

Выбросы Outliers

это наблюдение в анализируемых данных, значение которого сильно отличается от других.

Как опознать выброс? Эвристики.
Например, расстояние три стандартных отклонения от среднего.

С точки зрения исследователя:
принять решение, как их
“нейтрализовать”.

С точки зрения журналиста:
что-то, что нужно расследовать!

Сила выбросов

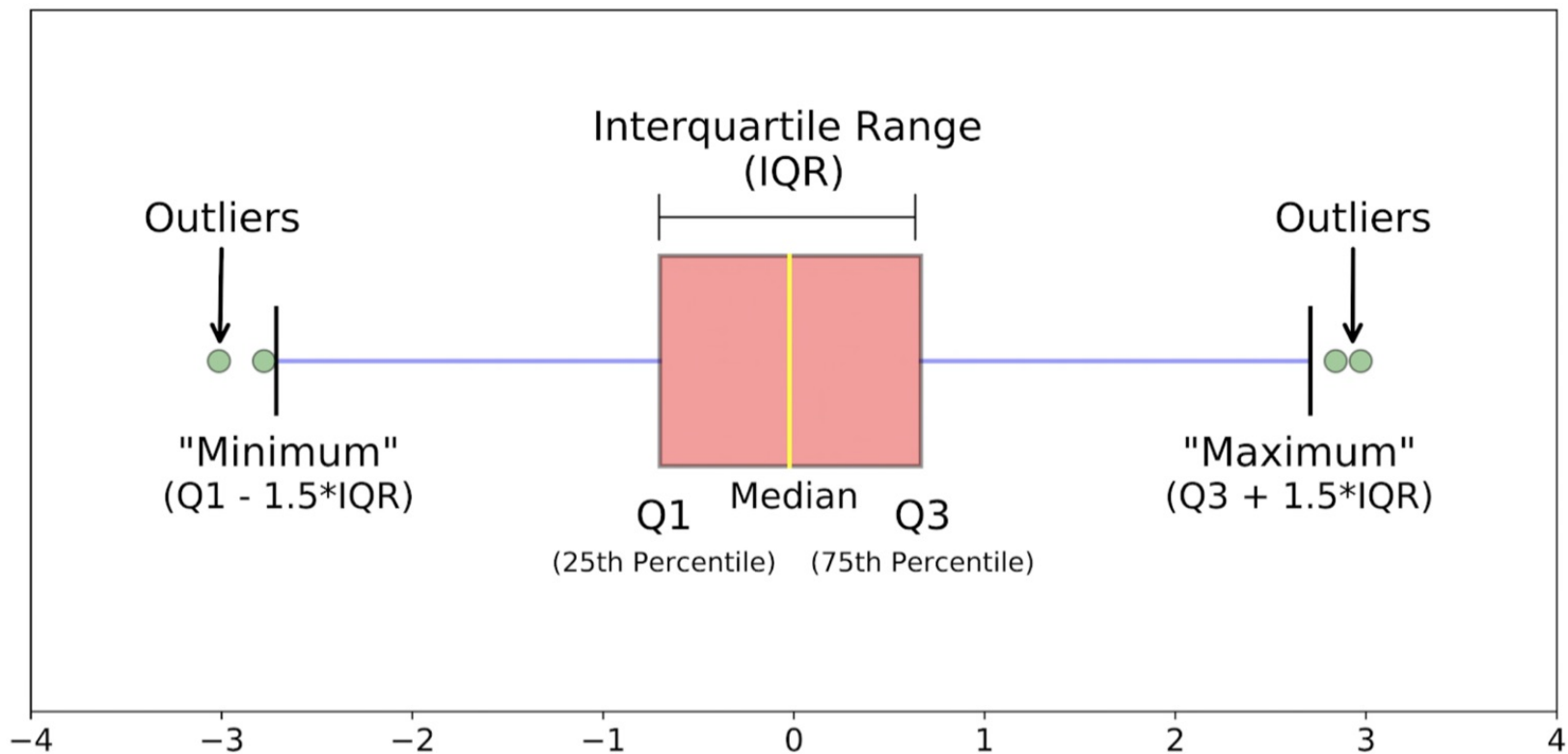
Слабые выбросы

- Те значения, которые меньше 25% перцентиля минус $1,5 \cdot \text{ИКР}$ (интерквартильный размах) или больше 75% персентиля плюс $1,5 \cdot \text{ИКР}$.
- Или находятся от среднего на расстоянии от двух до трех стандартных отклонений.

