

1.1 пропуски

метрики cameo:

эфф-сть источников трафика

конв-я

-доля посет

-доля целевых д-й

веб cameo:

счетчик д-ий отпр инфу в я.метрику

ID - куки

запись инфы в логи

<https://loginom.ru/blog/missing>

почему пропуски?

типы пропусков:

1. полностью случ

независимая и const вер-сть их встретить

(зам медианой)

(можно исключить)

2. случ

P freq зав от др столбцов/данных

(медиана с группировкой)

(а можно исключить, но нежелательно)

3. неслуч

P зав от атрибутов ктрых нету в данных

(их не стоит игнорить или искл)

замены пропусков:

1. дефолт знач-е

2. характерные

среднее, медиана

когда что? (чув-сть к выбросам)

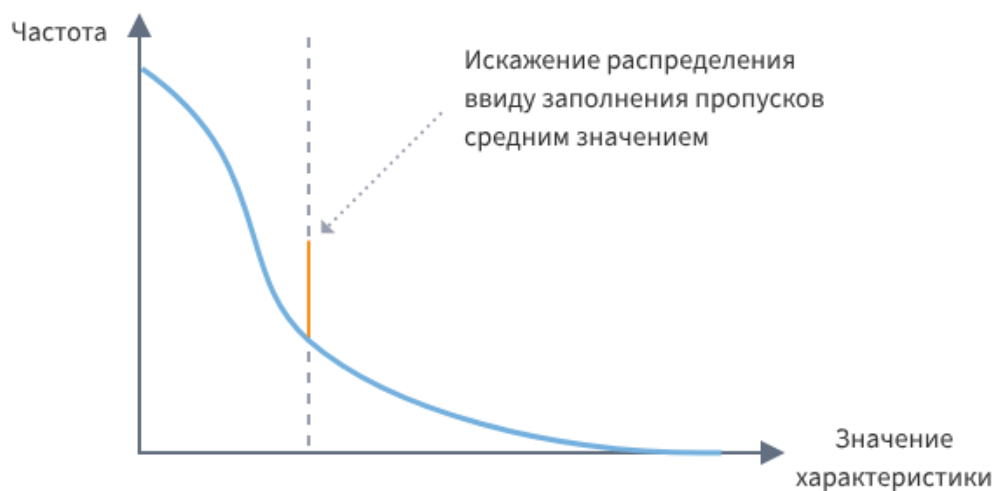
исключение пропусков

- рост std
- падение репр-сти
- особенно при *случ* и еще более при *неслуч*

игнор пропусков

- коэф корр abs>1
в отличие от исключения

зам средним



- изменение распр-я неестественно(появ мода)
- искажение статистик

зам модой

- при катег дискр перемен
- искажение статистик

зам медианой

- лучше если есть выбросы

зам посл значением

- врем ряды

индикация пропусков 0-1 столбцом

регрессионные модели

