Лекция 8

# Анализ вариационных рядов

# Математическая статистика

- *Основная цель математической статистики* - это получение и обработка данных для статистически значимой поддержки процесса принятия решения, например, при решении задач планирования, управления, прогнозирования.
- Методы математической статистики можно разделить на **описательные** и **аналитические**.

# Математическая статистика

***Описательные методы позволяют описать реальные наблюдения*** с помощью таблиц, графиков, *характеристик положения* (среднее арифметическое, мода, медиана), *характеристик рассеяния* (среднее линейное отклонение, среднее квадратическое отклонение, дисперсия, коэффициент вариации) и т. д.

***Аналитические методы позволяют*** на основании выборочных наблюдений сделать статистически значимые выводы о наличии закономерностей для всей совокупности.

Две группы методов:

- методы *параметрической статистики*
- методы *непараметрической статистики.*

# Вариационные ряды распределения

В реальных социально-экономических системах нельзя проводить активные эксперименты, поэтому данные обычно представляют собой наблюдения за происходящим процессом.

*Результаты наблюдений* – это, в общем случае, ряд чисел, расположенных в беспорядке, который для изучения необходимо упорядочить (проранжировать).

- Операция, заключенная в расположении значений признака по возрастанию, называется *ранжированием* опытных данных.

# Вариационные ряды распределения

- После операции ранжирования опытные данные можно сгруппировать так, чтобы в каждой группе признак принимал одно и то же значение, которое называется *вариантом* ($X_i$). Число элементов в каждой группе называется *частотой* варианта ($n_i$).

Размахом вариации называется число

$$W = x_{max} - x_{min},$$

где $x_{max}$ – наибольший вариант, $x_{min}$ – наименьший вариант.

# Вариационные ряды распределения

- Сумма всех частот равна определенному числу *n*, которое называется объемом совокупности:

$$\sum_{i=1}^{k} n_i = n_1 + n_2 + \cdots + n_k = n\,. \qquad (1)$$

- Отношение частоты данного варианта к объему совокупности называется *относительной частотой* ($\widehat{p}_i$), или *частостью* этого варианта:

$$\widehat{p}_i = \frac{n_i}{n}, \qquad (2)$$

$$\sum_{i=1}^{k} \widehat{p}_i = \sum_{i=1}^{k} \frac{n_i}{n} = \frac{\sum_{i=1}^{k} n_i}{n} = \frac{n}{n} = 1 \qquad (3)$$

# Вариационные ряды распределения

- Последовательность вариантов, расположенных в возрастающем порядке, называется *вариационным рядом* (вариация - изменение).
- Вариационные ряды бывают дискретными и непрерывными. *Дискретным вариационным рядом* называется ранжированная последовательность вариантов с соответствующими частотами и (или) частостями.

# Вариационные ряды распределения

**Пример 1.** В результате тестирования группа из 24 человек набрала баллы: 4, 0, 3, 4, 1, 0, 3, 1, 0, 4, 0, 0, 3, 1, 0, 1, 1, 3, 2, 3, 1, 2, 1, 2. Построить дискретный вариационный ряд.

**Решение.** Проранжируем исходный ряд, подсчитаем частоту и частость вариантов:

0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4.

В результате получим дискретный вариационный ряд

| Балл, $x_i$ | Число студентов, $n_i$ | Относительная частота, $\widehat{p_i}$ |
|---|---|---|
| 0 | 6 | 6/24 |
| 1 | 7 | 7/24 |
| 2 | 3 | 3/24 |
| 3 | 5 | 5/24 |
| 4 | 3 | 3/24 |
| $\sum$ | 24 | 1 |

# Вариационные ряды распределения

- *Построение дискретного вариационного ряда нецелесообразно, если число значений признака велико*.
- В этом случае следует построить *интервальный вариационный ряд* (промежуток изменения признака разбивается на ряд отдельных интервалов и подсчитывается количество значений величины в каждом из них).

# Вариационные ряды распределения

- Будем считать, что отдельные (частичные) интервалы имеют одну и ту же длину. Число интервалов ($k$) в случае нормально распределенной совокупности можно определить по ***формуле Стерджесса***:

$$k = 1+3{,}322\ \lg n \qquad (4)$$

или приближенно: $k \in [6;12]$.

