

# 1.1 пропуски - причины, типы, замены

---

1 метрики и веб cameo

2 причины пропусков

3 3 типа пропусков

4 что делать с пропусками

медиана, медиана с группировкой(когда не работает)

---

## 1 метрики cameo:

эфф-сть источников трафика

конв-я

-доля посет

-доля целевых д-й

---

## 1 веб cameo:

счетчик д-ий отпр инфу в я.метрику

ID - куки

запись инфы в логи

---

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

## ключевой вопрос - почему пропуски?

1 типы

2 единицы измерения

## 2 типы пропусков:

1. *полностью случ*

независимая и const вер-сть их встретить

**(можно исключить)**

2. *случ*

P freq зав от др столбцов/данных

**(а можно исключить, но нежелательно)**

3. *неслуч*

P зав от атрибутов ктрых нету в данных

**(их не стоит игнорить или искл)**

сгруппируем их по столбцу опр-щему вер-сть в случае 2:

1 по группам одинак колво

2 по группам разн колво

3 рандом колво

---

### 3 замены пропусков:

1. дефолт знач-е
2. характерные  
среднее, медиана  
когда что? (чув-сть к выбросам)

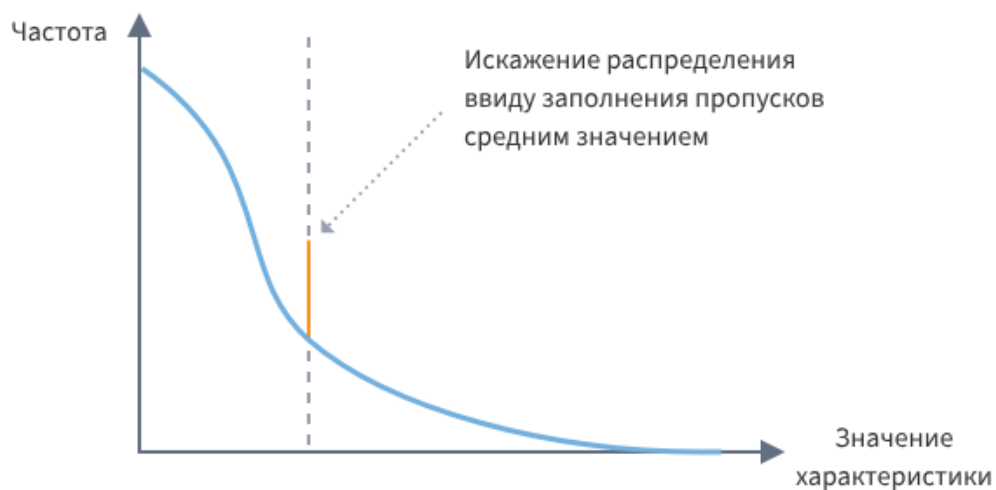
### исключение пропусков

- рост std
- падение репр-сти
- особенно при 2 случ и еще более при 3 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 МЕДИАННОЙ С ГРУППИРОВКОЙ**

по столбцу ктрый коррелирует с нашим

а) плохо работает если большая часть группы - пропуски

б) не работает если вся группа - пропуски

## **зам посл значением**

- врем ряды

## **индикация пропусков 0-1 столбцом**

## **регрессионные модели**

---

<https://www.statisticssolutions.com/dissertation-resources/multiple-imputation-for-missing-data/>