ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

Кейс «Ковчег»

Уваров Кирилл

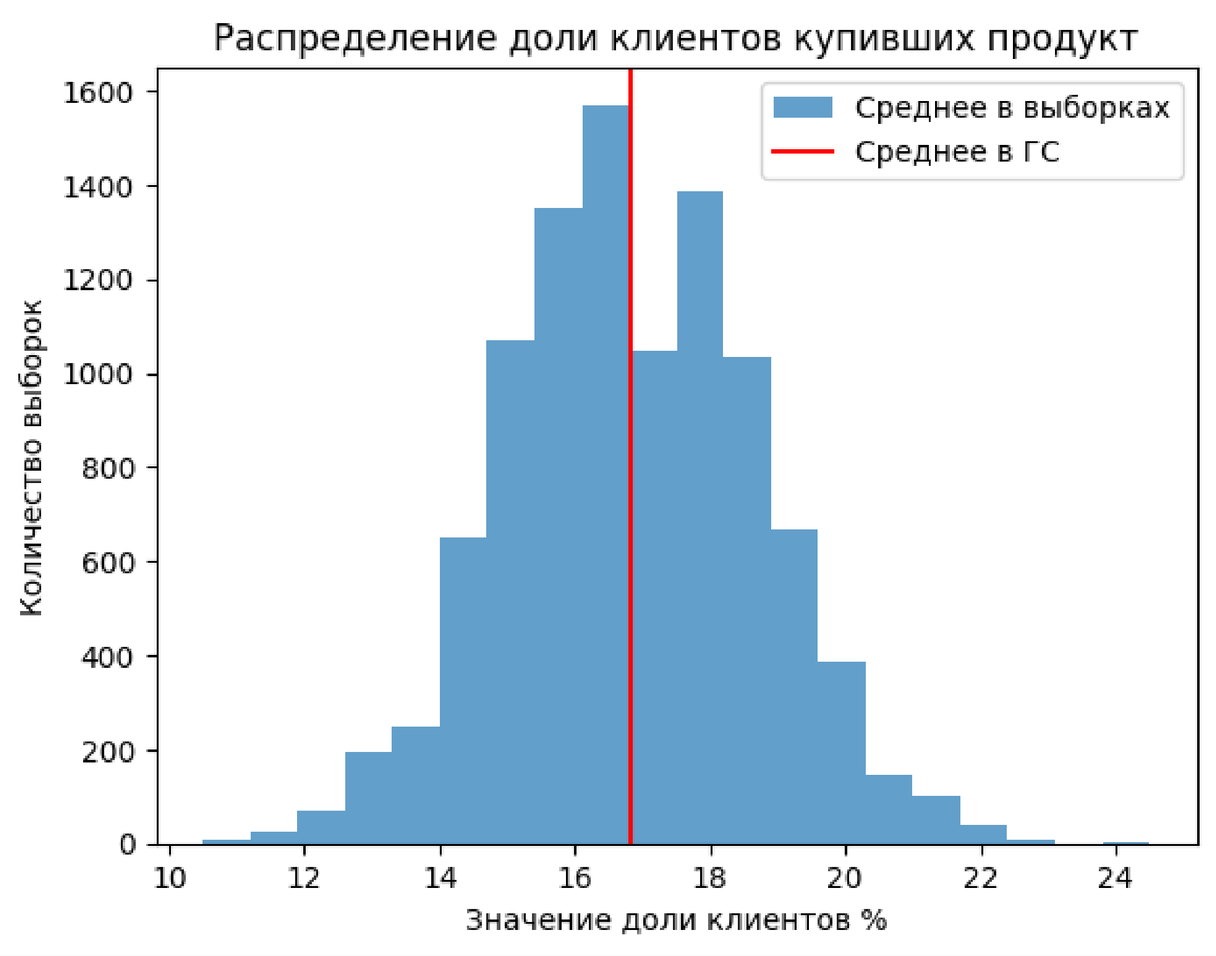
ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

Синий уровень

Как получать выборки?

Способ 1

Будем получать выборки случайным образом выбирая из всевозможных выборок одну.



Результат 1 способа

Наблюдаем небольшое занижение, но в целом выборки получаются более менее репрезентативными.
Так же видим небольшое количество выборок(около 1000) с завышением/занижением более 3%, т.е. около 10% полученных выборок дают достаточно большую ошибку.