- *Длина частичного интервала* определяется по формуле:

$$h = \frac{W}{k} = \frac{x_{max} - x_{min}}{k} = \frac{15-4}{7} \approx 1{,}6. \qquad (5)$$

# Вариационные ряды распределения

**Пример 2.** Пусть дан ряд распределения хозяйств по количеству рабочих на 100 га сельскохозяйственных угодий ($n$=60)

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 6 | 8 | 6 | 10 | 11 | 7 | 10 | 12 | 8 | 7 | 7 | 6 | 7 | 8 | 6 | 11 | 9 | 11 |
| 9 | 10 | 11 | 9 | 10 | 7 | 8 | 8 | 8 | 11 | 9 | 8 | 7 | 5 | 9 | 7 | 7 | 14 | 11 |
| 9 | 8 | 7 | 4 | 7 | 5 | 5 | 10 | 7 | 7 | 5 | 8 | 10 | 10 | 15 | 10 | 10 | 13 | 12 |
| 11 | 15 | 6 | | | | | | | | | | | | | | | | |

Построить интервальный вариационный ряд.

**Решение**. Для определения числа групп подставим значение $n$=60 в формулу Стерджесса: $k$=l + 3,322lg60 $\approx$ 6,907; $k$ = 7.

| Группы хозяйств по численности работников на 100 га с/х угодий | Число хозяйств в группе ($n_i$) | Накопленное число хозяйств (Si) | Относительная частота ($\widehat{p}_i$) |
|---|---|---|---|
| 4-5,6 | 5 | 5 | 5/60. |
| 5,61-7,2 | 17 | 22 | 17/60 |
| 7,21-8,8 | 9 | 31 | 9/60 |
| 8,81-10,4 | 15 | 46 | 15/60 |
| 10,41 -12,0 | 10 | 56 | 10/60 |
| 12,01 -13,6 | 1 | 57 | 1/60 |
| 13,61-15,2 | 3 | 60 | 3/60 |
| Итого: | 60 | - | 1 |

# Графическое изображение вариационных рядов

*Вариационные ряды изображают графически с помощью полигона и гистограммы.*

- *Полигон частот* - это ломаная, отрезки которой соединяют точки $(x_1; n_1)$, $(x_2; n_2)$,…, $(x_k; n_k)$.
- *Полигон относительных частот* - это ломаная, отрезки которой соединяют точки: $(x_1; \frac{n_1}{n}), (x_2; \frac{n_2}{n}), \ldots, (x_k; \frac{n_k}{n})$.
- *Гистограммой частот* называется фигура, состоящая из прямоугольников с основанием $h$ и высотами $n_i$. Для *гистограммы относительных частот* в качестве высоты рассматривают $n_i/n$.
- Гистограмма относительных частот является аналогом дифференциальной функции случайной величины.

# Числовые характеристики вариационных рядов

- Вариационные ряды позволяют получить первое представление об изучаемом распределении.
- Далее необходимо исследовать ***числовые характеристики распределения*** (аналогичные характеристикам распределения теории вероятностей):
  - *характеристики положения* (средняя арифметическая, мода, медиана);
  - *характеристики рассеяния* (дисперсия, среднее квадратическое отклонение, коэффициент вариации);
  - *характеристики меры скошенности* (коэффициент асимметрии) и островершинности (эксцесс) распределения.

# Числовые характеристики вариационных рядов

## *Характеристики положения вариационного ряда*

- ***Средней арифметической*** *(X)* дискретного вариационного ряда называется отношение суммы произведений вариантов на соответствующие частоты к объему совокупности:

$$\bar{X} = \frac{\sum x_i n_i}{\sum n_i} = \frac{\sum x_i n_i}{n}.$$

- Средняя арифметическая имеет те же единицы измерения, что и варианты.

# Числовые характеристики вариационных рядов

## *Характеристики положения вариационного ряда*

- ***Свойства средней арифметической***

1) Средняя арифметическая суммы соответствующих друг другу значений, принадлежащих двум группам наблюдений, равна алгебраической сумме средних арифметических этих групп:

$$\overline{X \pm Y} = \bar{X} \pm \bar{Y}.$$

2) Если ряд наблюдении состоит из двух непересекающихся групп наблюдений, то средняя арифметическая $\bar{Z}$ всего ряда наблюдений равна взвешенной средней арифметической групповых средних $\bar{\mathrm{X}}$ и $\bar{Y}$, причем весами являются объемы групп $n_1 \ = \sum n_i, n_2 = \sum m_j$ соответственно

