

## ПРАКТИЧЕСКАЯ РАБОТА №5

**Тема:** Выборочный метод

**Цель:** Научиться рассчитывать ошибки выборки и доверительные интервалы.

**Оборудование:** вычислительная техника, чертежные инструменты.

### *Теоретическое обоснование*

Выборочное наблюдение в настоящее время находит достаточно широкое применение в обследованиях промышленных и сельскохозяйственных предприятий, изучении цен на потребительском рынке, в обследованиях бюджетов и занятости населения. Выборочный метод является важнейшим источником информации в контроле качества продукции, в маркетинговых и социологических исследованиях.

**Выборочным наблюдением** называется такое несплошное обследование, при котором признаки регистрируются у отдельных единиц изучаемой статистической совокупности, отобранных с использованием специальных методов, а полученные в процессе обследования результаты с определенным уровнем вероятности распространяются на всю исходную совокупность.

Совокупность единиц, из которых производится отбор, принято называть **генеральной совокупностью**. Совокупность отобранных единиц из генеральной совокупности называется **выборочной совокупностью**.

**Метод отбора** определяет возможность продолжения участия отобранной единицы в процедуре отбора.

**Бесповторным** называется такой отбор, при котором попавшая в выборку единица не возвращается в совокупность, из которой осуществляется дальнейший отбор. **При повторном отборе** попавшая в выборку единица после регистрации наблюдаемых признаков возвращается в исходную (генеральную) совокупность для участия в дальнейшей процедуре отбора.

**Выборка должна быть репрезентативной**, т. е. правильно отражать пропорции генеральной совокупности. Это достигается **случайностью отбора**, когда все объекты генеральной совокупности имеют одинаковую вероятность быть отобранными.

Однако вычисленные по материалам выборочного наблюдения статистические показатели не будут точно совпадать с соответствующими характеристиками для всей совокупности (генеральной совокупности). Величина этих отклонений называется **ошибкой наблюдения**.

Величина случайной ошибки определяет надежность данных выборочного наблюдения, их пригодность для суждения о генеральной совокупности.

Ошибки выборки рассчитываются по известным из теории вероятностей формулам. Бывают **средние ошибки и предельные ошибки**.

**Условные обозначения:**

$N$  - объем генеральной совокупности (число входящих в нее единиц);

$n$  - объем выборочной совокупности (число единиц, попавших в выборку);

$\bar{x}$  - генеральная средняя (среднее значение признака в генеральной совокупности);

$\tilde{x}$  - выборочная средняя (среднее значение признака в выборочной совокупности);

$p$  - генеральная доля (доля единиц, обладающих данным признаком в генеральной совокупности);

$w$  - выборочная доля (доля единиц, обладающих данным признаком в выборочной совокупности);

$S^2$  - выборочная дисперсия (дисперсия признака в выборочной совокупности).

|   | Способ отбора единиц                        |  |
|---|---|--|
|   | повторный                                   | бесповторный   |
| <b>Средняя ошибка <math>\mu</math></b><br>для средней       | $\mu_{\bar{x}} = \sqrt{\frac{S^2}{n}}$      | $\mu_{\bar{x}} = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$                        |
| для доли  | $\mu_p = \sqrt{\frac{w(1-w)}{n}}$           | $\mu_p = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$                             |
| <b>Предельная ошибка <math>\Delta</math></b><br>для средней | $\Delta_{\bar{x}} = t \sqrt{\frac{S^2}{n}}$ | $\Delta_{\bar{x}} = t \mu_{\bar{x}} = t \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$ |
| для доли  | $\Delta_p = t \sqrt{\frac{w(1-w)}{n}}$      | $\Delta_{\bar{x}} = t \mu_p = t \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$      |

При обобщении результатов выборочного наблюдения наиболее часто используются следующие уровни вероятности ( $P$ ) и соответствующие им значения  $t$ :

|     |       |       |       |       |
|-----|-------|-------|-------|-------|
| $P$ | 0,683 | 0,950 | 0,954 | 0,997 |
| $t$ | 1     | 1,96  | 2     | 3     |

Например, если при расчете предельной ошибки выборки мы используем значение  $t=2$ , то с вероятностью 0,954 можно утверждать что расхождение между выборочной средней и генеральной средней не превысит двукратной величины средней ошибки выборки.

Определение пределов генеральных характеристик с заданной степенью надежности (доверительной вероятностью) на основе показателей, полученных по данным выборки, осуществляется следующим образом:

Доверительные интервалы для генеральной средней -

$$\bar{x} = \tilde{x} \pm \Delta_{\bar{x}} \quad \text{т.е.} \quad \tilde{x} - \Delta_{\bar{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\bar{x}}$$

Доверительные интервалы для генеральной доли -

$$p = w \pm \Delta_p \quad \text{т.е.} \quad w - \Delta_p \leq p \leq w + \Delta_p$$

### **Порядок выполнения**

- 1) Ознакомиться с теоретическим обоснованием.
- 2) Записать тему, цель, оборудование, задание в отчет.
- 3) Выполнить задание.
- 4) Сделать анализ полученных результатов.
- 5) Ответить на контрольные вопросы в отчете.
- 6) Сделать вывод по работе.

