

**ФГБОУ ВО Тюменский ГМУ Минздрава РФ  
Кафедра общественного здоровья и здравоохранения ИНПР**

# **Санитарная статистика**

## **Средние величины**

### **Занятие 1.1.3**

**Тюмень 2018**

**ТЕМА:**

**«СРЕДНИЕ ВЕЛИЧИНЫ.  
ОЦЕНКА ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ  
ИССЛЕДОВАНИЯ »**

**Средняя величина** - это величина, которая средним цифровым значением характеризует изучаемое явление или признак.

**Средняя величина является сводной, обобщающей характеристикой статистической совокупности по определенному изменяющемуся количественному признаку** (*средний рост, средний вес, средний возраст умерших*).

**Средняя величина отражает общее определяющее свойство всей статистической совокупности в целом, заменяя его одним числом с типичным значением данного признака.**

Средняя величина нивелирует, ослабляет случайные отклонения индивидуальных наблюдений в ту или иную сторону и характеризует постоянное свойство явлений.

В медицине средние величины могут использоваться для характеристики физического развития, основных антропометрических признаков (морфологических и функциональных: рост, вес, динамометрия и др.) и их динамики (средние величины прироста или убыли признака).

Разработка этих показателей и их сочетаний в виде стандартов имеет большое практическое значение для анализа здоровья населения (в особенности детей, спортсменов).

Эпидемиологи рассчитывают среднее число заболеваний в очаге, распределение очагов по срокам и средние сроки производства дезинфекции.

В демографических и медико-социальных исследованиях рассчитываются: средняя продолжительность предстоящей жизни, средний возраст умерших, средняя численность населения и т.д.

**И статистические коэффициенты, и средние величины представляют собой вероятностные величины, однако между ними существуют значительные различия:**

**1) Статистические коэффициенты характеризуют** признак, встречающийся только у некоторой части совокупности (так называемый альтернативный признак), который может наступить, но может и не наступить (рождение, смерть, заболевание).

*Средние величины характеризуют, признаки, присущие всей совокупности, но в разной степени* (вес, рост, дни лечения).

**2) Статистические коэффициенты** применяются для измерения качественных (атрибутивных или описательных) признаков, а средние - для варьирующих количественных признаков, где речь идет об отличиях в числовых размерах признака, а не о факте его наличия или отсутствия.

**Основное достоинство средних величин их типичность - средняя сразу дает общую характеристику явления.**

В связи с этим можно выделить *два основных требования* для вычисления средних величин:

- *однородность совокупности;*
- *достаточное число наблюдений.*

Любое распределение случайной величины, не обязательно подчиняющееся определенному закону распределения вероятностей, характеризуется параметрами распределения:

*□ средняя величина ( $M$ ),*

*□ среднее квадратическое отклонение ( $\sigma$ ),*

*□ коэффициент вариации ( $Cv$ ) и др.*

Например, при изучении распределения 10 больных по срокам лечения, мы получим ряд числовых значений: 38, 13, 17, 20, 14, 18, 25, 32, 23, 25 - *неупорядоченный ряд*.

Рассчитать параметры распределения можно, пользуясь и таким рядом. Однако охарактеризовать ряд несколькими параметрами еще недостаточно, необходимо исследовать, есть ли в статистическом ряду какая-либо устойчивая закономерность. Но, пользуясь неупорядоченным рядом, возможную закономерность обнаружить сложно, поэтому строят ранжированные ряды.

**Ряд, в котором дается распределение единиц изучаемой совокупности по значениям варьирующего признака, называется вариационным.**



## **Основные требования к составлению вариационных рядов:**

1. Расположить все варианты по порядку.
2. Суммировать единицы, имеющие одинаковый признак, т.е. найти частоту каждой варианты.
3. Определить число групп и размер интервала.
4. Разбить весь ряд на группы, используя выбранный интервал и строго соблюдая непрерывность сгруппированного ряда.
5. Дать графическое изображение.

**Все пять требований** выполняются при составлении **сгруппированных** вариационных рядов.

Требования **-1, 2 и 5** – при составлении **не сгруппированных** вариационных рядов.

В медицинской статистике приняты следующие условные обозначения для вариационного ряда:

**V** - варианты, числовые значения признака

Меняющийся, варьирующий признак изучаемого явления (рост, вес и др.), его числовое значение называется **вариантой (V)**.

**P** - частота признака.

Числа случаев наблюдения данного признака, указывающие сколько раз встречается данная варианта, называются **частотами (p)**.

**n** - **сумма всех частот** (общее число вариантов).

**V max** и **V min** – **крайние варианты**, ограничивающие вариационный ряд.

**A** – **амплитуда ряда**, (т.е. разность между максимальной и минимальной вариантами

$$A = V \max - V \min$$

## Вариационные ряды могут быть:

**1) в зависимости от изучаемого явления:**

- **дискретные (прерывные)** – образуются на основе прерывно меняющихся признаков, значения которых выражаются только в целых числах (частота пульса, количество студентов в группе и т.д.);
- **интервальные (непрерывные)** – образуются обычно на основе признаков, которые могут принимать любые значения и выражаются любым числом (рост, вес и т.д.)

## Вариационные ряды могут быть: (продолжение)

### 2) в зависимости от числа наблюдений:

- **простые** – варианта представлена одним числовым значением;
- **сгруппированные** – варианты группируются по определенному признаку. Например, при изучении физического развития может производиться группировка по весу: 40-44 кг; 45-49 кг. и т.д.

### 3) в зависимости от порядка расположения вариантов:

- **возрастающие** – варианты располагаются в порядке возрастания;
- **убывающие** – варианты располагаются в порядке убывания

*Отдельный вариационный ряд* может одновременно включать в себя *несколько характеристик*.

*Например:* простой, убывающий, прерывный; или – сгруппированный, возрастающий, непрерывный.

