

Steel Supplier Analysis

Регрессионный анализ

Загружаем данные в объект Data.Frame и добавляем количество несломанных мечей для каждой партии, а также процент сломавшихся в последний месяц мечей к количеству несломанных.

```
steel <- read.csv("production-data.csv")
steel <- steel %>%
  group_by(unsullen.id, production.date) %>%
  mutate(
    unbroken = lag(cumsum(produced) - cumsum(defects)),
    broke_perc = defects / unbroken
  )
```

Найдем средние значения процента сломавшихся мечей по кузнецам.

```
steel.sum <- steel %>%
  group_by(supplier, production.date, report.date) %>%
  filter(production.date != report.date) %>%
  summarize(
    mean = mean(broke_perc)
  )
```

Найдем среднее геометрическое значения процента сломавшихся мечей по времени использования меча, а также для каждой партии найдем отклонение от этого среднего.

```
steel.sum <- steel.sum %>%
  group_by(supplier, report.date - production.date) %>%
  mutate(
    memean = exp(mean(log(mean))),
    deviation = mean/memean
  )
```

Выведем наглядную зависимость процента сломавшихся мечей от месяца репорта и месяца производства.

```
steel.time <- steel.sum %>%
  ungroup() %>%
  select(supplier, report.date, production.date, deviation) %>%
  spread(report.date, deviation)
options(knitr.kable.NA = '')
kable(steel.time, format="markdown")
```

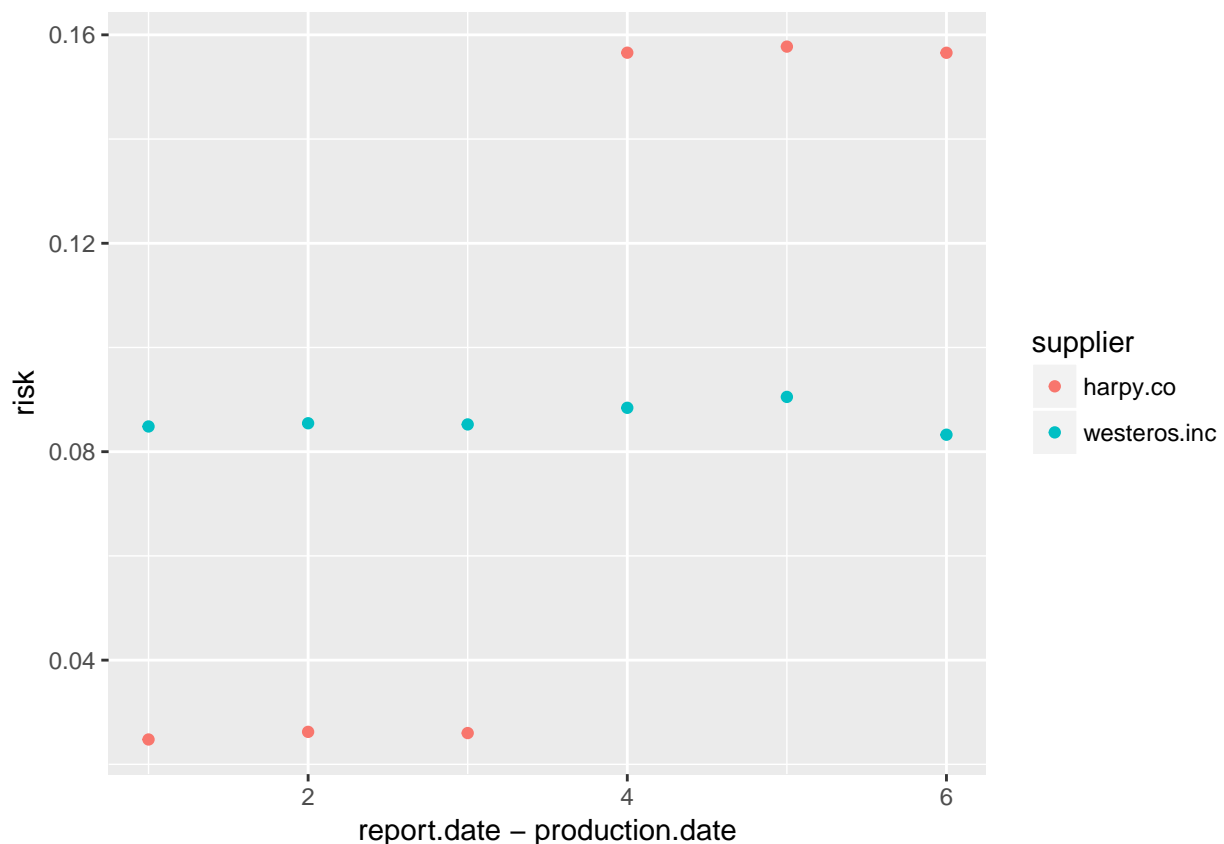
supplier	production.date	2	3	4	5	6	7
harpy.co	1	0.9696078	1.0046407	1.019789	1.0221937	1.0075033	1.0000000
harpy.co	2		1.0200700	1.032069	1.0619275	0.9837887	0.9925526
harpy.co	3			1.009517	1.0084505	0.8642836	0.9944089
harpy.co	4				1.0028517	1.0224166	1.0684119
harpy.co	5					1.0235991	0.9354019
harpy.co	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