$$\overline{X \pm Y} = \frac{\sum x_i n_i + \sum y_j m_j}{n_1 + n_2}$$

# Числовые характеристики вариационных рядов

## *Характеристики положения вариационного ряда*

***Свойства средней арифметической***

3) Средняя арифметическая постоянной равна самой постоянной

$$\bar{\mathrm{C}} = \mathrm{C}.$$

4) Если все результаты наблюдений умножить на одно и то же число, то имеет место равенство:

$$\bar{Z} = \overline{CX} = C\bar{X}$$

5) Сумма отклонений результатов наблюдений от их средней, взвешенная с соответствующими частотами, равна нулю

$$\sum (x_i - \bar{X}) n_i = 0.$$

# Числовые характеристики вариационных рядов

## *Характеристики положения вариационного ряда*

***Свойства средней арифметической***

6) Если все результаты наблюдений увеличить (уменьшить) на одно и то же число, то средняя арифметическая увеличится (уменьшится) на то же число, т. е.:

$$\bar{Z} = \overline{X \pm C} = \bar{X} \pm C.$$

7) Если все частоты вариантов умножить на одно и то же число, то средняя арифметическая не изменится.

# Числовые характеристики вариационных рядов

## *Характеристики положения вариационного ряда*

- *Модой* $(\mathrm{M}_0^*\,(X))$ дискретного вариационного ряда называется вариант, имеющий наибольшую частоту.

- *Медианой* $(\mathrm{M}_e^*(\mathrm{X}))$ дискретного вариационного ряда называется вариант, делящий ряд на две равные части.

  - Если дискретный вариационный ряд имеет $2n$ членов в ранжированной совокупности: $x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_{2n}$, то

$$M_e^*(X) = \frac{x_n + x_{n+1}}{2} \qquad (6)$$

# Числовые характеристики вариационных рядов

## *Характеристики положения вариационного ряда*

- Если дискретный вариационный ряд в ранжированной совокупности имеет $2n+1$ членов: $x_1, x_2, \dots, x_{n-1}, x_n, x_{n+1}, \dots, x_{2n+1}$, то

$$M_e^*(X) = x_{n+1} \qquad (7)$$

В примере 1:

$$\bar{X} = \frac{0 \cdot 6 + 1 \cdot 7 + 2 \cdot 3 + 3 \cdot 5 + 4 \cdot 3}{24} = \frac{40}{24} = \frac{5}{3} = 1{,}67;$$

$$M_0^*(X) = 1, \qquad M_e^*(X) = \frac{1+1}{2} = 1.$$

# Числовые характеристики вариационных рядов

## *Характеристики положения вариационного ряда*

Для **интервальных вариационных рядов** имеют место формулы:

а) медианы: $M_e^*(X) = x_{Me} + h \cdot \frac{0{,}5n - S_{Me-1}}{n_{Me}}$, (8)

где $x_{Me}$ - начало медианного интервала,

$h$ - длина частичного интервала, $n$ - объем совокупности,

$S_{Me\text{-}1}$ - накопленная частота интервала, предшествующего медианному,

$n_{Me}$ - частота медианного интервала;

б) моды: $M_0^*(X) = x_{Mo} + h \cdot \frac{(n_{Mo} - n_{Mo-1})}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} - n_{Mo+1})}$, (9)

где $x_{Mo}$ - начало модального интервала,

$h$ - длина частичного интервала, $n_{Mo}$ - частота модального интервала,

$n_{Mo\text{-}1}$ - частота предмодального интервала,

$n_{Mo+1}$ - частота послемодального интервала;

# Числовые характеристики вариационных рядов

*Характеристики положения вариационного ряда*

в) средней арифметической, совпадающей с формулой (6) для дискретного вариационного ряда, причем в качестве вариантов $x_i$ принимаются середины соответствующих интервалов.

**Мода и медиана используются в качестве характеристики среднего положения в случае, если границы ряда нечеткие или если ряд не симметричен.**

# Числовые характеристики вариационных рядов

## *Показатели вариации*

Показатели центральной тенденции ($M_0$, $M_e$, $\overline{X}$) не исчерпывают всех свойств распределения. В одних случаях значения признака концентрируются тесно около среднего значения, в других наблюдается значительное рассеяние.

Для изучения степени изменчивости признака вводят *показатели вариации*:

1) $W = x_{max} - x_{min}$ - *размах вариации*;

2) значения $x_i$ имеют свойство концентрироваться около $\bar{X}$, поэтому вводят следующие характеристики

(т. к $\sum(x_i - \bar{X})n_i = 0$):

# Числовые характеристики вариационных рядов

## *Показатели вариации*

- *Дисперсия* дискретного ряда распределения:

$$D^* = \frac{\sum(x_i - \bar{X})^2 n_i}{n} \qquad (10)$$

характеризует средний квадрат отклонения $x_i$, от $\bar{x}$.