**Виды средних величин, которые обычно используются в медицинской статистике:**

□ медиана,

□ мода,

□ средняя арифметическая.

Другие виды средних: средняя гармоническая, средняя квадратическая, средняя кубическая, средняя геометрическая и другие - применяются лишь в специальных исследованиях.

**Медиана (Me)** - это срединная, центральная варианта, делящая вариационный ряд пополам на две равные части.

*Например*, если число наблюдений составляет 33, медианой будет варианта, занимающая **17-е** ранговое место, так как в обе стороны от неё находится по 16 наблюдений.

**В ряде с чётным числом наблюдений в центре** находятся **две величины**.

Если они одинаковы по своему значению, не возникает затруднений в приближенном определении медианы, если же числовые значения двух величин различны, то **за медиану принимается их полусумма**.

**Мода ( $M_o$ ) – это чаще всего встречающаяся или наиболее часто повторяющаяся величина признака.** При приближенном нахождении моды в простом (не сгруппированном) ряде, она определяется как варианта с наибольшим количеством частот.

*Отличие медианы и моды от средней арифметической* заключается в том, что при упрощенном, ориентировочном определении *эти величины легко и быстро найти по их положению в вариационном ряду (позиционные средние)*, кроме того, *они не зависят от значений крайних вариантов или от степени рассеяния ряда*

Чаще всего используется в медицинской статистике **средняя арифметическая величина** (*M* - от латинского *Media*).

Средняя арифметическая может быть **простая и взвешенная**.

*Примером* средней арифметической простой может служить результат измерения веса, например, 6 человек:

|               |    |    |    |    |    |    |                    |
|---------------|----|----|----|----|----|----|--------------------|
| <i>V</i> (кг) | 59 | 60 | 61 | 62 | 63 | 64 | $\Sigma = 369$     |
| <i>p</i>      | 1  | 1  | 1  | 1  | 1  | 1  | $\Sigma p = n = 6$ |

Расчет производится по формуле:

$$M = \frac{\Sigma V}{n}$$

Сумма этих измерений, деленная на число наблюдений, и дает среднюю величину веса:

$$M = \frac{369}{6} = 61,5 \text{ кг.}$$

Таким образом, **средняя арифметическая простая** получается как **сумма величин (вариант), деленная на их число.**

*Среднюю арифметическую простую* можно вычислить лишь в тех случаях, когда каждая величина (варианта) представлена **единичным наблюдением**, т. е. *когда частоты равны единице.*

**Если частоты вариант больше единицы**, простая средняя неприменима - здесь *надо вычислять среднюю арифметическую взвешенную*, которая получается как *сумма произведений вариант на соответствующие частоты, деленная на общее число наблюдений:*

$$M = \frac{\sum V \cdot p}{n}$$

Например: частота пульса (число ударов в минуту) у 18 студентов после проведения атропиновой пробы составила: 86, 92, 100, 96, 90, 102, 88, 92, 80, 92, 96, 100, 86, 84, 102, 90, 86, 92.

|               |    |    |     |    |     |     |     |     |     |                   |
|---------------|----|----|-----|----|-----|-----|-----|-----|-----|-------------------|
| V<br>(уд/мин) | 80 | 84 | 86  | 88 | 90  | 92  | 96  | 100 | 102 |                   |
| p             | 1  | 1  | 3   | 1  | 2   | 4   | 2   | 2   | 2   | $\sum p = n = 18$ |
| Vp            | 80 | 84 | 258 | 88 | 180 | 368 | 192 | 200 | 204 | $\sum Vp = 1654$  |

$$M = \frac{1654}{18} = 91,9 \text{ уд/мин.}$$

**Средняя арифметическая простая** - это частный случай средней арифметической взвешенной, поэтому формула средней арифметической взвешенной может использоваться и для расчета средней арифметической простой. В последнем случае частоты равны единице и умножение излишне.

**Все три средние величины ( $M_o$ ,  $M_e$ ,  $M$ ) совпадают (либо практически очень близки) в симметричном вариационном ряду:**

- ❖ средняя арифметическая соответствует середине ряда (в симметричном ряду отклонения в сторону увеличения и в сторону уменьшения вариант соответственно уравниваются);
  - ❖ медиана (как центральная величина) также соответствует середине ряда;
  - ❖ мода (как наиболее насыщенная величина) приходится на наивысшую точку ряда, также находящуюся в его центре.
- Поэтому для всех симметричных рядов нет необходимости вычислять другие средние величины, кроме средней арифметической.*

## Свойства средней арифметической величины:

1. Средняя величина является обобщающей характеристикой статистической совокупности по определенному изменяющемуся количественному признаку, отражает общее определяющее свойство всей статистической совокупности в целом, заменяя его одним числом с типичным значением данного признака. Средняя величина нивелирует, ослабляет случайные отклонения индивидуальных наблюдений в ту или иную сторону и характеризует постоянное свойство явлений.

2. Сумма отклонений вариант от средней арифметической величины равна 0.

3. В строго симметричном вариационном ряду средняя арифметическая занимает срединное положение и равна  $M_0$ ,  $M_e$ .

Средние арифметические величины, взятые сами по себе без дополнительных приемов оценки, часто имеют ограниченное значение, так как они не отражают степени рассеяния (разнообразия) ряда.

Одинаковые по размеру средние величины могут быть получены из рядов с различной степенью рассеяния.

**Средние** - это величины, вокруг которых рассеяны различные варианты, и чем ближе друг к другу отдельные варианты, чем меньше рассеяние ряда, тем типичнее средняя величина.

Приближенным методом оценки разнообразия ряда может служить определение амплитуды.

