Статистика. Семинар №1.

Трубицын Александр.

**Вопрос 0.**

Что такое генеральная совокупность, что такое выборка? Приведите примеры одного и второго. Может ли одно быть другим и в каких случаях?

У вас есть 2 таблицы с данными ([Тинькоф](https://docs.google.com/spreadsheets/d/1SEjFNHPU6j3i1idmF7MWCozGRiKharWs/edit#gid=1433400990) и [Альфа](https://docs.google.com/spreadsheets/d/1L4R_JhetPrqyTJv2PkHjs1G4k8vPoAlo/edit#gid=897316620)). Выберите ту, которая относится к БД, с которой вы работаете и откройте.

**Вопрос 1.**

Какая из переменных к какому типу относится?

**Вопрос 2.**

Какие меры центра (мода, медиана, среднее) и разброса (размах, межквартильный размах, дисперсия, стандартное отклонение, коэффициент вариации) можно посчитать для этих переменных?

**Задание 1.**

Выберите 2 качественные (номинальную и порядковую) и 1 количественную переменные. Рассчитайте подходящие меры центра и меры разброса для каждой из них. Проинтерпретируйте полученные результаты. (Для положительных переменных сравните с 0 значение меры центра в терминах меры вариации).

**Задание 2.**

Возьмите 2 переменные из БД. Визуализируйте одну в виде гистограммы, а другую в виде таблицы частот. Проинтерпретируйте результаты с точки зрения симметричности и равномерности распределения.

**Задание 3.**

Возьмите переменную, по которой выборку можно разделить на подгруппы (например, по переменной доход можно разделить индивидов на подгруппы “бедные”, “богатые” и “средние”).

a) Посчитайте по другой количественной переменной меру центра, меру разброса и нарисуйте гистограмму для каждой из подгрупп и по выборке в целом. Сравните получившиеся результаты в подгруппах и по выборке в целом.

b)\* Сравните распределение частот по качественной переменной. Сравните получившиеся результаты в подгруппах и по выборке в целом.

**Задание 4.**

1. Определите по 2 признака (переменных) в вашем наборе данных (подходящих к вашему кейсу), которые являются качественными (2 шт), количественными (2 шт).
2. Определите, какие из качественных переменных являются номинальными, а какие порядковыми

Запишите ваши ответы в таблицу

| Качественные | | | | |
| --- | --- | --- | --- | --- |
| Номинальные |  |  |  |  |
| Порядковые |  |  |  |  |
| Количественные | | | | |
|  |  |  |  |  |

1. Какие из этих переменных стоит использовать для исследования выборки по категориям и для чего?
2. Постройте сводную таблицу на основе двух качественных переменных.Сколько подгрупп у вас получилось? Выберите интересующую вас количественную переменную. По ней сравните средние и вариации между подгруппами.
3. Для описания типичного представителя вашей выборки отметьте признаки, которые будете описывать. Какие меры центра и разброса будете использовать для каждой переменной?
4. Вычислите показатели центра для выбранных переменных.
5. Вычислите показатели разброса для выбранных переменных.
6. Какую из переменных имеет смысл перекодировать в другую шкалу? Какую новую переменную имеет смысл составить (Например, количественную переменную возраст можно переделать в категориальную).

Подумайте, каким образом можно перекодировать эти переменные.

* 1. Возьмите одну непрерывную переменную в вашем наборе данных и сгенерируйте на ее основе дискретную переменную. Приведите пример, когда это нужно.
  2. Возьмите одну количественную переменную в вашем наборе данных и сгенерируйте качественную переменную на ее основе. Приведите пример, когда это нужно.
  3. Для каждой подгруппы посчитать меры центра, меры разброса.
  4. Сформулируйте выводы, которые вам удалось сделать из анализа.