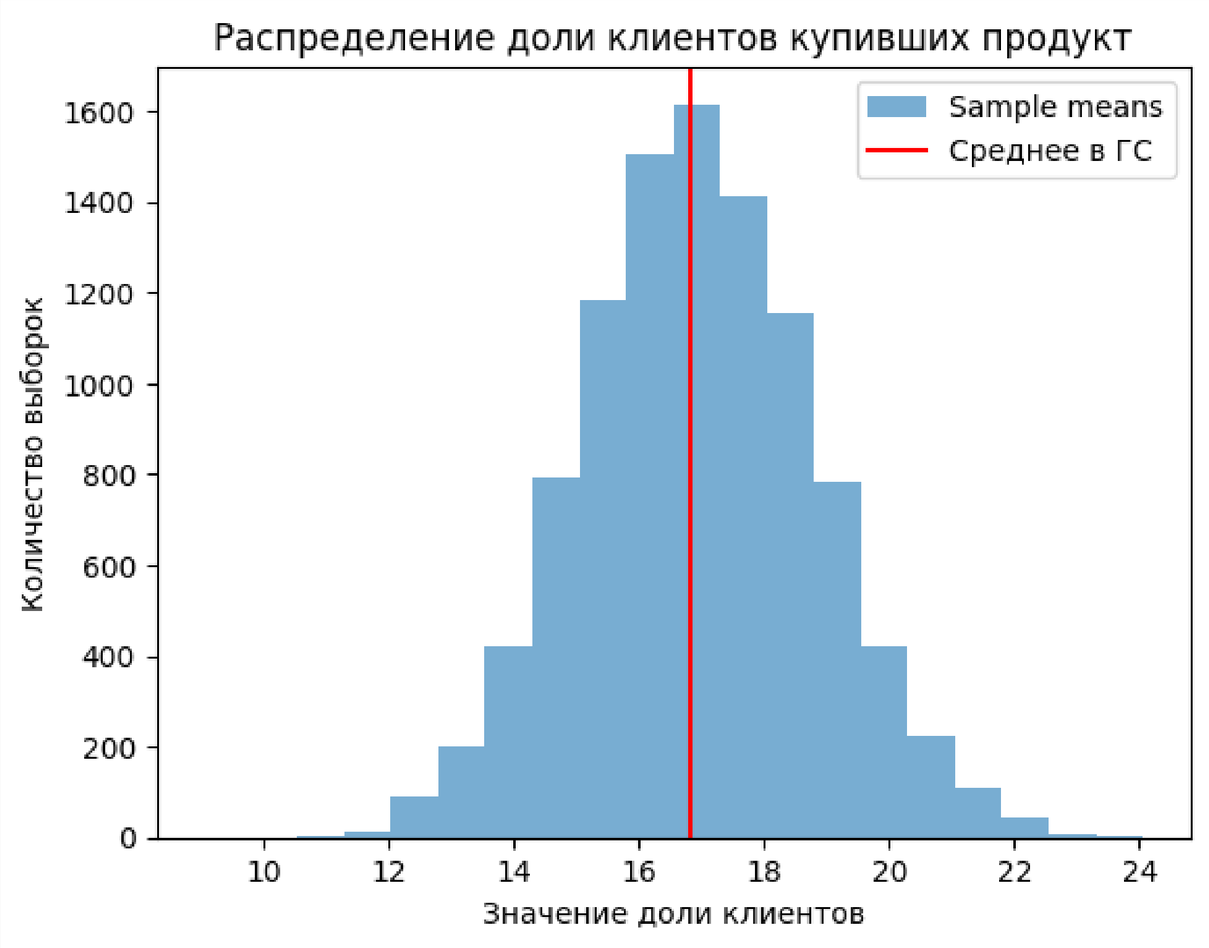
ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

Красный уровень

Как получать выборки?

Способ 2

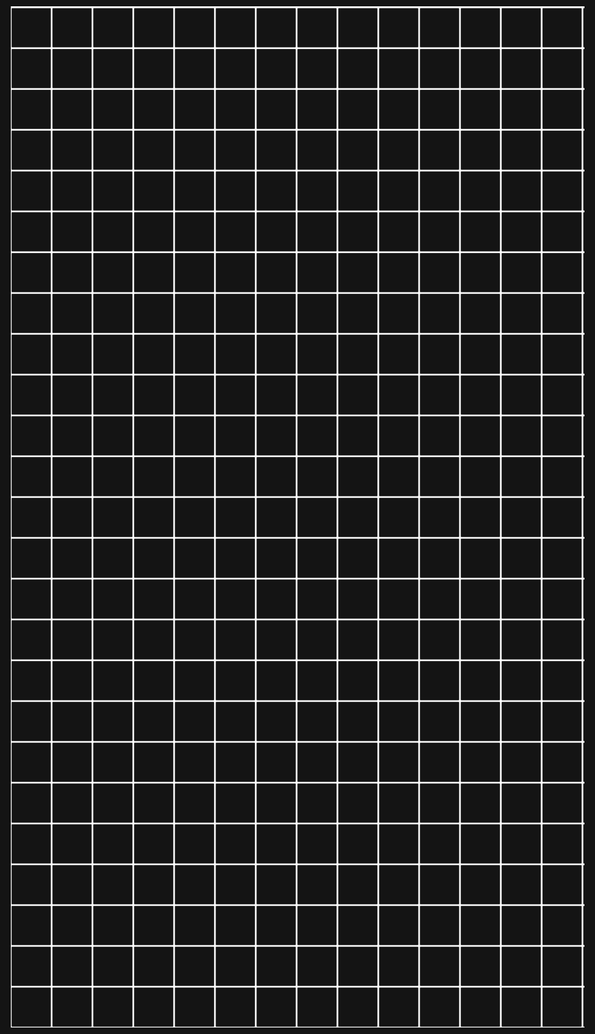
Случайным образом будем выбирать выборку из всевозможных выборок, таких, что соотношение в них возрастных групп и полов будет таким же как и в генеральной совокупности.