### **Задание:**

Рассчитать среднюю и предельную ошибки выборки и определить пределы генеральных характеристик с заданной степенью надежности (0,954) на основе показателей, полученных по данным выборки.

**Исходные данные:** В таблице сгруппированы данные о возрасте студентов ВУЗа при бесповторной 5%-ной выборке.

| Вариант | Количество студентов<br>(чел.) | Возраст (лет) |       |       |       |       |
|---------|--------------------------------|---------------|-------|-------|-------|-------|
|         |                                | 14-16         | 16-18 | 18-20 | 20-22 | 22-24 |
| 0       |                                | 13            | 15    | 24    | 18    | 7     |
| 1       |                                | 10            | 18    | 32    | 23    | 17    |
| 2       |                                | 8             | 10    | 14    | 20    | 22    |
| 3       |                                | 12            | 18    | 20    | 16    | 15    |
| 4       |                                | 16            | 23    | 14    | 18    | 24    |
| 5       |                                | 14            | 16    | 21    | 24    | 19    |

### **Ход работы:**

1. Выпишите данные своего варианта.
2. Составьте расчетную таблицу вида:

| Возраст, лет, $x$ | Количество студентов, $f$ | Середина интервала, $x'$ | $x' \cdot f$ | $(x'_i - \bar{x})^2$ | $(x'_i - \bar{x})^2 \cdot f_i$ |
|-------------------|---------------------------|--------------------------|--------------|----------------------|--------------------------------|
|                   |                           |                          |              |                      |                                |
| Итого:            | $\Sigma$                  |                          | $\Sigma$     |                      | $\Sigma$                       |

3. Найдите выборочную среднюю, используя формулу средней арифметической взвешенной.
4. Найдите дисперсию.
5. Найдите объем генеральной совокупности.
6. Найдите среднюю и предельную ошибки выборки с доверительной вероятностью 0,954.
7. Установите границы генеральной средней и сделайте вывод.
8. Найдите выборочную долю несовершеннолетних студентов.
9. Определите границы для выборочной доли несовершеннолетних студентов с той же вероятностью и сделайте вывод.

**Пример выполнения задания (рассмотрен вариант 0)**

1-2)

| Возраст, лет, x | Количество студентов, f | Середина интервала, x' | x' · f | (x <sub>i</sub> ' - $\bar{x}$ ) <sup>2</sup> | (x <sub>i</sub> ' - $\bar{x}$ ) <sup>2</sup> · f <sub>i</sub> |
|-----------------|-------------------------|------------------------|--------|--|---|
| 14-16           | 13                      | 15                     | 195    | 14,21  | 184,73  |
| 16-18           | 15                      | 17                     | 255    | 3,13   | 46,95   |
| 18-20           | 24                      | 19                     | 456    | 0,05   | 1,2   |
| 20-22           | 18                      | 21                     | 378    | 4,97   | 89,46   |
| 22-24           | 7                       | 23                     | 161    | 17,89  | 125,23  |
| Итого:          | 77                      |                        | 1445   |  | 447,57  |

$$3) \quad \tilde{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i} = \frac{1445}{77} = 18,77 \text{ лет} - \text{средний возраст студентов в выборке}$$

$$4) \quad S^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i} = \frac{447,57}{77} = 5,81 \text{ лет}$$

5) Т.к. количество студентов, попавших в 5%-ную выборку – 77 чел., следовательно, общее количество студентов, обучающихся в ВВУЗе –  $77 \times 100 / 5 = 1540$  человек.

6)

$$\mu_{\bar{x}} = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{5,81}{77} \left(1 - \frac{77}{1540}\right)} = 0,27 \text{ лет}$$

$$\Delta_{\bar{x}} = t \mu_{\bar{x}} = 2 \times 0,27 = 0,54 \text{ лет}$$

7)

$$\tilde{x} - \Delta_{\bar{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\bar{x}}$$

$$18,77 - 0,54 \leq \bar{x} \leq 18,77 + 0,54$$

$$18,23 \leq \bar{x} \leq 19,31$$

Т.о., средний возраст всех студентов ВВУЗа находится в пределах от 18,2 лет до 19,3 лет.

8) Т.к. в выборку попали 28 несовершеннолетних студентов (14-18 лет), то выборочная доля составит:  $28 / 77 \times 100 = 36\%$ . Т.е.  $w = 0,3636$ .

$$9) \quad \mu_p = \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{0,3636(1-0,3636)}{77}} = 0,0548$$

$$\Delta_p = 2 \times \mu_p = 2 \times 0,0548 = 0,2192$$

$$w - \Delta_p \leq p \leq w + \Delta_p$$

$$0,3636 - 0,2192 \leq p \leq 0,3636 + 0,2192$$

$$0,1444 \leq p \leq 0,5828$$

Т.о. количество несовершеннолетних студентов в ВВУЗе может быть от 14,4 до 58,3%.