

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>

пропуски/их причины:

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)

3. *неслуч*

P зав от атрибутов ктрых нету в данных

(их нельзя игнорить или искл)

замены пропусков:

1. расчетн

2. дефолт

3. характерные

среднее, медиана

когда что? (чув-сть к выбросам)

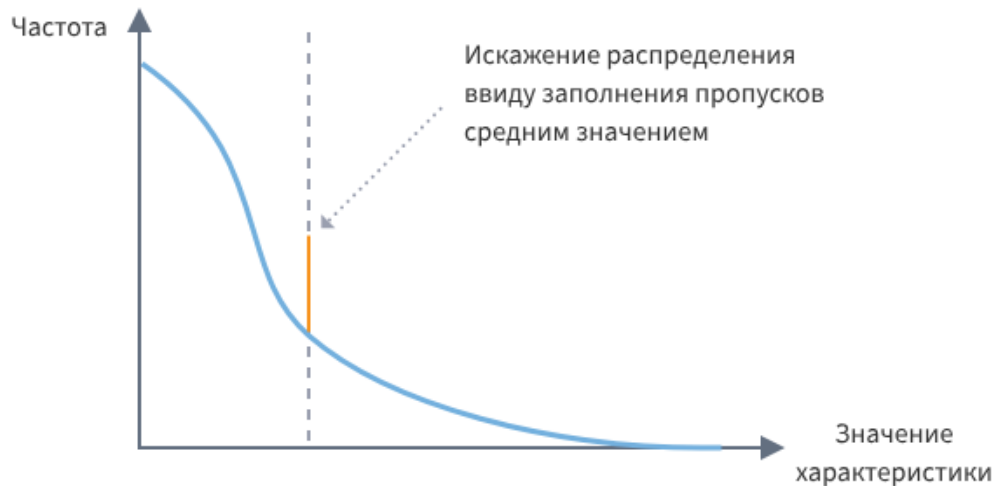
исключение пропусков

- пост 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

- какие преимущества??

зам посл значением

- врем ряды

индикация пропусков 0-1 столбцом

регрессионные модели
