

## ЭТАПЫ ОПИСАНИЯ (ОБОБЩЕНИЯ) КОЛИЧЕСТВЕННОГО ПРИЗНАКА

Описание (обобщение) количественного признака проводится по следующим этапам:

1. Определение вида распределения признака.
2. Оценка центральной тенденции изучаемой совокупности..
3. Оценка ее разнообразия (разброса).

Анализ совокупности начинается с установления вида распределения изучаемого признака. Для этого полученные данные представляются в виде вариационного ряда, изображаются графически и делаются соответствующие расчеты.

В случае распределения близкого к нормальному мы вправе для дальнейшего статистического анализа применять параметрическую статистику, если распределение отлично от нормального или при неизвестном распределении рекомендуется применять непараметрическую статистику.

Непараметрические методы:

- -не требуют предварительного знания вида распределения;
- -не требуют предварительного расчета параметров распределения(средних величин, стандартного отклонения и др.);
- -позволяют сравнивать совокупности с номинальными и порядковыми признаками;
- -просты в применении.

Отрицательные стороны непараметрических методов:

- -обладают меньшей мощностью, чем параметрические;
- -имеют существенные ограничения в применении по числу наблюдений.

**Вариационный ряд (frequency table)**- ранжированный ряд распределения по величине какого-либо признака. Этот признак носит название варьирующего, а его отдельные числовые значения называются вариантами и обозначаются через "х". Число, показывающее, сколько раз данная варианта встречается в вариационном ряду, называется частотой и обозначается через "р"

Вариационный ряд можно разбивать на отдельные (по возможности равные) части, которые называются **квантилями (quantile)**. Наиболее часто употребляемые квантили представлены в таблице.

Название квантилей	Число частей, на которые разбивается ряд
Медиана	2
Терциль	3
Квартиль	4
Дециль	10
Процентиль	100

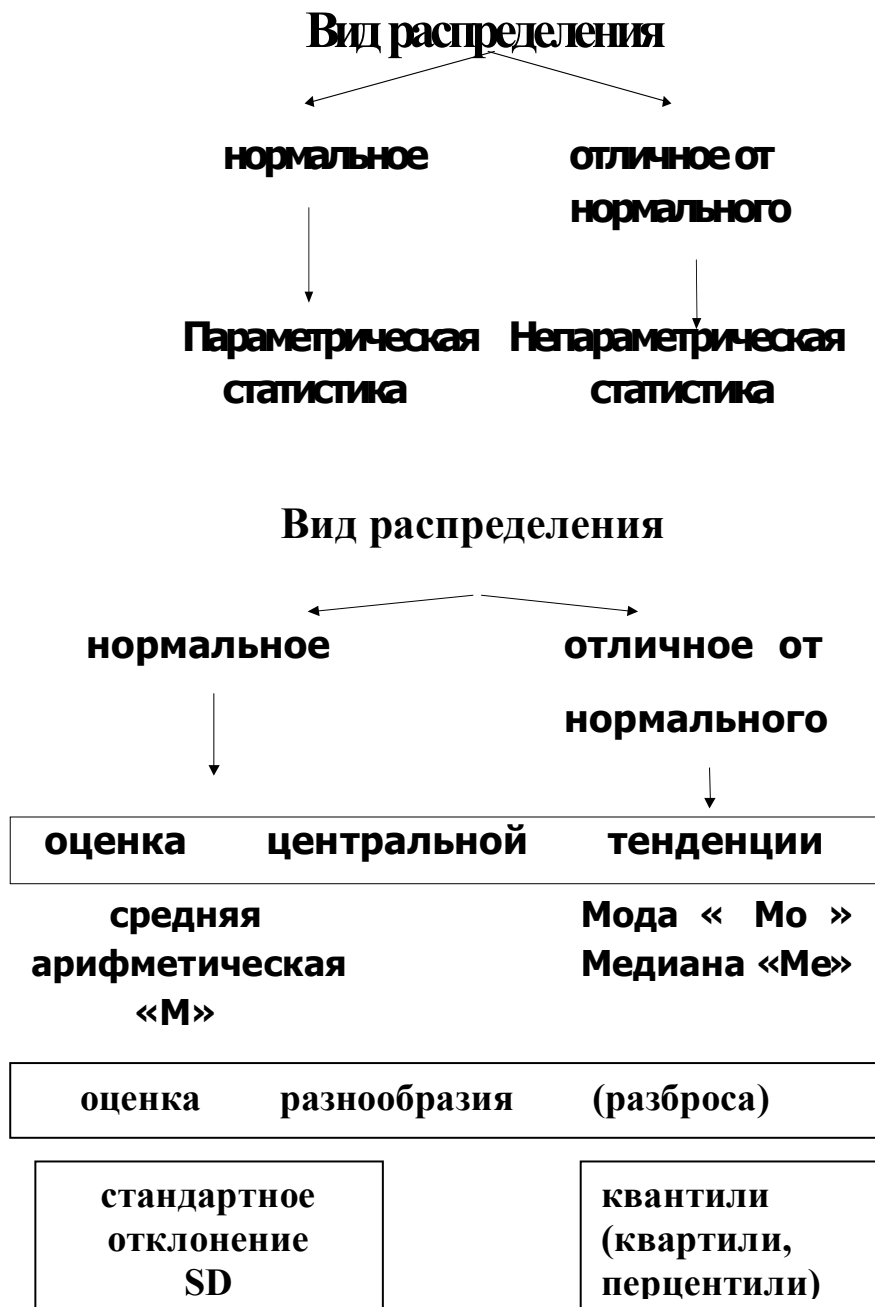
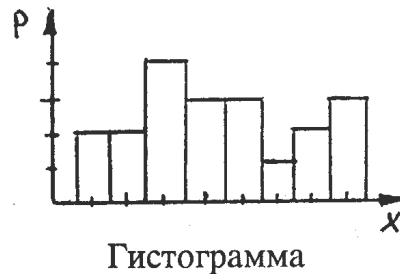