**Амплитуда** - разность между наибольшим и наименьшим значением вариант:

$$A = V_{max} - V_{min}$$

Но амплитуда не учитывает промежуточные значения вариант внутри ряда, кроме того, её размеры могут зависеть и от числа наблюдений.

**Основной мерой оценки разнообразия ряда является среднее квадратическое отклонение ( $\sigma$ ).**

Вычисление точного значения среднего квадратического отклонения производится по формуле:

$$\sigma = \pm \sqrt{\frac{\sum d^2 p}{n}}$$

**Если число наблюдений меньше 30 (малая выборка), то расчет производится по формуле:**

$$\sigma = \pm \sqrt{\frac{\sum d^2 p}{n - 1}}$$

Для вычисления сигмы необходимо:

- 1) определить отклонения ( $d$ ) от средней ( $V - M$ );
- 2) возвести отклонения в квадрат ( $d^2$ );
- 3) перемножить квадраты отклонений на частоты ( $d^2 p$ );
- 4) суммировать произведения квадратов отклонений на частоты;
- 5) разделить эту сумму на число наблюдений;
- 6) извлечь из частного квадратный корень.

При помощи сигмы можно установить степень типичности средней, пределы рассеяния ряда, пределы колебаний вокруг средней отдельных вариантов.

Чем меньше сигма, тем меньше рассеяние ряда, тем точнее и типичнее получается вычисленная для этого ряда средняя величина.

Применение сигмы дает возможность оценки и сравнения разнообразия нескольких однородных рядов распределения, так как  $\sigma$  - величина именная, выражается абсолютным числом в единицах изучаемой совокупности (см, кг, мг/л и т.д.).

В этом случае принимаются во внимание абсолютные размеры сигмы. Например, при сравнении двух рядов распределения по признаку веса, при условии, что средние будут близки по уровню, но сигма в одном ряду будет  $\pm 5,6$  кг., а в другом  $\pm 2,1$  кг. - второй ряд менее рассеян, и его средняя более типична.

При оценке разнообразия неоднородных рядов (например, таких признаков как вес и рост), непосредственное сравнение размеров сигмы невозможно.

В этом случае, для установления степени относительного разнообразия рядов, прибегают к производной величине - **коэффициенту изменчивости (вариации), который является относительной величиной, выражается в % и обозначаемому буквой Cv (V).**

Коэффициент изменчивости получается из процентного отношения сигмы к средней:

$$Cv = \frac{\sigma}{M} \cdot 100\%$$

Например, при изучении физического развития студентов – мужчин 1 курса получены следующие показатели:

$M$  (вес) = 67,5 кг.;  $M$  (рост) = 178,1 см.

Соответственно  $\sigma = \pm 2,8$  кг. и  $\pm 6,2$  см.

Среднее квадратическое отклонение по росту более чем в 2 раза превышает сигму по весу.

Коэффициент вариации  $C_v$  равен:

$$C_v \text{ (по весу)} = \frac{2,8 \text{ кг}}{67,5 \text{ кг}} \cdot 100\% = 4,1\%$$

$$C_v \text{ (по росту)} = \frac{6,2 \text{ см}}{178,1 \text{ см}} \cdot 100\% = 3,5\%$$

*Коэффициент вариации по росту меньше, чем по весу, то есть рост оказался более устойчивым признаком, чем вес.*

Различают **три степени разнообразия коэффициентов вариации:**

- ✓ **до 10%** - **слабое разнообразие;**
- ✓ **10 – 20 %** - **среднее разнообразие;**
- ✓ **более 20 %** - **сильное разнообразие.**

Этот же метод вычисления коэффициента разнообразия пригоден и при анализе однородных рядов, у которых средние величины очень разнятся по размеру, а также для оценки изолированного, единичного ряда.

**Пример** вычисления средней арифметической ( $M$ ); среднего квадратического отклонения ( $\sigma$ ); коэффициента вариации ( $Cv$ ):

Длительность лечения ангины у 45 больных составила: 20, 20, 19, 16, 19, 16, 14, 13, 15, 13, 12, 13, 13, 3, 12, 11, 12, 11, 10, 12, 11, 10, 11, 8, 7, 11, 11, 10, 10, 10, 9, 8, 8, 9, 5, 5, 6, 9, 5, 5, 9, 6, 7, 7, 14, и 15 дней.

**Первый этап:**

Строим вариационный ряд, с учетом частоты встречаемости каждой варианты; даем характеристику ряда; находим произведения вариант на соответствующую частоту, суммируем полученные произведения и рассчитываем среднюю арифметическую:

$$M = \frac{\sum V \cdot p}{n} = \frac{495}{45} = 11 \text{ дней}$$

**Первый этап****Второй этап**

| Длительность<br>лечения<br>(в днях) $V$ | Число<br>больных<br>$p$ | $Vp$              | $d (V-M)$ | $d^2$ | $d^2p$               |
|---|-------------------------|-------------------|-----------|-------|----------------------|
| 20                                      | 2                       | 40                | 9         | 81    | 162                  |
| 19                                      | 2                       | 38                | 8         | 64    | 128                  |
| 16                                      | 2                       | 32                | 5         | 25    | 50                   |
| 15                                      | 2                       | 30                | 4         | 16    | 32                   |
| 14                                      | 2                       | 28                | 3         | 9     | 18                   |
| 13                                      | 4                       | 52                | 2         | 4     | 16                   |
| 12                                      | 4                       | 48                | 1         | 1     | 4                    |
| 11                                      | 6                       | 66                | 0         | 0     | 0                    |
| 10                                      | 5                       | 50                | -1        | 1     | 5                    |
| 9                                       | 4                       | 36                | -2        | 4     | 16                   |
| 8                                       | 3                       | 24                | -3        | 9     | 27                   |
| 7                                       | 3                       | 21                | -4        | 16    | 48                   |
| 6                                       | 2                       | 12                | -5        | 25    | 50                   |
| 5                                       | 3                       | 15                | -6        | 36    | 108                  |
| 3                                       | 1                       | 3                 | -8        | 64    | 64                   |
|   | $\Sigma p = n = 45$     | $\Sigma Vp = 495$ |           |       | $\Sigma d^2 p = 728$ |