- *Среднее квадратическое отклонение* дискретного ряда распределения

$$\sigma^* = \sqrt{\frac{\sum(x_i - \bar{X})^2 n_i}{n}} \qquad (11)$$

выражается в тех же единицах, что и $x_i$.

# Числовые характеристики вариационных рядов

## *Показатели вариации*

- *Среднее линейное отклонение:*

$$L(X) = \frac{\sum |x_i - \bar{X}| n_i}{n} \qquad (12)$$

- *Коэффициент вариации:*

$$V^* = \frac{\sigma^*}{\bar{X}} \cdot 100\%, \qquad (13)$$

характеризует относительное значение среднего квадратического отклонения и обычно служит для сравнения колеблемости несоизмеримых показателей.

# Числовые характеристики вариационных рядов

## *Показатели вариации*

***Свойства дисперсии:***

1) Дисперсия постоянной величины равна 0

$$D^*(C) = 0.$$

2) Если все результаты наблюдений увеличить (уменьшить) на одно и то же число С, то дисперсия и среднее квадратическое отклонение не изменятся, т. е.

$$D^*(X \pm C) \ = \ D^*(X), \sigma^*(\mathrm{X} \pm \mathrm{C}) \ = \ \sigma^*(\mathrm{X}).$$

3) Если все результаты наблюдений умножить на одно и то же число, то имеет место равенство:

$$D^*(CX) \ = \ C^2 D(X), \qquad \sigma^*(\mathrm{CX}) \ = |\mathrm{C}|\sigma^*(\mathrm{X}).$$

# Числовые характеристики вариационных рядов

## *Показатели вариации*

***Свойства дисперсии:***

4) Если все частоты вариантов умножить на одно и то же число, то дисперсия и среднее квадратическое отклонение не изменятся.

5) Свойство минимальности дисперсии.

$$\frac{\sum (x_i - C)^2 n_i}{n} \to \min \ \text{при}\ C = \bar{X}.$$

- *Следствие 1.* Средний квадрат отклонений значений $x_i$ от их средней арифметической равен среднему квадрату отклонений $x_i$ от произвольной постоянной $a$ минус квадрат разности между средней арифметической $(\bar{X})$ и этой произвольной постоянной.

Пусть $\sigma^{*2} = \frac{\sum (x_i - \bar{X})^2 n_i}{n}$, $\quad \sigma_a^{*2} = \frac{\sum (x_i - a)^2 n_i}{n}$, тогда

$$\sigma_x^{*2} = \sigma_a^{*2} - (\bar{X} - a)^2.$$

## *Показатели вариации*

- *Следствие 2.* Дисперсия равна средней арифметической из квадратов значений признака минус квадрат средней арифметической:

$$\sigma_x^{*2} = \overline{X^2} - (\bar{X})^2.$$

6) *Правило сложения дисперсий.*

Если объединяются несколько распределений в одно, то общая дисперсия $\sigma_0^{*2}$ нового распределения равна средней арифметической из дисперсий объединяемых распределений, сложенной с дисперсией частных средних относительно общей средней нового распределения.

# Числовые характеристики вариационных рядов

## *Показатели вариации*

***Правило сложения дисперсий***

Иначе говоря, общая дисперсия равна сумме внутригрупповой и межгрупповой дисперсий:

$$\sigma_0^{*2} = \overline{\sigma^{*2}} + \delta_0^{*2}, \ (14) \quad \text{или} \quad \sigma_0^{*2} = \frac{\sum_{i,j}\left(x_{ij} - \overline{X_0}\right)^2 \cdot n_{ij}}{N} = \frac{\sum x_j^2 N_j}{N} - \left(\overline{X_0}\right)^2,$$