Рис. Этапы описания (обобщения) количественного признака

### Виды вариационных рядов:

1. В зависимости от вида случайной величины :
  - дискретный;
  - непрерывный .
2. В зависимости от группировки вариант:
  - несгруппированный;
  - сгруппированный (интервальный):
3. В зависимости от частоты, с которой каждая варианта встречается в вариационном ряду:
  - простой ( $p = 1$ );
  - взвешенный ( $p > 1$ ).

### Графическое изображение вариационных рядов



где  
 $x$  - варианты;  
 $p$  - частоты.

### ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ВАРИАЦИОННОГО РЯДА:

- 1). Показатели, характеризующие **центральную тенденцию (central tendency)** или уровень ряда: средние величины или меры расположения (собственно средние и структурные средние).
- 2). Показатели, характеризующие **разнообразие (рассеяние, вариацию, разброс) (spread)** признака: стандартное отклонение, дисперсия, размах.

Выбор характеристик центральной тенденции и разнообразия признака прежде всего зависит от вида распределения. В случае нормального распределения используют показатели параметрической статистики, в случае распределения, отличного от нормального и при неизвестном виде распределения применяют показатели непараметрической статистики.

### Средние величины

Средняя величина - обобщающий коэффициент, который характеризует наиболее типичный размер определенного признака в целом для совокупности или для отдельных ее частей. Расчет средних величин имеет смысл только для качественно однородной совокупности, в связи с этим в одной совокупности может быть столько средних, на сколько однородных групп она может быть разбита.

### Виды средних величин

**Средняя арифметическая (mean)** - применяется, если варианты возрастают (убывают) в арифметической прогрессии.

$$\bar{x} = \frac{\sum x_i * p_i}{n}$$

$\bar{x}$  - средняя арифметическая;

$x_i$  - варианты;

$p$  - частота встречаемости варианты;

$n$  - число наблюдений

### Свойства средней арифметической

- носит обобщающий характер;
- имеет абстрактное значение;
- алгебраическая сумма отклонений отдельных вариантов от средней равна 0: (сущность средней и способ проверки правильности расчета средней);
- сумма квадратов отклонений отдельных вариантов от средней меньше суммы квадратов отклонений вариант от любой другой величины, неравной средней;
- сумма произведений отдельных вариантов на свои частоты равна произведению средней на число наблюдений (единство суммарного действия и способ проверки правильности расчета средней);
- если каждую из вариантов увеличить или уменьшить на определенное число (в определенное число раз), то средняя арифметическая увеличится или уменьшится на столько же (во столько же);
- если частоту всех вариантов пропорционально изменить, то средняя арифметическая от этого не изменится.