Сильные выбросы

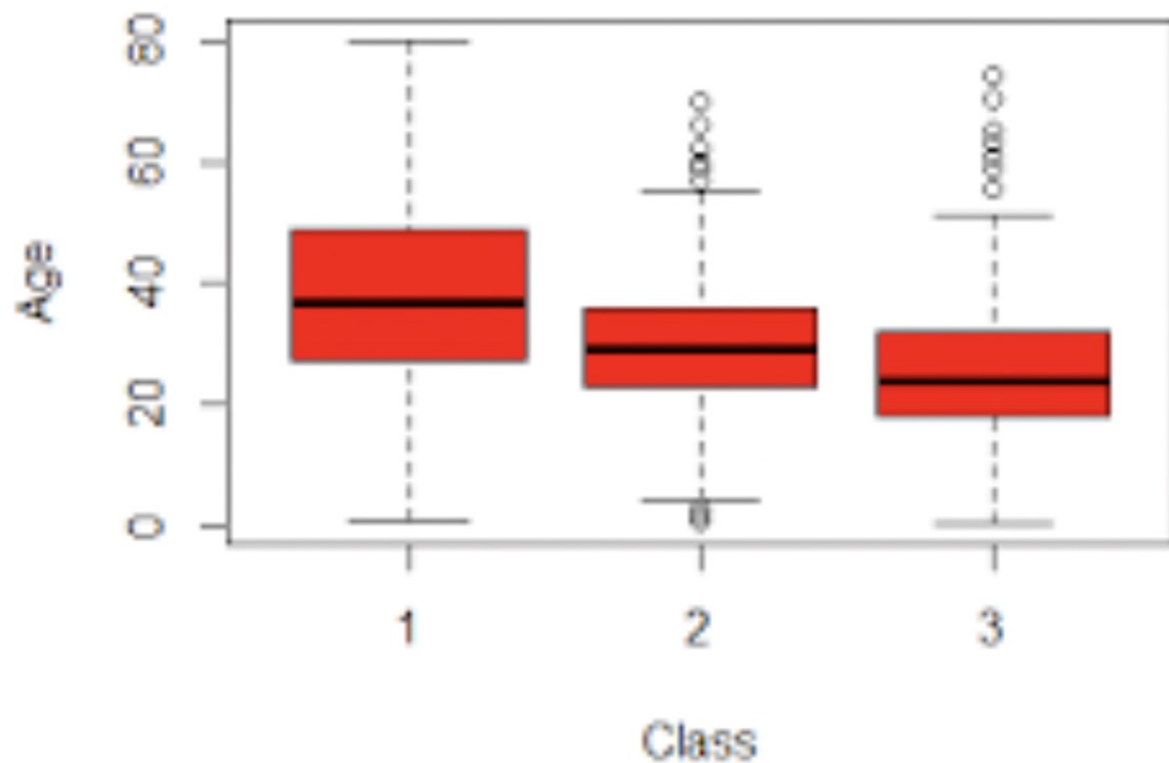
- Те значения, которые меньше 25% перцентиля минус $3 \cdot \text{ИКР}$ (межквартильный размах) или больше 75% персентиля плюс $3 \cdot \text{ИКР}$.
- Или дальше от среднего чем три стандартных отклонения.

Выбросы



В ящике-с-усами, усы как раз отграничивают выборку от выбросов.

Выбросы



Найдите выбросы
на графике.

- Регулярные выражения (англ. regular expressions) — формальный язык для работы с подстроками в тексте.
- валидация и поиск
- Операторы:
- `.` - любой символ (`3.` -- `за`, `зо`, но не `з`)
- `?` - символ может быть, а может не быть (`п?ол` -- `пол`, `ол`, но не `пл`)
- `|` - либо один, либо другой (`сон|ром`) соответствует `сон`, `ром`, но не `сор`)
- `()` - группировка (`пл(ов|аст)` соответствует `плов` `пласт`, но не `плав`)

Операторы

- $*$ — предыдущий символ или группа, повторённые любое количество раз

(включая 0)

- $+$ — предыдущий символ или группа, повторённые любое положительное

количество раз

соответственно, $.*$ — любое количество любых символов

Пример

ко

Кошка, или домашняя кошка (лат. *Felis silvestris catus*)

— домашнее животное, одно из наиболее популярных

(наряду с собакой) «животных-компаньонов»

Пример

ко

Кошка, или домашняя кошка (лат. *Felis silvestris catus*)

— домашнее животное, одно из наиболее популярных

(наряду с собакой) «животных-компаньонов»

Пример

и(б|в)?ол

Кошка, или домашняя кошка (лат. *Felis silvestris catus*)

— домашнее животное, одно из наиболее популярных

(наряду с собакой) «животных-компаньонов»

Пример

и(б|в)?ол

Кошка, или домашняя кошка (лат. *Felis silvestris catus*)

— домашнее ж**ИВОТ**ное, одно из на**ИБО**лее популярных

(наряду с собакой) «ж**ИВОТ**ных-компаньонов»

Пример

до.*е+

Кошка, или домашняя кошка (лат. *Felis silvestris catus*)

— домашнее животное, одно из наиболее популярных

(наряду с собакой) «животных-компаньонов»

Пример

до.*е+

Кошка, или домашняя кошка (лат. *Felis silvestris catus*)

— домашнее животное, одно из наиболее популярных

(наряду с собакой) «животных-компаньонов»