*Ряд простой, убывающий, прерывный*

**Второй этап:** рассчитываем  $d(V-M)$ ;  $d^2$ ;  $d^2p$ .

**Третий этап:** рассчитываем среднее квадратическое отклонение ( $\sigma$ ); коэффициент вариации ( $Cv$ ):

$$\sigma = \pm \sqrt{\frac{\sum d^2 p}{n}} = \pm \sqrt{\frac{728}{45}} = \pm \sqrt{16,2} = \pm 4,02 \text{ дня}$$

$$Cv = \frac{\sigma}{M} \cdot 100\% = \frac{4,02}{11} \cdot 100\% = 36,5\%$$

**Заключение:**

Средняя длительность лечения ангины в поликлинике составила **11 дней**. Средняя является недостаточно типичной для данного ряда, о чем свидетельствует коэффициент вариации, равный **36,5%** (большая степень разнообразия признака).



**САМОСТОЯТЕЛЬНАЯ РАБОТА СТУДЕНТОВ**  
**УЧЕБНОЕ ЗАДАНИЕ 5**

Средний уровень признака определяется средними величинами:

- 1.Мода (Mo)** – самая частая варианта
- 2.Медиана (Me)** – варианта, расположенная по середине ряда
- 3.Средняя арифметическая (M)**

а) **Средняя арифметическая простая** рассчитывается из вариационного ряда, в котором каждая варианта встречается один раз.

Она рассчитывается по формуле:

$$M = \frac{\sum v}{n} = \frac{242}{4} = 61$$

| Вес в кг<br>(V) | Число<br>подростков |
|-----------------|---------------------|
| 59              | 1                   |
| 60              | 1                   |
| 61              | 1                   |
| 62              | 1                   |
| $\Sigma = 242$  | $n = 4$             |

б) **Средняя арифметическая взвешенная** высчитывается из вариационного ряда, в котором каждая варианта встречается один раз и более.

Она рассчитывается:

- **непосредственным способом**
- **по способу моментов**

Средняя арифметическая взвешенная **непосредственным способом** рассчитывается по формуле:

$$M = \frac{\sum V \cdot P}{n}$$

**ЗАДАЧА 1.** Определите средний вес 26 подростков (непосредственным способом).

| Вес в кг (V) | Число подростков ( P) | VP |
|--------------|-----------------------|----|
| 59           | 1                     |    |
| 60           | 4                     |    |
| 61           | 6                     |    |
| 62           | 9                     |    |
| 63           | 3                     |    |
| 64           | 2                     |    |
| 65           | 1                     |    |

**n = 26**

# ЗАДАЧА 1. Определение среднего веса 26 подростков (непосредственным способом).

| Вес в кг ( V ) | Число подростков ( P ) | VP          |
|----------------|------------------------|-------------|
| 59             | 1                      | 59          |
| 60             | 4                      | 240         |
| 61             | 6                      | 366         |
| 62             | 9                      | 558         |
| 63             | 3                      | 189         |
| 64             | 2                      | 128         |
| 65             | 1                      | 65          |
| <b>Сумма</b>   | <b>26</b>              | <b>1605</b> |

$$M = \frac{\sum V \cdot P}{n} = \frac{1605}{26} = 62$$

**в) Средняя арифметическая взвешенная по способу моментов** рассчитывается по формуле:

$$M = M_1 + \frac{\sum d \cdot P}{n}$$

В процессе работы необходимо дать характеристику изучаемого явления (характеристику разнообразия признака).  
Разнообразию (или похожесть) характеризуют:

- амплитуда
- лимит
- среднеквадратическое отклонение
- коэффициент вариации

**г) Среднеквадратическое отклонение (сигма)** характеризует колеблемость (вариабельность) вариационного ряда и является его абсолютной мерой. Чем большее число наблюдений будет находиться в пределах  $\pm$  одна сигма от средней арифметической, тем однообразнее будет вариационный ряд.

Сигма для средней арифметической, вычисленной по способу моментов, рассчитывается по формуле:

$$\sigma = \pm \sqrt{\frac{\sum d^2 \cdot P}{n} - \left( \frac{\sum d \cdot P}{n} \right)^2}$$

д) Коэффициент вариации характеризует колеблемость (вариабельность) вариационного ряда и является его относительной мерой. Чем меньше показатель, тем однороднее вариационный ряд.

Различают степени градации вариации:

- ✓ низкая (меньше 10,0 %)
- ✓ средняя (10 - 20,0 %)
- ✓ высокая ( больше 20,0 %)

е) Кроме характеристики разнообразия признака оценивается **достоверность результатов исследования** - рассчитываются:

- ошибка средней арифметической
- доверительные границы средней арифметической
- достоверность разности результатов исследования

## Оценка достоверности результатов исследования

*Под выборочным методом в статистике понимается такой метод наблюдения, при котором для отыскания типичных черт характеристик какой-либо совокупности изучаются не все единицы этой совокупности, а лишь часть их.*

Как бы тщательно ни производилась выборка, какой репрезентативной ни была бы выборочная совокупность (отобранная часть наблюдений), она неизбежно будет отличаться от всей генеральной (общей) совокупности.

Таким образом, полного тождества достичь не удастся, и некоторая неточность встречается неизбежно. Однако имеются методы установления степени различий числовых характеристик обеих совокупностей и пределов возможных колебаний показателей при данном числе наблюдений. *Число наблюдений играет значительную роль - чем больше число наблюдений, тем точнее отображается генеральная совокупность и тем меньше размеры ошибки.*

Так называемые средние ошибки являются *мерой точности и достоверности* любых статистических величин.