**Средняя геометрическая** - вычисляется, если варианты возрастают (убывают) в геометрической прогрессии.

$$\bar{x}_g = \sqrt[n]{x_1^{p_1} * x_2^{p_2} * ... * x_i^{p_i}} = \sqrt[n]{\prod x_i^{p_i}}$$

На практике используют логарифмированную формулу:

$$\log \bar{x}_g = 1/n * (\log x_1 p_1 + \log x_2 p_2 + \log x_3 p_3 + ... + \log x_i p_i)$$

## Структурные средние

**Мода (Mo) (mode)**- наиболее часто встречающаяся в вариационном ряду варианта.

Мода используется:

- при малом числе наблюдений, когда велико влияние состава совокупности на среднюю ;
- для характеристики центральной тенденции при ассиметричных распределениях, когда велико влияние на среднюю крайних вариантов;
- 

**Медиана (Me)(median)** - варианта, которая делит вариационный ряд на две равные части.

Медиана используется:

- при необходимости знать, какая часть вариантов лежит выше и ниже среднего значения ;
- для характеристики центральной тенденции при ассиметричных распределениях .

## Характеристики разнообразия вариационного ряда

1. Размах вариации (амплитуда) (range):  $A = X_{\max} - X_{\min}$

2. Стандартное отклонение (среднее квадратическое отклонение) (standard deviation, SD)

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot p}{n}}$$

- приблизительный расчет стандартного отклонения по амплитуде:

$$\sigma = \frac{X_{\max} - X_{\min}}{K}$$

где K - коэффициент Ермолаева, рассчитывается по специальной таблице с учетом числа наблюдений.

При числе наблюдений больше 30:

$$\sigma = \frac{X_{\max} - X_{\min}}{6}$$

Стандартное отклонение наиболее часто используется при определении нормы и патологии, в основе которого лежит "правило трех сигм", справедливое только для нормального распределения.

### "Правило трех сигм"

68.3 % всех вариантов отклоняются от своей средней не более, чем на  $\sigma$

95.4% вариантов находятся в пределах  $X \pm 2\sigma$

99.7% вариантов находятся в пределах  $X \pm 3\sigma$

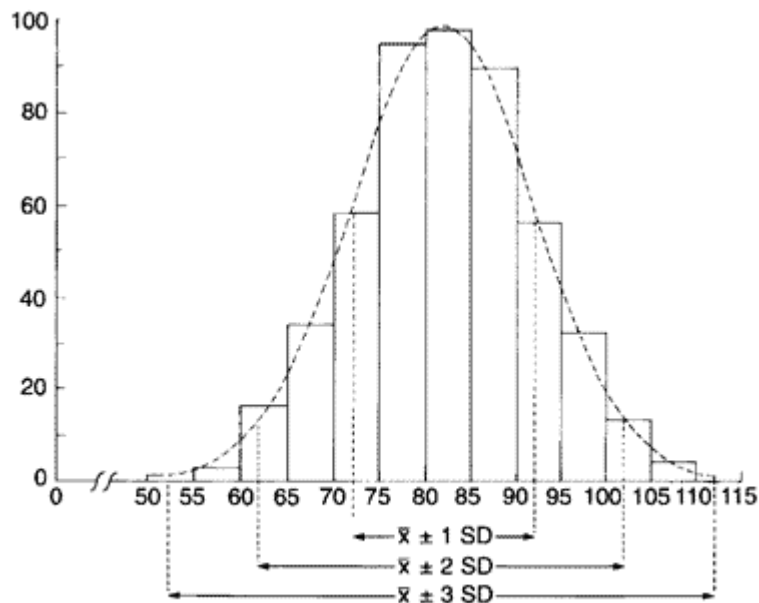


Рис. Правило «трех сигм» ( SD – стандартное отклонение).

