

Лекция 4. Средние величины

К.и.н., доцент кафедры Истории РБ,
археологии и этнологии
Р.Р.Газизов

- Категория средней величины имеет одну из самых древних историй. Теоретическое осмысление средних можно найти в трудах античных философов. В произведениях Аристотеля, Гераклита, Архимеда, Пифагора и других содержится понимание средней как равнодействующей всех определенных условий, которые участвуют в образовании рассматриваемой совокупности индивидуальных величин.
- В.Петти (1623-1687 гг.)
- А.Кетле (1796-1874 гг.)

- Главное значение средних состоит в их обобщающей функции, т.е. в замене множества различных индивидуальных значений признака средней величиной, характеризующей всю совокупность явлений. Средняя отражает совокупный результат развития и является равнодействующей различных причин и сил, воздействующих на эти явления.

Использование средних величин предполагает следование определенным правилам.

- 1. До вычисления средних необходимо обеспечить качественную однородность совокупности.
- 2. Средние величины вычисляются по массовым данным, т.е. по данным достаточно большого числа единиц наблюдения.
- 3. Нельзя ограничиваться вычислением средней в целом по совокупности, не меньшее значение имеют средние характеристики и для каждого отдельного типа.

Средняя арифметическая (\bar{X})

- - является самым распространенным видом средних величин. Она исчисляется путем отношения суммы всех значений признака к общему числу наблюдений.
- $$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n},$$
- где $X_1, X_2, X_3 \dots X_n$ – варианты признака;
- n – число единиц наблюдения.

Пример 1. Даны сведения о заработной плате шести работников (в условных единицах) – 90, 120, 108, 206, 160, 184. Определить средний размер заработной платы данной совокупности работников.

$$\bullet \bar{X} = \frac{90+120+108+206+160+184}{6} = 144,67$$

- Если значения изучаемого признака в совокупности не повторяются, то любое значение этого признака оказывает на величину \bar{X} одинаковое влияние, т.е. имеет одинаковый «вес».

- Если изучаемая совокупность велика, исходная информация чаще всего представлена группировкой, где значения усредняемого признака встречаются по несколько раз и частота их различна. Это значит, что любая варианта этого признака оказывает неодинаковое влияние на среднюю величину, которая должна представлять собой результат равномерного распределения значений признака.

- Для уравнивания указанных влияний используют средневзвешенную величину, равную сумме произведений каждого значения признака на его частоту, деленной на сумму всех частот.
- $\bar{X} = \frac{\sum xp}{\sum p}$, где
- X – варианты признака;
- P – частота вариант (в статистической литературе «f»)

Пример 2. Распределение футбольных матчей высшей лиги России по числу забитых мячей за игру в 1992 г.

Число забитых мячей	0	1	2	3	4	5	6	7
Число матчей	21	41	42	37	19	10	6	3

Определить среднее число забитых голов за одну игру.

$$\frac{0 * 21 + 1 * 41 + 2 * 42 + 3 * 37 + 4 * 19 + 5 * 10 + 6 * 6 + 7 * 3}{21 + 41 + 42 + 37 + 19 + 10 + 6 + 3} \approx 2,34$$

- Если в группировке значения усредняемого признака заданы интервальным рядом, то при исчислении средней арифметической в качестве значения признака берутся середины интервалов. Условно предполагается, что единицы совокупности распределены равномерно по интервалу.
- Для открытых интервалов значения признака определяются экспертным путем, качественным анализом, исходя из сущности и свойств природы признака. Можно также использовать формальный способ прибавления единицы к максимальному определенному значению и вычитания единицы из минимального заданного значения признака.

Пример 3. Распределение рабочих N-ского предприятия по возрасту.

Возраст	До 20	20-30	30-40	40-50	Старше 50
Число рабочих	48	120	75	62	54

$$\frac{18,5 * 48 + 25 * 120 + 35 * 75 + 45 * 62 + 57,5 * 54}{48 + 120 + 75 + 62 + 54} \approx 34,56$$

$$\frac{19 * 48 + 25 * 120 + 35 * 75 + 45 * 62 + 51 * 54}{48 + 120 + 75 + 62 + 54} \approx 33,65$$

Пример 4.