где $n_{ij}$ – частота *j*-го варианта *i*-го частного распределения

(*j*=1,…, *m*; *i*=1,2,…,*k*),

$x_{ij}$ - *j-й* вариант *i*-го частного распределения

(*j*=1,…, *m; i*=1, 2,…, *k*),

$n_i$ - объем *i*-го частного распределения,

$N_j = \sum_i n_{ij}$ - частота *j*-го варианта нового распределения,

*N* - объем нового распределения,

$\overline{X_i} = \frac{\sum_j n_{ij} x_i}{n_i}$ - средняя арифметическая

(*i*-го частного распределения, (*i*=1,…, к),

$\overline{X_0} = \frac{\sum x_j N_j}{N}$ - средняя арифметическая нового распределения,

| j<br>i | $x_1$ | $x_2$ | … | $x_m$ | $\Sigma$ |
|---|---|---|---|---|---|
| 1 | $n_{11}$ | $n_{12}$ | … | $n_{1m}$ | $n_1$ |
| 2 | $n_{21}$ | $n_{22}$ | … | $n_{2m}$ | $n_2$ |
| 3 | $n_{31}$ | $n_{32}$ | … | $n_{3m}$ | $n_3$ |
| … | … | … | … | … | … |
| k | $n_{k1}$ | $n_{k2}$ | | $n_{km}$ | $n_m$ |
| $\Sigma$ | $N_{11}$ | $N_2$ | | $N_m$ | N |

$\sigma_i^2 = \frac{\sum_j x_{ij}^2 n_{ij}}{n_i} - \left(\overline{X_i}\right)^2$ - дисперсия *i*-го частного распределения

$\overline{\sigma^{*2}} = \frac{\sum \sigma_i^2 n_i}{N}$ - внутригрупповая дисперсия,

$\delta^{*2} = \frac{\sum\left(\overline{X_i} - \overline{X_0}\right)^2}{N}$ - межгрупповая дисперсия.

# Числовые характеристики вариационных рядов

- *Моменты* для вариационных рядов в математической статистике находятся по формулам, аналогичным формулам для ДСВ:

$a_s^* = \frac{\sum x_i^s n_i}{n}$ - начальный момент *s*-го порядка,

$\mu_s^* = \frac{\sum (x_i - \bar{x})^s n_i}{n}$ - центральный момент *s*-го порядка,

$r_s^* = \frac{\sum (x_i - \bar{x})^s n_i}{n \sigma_x^{*s}}$ - основной момент *s*-го порядка,

$r_{s,h}^* = \frac{\sum (x_i - \bar{x})^s \cdot (y_i - \bar{y})^h n_i}{n \sigma_x^{*s} \cdot \sigma_y^{*h}}$ - основной момент порядка *s,h.*

- Соотношения между начальными и центральными моментами в математической статистике соответствуют таким формулам для ДСВ.

- *Коэффициент асимметрии:* $Sk^* = \frac{\sum (x_i - \bar{X})^3 n_i}{n \sigma^{*3}}$.

- *Эксцесс:* $Ex^* = \frac{\sum (x_i - \bar{X})^4 n_i}{n \sigma^{*4}} - 3$.

# Числовые характеристики вариационных рядов

- Рассчитаем среднюю арифметическую, дисперсию, коэффициенты асимметрии и эксцесса для примера 2.

Среднее значение признака:

$$\bar{X} = \frac{\sum x_i n_i}{\sum n_i} = \frac{516{,}18}{60} = 8{,}613.$$

Дисперсия и среднее квадратическое отклонение:

$$D^* = \frac{\sum (x_i - \bar{X})^2 n_i}{n} = \frac{358{,}869}{60} = 5{,}981,$$

$$\sigma^* = \sqrt{\frac{\sum (x_i - \bar{X})^2 n_i}{n}} = \sqrt{5{,}981} = 2{,}446.$$

Коэффициент вариации:

$$V^* = \frac{\sigma^*}{\bar{X}} \cdot 100\% = \frac{2{,}446}{8{,}613} \cdot 100\% = 28{,}4\%.$$

# Числовые характеристики вариационных рядов

- Рассчитаем среднюю арифметическую, дисперсию, коэффициенты асимметрии и эксцесса для примера 2.

**Вспомогательная таблица для расчета числовых характеристик ряда распределения**

| Группы предприятий по численности работников на 100 га сельхозугодий, чел | Среднее значение интервала ($X_i$) | Число хозяйств в группе ($n_i$) | $x_i n_i$ | $x_i - \bar{X}$ | $(x_i - \bar{X})^2 n_i$ | $\frac{x_i - \bar{X}}{\sigma^*}$ | $\left(\frac{x_i - \bar{X}}{\sigma^*}\right)^3 n_i$ | $\left(\frac{x_i - \bar{X}}{\sigma^*}\right)^4 n_i$ |
|---|---|---|---|---|---|---|---|---|
| 4-5,6 | 4,8 | 5 | 24 | -3,813 | 72,708 | -1,559 | -18.954 | 29,554 |
| 5,61 - 7,2 | 6,4 | 17 | 108,8 | -2,213 | 83,280 | -0,905 | -12,601 | 11,404 |
| 7,21-8,8 | 8 | 9 | 72 | -0,613 | 3,386 | -0,251 | -0,142 | 0,036 |
| 8,81-10,4 | 9,6 | 15 | 144 | 0,987 | 14,603 | 0,403 | 0,985 | 0,397 |
| 10,41-12 | 11,2 | 10 | 112 | 2,587 | 66.908 | 1,058 | 11,832 | 12,514 |
| 12,01-13,6 | 12,8 | 1 | 12,8 | 4,187 | 17,528 | 1,712 | 5,017 | 8,588 |
| 13,61-15,2 | 14,4 | 3 | 43,2 | 5,787 | 100,457 | 2,366 | 39.740 | 94,030 |
| **итого** | **-** | **60** | **516,8** | **-** | **358,869** | **-** | **25,876** | **156,523** |