**Под достоверностью статистических показателей** (синонимы: **существенность, значимость, надежность**) **понимают доказательность, то есть право на обобщение явления, правомерность распространения выводов и на другие аналогичные явления. Или - степень их соответствия отображаемой ими действительности.**

Достоверными результатами считаются те, которые не искажают и правильно отражают объективную реальность.

*Оценить достоверность результатов исследования означает определить, с какой вероятностью возможно перенести результаты, полученные на выборочной совокупности, на всю генеральную совокупность.*

В большинстве медицинских исследований врачу приходится, как правило, иметь дело с частью изучаемого явления, а выводы по результатам такого исследования переносить на все явление в целом - на генеральную совокупность.

**Оценка достоверности результатов исследования предусматривает определение:**

- 1) *ошибок репрезентативности (средних ошибок средних арифметических и относительных величин) -  $m$ ;*
- 2) *доверительных границ средних (или относительных) величин;*
- 3) *достоверности разности средних (или относительных) величин (по критерию  $t$  - **Стьюдента**).*

## **1.Определение средней ошибки средней (или относительной) величины (ошибка репрезентативности – $m$ ).**

Теория выборочного метода, наряду с обеспечением репрезентативности, практически сводится к оценке расхождений между числовыми характеристиками генеральной и выборочной совокупности, т. е. к *определению средних ошибок и так называемых доверительных границ или интервалов.*

**Средняя ошибка** позволяет установить тот интервал, в котором заключено действительное значение производной величины при данном числе наблюдений, т. е. *средняя ошибка всегда является конкретной.*

**Ошибка репрезентативности является важнейшей статистической величиной, необходимой для оценки достоверности результатов исследования.**

Эта ошибка возникает в тех случаях, когда требуется по части охарактеризовать явление в целом. Эти ошибки неизбежны. Они «вытекают» из сущности выборочного исследования.

Генеральная совокупность может быть охарактеризована по выборочной совокупности только с некоторой погрешностью, измеряемой ошибкой репрезентативности.

Ошибки репрезентативности не тождественны обычным представлением об ошибках: методических, точности измерения, арифметических и др.

По величине ошибки репрезентативности определяют, насколько результаты, полученные при выборочном исследовании, отличаются от результатов, которые могли бы быть получены при проведении сплошного исследования без исключения всех элементов генеральной совокупности.

Это единственный вид ошибок, учитываемых статистическими методами, которые не могут быть устранены, если не проведено сплошное исследование.

Ошибки репрезентативности можно свести к достаточно малой величине, т.е. к *величине допустимой погрешности*. Делается это путем увеличения числа наблюдений (**n**).

Каждая средняя величина -  $M$  (средняя длительность лечения, средний рост, средняя масса тела и др.), а также относительная величина -  $P$  (уровень летальности, заболеваемости и др.) должны быть представлены со своей средней ошибкой -  $m$ .

Средняя арифметическая величина выборочной совокупности ( $M$ ) имеет ошибку репрезентативности, которая называется **средней ошибкой средней арифметической** ( $m_m$ ) и определяется по формуле:

$$m_m = \pm \frac{\sigma}{\sqrt{n}}$$

Как видно из этой формулы, между размерами сигмы (отражающей разнообразие явления) и размерами средней ошибки существует прямая связь.

Между числом наблюдений и размерами средней ошибки существует обратная связь (пропорциональная не числу наблюдений, а квадратному корню из этого числа).

Следовательно, уменьшение величины этой ошибки при определении степени разнообразия ( $\sigma$ ) возможно путем увеличения числа наблюдений. При числе наблюдений менее 30 в знаменателе следует взять  $(n - 1)$ .

$$m_m = \pm \frac{\sigma}{\sqrt{n - 1}}$$

На этом принципе основан метод определения достаточного числа наблюдений для выборочного исследования.

Относительные величины ( $P$ ), полученные при выборочном исследовании, также имеют свою ошибку репрезентативности, которая называется средней ошибкой относительной величины и обозначается  $m_p$ .

Для определения средней ошибки относительной величины ( $P$ ) используется следующая формула:

$$m_p = \pm \sqrt{\frac{P \cdot q}{n}}$$

Где:  $P$  - относительная величина.;

$q$  – разность между основанием, на которое рассчитана относительная величина и самой относительной величиной. Если показатель выражен в процентах, то  $q = 100 - P$ ; если  $P$  - в промиллях, то  $q = 1000 - P$ , если  $P$  - в продецимиллях, то  $q = 10.000 - P$ , и т.д.;

$n$  - число наблюдений. При числе наблюдений менее 30 в знаменатель следует взять  $(n - 1)$ .

$$m_p = \pm \sqrt{\frac{P \cdot q}{n - 1}}$$

• Каждая средняя арифметическая или относительная величина, полученная на выборочной совокупности, должна быть представлена со своей средней ошибкой.

Это дает возможность рассчитать доверительные границы средних и относительных величин, а также определить достоверность разности сравниваемых показателей (результатов исследования).

## **2. Определение доверительных границ.**

Определяя для средней арифметической (или относительной) величины два крайних значения: минимально возможное и максимально возможное, находят пределы, в которых может быть искомая величина **генерального параметра**. Эти пределы называют доверительными границами.

**Доверительные границы - границы средних (или относительных) величин, выход за пределы которых вследствие случайных колебаний имеет незначительную вероятность.**

Вероятность попадания средней или относительной величины в доверительный интервал называется **доверительной вероятностью**.