supplier	production.date	2	3	4	5	6	7
westeros.inc	5				0.9822964	0.9558852	
westeros.inc	6					0.9859009	

Все значения близки к единице. Получили, что процент сломавшихся мечей не зависит от месяца ведения войны и не зависит от месяца поступления стали. В такой случае достаточно наблюдать зависимость процента сломанных мечей от времени их использования отдельно по каждому производителю.

```
steel.sword.life <- steel %>%
  group_by(supplier, report.date - production.date) %>%
  filter(report.date - production.date != 0) %>%
  summarize(
    defects.sum = sum(defects),
    swords.sum = sum(unbroken),
    risk = defects.sum/swords.sum
  )

ggplot(steel.sword.life, aes(x = `report.date - production.date`, y = risk)) +
  geom_point(aes(colour = supplier))
```



По полученному графику можно сделать предположение о том, что в Westeros.Inc вероятность поломки меча постоянная во времени, а в Harpy.Co первые три месяца низкая, а потом резко возрастает и остается постоянной. Вычислим эти вероятности, а также среднюю производительность кузнецов.

```
risk.west <- mean(
  (steel.sword.life %>%
```

```

    filter(supplier == "westeros.inc")
  )$risk
)

risk.harp.1 <- mean(
  (steel.sword.life %>%
    filter(supplier == "harpy.co") %>%
    slice(1:3)
  )$risk
)

risk.harp.2 <- mean(
  (steel.sword.life %>%
    filter(supplier == "harpy.co") %>%
    slice(4:6)
  )$risk
)

swords.mean <- mean(
  (steel %>%
    filter(produced != 0)
  )$produced
)