# Числовые характеристики вариационных рядов

- *Коэффициент асимметрии:* $Sk^* = \frac{\sum(x_i - \bar{X})^3 n_i}{n\sigma^{*3}} = \frac{25,876}{60} = 0,43.$
- *Эксцесс:* $Ex^* = \frac{\sum(x_i - \bar{X})^4 n_i}{n\sigma^{*4}} - 3 = \frac{156,523}{60} - 3 = -0,39.$

**Выводы:** Плотность работников -

$\bar{X} \pm \sigma^* = 8,61 \pm 2,45,$ то есть от 6,16 до 11,06 чел. на 100га с/х угодий.

Коэффициент асимметрии недостаточно близко к нулю => распределение не симметрично.

Эксцесс $Ex^* \neq 0$ => возможно распределение отлично от нормального.

# Выборочный метод

- В реальных условиях обычно бывает трудно или экономически нецелесообразно, а иногда и невозможно исследовать всю совокупность, характеризующую изучаемый признак *(**генеральную совокупность**).*
- На практике широко применяется выборочное наблюдение, когда обрабатывается часть генеральной совокупности *(**выборочная совокупность**).*

# Выборочный метод

- Свойства (закон распределения и его параметры) генеральной совокупности неизвестны, поэтому возникает задача их оценки по выборке.
- Для получения хороших оценок характеристик генеральной совокупности необходимо, чтобы выборка была ***репрезентативной*** (представительной).
- Репрезентативность, в силу закона больших чисел, достигается случайностью отбора.

# Выборочный метод

**Различают 5 основных типов выборок.**

**1)** ***Собственно случайная***:

- *повторная* (элементы после выбора возвращаются обратно);
- *бесповторная* (выбранные элементы не возвращаются).

**2)** ***Типическая*** - генеральная совокупность предварительно разбивается на группы типических элементов, и выборка осуществляется из каждой.

Следует различать:

а) ***равномерные* выборки** (при равенстве объемов исходных групп в генеральной совокупности выбирается одинаковое количество элементов из каждой);

б) ***пропорциональные*** (численность выборок формируют пропорционально численностям или средним квадратическим отклонениям групп генеральной совокупности);

в) ***комбинированные*** (численность выборок пропорциональна и средним квадратическим отклонениям, и численностям групп генеральной совокупности).

# Выборочный метод

**Различают 5 основных типов выборок.**

**3)** ***Механическая*** - отбор элементов проводится через определенный интервал.

**4)** ***Серийная*** - отбор проводится не по одному элементу, а сериями для проведения сплошного обследования.

**5)** ***Комбинированная*** - используются различные комбинации вышеуказанных методов, например, типическая выборка сочетается с механической и собственно случайной.

## Выборочный метод

- После осуществления выборки возникает задача оценки числовых характеристик генеральной совокупности по элементам выборочной совокупности.
- Различают *точечные* и *интервальные оценки*.
- ***Точечная оценка*** характеристики генеральной совокупности - это число, определяемое по выборке.
- Пусть $\hat{\theta} = \widehat{\theta_n}$ - выборочная характеристика, вычисленная по результатам *n* наблюдений величины *X,* используемая в качестве оценки θ - характеристики генеральной совокупности (в качестве θ может быть *M(X), D(X)* и т. д.).

# Выборочный метод

- Качество оценки $\hat{\theta}$ устанавливается по трем свойствам: **состоятельность, несмещенность, эффективность**.

***1) Состоятельность.*** Оценка $\widehat{\theta_n}$ является состоятельной оценкой генеральной характеристики θ, если для любого ε > 0 выполняется следующее равенство

$$\lim_{n\to\infty} P\left(\left|\widehat{\theta_n} - \theta\right| < \varepsilon\right) = 1$$

Это означает, что при увеличении объема выборки *n* выборочная характеристика $\widehat{\theta_n} \to \theta$.