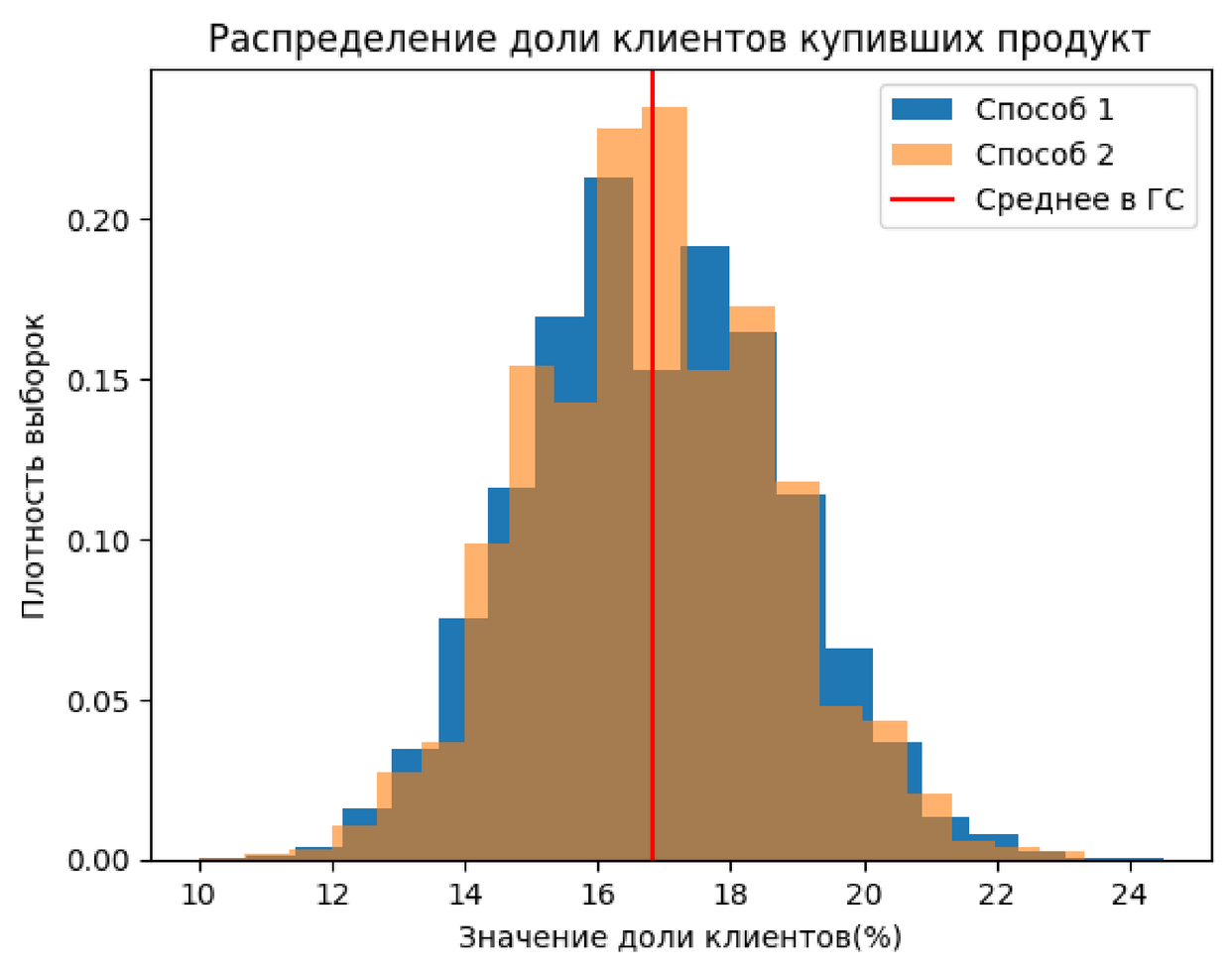
Результат 2 способа

В данном способе систематическое завышение/занижение отсутствует, но доля выборок с большой ошибкой сохранилась (примерно 10%).

ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ



Чёрный уровень



Сравнение результатов

Видим, что второй метод получения выборок оказывается точнее первого.

Сравнение результатов

Почему второй метод точнее?

Потому, что нем мы берем выборки максимально схожие с генеральной совокупностью по соотношению категорий наблюдений. Поэтому мы охватываем все категории из генеральной совокупности и, т.к. наблюдения в каждой категории могут иметь свои тенденции, получаем репрезентативные выборки.

Почему нельзя взять по одной выборке, даваемой каждым методом и сравнить их?

Потому, что нам не важно дает ли метод только «хорошие» или только «плохие» выборки, нам важно как часто и как сильно он ошибается, поэтому необходимо оценивать работу метода на как можно большем количестве выборок.

ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

Ссылка на расчеты в Google Colab