### 3. Дисперсия (варианса) (variance)

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 \cdot p}{n}$$

При распределении Пуассона дисперсия равна средней:  $\sigma^2 = x$  . \_\_\_\_

### 4. Коэффициент вариации (variation coefficient):

$$C_v = \frac{\sigma}{\bar{x}} \cdot 100\%$$

Вариационный ряд считается однородным при  $C_v < 10\%$  , обладающим средней вариабельностью (разнообразием) при  $C_v = 10-15\%$  и обладающим значительной вариабельностью при  $C_v > 15\%$  .

Коэффициент вариации используется при сравнении вариационных рядов, имеющих различную размерность, или одной размерности, но обладающими резкими различиями в своих значениях, затрудняющими их сопоставление.

### 5. Интерквартильный интервал (inter-quartile range, IQR)

Вариационный ряд разбивают на четыре интервала, получая, соответственно, 25%, 50% и 75% квантили; 25% и 75% квантили называют также нижним (low quartile) и верхним квартилями (high quartile). 50% квантиль – это медиана. Внутри интерквартильного интервала (между 25% и 75% квантилями) лежат 50% наиболее типичных (близких к центральному) значений.

Таким образом, в случае нормального распределения вариационный ряд описывается средней величиной и стандартным отклонением, если распределение неизвестно или оно отлично от нормального центральную тенденцию и разброс можно описать с помощью медианы, нижнего и верхнего квартиля (интерквартильным интервалом).

Соответствие экспериментального распределения нормальному проверяется следующими способами:

1. По числам Вестергарда при нормальном распределении в пределах:  
 $x \pm 0.3 \sigma$  находится 25 % всех единиц наблюдения;  
 $x \pm 0.7 \sigma$  находится 50 % всех единиц наблюдения;  
 $x \pm 1.1 \sigma$  находится 75 % всех единиц наблюдения;  
 $x \pm 3.0 \sigma$  находится 99 % всех единиц наблюдения.
2. По соотношению средней арифметической и структурных средних:  
 - при нормальном распределении, которое обладает симметричностью:

$$\bar{x} = Me = Mo, \text{ или } \bar{x} = \frac{3Me - Mo}{2}$$

- правило "двух третей" Юла:

$$Mo = 3(Me - 2/3\bar{x})$$

- а). если распределение симметрично:  $Me = Mo$ ;
- б). если распределение обладает правосторонней асимметрией:  $Me > Mo$ ;
- в). если распределение имеет левостороннюю асимметрию  $Me < Mo$

3. По коэффициенту асимметрии (skewness):

$$As = \frac{x - Mo}{\sigma}; \quad -3 \leq As \leq +3$$

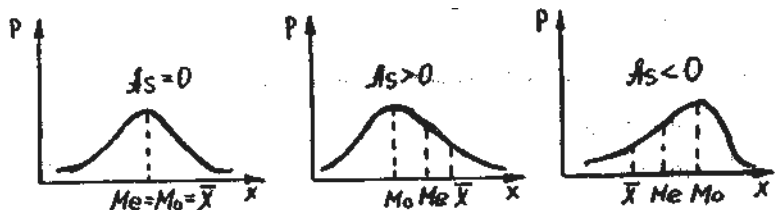
а), если распределение симметрично:  $As = 0$

б), при правосторонней асимметрии:  $As > 0$

в). при левосторонней асимметрии:  $As < 0$

4. Если  $Me$  занимает срединное положение между 25-м и 75-м процентилем, то распределение близко к нормальному.