Доверительные границы *средней арифметической генеральной совокупности* определяют по формуле:

$$M_{ген} = M_{выб} \pm t \cdot mM$$

Доверительные границы относительной величины в генеральной совокупности определяют по следующей формуле:

$$P_{ген} = P_{выб} \pm t \cdot mP$$

Где:  $M_{ген}$  и  $P_{ген}$  - значения средней и относительной величин, полученных для генеральной совокупности;

$M_{выб}$  и  $P_{выб}$  - значения средней и относительной величин, полученных для выборочной совокупности;

$mM$  и  $mP$  - ошибки репрезентативности выборочных величин;

$t$  - доверительный критерий, который зависит от величины безошибочного прогноза, устанавливаемого при планировании исследования.

Произведение  $t \cdot m$  ( $\Delta$ ) - предельная ошибка показателя, полученного при данном выборочном исследовании.

Размеры предельной ошибки зависят от коэффициента  $t$ , который избирает сам исследователь, исходя из заданной вероятности безошибочного прогноза.

Величина критерия  $t$  связана с вероятностью безошибочного прогноза ( $P$ ) и числом наблюдений в выборочной совокупности (табл. 4.1).

Таблица 4.1

Зависимость доверительного критерия  $t$  от степени вероятности безошибочного прогноза  $P$  (при  $n > 30$ )

| Степень вероятности безошибочного прогноза ( $P$ %) | Доверительный критерий $t$ |
|---|----------------------------|
| 95,0  | 2                          |
| 99,0  | 2,6                        |
| 99,9  | 3,3                        |

Для большинства медико-биологических и социальных исследований достоверными считаются доверительные границы, установленные с вероятностью безошибочного прогноза = **95% и более**.

Чтобы найти критерий  $t$  при числе наблюдений  $(n) < 30$ , необходимо пользоваться специальной таблицей Н.А. Плохинского в которой слева показано число наблюдений - единица  $(n - 1)$ , а сверху  $(P)$  - степень вероятности безошибочного прогноза.

При определении доверительных границ сначала надо решить вопрос о том, с какой степенью вероятности безошибочного прогноза необходимо представить доверительные границы средней или относительной величины. Избрав определенную степень вероятности, соответственно этому находят величину доверительного критерия  $t$  при данном числе наблюдений. Таким образом, доверительный критерий устанавливается заранее, при планировании исследования.

| <b>P n = n-1</b> | <b>95%</b>  | <b>99%</b>  | <b>99,9%</b> |
|------------------|-------------|-------------|--------------|
| <b>1</b>         | <b>12,7</b> | <b>63,7</b> | <b>37,0</b>  |
| <b>2</b>         | <b>4,3</b>  | <b>9,9</b>  | <b>31,6</b>  |
| <b>3</b>         | <b>3,2</b>  | <b>5,8</b>  | <b>12,9</b>  |
| <b>4</b>         | <b>2,8</b>  | <b>4,6</b>  | <b>8,6</b>   |
| <b>5</b>         | <b>2,6</b>  | <b>4,0</b>  | <b>6,9</b>   |
| <b>6</b>         | <b>2,4</b>  | <b>3,7</b>  | <b>6,0</b>   |
| <b>7</b>         | <b>2,4</b>  | <b>3,5</b>  | <b>5,3</b>   |
| <b>8</b>         | <b>2,3</b>  | <b>3,4</b>  | <b>5,0</b>   |
| <b>9</b>         | <b>2,3</b>  | <b>3,3</b>  | <b>4,8</b>   |
| <b>10</b>        | <b>2,2</b>  | <b>3,2</b>  | <b>4,6</b>   |
| <b>11</b>        | <b>2,2</b>  | <b>3,1</b>  | <b>4,4</b>   |
| <b>12</b>        | <b>2,2</b>  | <b>3,1</b>  | <b>4,3</b>   |
| <b>13</b>        | <b>2,3</b>  | <b>3,0</b>  | <b>4,1</b>   |
| <b>14-15</b>     | <b>2,1</b>  | <b>3,0</b>  | <b>4,1</b>   |
| <b>16-17</b>     | <b>2,1</b>  | <b>2,9</b>  | <b>4,0</b>   |
| <b>18-20</b>     | <b>2,1</b>  | <b>2,9</b>  | <b>3,9</b>   |
| <b>21-24</b>     | <b>2,1</b>  | <b>2,8</b>  | <b>3,8</b>   |
| <b>25-29</b>     | <b>2,0</b>  | <b>2,8</b>  | <b>3,7</b>   |

Любой параметр (средняя или относительная величина) может оцениваться с учетом доверительных границ, полученных при расчете.

**Например:** требуется определить доверительные границы среднего уровня пепсина у больных гипертериозом с 95% вероятностью безошибочного прогноза. Если известно, что:

$$n = 49;$$

$$M_{\text{выб}} = 1\text{г}\%;$$

$$mM = \pm 0,05\text{г}\%$$

1. Определение доверительных границ средней величины в генеральной совокупности:

$$M_{\text{ген}} = M_{\text{выб}} \pm t \cdot mM = 1\text{г}\% \pm 2 \cdot 0,05\text{г}\% \\ 1\text{г}\% + 0,1\text{г}\% = 1,1\text{ г}\%$$

$$M_{\text{ген}} = \\ 1\text{г}\% - 0,1\text{г}\% = 0,9\text{ г}\%$$

**Заключение:** установлено с вероятностью безошибочного прогноза 95%, что средний уровень пепсина в генеральной совокупности у больных гипертериозом находится в пределах от 1,1 г% до 0,9 г%.

Как видно, ***доверительные границы зависят от размера доверительного интервала.***

Анализ доверительных интервалов указывает, что при заданных степенях вероятности и  $n > 30$  -  $t$  имеет неизменную величину и при этом доверительный интервал зависит от величины ошибки репрезентативности.

С уменьшением величины ошибки суживаются доверительные границы средних и относительных величин, полученных на выборочной совокупности, т.е. уточняются результаты исследования, которые приближаются к соответствующим величинам генеральной совокупности.