***2) Несмещенность.*** Оценка $\hat{\theta}$ генеральной характеристики θ называется несмещенной, если для любого фиксированного числа наблюдений и выполняется равенство $M(\widehat{\theta_n})=\theta$.

***3) Эффективность***. Несмещенная оценка $\hat{\theta} = \widehat{\theta_n}$ генеральной характеристики $\theta$ называется несмещенной эффективной, если среди всех подобных оценок той же характеристики она имеет наименьшую дисперсию: $D(\widehat{\theta_n}) \to min$.

# Выборочный метод

- Можно показать, что статистики $\bar{X}$, $\hat{p}$ являются состоятельными, несмещенными и эффективными характеристиками математического ожидания *M(X)* и вероятности *p* соответственно.

- Выборочная дисперсия $\widehat{D}$ (далее $\widehat{D} = \sigma^2$) не обладает свойством *несмещенности*. На практике используют исправленную выборочную дисперсию $S^2$, которая является несмещенной оценкой дисперсии генеральной совокупности:

$$S^2 = \frac{n}{n-1}\sigma^2(x) = \frac{n}{n-1} \cdot \frac{\sum (x_i - \bar{X})^2 \cdot n_i}{n} = \frac{\sum (x_i - \bar{X})^2 \cdot n_i}{n-1}, \quad (15)$$

где S - стандартное отклонение.

# Выборочный метод

- Кроме того, в расчетах используют стандартную ошибку выборки:

$$S_x = \frac{S}{\sqrt{n}}. \qquad (16)$$

- Точечные оценки получают обычно с помощью ***метода моментов*** и ***метода максимального правдоподобия.***
- ***Интервальной*** называют оценку, которая определяется двумя числами - границами интервала. Она позволяет ответить на вопрос: <u>внутри какого интервала и с какой вероятностью находится неизвестное значение оцениваемого параметра $\vartheta$ генеральной совокупности</u>?

# Выборочный метод

- Пусть $\widehat{\theta}$ - точечная оценка параметра $\theta$. Чем меньше разность $\hat{\theta}$ и $\theta$, тем точнее и лучше оценка.
- Обычно говорят о *доверительной вероятности* (надежности оценки) $p = 1 - \alpha$, с которой $\theta$ будет находиться в интервале

$$\hat{\theta} - \Delta < \theta < \hat{\theta} + \Delta,$$

где: $\Delta(\Delta > 0)$ - предельная ошибка выборки, которая может быть либо задана наперед, либо вычислена; α - риск или уровень значимости (вероятность того, что неравенство будет неверным).

# Выборочный метод

- Оценка указанного доверительного интервала может быть получена (с наименьшей вероятностью) с помощью неравенства Чебышева (при $\varepsilon = \Delta$).
- В качестве 1-$\alpha$ принимают значения 0,90; 0,95; 0,99; 0,999.
- Доверительная вероятность показывает, что в (1-$\alpha$)100% случаев оценка $\theta$ будет накрываться указанным интервалом.

# Выборочный метод

- Точечная оценка математического ожидания $M(X)=a$ определяется как средняя арифметическая:

$$\bar{X} = \frac{1}{n}\sum x_i n_i \qquad (17)$$

- Точечная оценка вероятности $p_i$ определяется как относительная частота:

$$\widehat{p_i} = \frac{n_i}{n}. \qquad (18)$$

# Выборочный метод

- Для построения доверительного интервала параметра *a* - математического ожидания нормального распределения составляют выборочную характеристику *(статистику),* функционально зависимую от наблюдений и связанную с *a,* например, для повторного отбора:

$$u = \frac{\bar{X}-a}{\sigma(\bar{X})} = \frac{\bar{X}-a}{\frac{\sigma}{\sqrt{n}}} \qquad (19)$$

# Выборочный метод

- Статистика $u$ распределена по нормальному закону распределения с математическим ожиданием $a = 0$ и средним квадратическим отклонением $\sigma=1$.

Отсюда

$$P\left(|u| < u_{\alpha/2}\right) = 1 - \alpha,$$

$$\text{или } 2\Phi\left(u_{\alpha/2}\right) = 1 - \alpha,$$

где $\Phi$ - функция Лапласа, $u_{\alpha/2}$ - *квантиль* нормального закона распределения, соответствующая уровню значимости $\alpha$.