**Графическое изображение симметричного и асимметричного распределений:**



Так как значительная часть статистических методов (параметрическая статистика) основана на предположении, что распределение близко к нормальному, то, если экспериментальные данные не ложатся на кривую нормального распределения, их пытаются преобразовать таким образом, чтобы полученная кривая соответствовала нормальному распределению. Наиболее часто используются следующие способы "нормализующего преобразования" (**transformation to normality**) данных :

- гармоническое преобразование:  $1/x$ ;
- извлечение квадратного корня:  $\sqrt{x_i}$
- логарифмирование {дает наиболее точное приближение}:  $\log x_i$

Успешность преобразования данных оценивают по коэффициенту асимметрии: чем ближе он к 0, тем ближе экспериментальное распределение к нормальному.

### Исключение "выскакивающих" вариант.

Иногда в небольших совокупностях встречаются варианты резко отличающиеся по своему значению от других, так называемая «**выскакивающая**» **варианта (outlying case)**. Если данное отличие обусловлено случайными колебаниями изучаемой величины, то такие варианты оставляют в совокупности и включают в общее число наблюдений. Если отличие обусловлено ошибками в исследовании или ее причину точно нельзя установить, то "выскакивающие" варианты необходимо исключить из исследования. Методика исключения вариант:

- рассчитываются средняя величина и стандартное отклонение без учета "выскакивающих" вариант;
- анализируется соотношение:
  - если  $X_{\text{выск}} - \bar{x} > \sigma * f$ , то "выскакивающая" варианта исключается из исследования;
  - если  $X_{\text{выск}} - \bar{x} < \sigma * f$  то "выскакивающая" варианта должна быть оставлена в общем числе наблюдений.

При этом  $f$  - коэффициент Романовского, который определяется по специальной таблице с учетом числа наблюдений и вероятностью исключения варианты.

### Задание для самостоятельной работы по средним величинам и мерам рассеяния:

1. Сформируйте вариационный ряд из предложенных данных.
2. Вычислите среднюю величину, стандартное отклонение, моду, медиану, коэффициент вариации (оцените его).
3. Можно ли считать, что предложенный для анализа признак имеет нормальное распределение?

### Задача – эталон

Приведены результаты измерения частоты пульса у некурящих студентов-медиков в возрасте 20 лет:

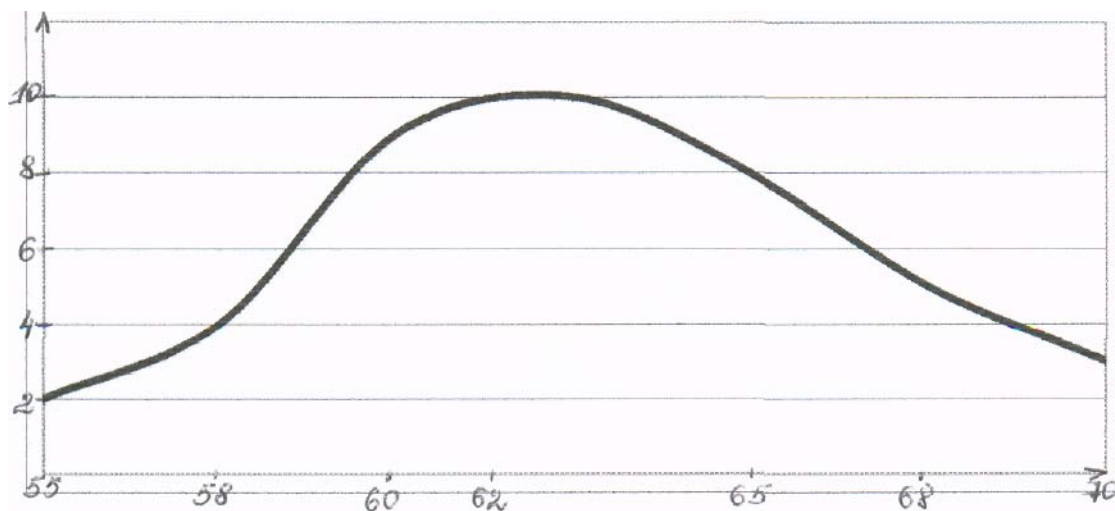
68,58,65,55,70,62,60,65,70,58,62,58,62,60,60,65,62,55,62,58,60,70,62,65,60,68,65,62,68,65,60,62,60,68,65,60,62,60,65,62,68