Если ошибка большая, то получают для выборочной величины большие доверительные границы, которые могут противоречить логической оценке искомой величины в генеральной совокупности.

В подобном случае надо искать резервы сокращения размаха доверительных границ в размере величины ошибки репрезентативности.

Доверительные границы ***Mвыб*** и ***Rвыб*** зависят не только от средних ошибок этих величин, но и от избранной исследователем степени вероятности безошибочного прогноза.

При большой степени вероятности размах доверительных границ увеличивается.

### **3. Определение достоверности разности средних (или относительных) величин (по критерию $t$ - Стьюдента).**

В медицине и здравоохранении по разности параметров оценивают средние и относительные величины, полученные для разных групп населения по полу, возрасту, а также групп больных и здоровых и т.д. Во всех случаях при сопоставлении двух сравниваемых величин возникает необходимость не только определить их разность, но и оценить ее достоверность.

Достоверность разности величин, полученных при выборочных исследованиях, означает, что вывод об их различии может быть перенесен на соответствующие генеральные совокупности.

Достоверность разности выборочной совокупности измеряется доверительным критерием, который рассчитывается по специальным формулам для средних и относительных величин.

Формула оценки достоверности разности сравниваемых средних величин:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

Для относительных величин:

$$t = \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}}$$

Где:  $M_1; M_2 ; P_1; P_2$  - параметры, полученные при выборочных исследованиях;

$m_1; m_2$  - их средние ошибки;

$t$  - критерий достоверности (Стьюдента).

Разность статистически достоверна при  $t \geq 2$ , что соответствует вероятности безошибочного прогноза, равной 95% и более.

Для большинства исследований, проводимых в медицине и здравоохранении, такая степень вероятности является вполне достаточной.

При величине критерия достоверности  $t < 2$  степень вероятности безошибочного прогноза составляет  $P < 95\%$ . При такой степени вероятности нельзя утверждать, что полученная разность показателей достоверна с достаточной степенью вероятности. В этом случае необходимо получить дополнительные данные, увеличив число наблюдений.

Иногда при увеличении численности выборки разность продолжает оставаться не достоверной. Если при повторных исследованиях разность остается недостоверной, можно считать доказанным, что между сравниваемыми совокупностями не обнаружено различий по изучаемому признаку.

**Например:** требуется определить, достоверны ли различия в уровне пепсина в желудочном соке больных гипертериозом и здоровых лиц. Обследуются на пепсин две группы: 49 больных гипертериозом и 50 здоровых людей (контрольная группа). Результаты представлены в таблице 4.3.

Таблица 4.3

Сравнение среднего уровня пепсина в желудочном соке больных гипертериозом и здоровых лиц

| Сравниваемые группы           | <i>N</i> | <i>M</i><br>(г%) | <i>m</i><br>(г%) | <i>t</i> | Уровень вероятности безошибочного прогноза ( <i>P</i> ) |
|-------------------------------|----------|------------------|------------------|----------|---|
| Больные гипертериозом         | 49       | 1,0              | ± 0,3            | 10,0     | < 99,9  |
| Здоровые (контрольная группа) | 50       | 4,0              | ± 0,1            |          |   |

$$t = \frac{M_1 - M_2}{m_1^2 + m_2^2}$$

$$t = \frac{4 - 1}{0,3^2 + 0,1^2} = 10,0$$

**Заключение:** при гипертериозе наблюдается снижение уровня пепсина, что подтверждается с большой степенью вероятности безошибочного прогноза ( $P > 99,9\%$ ). Следовательно, снижение уровня пепсина может быть использовано в качестве одного из симптомов для подтверждения диагностики гипертериоза.

Подобным же образом оценивают достоверность разности сравниваемых относительных величин.

Указанная методика оценки достоверности и разности результатов исследования позволяет проводить только сравнение групп по парам, при обязательном наличии обобщающих параметров - средних арифметических или относительных величин и их средних ошибок.

# ТЕСТОВЫЙ ЭКЗАМЕН К ЗАНЯТИЮ

## «Средние величины»

Факультет

№ группы

ФИО студента

Вопросы

01 02 03 04 05 06 07 08 09 10 11 12 13 14 15

Ответы

Дата

Подпись тестируемого

Подпись преподавателя



**ТЕСТОВЫЕ ЗАДАНИЯ К ЗАНЯТИЮ  
«СРЕДНИЕ ВЕЛИЧИНЫ»**

01. Средние величины применяются для характеристики признаков:

- а) качественных
- б) количественных
- в) относительных

## 02. Вариационный ряд - это:

- а) ряд любых измерений
- б) измерения, расположенные в ранговом порядке
- в) однородные показатели, изменяющиеся во времени
- г) ряд отвлеченных величин
- д) ряд элементов статистической совокупности

### 03. Варианта - это:

- а) числовое выражение признака
- б) средняя величина
- в) относительный показатель
- г) абсолютная величина
- д) качественная характеристика признака

04. Обобщающей характеристикой  
вариационного ряда является:

- а) число наблюдений
- б) среднее квадратическое отклонение
- в) только средняя арифметическая величина
- г) любая средняя величина
- д) варианта

05. Критериями разнообразия признака являются:

а) амплитуда

б) среднее квадратическое отклонение

в) коэффициент вариации

г) все вышеперечисленное

06. Характеристикой разнообразия изучаемого признака в вариационном ряду является:

- а) абсолютный прирост
- б) среднее квадратическое отклонение
- в) коэффициент корреляции
- г) число наблюдений
- д) средняя величина

07. Средняя ошибка средней арифметической величины и относительного показателя - это:

а) мера достоверности результата

б) разность между минимальным и максимальным значением признака

в) число степеней свободы

08. Средняя ошибка средней арифметической величины прямопропорциональна:

а) числу наблюдений

б) частоте встречаемости признака

в) коэффициенту достоверности

г) показателю разнообразия изучаемого признака

д) величине изучаемого признака

09. Средняя ошибка средней арифметической величины обратнопропорциональна:

- а) числу наблюдений
- б) показателю разнообразия изучаемого признака
- в) ошибке репрезентативности
- г) показателю достоверности
- д) величине изучаемого признака

10. С увеличением объема наблюдений  
ошибка репрезентативности:

- а) увеличивается
- б) остается без изменений
- в) уменьшится

11. При коэффициенте достоверности, равном 2 доверительная вероятность составляет:

- а) 68 %
- б) 95,5 %
- в) 99 %
- г) 99,5 %
- д) 99,9 %

## 12. Коэффициент Стьюдента - это:

- а) стандартизованный показатель
- б) средняя величина
- в) коэффициент корреляции
- г) коэффициент достоверности
- д) характеристика разнообразия признака

13. Показателем достоверности различия средних величин является:

- а) средняя ошибка
- б) коэффициент корреляции
- в) коэффициент достоверности (t)
- г) интенсивный показатель
- д) темп роста

14. Минимальная величина коэффициента Стьюдента, при которой различие между сравниваемыми величинами считается достоверным:

а) 1,0

б) 1,8

в) 2,0

г) 2,5

д) 3,0

15. Нормой распределения изучаемого признака является сигмальная зона:

- а) первая
- б) вторая
- в) третья
- г) за пределами 3-х зон.

## Ответы на тестовые задания к занятию №3

1. - б

2. - б

3. - а

4. - г

5. - г

6. - б

7. - а

8. - г

9.- а

10.- а

11.- б

12.- г

13.- в

14.- в

15.- а

**ЗАДАЧА 2.** Определите средний рост девушек 12-летнего возраста по способу моментов, среднее квадратическое отклонение, коэффициент вариации и ошибку средней арифметической

| Рост в см (V) | Число девушек (P) | d | dp | d <sup>2</sup> p |
|---------------|-------------------|---|----|------------------|
| 136           | 15                |   |    |                  |
| 137           | 17                |   |    |                  |
| 138           | 22                |   |    |                  |
| 139           | 20                |   |    |                  |
| 140           | 19                |   |    |                  |
| 141           | 18                |   |    |                  |

n = 111

## УЧЕБНОЕ ЗАДАНИЕ 6

Произвести группировку вариационного ряда и вычислить среднюю арифметическую по способу моментов.

Группировку провести в следующей последовательности:

1. Определить амплитуду вариационного ряда:

$$A = V_{\text{макс.}} - V_{\text{мин.}}$$

2. Выбрать интервал (классовый промежуток - K)

3. Определить число вариантов в сгруппированном интервальном ряду по формуле:

$$V = A/K + 1$$

## УЧЕБНОЕ ЗАДАНИЕ 7

Доверительные границы средних величин показывают в каких пределах можно доверять результатам выборочного исследования ( $M$ ), или как он будет колебаться при переносе его на всю генеральную совокупность ( $M$ ) с определенной долей вероятности ( $t$ ). Коэффициент вероятности (достоверности) может выражаться в единицах или процентах:

$$t - 1 = 68,3 \%$$

$$t - 2 = 95,5 \%$$

$$t - 3 = 99,7 \%$$

Для большинства медико-биологических исследований достаточной считается вероятность равная 95,5 %.

Доверительные границы рассчитываются по формуле:

$$M = M \pm tm$$

**ЗАДАЧА 1.** Группа больных коронарным атеросклерозом исследовалась на содержание холестерина в сыворотке крови. Среднее содержание холестерина у всех больных составляло  $231 \pm 4,0$  мг%. Определить доверительные границы средней арифметической с вероятностью безошибочного прогноза  $P = 95,5\%$  и  $P = 99,7\%$ .

**ЗАДАЧА 2.** Произведено обследование роста семилетних мальчиков ( $n=112$ ) города Тобольска. Средний рост составил  $123,96 \pm 0,50$  см. Определить доверительные границы средней арифметической с вероятностью безошибочного прогноза  $P=95,5\%$ .

## УЧЕБНОЕ ЗАДАНИЕ 8

Достоверность разности результатов исследования ( $t$ ) рассчитывается для определения различия между несколькими показателями. При этом решается вопрос, является ли различие показателей существенным (обусловленным действием каких-то фактов), или вызвано случайными колебаниями.

Достоверность разности рассчитывается по формуле:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

Различия считаются существенными, если  $t = 2$  или больше 2. При этом разность соответствует вероятности безошибочного прогноза равной 95,5 %.

**ЗАДАЧА 1.** Изучалась обсемененность кожных покровов патогенной и условно-патогенной флорой с одновременным анализом других иммунологических показателей у онкологических больных по сравнению с группой здоровых.

Определить существенны ли различия показателей обсемененности кожных покровов патогенной и условнопатогенной флорой в сравниваемых группах женщин.

|                           | Число<br>обследованных<br>(n) | Среднее число<br>микробных клеток на<br>коже предплечья (M) | m | t |
|---------------------------|-------------------------------|---|---|---|
| Больные раком<br>яичников | 28                            | 78,5  | 3 |   |
| Здоровые<br>женщины       | 20                            | 29,7  | 2 |   |

**ЗАДАЧА 2.** Студентов-медиков исследовалось артериальное давление до и после сдачи экзаменов. Максимальное артериальное давление до сдачи экзаменов в среднем составило  $127,2 \pm 3,0$ . После сдачи -  $117,0 \pm 4,0$  мм.рт.ст. Можно ли на основании этих данных считать, что действительно до сдачи экзаменов у студентов отмечалось некоторое повышение максимального артериального давления?