Раздел 6. Математическая статистика и её роль в медицине.

Тема 6.1. Выборки и выборочные распределения

План

1. **Задачи математической статистики.**
2. **Генеральная совокупность и выборка.**
3. **Вариационный ряд. Гистограмма. Полигон.**
4. **Характеристики положения и рассеяния статистического распределения.**
5. **Оценка параметров генеральной совокупности по её выборке.**

Задачи математической статистики

- **Математическая статистика** – это раздел математики, изучающий методы сбора, систематизации и обработки результатов наблюдений массовых случайных явлений с целью выявления существующих закономерностей. Единственный способ получения информации о случайной величине – это проведение экспериментов. Все характеристики должны быть получены по экспериментальным данным.
- Одна из основных задач математической статистики состоит в том, чтобы по экспериментальным данным сделать выводы о параметрах распределения. Например, определить их приближенные значения (оценки) и указать ошибку их определения.

- Важным разделом является статистическая проверка предположений (гипотез) о законах распределения случайных величин, равенстве математических ожиданий, дисперсии и т.п.

Основные задачи математической статистики:

1. статистическое оценивание параметров законов распределения.
2. статистическая проверка гипотез.

Генеральная совокупность и выборка

- Обычно исследования проводятся не на единичных, а на групповых объектах объединённых по какому-либо признаку. Совокупность таких относительно однородных, но индивидуально различных единиц наблюдения, объединяемых по некоторым качественным или количественным признакам, характеризующим эти объекты, называется **совокупностью**.
- **Опр.** Совокупность всех мыслимых наблюдений или мысленно возможных объектов исследования называется **генеральной**.

- Генеральная совокупность есть понятие условно математическое или абстрактное, а на практике обычно используется часть членов генеральной совокупности, которая носит название **выборки** или **выборочной совокупности**.
- Например, чтобы дать ответ об эффективности некоторого препарата для лечения гриппа, необходимо его проверить в отношении всех больных страдающих этим заболеванием на земном шаре. Такая группа больных относится к генеральной совокупности. На практике клиническая апробация препаратов проводится на ограниченном контингенте больных (выборочной совокупности).

- Сущность выборочного метода заключается в том, чтобы по свойствам части (выборки) судить о численных характеристиках целого (генеральной совокупности).

Ввиду неполного отображения выборкой статистических характеристик генеральной совокупности необходимо:

1. организовать получение выборки так, чтобы она наиболее полно характеризовала свойства и особенности генеральной совокупности (репрезентативность выборки);
2. в каждом конкретном случае устанавливать, с какой уверенностью можно перенести результаты выборочного наблюдения на всю генеральную совокупность.

- Для выполнения первого условия необходимо, чтобы выборка была типичной и объективной, что достигается использованием принципа случайного отбора объектов исследования из генеральной совокупностью.
- Выделяют два метода проведения исследования: повторный и бесповторный. В первом случае все объекты после проведения наблюдений над ними возвращаются обратно в генеральную совокупность. При бесповторном отборе выбранный объект обратно в генеральную совокупность не возвращается.

Вариационный ряд. Гистограмма. Полигон

- В ходе экспериментов, исследователь получает набор числовых данных, отражающих результаты измерений или наблюдений исследуемых объектов. Совокупность этих числовых данных представленных в виде последовательности результатов наблюдений x_1, x_2, x_3, ….., x_n - есть выборка из генеральной совокупности. Основная задача первичного статистического анализа состоит в том, чтобы по имеющимся экспериментальным данным охарактеризовать исследуемую генеральную совокупность небольшим числом параметров.

- Если полученные данные расположить в порядке убывания или возрастания числовых значений исследуемого признака, то такой ряд чисел будет называться **вариационным рядом**.
- В том случае, когда среди числовых данных есть одинаковые значения, их можно представить в виде таблицы. В первой строке таблицы указываются значения признака (варианты), а во второй – абсолютные или относительные частоты их встречаемости. Такое представление вариационного ряда называют **статистическим распределением**.
- **Опр.** Статистическим распределением выборки называют перечень вариант и соответствующих им частот или относительных частот.

- **Пример 3.1.** Ежедневное количество студентов, посещающих методический кабинет на протяжении ряда дней: 15,17,16,18,20,21,18,17,20,15,18,17,16,19,17,16,18, 19,18,19. Составить статистическое распределение выборки.