Построим вариационный ряд:

<b>x</b>	55	58	60	62	65	68	70
<b>n</b>	2	4	9	10	8	5	3



## Графическое изображение распределения результатов измерения частоты пульса у студентов-медиков



Средняя величина – средняя частота пульса у некурящих студентов-медиков

$$\bar{x} = \frac{\sum(x * p)}{n}$$

$$\bar{x} = \frac{2 \times 55 + 4 \times 58 + 8 \times 60 + 10 \times 62 + 8 \times 65 + 4 \times 68 + 2 \times 70}{41} = \frac{2572}{41} = 62,73 \text{ ударов в минуту}$$

$$M_0 = 62 \text{ удара в минуту}$$

$$M_e = 62 \text{ удара в минуту}$$

Стандартное отклонение

$$\delta = \sqrt{\frac{\sum(x_i - \bar{x})^2 * p}{n}} = \sqrt{\frac{711,29}{41}} = 4,2 \text{ уд. в минуту}$$

Коэффициент вариации

$$CV = \frac{\delta}{\bar{x}} \times 100 = \frac{4,2}{62,7} \times 100 = 6,7\%$$

Проверка нормальности распределения

$$g = \frac{\bar{x} - M_0}{\delta} = \frac{62,7 - 62,0}{4,2} = 0,16$$

**Вывод:** Так как коэффициент асимметрии близок к 0, можно предположить нормальное распределение. Тогда центральная тенденция оценивается по средней величине,  $\bar{x} = 62,7$  уд. в минуту (частота пульса у некурящих студентов-медиков), а разброс характеризует стандартное отклонение  $\delta = 4,2$  уд. в минуту. Так как коэффициент вариации  $CV < 10\%$  ( $CV = 6,7$ ), то изучаемую совокупность можно считать однородной.

### Контрольные вопросы:

1. Приведите примеры использования в здравоохранении абсолютных чисел для анализа явлений и процессов.
2. Почему при анализе результатов исследования возникает потребность в обобщении полученных данных и каковы основные виды этого обобщения?
3. Назовите виды относительных величин и приведите примеры их использования в анализе общественного здоровья.

4. Назовите виды средних величин, условия их использования и приведите примеры их практического применения.
5. Что такое вид распределения случайной величины и какие виды распределения знаете (примеры)?
6. Проведите сравнительный анализ параметрической и непараметрической статистики.
7. Для чего необходимо формирование вариационного ряда и каковы его правила?
8. Покажите на примерах виды вариационных рядов.
9. Расскажите об основных направлениях анализа вариационного ряда.
10. Какие характеристики относятся к показателям центральной тенденции ряда и каково их практическое применение?
11. Что такое «меры рассеяния» и какие показатели к ним относятся?
12. Опишите достоинства и недостатки критериев разнообразия вариационного ряда.
13. В чем заключается правило «трех сигм» на конкретных примерах здравоохранения?
14. Как проверить нормальность распределения?
15. В чем заключается правило исключения «выскакивающих» вариант?

## **Тесты**

(Выберите один или несколько правильных ответов)

**1. Модой называется:**

- А. Варианта с наибольшей частотой
- Б. Варианта с наименьшей частотой
- В. Варианта, находящаяся в середине ряда
- Г. Выскакивающая варианта

Ответ: А

**2. Медианой называется:**

- А. Варианта с наибольшей частотой
- Б. Варианта с наименьшей частотой
- В. Варианта, находящаяся в середине ряда
- Г. Выскакивающая варианта

Ответ: В

**3. Какая зависимость между степенью разнообразия вариационного ряда и значением среднего квадратического отклонения:**

- А. Прямая
- Б. Обратная

Ответ: А

**4. Коэффициент вариации применяется в целях:**

- А. Определения разности между наибольшей и наименьшей вариант
- Б. Определения частоты вариант в вариационном ряду
- В. Сравнения признаков, выраженных в разных единицах измерения

Ответ: В

**5. Средняя арифметическая величина применяется для:**

- А. Обобщения качественных признаков
- Б. Обобщения числовых значений варьирующего признака
- В. Выявления взаимосвязи между явлениями