- Известно, что с площади 145 десятин собран урожай в 2595,5 т какой-то продукции. Отношен $\frac{2595,5}{145} = 17,9$ тонн (средняя урожайность данной культуры с одной десятины). Этот вид средней называется в статистике **неявной формой средней**.

Мода (Mo).

- - представляет наиболее часто встречающееся значение признака в упорядоченной совокупности, наиболее типичное среднее. В дискретном ряду Mo определяется без вычислений как значение признака с наибольшей частотой. (См. пример 2.)
- Если в вариационном ряду (в группировке) равная максимальная частота встречается у двух или нескольких значений признака, то он считается бимодальным или мультимодальным. Это говорит о неоднородности совокупности и, следовательно, надо проверить правильно ли составлена группировка.

- Для вычисления моды в интервальном ряду сначала определяется класс, т.е. интервал с наибольшей частотой. Затем МО вычисляется по формуле $X_0 + K \frac{P_2 - P_1}{2P_2 - P_1 - P_3}$, где

- X_0 – нижняя граница модально интервала;
- K – величина интервала;
- P_1 – частота интервала предшествующая модальному;
- P_2 - частота модально интервала;
- P_3 - частота интервала, последующего за модальным.

Вычислим МО по данным примера 3.

$$20 + 10 \frac{120 - 48}{2 * 120 - 48 - 75} = 26,15$$

- Получается, что наиболее типичный возраст рассматриваемой группы рабочих – 26, 15 лет. Этот возраст наиболее часто встречается в данной группе рабочих.

Приближенное значение моды можно определить по графику. Для этого надо построить гистограмму распределения.