# Выборочный метод

- *Доверительный интервал* для параметра *a*:

$$\bar{X} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}} < a < \bar{X} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \qquad (20)$$

где $\Delta_{\bar{X}} = u_{\alpha/2}\sigma(\bar{X})$ - предельная ошибка выборочной средней.

# Выборочный метод

## Формулы предельной ошибки и необходимого объема выборки для различных случаев отбора

| Выборка | | Собственно-случайная | | Типическая | | Серийная | |
|---|---|---|---|---|---|---|---|
| | | повторная | бесповторная | повторная | бесповторная | повторная | бесповторная |
| Предельная ошибка, Δ | Средней, $\bar{x}$ | $t\sqrt{\frac{\sigma^2}{n}}$ | $t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$ | $t\sqrt{\frac{\overline{\sigma^2}}{n}}$ | $t\sqrt{\frac{\overline{\sigma^2}}{n}\left(1-\frac{n}{N}\right)}$ | $t\sqrt{\frac{\delta^2_{\text{м.с}}}{n_{\text{с}}}}$ | $t\sqrt{\frac{\delta^2_{\text{м.с}}}{n_{\text{с}}}\left(1-\frac{n_{\text{с}}}{N_{\text{с}}}\right)}$ |
| | Доли, $\hat{P}$ | $t\sqrt{\frac{pq}{n}}$ | $t\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}$ | $t\sqrt{\frac{\overline{pq}}{n}}$ | $t\sqrt{\frac{\overline{pq}}{n}\left(1-\frac{n}{N}\right)}$ | $t\sqrt{\frac{pq_{\text{м.с}}}{n_{\text{с}}}}$ | $t\sqrt{\frac{pq_{\text{м.с}}}{n_{\text{с}}}\left(1-\frac{n_{\text{с}}}{N_{\text{с}}}\right)}$ |
| Необходимая численность, n | Средней, $\bar{x}$ | $\frac{t^2\sigma^2}{\Delta^2}$ | $\frac{t^2\sigma^2 N}{t^2\sigma^2+\Delta^2 N}$ | $\frac{t^2\overline{\sigma^2}}{\Delta^2}$ | $\frac{t^2\overline{\sigma^2}N}{t^2\overline{\sigma^2}+\Delta^2 N}$ | $\frac{t^2\delta^2_{\text{м.с}}}{\Delta^2}$ | $\frac{t^2\delta^2_{\text{м.с}}N_{\text{с}}}{t^2\delta^2_{\text{м.с}}+\Delta^2 N_{\text{с}}}$ |
| | Доли, $\hat{P}$ | $\frac{t^2 pq}{\Delta^2}$ | $\frac{t^2 Npq}{t^2 pq+\Delta^2 N}$ | $\frac{t^2\overline{pq}}{\Delta^2}$ | $\frac{t^2 N\overline{pq}}{t^2\overline{pq}+\Delta^2 N}$ | $\frac{t^2 pq_{\text{м.с}}}{\Delta^2}$ | $\frac{t^2 N_{\text{с}}pq_{\text{м.с}}}{t^2 pq_{\text{м.с}}+\Delta^2 N_{\text{с}}}$ |

# Выборочный метод

**Формулы предельной ошибки и необходимого объема выборки для различных случаев отбора** (пояснения к таблице)

- t - квантиль распределения, соответствующая уровню значимости α,

а) при $n \geq 30$ $t = u_{\alpha/2}$ - квантиль нормального закона распределения (прил.1),

б) при $n < 30$ t - квантиль распределения Стьюдента с $v = n - l$ степенями свободы для двусторонней области;

- $\sigma^2$ - выборочная дисперсия,

а) при $n \geq 30$ $\sigma^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{n}$,

б) при $n < 30$ вместо $\sigma^2$ берут $S^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{n-1}$;

- pq - дисперсия относительной частоты в схеме повторных независимых испытаний;

# Выборочный метод

**Формулы предельной ошибки и необходимого объема выборки для различных случаев отбора** (пояснения к таблице)

- N - объем генеральной совокупности;
- n- объем выборки;
- $\overline{\sigma^2}$ - средняя арифметическая групповых дисперсий (внутригрупповая дисперсия);
- $\overline{pq}$- средняя арифметическая дисперсий групповых долей;
- $\delta^2_{\text{м.с}}$ -межсерийная дисперсия;
- Pqм.с - межсерийная дисперсия доли;
- Nc - число серий в генеральной совокупности;
- nc - число отобранных серий (объем выборки);