```

Построение предсказания

В рамках принятой модели найдем количество несломанных мечей в каждый месяц при выборе одного из поставщиков.

```

production.date <- (9:17)

unbroken <- rep(swords.mean, 9)
unbroken <- unbroken*(1-risk.west)^(1:9)
unbroken <- cumsum(unbroken)

west.model <- data.frame(production.date, unbroken)

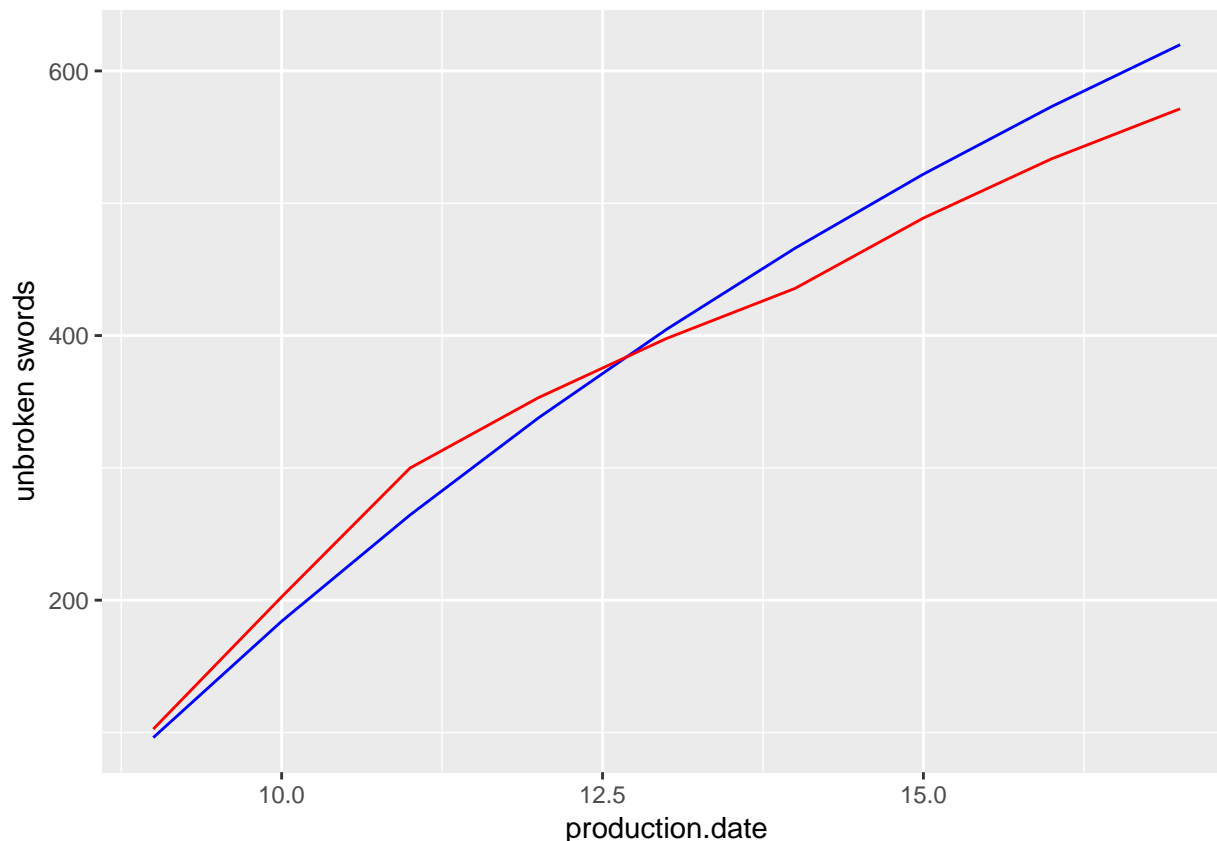
unbroken <- rep(swords.mean, 9)
unbroken <- c(unbroken[1:3]*(1-risk.harp.1)^(1:3),
             unbroken[4:9]*(1-risk.harp.2)^(4:6))
unbroken <- cumsum(unbroken)

harp.model <- data.frame(production.date, unbroken)

models <- data.frame(production.date, west.model$unbroken, harp.model$unbroken)

ggplot(models, aes(x = production.date, color=condition)) +
  geom_line(aes(y=west.model.unbroken, color="blue")) +
  geom_line(aes(y=harp.model.unbroken, color="red")) +
  ylab("unbroken swords")

```



По графику нельзя однозначно оценить правильность выбора конкретного поставщика. Для численного сравнения поставщиков введем численную характеристику их пользы. Выразим это как время работы всех мечей в течение горизонта планирования.

Для этого достаточно найти матожидание исправности меча для каждого из поставщиков. Выберем для этого распределение Бернулли с модификациями. Первая модификация – изменение вероятности поломки меча из стали Harpu.Со спустя 3 месяца эксплуатации. Вторая – мы условимся, что в конце горизонта планирования все мечи ломаются, так как их полезность становится равной нулю.

```
expect.west <- function(n) {
  prob.west <- rep(risk.west, n)
  prob.west <- c(prob.west*(1-risk.west)^(0:(n-1)), (1-risk.west)^(n))
  mean.west <- sum(prob.west*(1:(n+1)))
  return(mean.west)
}

expect.west.all = mean(sapply((1:9), expect.west))

expect.harp <- function(n) {
  prob.harp <- c(risk.harp.1, risk.harp.1*(1-risk.harp.1),
               risk.harp.1*(1-risk.harp.1)^2)
  temp <- rep(risk.harp.2*(1-risk.harp.1)^3, 6)
  prob.harp <- c(prob.harp, temp*(1-risk.harp.2)^(0:5))
  prob.harp <- head(prob.harp, n)
  if(n == 1) {
    prob.harp <- c(prob.harp, 1-risk.harp.1)
  } else if(n == 2) {
```

```

    prob.harp <- c(prob.harp, (1-risk.harp.1)^2)
  } else {
    prob.harp <- c(prob.harp, (1-risk.harp.1)^3*(1-risk.harp.2)^(n-3))
  }
  mean.harp <- sum(prob.harp*(1:(n+1)))
  return(mean.harp)
}

```

```
expect.harp.all = mean(sapply((1:9), expect.harp))
```

```
print(expect.harp.all)
```

```
## [1] 4.949132
```

```
print(expect.west.all)
```

```
## [1] 4.660751
```

Мы получили, что мечи из стали Harpu.Со долговечнее на ~0.3 месяца, а значит, именно эту компанию нужно выбрать в качестве эксклюзивного поставщика.