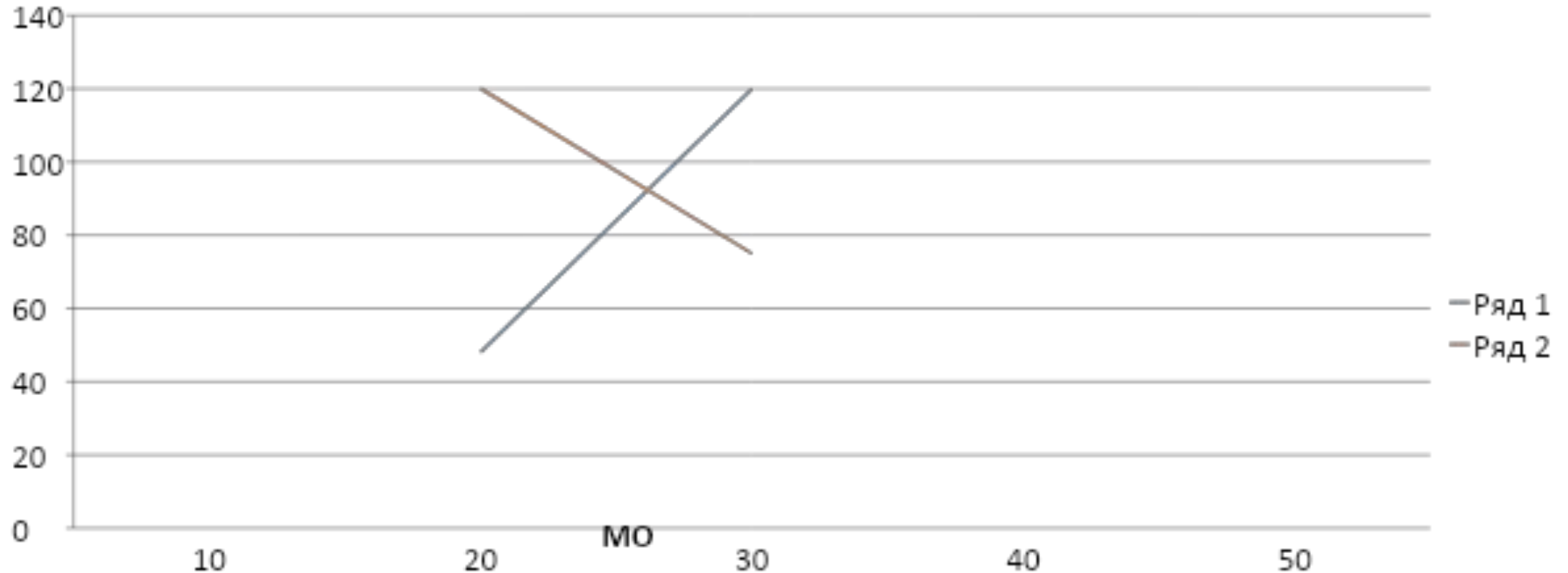


Рис.1. Гистограмма распределения рабочих N-ского предприятия по возрасту.

- Графическое определение моды применяется во всех случаях, когда в задачу исследования не входит обязательно получение точного значения наиболее распространенной величины признака. Например, для проверки рабочей гипотезы, когда точная величина принципиальной роли не играет, или для повышения наглядности материала. По нескольким графикам можно провести приблизительное сравнение мод различных признаков, чего невозможно сделать по таблицам.

Медиана (Me)

- - величина, определяющая значение признака, находящегося в середине упорядоченной совокупности. Медиана делит изучаемую совокупность так, что число единиц с большим и меньшим, чем медиана значением признака, одинаково.

- Чтобы определить M_e в дискретном ряду, надо построить ряд накопленных частот, затем поделить сумму всех частот пополам, а затем по накопленным частотам определить величину варианты, соответствующей той группе, в которой накопленная частота впервые превышает половину общей численности совокупности. В примере 2 ряд накопленных частот будет выглядеть так: 21, 62, 104, 141, 160, 170, 176, 179. Полусумма всех этих частот равна $179/2=89,5$. Эта величина входит в третью группу из накопленных частот, т.е. в данном примере третья из накопленных частот своей величиной превысила значение полусуммы всех частот. Следовательно медиана равна 2.

- В интервальной группировке для вычисления Me необходимо найти медианный интервал – интервал, которому соответствует первая из накопленных частот, превышающая половину суммы всех частот ряда распределения. Затем считают по формуле:

$$Me = x_0 + k \frac{\sum p/2 - \sum_{m-1}}{p_m}$$

- X_0 – нижняя граница медианного интервала;
- K – величина медианного интервала;
- \sum_{m-1} – сумма частот интервалов, предшествующих медианному;
- P_m – частота медианного интервала;
- $\sum P$ – сумма частот. P_m

- Определим Me по данным примера 3. Ряд накопленных частот принимает следующий вид: 48; 168; 243; 305; 359. Полусумма частот равна $359/2=179,5$. Полученные данные говорят о том, что медианным является третий интервал, т.е.

$$Me = 30 + 10 \frac{179,5 - 168}{75} \approx 31,53$$

- Величина Me свидетельствует, что половина рабочих рассматриваемой группы имеет средний возраст 31,53 года.

Примерное значение медианы можно определить по графику.

