

Пакеты прикладных программ. Весна 2017

Задание №2

Комиссаров Александр
Максимов Пётр
Облыгина Олеся

Московский Государственный Университет им. М.В. Ломоносова
Факультет Вычислительной Математики и Кибернетики
Кафедра Исследований Операций

Апрель 2017





Содержание

Постановка задачи

Извлечение данных

Регрессионный анализ

Построение предсказания

Заключение





Постановка задачи

Мы выступаем в качестве аналитика Дейнерис Бурерожденной из дома Таргариенов, Неопалимой, Кхалиси, Матери Драконов и пр. пр., и помогаем её советнику Тириону Ланистеру выбрать для армии Безупречных поставщика.





Постановка задачи

Необходимо по заданным данным (количество произведенных мечей и брака за каждый прошедший месяц) понять и обосновать, какая компания по производству стали лучше:

- Harpy&Co
- Westeros Inc.





Извлечение данных

Для начала нам необходимо разобраться, какие данные что означают, и как их обрабатывать:

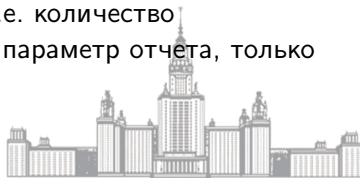
1	unsullen.id	production.date	report.date	produced	defects	supplier
2	1.0	1	1	103.0	0.0	harpy.co
3	1.0	1	2	0.0	2.0	harpy.co
4	1.0	1	3	0.0	4.0	harpy.co
5	1.0	1	4	0.0	5.0	harpy.co
6	1.0	1	5	0.0	13.0	harpy.co
7	1.0	1	6	0.0	11.0	harpy.co
8	1.0	1	7	0.0	11.0	harpy.co
9	1.0	2	2	105.0	0.0	harpy.co
10	1.0	2	3	0.0	2.0	harpy.co
11	1.0	2	4	0.0	2.0	harpy.co
12	1.0	2	5	0.0	2.0	harpy.co



Извлечение данных

Вполне очевидно, что первый столбец это номер кузнеца, а последний - компания поставщик.

Будем считать, что `production.date` это номер произведенной партии от данного кузнеца; `report.date` - месяц отчета по этой партии; `produced` - параметр отчета, т.е. количество произведенных мечей; `defects` - также параметр отчета, только уже количество сломанных мечей.





Извлечение данных

Для обработки данных перенесём эти значения в объект `Data.Frame` и добавим две дополнительные характеристики, которые понадобятся нам в дальнейшем:

unbroken

Количество целых мечей для каждой партии

broke_per

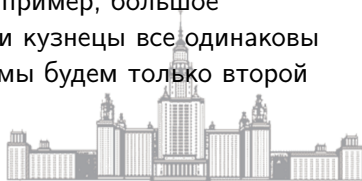
Процент сломавшихся за последний месяц мечей к количеству оставшихся целых





Анализ данных

Первоначально, для достоверного анализа мы должны убрать из рассмотрения эффекты, которые могут исказить представления о качестве поставляемой стали. Это могут быть личные качества каждого кузнеца, качества партий стали и какие-то временные происшествия (например, большое количество битв). Но по условию наши кузнецы все одинаковы и безупречны. Значит рассматривать мы будем только второй и третий эффекты.

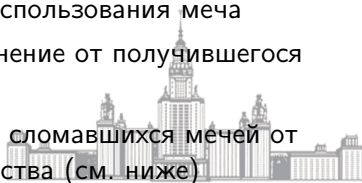




Анализ. Улучшение статистики

Для визуализации временных происшествий и оценки этого параметра на данные мы произвели следующие действия:

- 1 Нашли средние значения сломавшихся мечей по каждому кузнецу (использовали тот самый `broke_per`)
- 2 Нашли среднее геометрическое значения процента сломавшихся мечей по времени использования меча
- 3 Для каждой партии нашли отклонение от получившегося среднего
- 4 Построили таблицу зависимостей сломавшихся мечей от месяца отчета и месяца производства (см. ниже)





