:**

* Построить модель PD (Probability of default) в программе “R” на основе количественных факторов из финансовой отчетности банков-контрагентов по алгоритму, описанному ниже;
* Заполнить форму отчета, представленную ниже. Длина разделов и подробность изложения не регламентированы.

**Данные:** файл “Data.csv” содержит перечень количественных факторов (qn01,…qn10) и значения PD для выборки из 303 банков-контрагентов. Размерность факторов qn01-qn06 указана в столбцах “Unit qnXX”, факторы qn07-qn10 представлены в процентах.

**Алгоритм построения модели**

1. **Трансформация факторов**

1. Трансформация количественных факторов. Трансформация состоит из двух этапов:

* 1. К факторам qn01-qn10 применяется логистическое преобразование, в результате которого значения фактора заносятся в интервал (0,1):

, где

,

,

– 95-ый процентиль в выборке,

- 5-ый процентиль в выборке,

- значение фактора.

* 1. Нормализация значений факторов. Полученные в результате величины имеют среднее равное 0, и стандартное отклонение равное 50.Для каждого значения фактора преобразованный балл рассчитывался по следующей формуле:

, где

,

.

1. **Однофакторный анализ**
2. Оценка предстазательной силы факторов производится по двум критериям: APS и Availability.
   1. Скорректированный показатель статистической мощности (APS). **Отбираются факторы с абсолютным значением APS не менее 0,3.**

Расчет показателя представлен на Рис. 1. Кривая статистической мощности анализируемого фактора строится следующим образом: все наблюдения упорядочиваются по возрастанию в соответствии с числовыми значениями анализируемого фактора. Координаты *i*-го наблюдения:

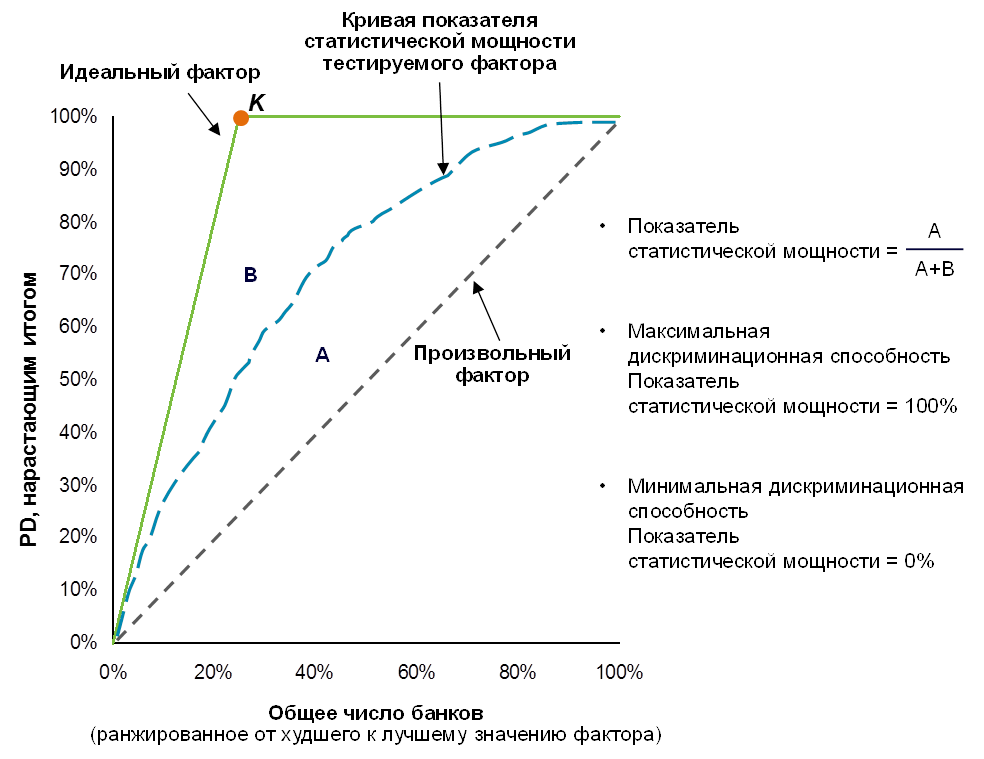
Таким образом, на оси Х нарастающим итогом представлено количество банков, ранжированных в соответствии с числовыми значениями анализируемого фактора (в порядке возрастания). На оси Y отражено отношение показателя подразумеваемого PD (нарастающим итогом) к сумме подразумеваемых PD всех банков в модельной выборке.

Идеальный фактор, позволяет разделить всю совокупность на худшие (дефолтные) и лучшие (недефолтные) банки. Для построения графика идеального фактора считается, что во всей совокупности банков доля дефолтных банков равна среднему значению подразумеваемого PD (например, выборка из 100 банков со средним значением подразумеваемого PD = 2% состоит из 2 дефолтных банков и 98 недефолтных). Таким образом, точка *K*, определяющая идеальный фактор для рассматриваемой совокупности банков, имеет коордитнаты

Скорректированный показатель статистической мощности представляет собой отношение площади А к сумме площадей A и B.

Идеальный фактор имеет показатель статистической мощности, равный 100%. Фактор, дающий в результате произвольное разделение, будет иметь показатель статистической мощности, равный 0%.

**Рис. 1.** График скорректированного показателя статистической мощности



* 1. Доступность данных (Availability) рассчитывается как количество доступных данных (не NA) к общему количеству наблюдений по фактору. **Доступность данных фактора должна быть не менее 80%.**
  2. На основании критериев, описанных в пп. 2.1. и 2.2., для многофакторного анализа отбираются факторы, которые в дальнейшем используются в многофакторном анализе.

1. **Многофакторный анализ**
2. Для каждой комбинации факторов,удовлетворяющих критериям однофакторного анализа строится многомерная регрессия по формуле:

, где

pi – подразумеваемый показатель PD,

β1, β2, β3, ..., βn – веса факторов,

x1 , x2, x3,…,xn – значения факторов (трансформированные).

Коэффициенты регрессии рассчитываются с помощью обощенной линейной модели (GLM). Стандартные ошибки являются состоятельными при гетероскедастичности.

* 1. Для оценки каждой полученной модели используются следующие критерии:
* P-values коэффициентов α, β1, β2, β3, ..., βn *состоятельные при гетероскедастичности* не должны превышать 5%.
* Скорректированный коэффициент детерминации должен быть максимальным. рассчитывается следующим образом:

, где

m – число наблюдений;

n – число независимых переменных;

– коэффициент детерминации (доля дисперсии зависимой переменной, объясняемая независимыми переменными).

* 1. **Скорректированный показатель статистической мощности должен составлять не менее 50% в абсолютном выражении.** Расчет APS выполняется аналогично п. 2.1., при этом ранжирование по оси x происходило в соответствии с предсказанными значениями вероятности дефолта :

, где

α, β1, β2, β3, ..., βn – расчетные коэффициенты оцениваемой регрессии;

x1 , x2, x3,…,xn – значения факторов (трансформированные).

* 1. На основании критериев из пп. 3.1. и 3.2. делается вывод о лучшей PD модели.

# Отчет о построении PD - модели

|  |  |
| --- | --- |
| **Раздел I. Трансформация факторов** |  |

|  |  |
| --- | --- |
| **Раздел II. Однофакторный анализ** |  |

|  |  |
| --- | --- |
| **Раздел II. Многофакторный анализ** |  |