

План лекции 3

«Меры центральной тенденции»

1. Среднее арифметическое и его свойства
2. Среднее гармоническое
3. Среднее квадратическое
4. Среднее кубическое
5. Медиана
6. Мода
7. Интерпретация моды, медианы и среднего
8. Выбор меры центральной тенденции

Зачем нужны меры центральной тенденции?

Это наиболее важная статистика больших массивов информации (статистика – это любая функция данных).

Средние значения обладают большей устойчивостью.

Средние значения – это наиболее репрезентативные значения.

Если нужно заменить весь массив одним числом – то нужно использовать среднее значение.

Разные виды средних обладают разными свойствами. Выбор вида среднего выбирается в каждой конкретной ситуации.

Меры центральной тенденции

1. Среднее арифметическое
2. Среднее гармоническое
3. Среднее квадратическое
4. Среднее кубическое
5. Среднее геометрическое
6. Мода
7. Медиана

Виды средних

Автомобиль движется из пункта А в пункт Б с постоянной скоростью 80 км/час, а из пункта Б в пункт А с постоянной скоростью 40 км/час.

Определить среднюю скорость движения автомобиля.

$$V = \frac{s_{\text{общ}}}{t_{\text{общ}}} = \frac{2s}{\frac{s}{v_1} + \frac{s}{v_2}} = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}} = \frac{2}{\frac{1}{80} + \frac{1}{40}} = 55,3$$

Виды средних

Диаметр одной корзины подсолнуха равен 10 см, диаметр другой корзины подсолнуха равен 30 см.

Определить средний диаметр корзин подсолнуха.

$$S_{\text{сум}} = 2\pi \left(\frac{\bar{d}}{2} \right)^2 = \pi \left(\frac{d_1}{2} \right)^2 + \pi \left(\frac{d_2}{2} \right)^2$$

$$\bar{d} = \sqrt{\frac{d_1^2 + d_2^2}{2}} = \sqrt{\frac{100 + 900}{2}} \approx 22,4$$

Виды средних

Диаметр одного яйца равен 5 см, диаметр другого яйца равен 3 см.

Определить средний диаметр яиц.

$$V_{\text{общ}} = \frac{4}{3} \pi \left(\frac{\bar{d}}{2} \right)^3 = \frac{4}{3} \pi \left(\frac{d_1}{2} \right)^3 + \frac{4}{3} \pi \left(\frac{d_2}{2} \right)^3$$

$$\bar{d} = \sqrt[3]{\frac{d_1^3 + d_2^3}{2}} = \sqrt[3]{\frac{125 + 27}{2}} \approx 4,24$$

Используемые обозначения

Точка (.) вместо индекса обозначает суммирование по этому индексу

$$x. = \sum_{i=1}^6 x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

Черточка над переменной ($\bar{x}.$) обозначает усреднение по индексам, по которым проводилось суммирование

$$\bar{x}. = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{N}$$

Среднее арифметическое и его свойства

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

1. Если каждое значение совокупности уменьшить или увеличить на одно и то же число, то среднее ?
2. Если каждое значение совокупности умножить или разделить на одно и то же число, то среднее ?
3. Среднее двух совокупностей является взвешенным средним этих совокупностей ?
4. Сумма отклонений значений совокупности от ее среднего равно ?
5. Сумма квадратов отклонений от их средней меньше суммы квадратов отклонений тех же значений от любой другой величины.

Сумма квадратов отклонений от их средней меньше суммы квадратов отклонений тех же значений от любой другой точки

$$\sum_{i=1}^n \left[x_i - \left(\bar{x} + c \right) \right]^2 = \sum_{i=1}^n \left[\left(x_i - \bar{x} \right) - c \right]^2 =$$
$$\sum_{i=1}^n \left(x_i - \bar{x} \right)^2 - 2c \sum_{i=1}^n \left(x_i - \bar{x} \right) + nc^2$$

Откуда

$$\sum_{i=1}^n \left(x_i - \bar{x} \right)^2 \leq \sum_{i=1}^n \left[x_i - \left(\bar{x} + c \right) \right]^2$$

Среднее, мода и медиана объединенных групп

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3}{n_1 + n_2 + n_3}$$

Для того, чтобы найти объединенное среднее, необходимо знать число элементов в подгруппах.

Для того, чтобы найти объединенную моду, необходимо знать какие элементы встречаются наиболее часто во всех подгруппах.

Для того, чтобы найти объединенную медиану, необходимо знать распределение всех подгруппах.

Структурные средние

Мода – это то значение, которое в выборке встречается наиболее часто.

Медиана – это то значение, относительно которого упорядоченная по возрастанию или по убыванию выборка делится пополам.

Как считать доход на душу населения? (как среднее или как медиану?)

Мода

Мода – это наиболее частое значение, а не частота этого значения.

1. Если все значения встречаются в массиве одинаково часто, то массив не имеет моды.

2. Если два соседних значения имеют одинаковую частоту и они больше частоты любого другого значения, то мода есть среднее этих двух значений

3. Если два несмежных значения в массиве имеют равные частоты и они больше частоты любого значения, то массив является бимодальным

Свойства моды

1. Мода вычисляется наиболее просто – ее можно определить на глаз.
2. Для очень больших массивов данных это достаточно стабильная мера центра распределения.
3. Во многих задачах мода близка к двум другим мерам – медиане и среднему.

Вычислить меры центральной тенденции

Диаметры корзинок подсолнухов:
15, 13, 11, 16, 8, 13, 15, 16, 17, 15

Вычислить

$M_0 =$

$M_e =$

$\bar{x} =$

Интерпретация моды, медианы и среднего

Интерпретация осуществляется в терминах ошибок, возникающих из-за того, что все значения в выборке заменяются одним значением (наиболее репрезентативным)

Мода – наиболее репрезентативное значение в том смысле, что совпадает с наибольшим числом значений в выборке.

Медиана – это такая точка на числовой оси, для которой сумма абсолютных разностей всех значений меньше суммы разностей для любой другой точки.

Среднее – обеспечивает минимальное значение суммы квадратов отклонений значений от среднего.

Критерии выбора меры центральной тенденции

1. В малых группах мода очень нестабильна (1,1,1,3,5,7,7,8) $M_o = 1$. Но если $1 \square 0$ и $1 \square 2$, то $M_o = 7$.

2. На медиану не влияют большие и малые (экстремальные) значения

3. На величину среднего влияет каждое значение. (Как?)

Для каких массивов среднее, мода и медиана совпадают?

Задача 1. Где строить дом?

	п.1		п.2			п.3	п.4	п.5
0	1	2	3	4	5	6	7	8

Задача 2. Какую меру центральной тенденции выбрать?

Доходы 5 мужчин:

1. 25 центов
2. 25 центов
3. 2 000 долларов
4. 15 000 долларов
5. 5 000 000 долларов

Как охарактеризовать их средний доход?

В США средний доход – это медиана, а не среднее

Литература

1. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2004, 479 с.
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 2004, 400 с.
3. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. Пер. с англ. – М.: Издательство «Прогресс», 1976. -496 с.
4. Маслак А.А. Основы планирования и анализа сравнительного эксперимента в педагогике и психологии. – Курск: РОСИ, 1998. – 167 с.