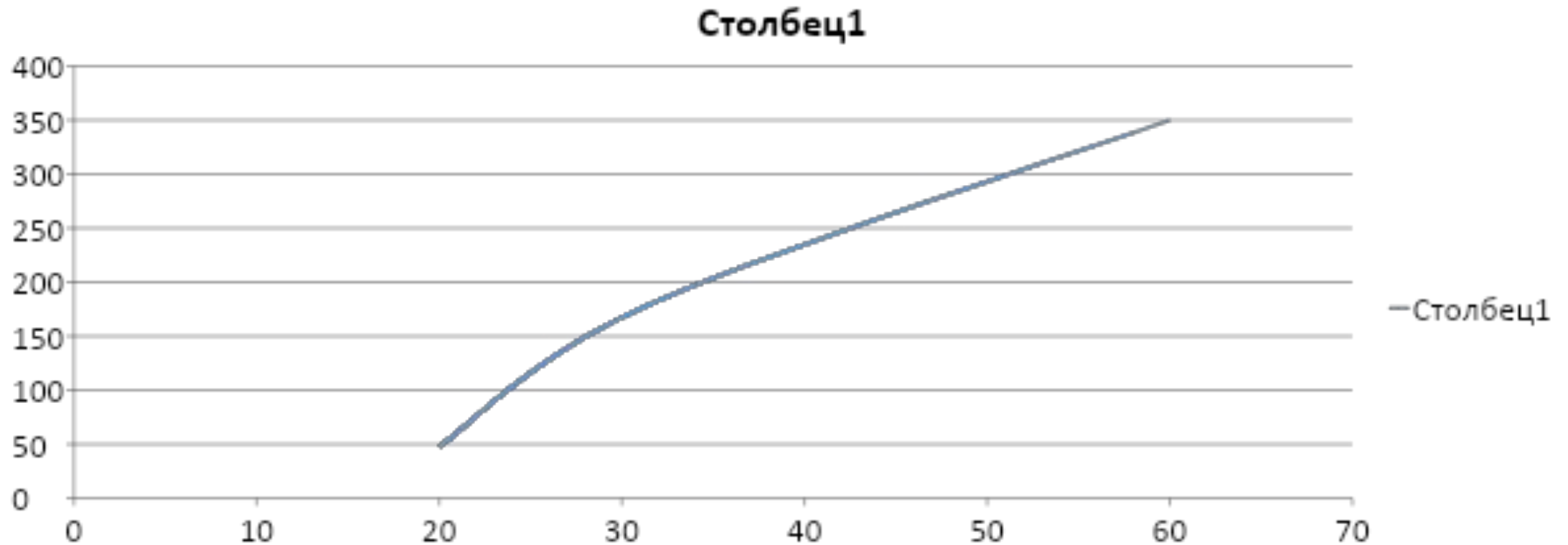


Рис.2. Кумулята распределения рабочих N-ского предприятия по возрасту.

- Обобщая три средние величины, рассчитанные по одним и тем же данным, видим существующую разницу. Средний возраст условной группы рабочих – 33,6 лет, наиболее распространенный, часто встречающийся средний возраст (M_o) – 26,2 лет, при этом половина рассматриваемой группы имеет средний возраст (M_e) – 31,5 лет. Какой величине следует отдать предпочтение? Какой показатель считать наиболее достоверным и точным?

При решении этих вопросов надо помнить:

- 1. Мода (M_o) имеет значение в том случае, когда её величина расходится и с медианой (M_e), и со средней арифметической (\bar{x}), им не следует пренебрегать. Это же можно сказать и о медиане. Так что для исследования полезно вычислять все три показателя.
- 2. Различие в значении величин обусловлено ассиметрическим распределением. Средняя арифметическая подвержена влиянию каждой варианты, поэтому она смещается в направлении наибольших значений признака. На моду (максимальные и минимальные) варианты влияния не оказывают. Медиана зависит только от числа вариантов, а не от их величины.

- 3. Медиана по своей математико-статистической природе является самой представительной средней. При больших колебаниях в значениях признаков или когда не определены крайние интервалы в группировках, лучше пользоваться медианой. При вычислении моды для интервальной группировки желательно, чтобы интервалы были равновеликими.
- 4. Мода чаще других величин применяется по отношению к качественным признакам. Если скопление частот возле моды составляет 10-15 % их общего числа, особое значение приобретает медиана, представляя более достоверное значение среднего показателя.

- Спасибо за внимание!

- 1. Какие условия необходимы для вероятностного описания событий? (2)
- 2. Что такое «закон больших чисел»? (1)
- 3. Дать определение термину «вариация». (1)
- 4. Какие правила необходимо соблюдать при составлении таблиц? (до 3-х б.)
- Какие группировки существуют в науке? (1) Раскройте их. (до 3-х б.)