

Регрессионный анализ для бинарных данных

Линейные модели...

Вадим Хайтов, Марина Варфоломеева



Мы рассмотрим

- ▶ Регрессионный анализ для бинарных зависимых переменных

Вы сможете

- ▶ Построить логистическую регрессионную модель, подобранную методом максимального правдоподобия
- ▶ Дать трактовку параметрам логистической регрессионной модели
- ▶ Провести анализ девиансы, основанный на логистической регрессии

Бинарные данные - очень распространенный тип зависимых переменных

- ▶ Вид есть - вида нет
- ▶ Кто-то в результате эксперимента выжил или умер
- ▶ Пойманное животное заражено паразитами или здорово
- ▶ Команда выиграла или проиграла

и т.д.



На каком острове лучше искать ящериц?



```
liz <- read.csv("data/polis.csv")  
head(liz)
```

| # | X.ISLAND | PARATIO | UTA | PA | PREDICT |
|-----|------------|---------|-----|----|---------|
| # 1 | Bota | 15.41 | P | 1 | 0.555 |
| # 2 | Cabeza | 5.63 | P | 1 | 0.915 |
| # 3 | Cerraja | 25.92 | P | 1 | 0.111 |
| # 4 | Coronadito | 15.17 | A | 0 | 0.568 |
| # 5 | Flecha | 13.04 | P | 1 | 0.678 |
| # 6 | Gemelose | 18.85 | A | 0 | 0.370 |

Зависит ли встречаемость ящериц от размера острова?

Обычную линейную регрессию подобрать можно,

Зависимая переменная: PA - (есть ящерицы "1" - нет ящериц "0")

Предиктор - PARATIO (отношение периметра к площади)

```
#  
# Call:  
# lm(formula = PA ~ PARATIO, data = liz`  
#
```

```
# Residuals:
```

```
#      Min       1Q   Median       3Q      Max  
# -0.649 -0.445  0.178  0.264  0.605
```

```
#
```

```
# Coefficients:
```

```
#              Estimate Std. Error t value Pr(>|t|)  
# (Intercept)  0.86881    0.14088   6.17  0.00001 ***  
# PARATIO     -0.01828    0.00557  -3.28  0.0044 **
```

```
# ---
```

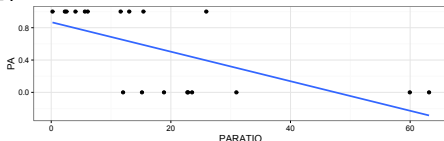
```
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#
```

```
# Residual standard error: 0.413 on 17 degrees of freedom
```

```
# Multiple R-squared:  0.388,    Adjusted R-squared:  0.352
```

```
# F-statistic: 10.8 on 1 and 17 DF,  p-value: 0.00438
```



но она категорически не годится