Ответ: Б

**6. Из всех видов распределения в медико-биологических исследованиях наиболее часто встречается:**

- А. Биномиальные
- Б. Нормальное
- В. Пуассона
- Г. Альтернативное
- Д. Все вышеперечисленные встречаются с одинаковой частотой

Ответ: Б

**7. Основным условием применения параметрических методов анализа является:**

- А. Формирование случайной выборки
  - Б. Наличие двух независимых выборок
  - В. Корреляционная связь между признаками
  - Г. Невозможность применения непараметрических методов
  - Д. Нормальное распределение признака
- Ответ: Д

**8. Вариационный ряд состоит из:**

- А. Набора вариант
  - Б. Набора ошибок репрезентативности
  - В. Набора частот
  - Г. Набора отклонений
- Ответ: А, В

**9. Укажите виды вариационных рядов:**

- А. Непрерывный
  - Б. Частотный
  - В. Полный
  - Г. Прерывный (дискретный)
  - Д. Интервальный (сгруппированный)
- Ответ: А, Г, Д

**10. К показателям разнообразия вариационного ряда относятся**

- А. Размах (амплитуда)
  - Б. Мода
  - В. Медиана
  - Г. Среднее квадратическое отклонение
  - Д. Коэффициент вариации
- Ответ: А, Г, Д

**11. Укажите виды средних арифметических величин:**

- А. Простая
  - Б. Взвешенная
  - В. Алгебраическая
  - Г. По способу моментов
  - Д. Квадратическая
- Ответ: А, Б, Г

**12. Для графического изображения структурных показателей следует применять:**

- А. Столбиковые диаграммы
  - Б. Секторные диаграммы
  - В. Линейные графики
  - Г. Внутрестолбиковые диаграммы
  - Д. Диаграммы рассеивания
- Ответ: Б, Г

**13. Для графического изображения динамики изучаемого явления следует применять:**

- А. Линейные графики
  - Б. Радиальные графики
  - В. Секторные диаграммы
  - Г. Внутрестолбиковые диаграммы
  - Д. Все вышеперечисленное
- Ответ: А, Б

**14. Статистические таблицы:**

- А. Являются рациональной формой представления сводных количественных данных
  - Б. Должны иметь четкое и краткое заглавие, отражающее содержание статистического материала
  - В. Не требуют итоговых граф/строк
  - Г. Используются для группировки материалов статистического наблюдения
  - Д. Содержат только абсолютные величины
- Ответ: А, Б, Г

**15. К статистической таблице можно отнести:**

- А. Таблицу умножения

- Б. Таблицу, содержащую показатели заболеваемости населения
  - В. Таблицу «Периодическая система элементов Д.И. Менделеева»
  - Г. Таблицу, характеризующую численность населения по полу и возрасту
  - Д. Табличную форму анкеты
- Ответ: Б, Г

**16. Перцентилями называют значения изучаемого количественного признака:**

- А. Повторяющиеся в вариационном ряду с наибольшей частотой
  - Б. Делящие вариационный ряд на десять равных частей
  - В. Находящиеся в центре вариационного ряда
  - Г. Делящие вариационный ряд на сто равных частей
  - Д. Делящие вариационный ряд на четыре равновеликие части
- Ответ: Г

**17. Квартили – это значения изучаемого количественного признака:**

- А. Повторяющиеся в вариационном ряду с наибольшей частотой
  - Б. Делящие вариационный ряд на десять равных частей
  - В. Находящиеся в центре вариационного ряда
  - Г. Делящие вариационный ряд на сто равных частей
  - Д. Делящие вариационный ряд на четыре равновеликие части
- Ответ: Д

**18. Децили – это значения изучаемого количественного признака:**

- А. Повторяющиеся в вариационном ряду с наибольшей частотой
  - Б. Делящие вариационный ряд на десять равных частей
  - В. Находящиеся в центре вариационного ряда
  - Г. Делящие вариационный ряд на сто равных частей
  - Д. Делящие вариационный ряд на четыре равновеликие части
- Ответ: Б