Анализ. Построенная таблица

supplier	production.date	2	3	4	5	6	7
harpy.co	1	0.9696078	1.0046407	1.019789	1.0221937	1.0075033	1.0000000
harpy.co	2		1.0200700	1.032069	1.0619275	0.9837887	0.9925526
harpy.co	3			1.009517	1.0084505	0.8642836	0.9944089
harpy.co	4				1.0028517	1.0224166	1.0684119
harpy.co	5					1.0235991	0.9354019
harpy.co	6						0.9756488
westeros.inc	1	1.0019465	0.9831676	1.070277	0.9394950	0.9451071	1.0000000
westeros.inc	2		0.9967793	1.031749	0.9554881	1.0135308	1.0580812
westeros.inc	3			1.028123	1.0407349	1.0498993	1.0501916
westeros.inc	4				1.0056238	0.9909514	0.9313887
westeros.inc	5					0.9822964	0.9558852
westeros.inc	6						0.9859009



Анализ. Выводы из таблицы

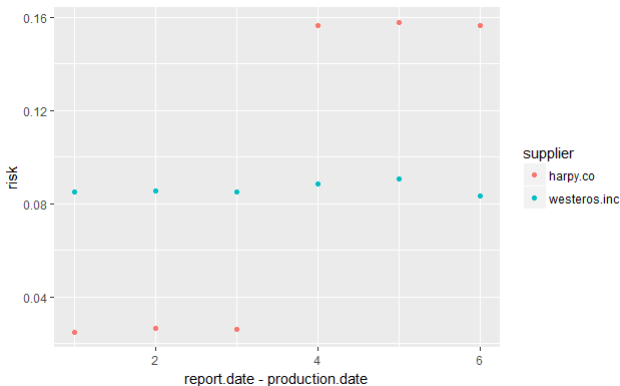
Невооруженным взглядом видно, что все значения близки к единице. Значит никаких ярких дефектов или происшествий не происходило за это время. Отлично, мы можем перейти к анализу данных без внешних возмутителей.





Прямой анализ

Для этого построим зависимость процента сломанных мечей от времени их использования по каждому производителю





Анализ. Вывод

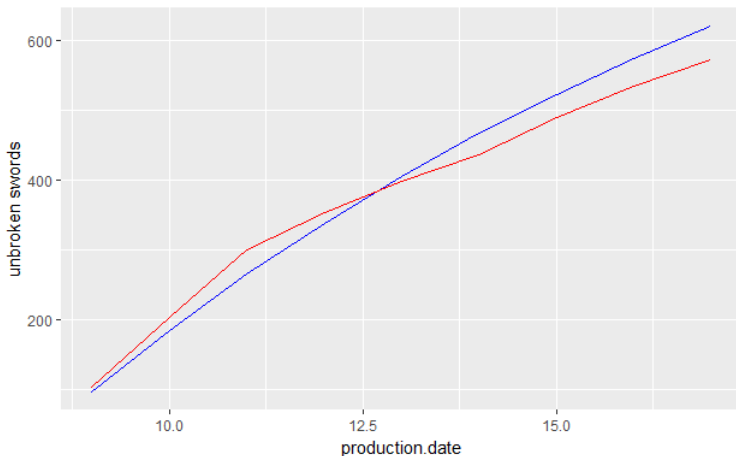
Получаем, что в Westeros Inc. вероятность поломки меча приблизительно постоянная по времени, а в Harry&Co первые три месяца низкая, а в следующие существенно возрастает и остается постоянной. Теперь мы можем перейти к прогнозу на следующие 11 месяцев.





Построение предсказания

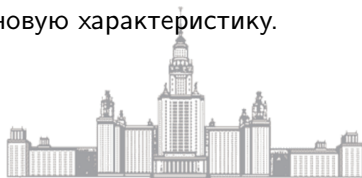
Используя полученную нами модель построим предположение и визуализируем его:





Предсказание. Вывод из графика

Красным обозначено Harpy&Co, синим - Westeros Inc.
Мы не можем точно определить, какая компания лучше, так как ни один график не превосходит другой на всем промежутке.
Для численного сравнения построим новую характеристику.





Предсказание. Численная характеристика

Новая характеристика представляет собой время работы всех мечей в течении горизонта планирования. Для каждой компании она равна:

- Harpy&Co - 4.949132
- Westeros Inc. - 4.660751





Заключение

Получаем, что мечи из стали Harpy&Co долговечнее в среднем на 0.3 месяца. Следовательно необходимо посоветовать именно эту компанию Тириону.