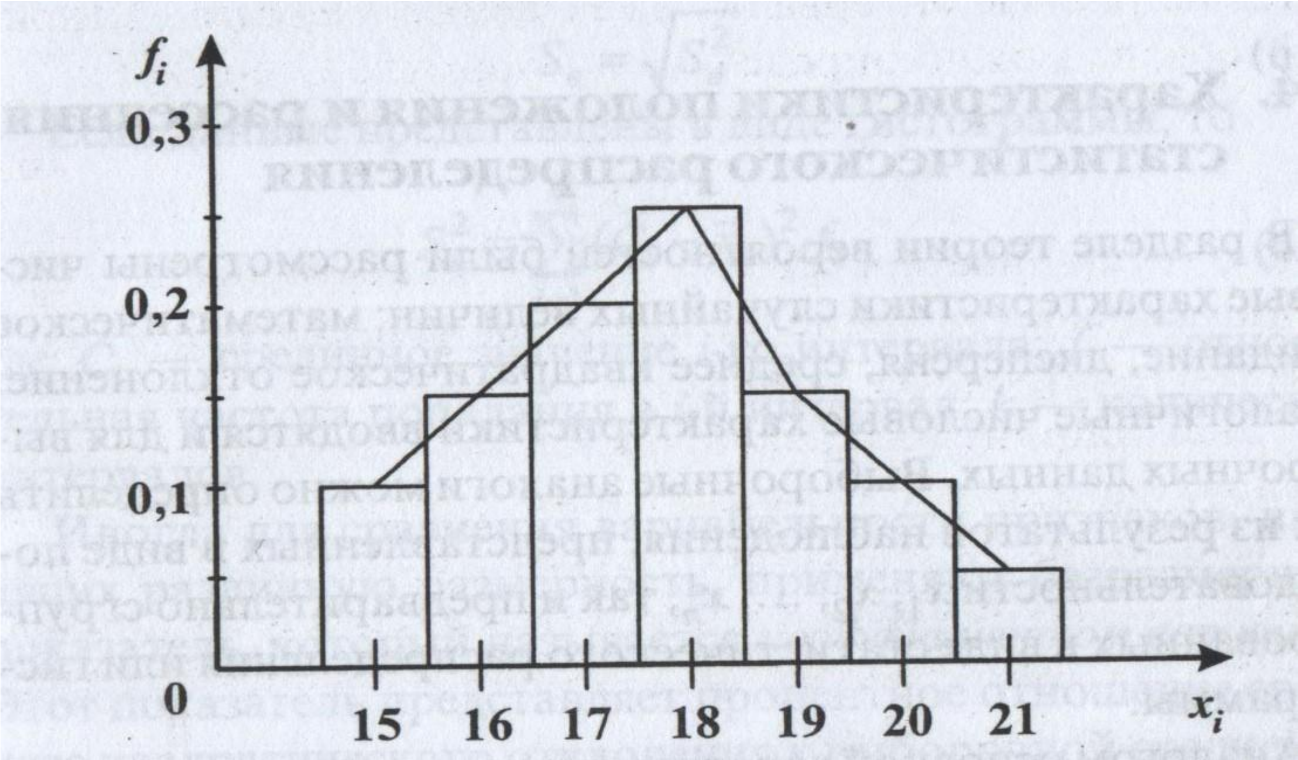
Значение признака x_i	15	16	17	18	19	20	21
Частота встречаемости m_i	2	3	4	5	3	2	1
	0,1	0,15	0,2	0,25	0,15	0,1	0,05

Для графического изображения статистического распределения строят полигоны и гистограммы.

- **<u>Опр.</u>** Гистограммой называется график, по оси абсцисс которого отложены границы классов, а по оси ординат – их частоты.
- Для построения гистограммы весь диапазон измеряемой величины разбивается на равные интервалы, называемые **классами**. Ширину интервала можно определить по формуле Стерджеса: $h = \frac{x_{max} - x_{min}}{1 + 3{,}32 \cdot \lg n}$, где h – ширина интервала, x_{max} – максимальное, x_{min} – минимальное значение выборочной величины, n – количество выборочных данных. Однако эта формула носит эмпирический характер и на практике количество интервалов выбирают в пределах 7-12. После выбора количества интервалов устанавливают границы классов (C_i) и срединные значения классов ($\overline{C}_i$), где $\overline{C}_i = \frac{C_i + C_{i+1}}{2}$ - середина i-го класса, i=1,2,3,…..,k – количество классов.

Пример 3.2. Построить гистограмму и полигон для примера 3.1.

Интервал	14,5-15,5	15,5-16,5	16,5-17,5	17,5-18,5	18,5-19,5	19,5-20,5	20,5-21,5
Частота встречаемости m_i	2	3	4	5	3	2	1
	0,1	0,15	0,2	0,25	0,15	0,1	0,05



Полигон частот можно получить из гистограммы путём соединения срединных значений классов. График полигона частот легко построить и по статистическому распределению. На оси абсцисс из точек x_i, проводятся перпендикуляры высотой $\frac{m_i}{n}$ и соединяются ломанной прямой.

Характеристики положения и рассеяния статистического распределения

- В разделе теории вероятностей были рассмотрены числовые характеристики случайных величин: математическое ожидание, дисперсия, среднее квадратическое отклонение. Аналогичные числовые характеристики вводятся и для выборочных данных.
- Аналогом основной характеристики положения математического ожидания случайной величины является **выборочная средняя**:
- $\overline{x_{в}} = \dfrac{\sum_{i=1}^{n} x_i}{n}.$

- Для характеристики рассеяния вариант относительно своей выборочной средней $\overline{x}_{\text{в}}$ вводят характеристику, называемую **выборочной дисперсией**, которая является аналогом дисперсии генеральной совокупности:
- $S_{\text{в}}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x_{\text{в}}})^2.$
- **Выборочное среднеквадратическое отклонение**: $S_{\text{в}} = \sqrt{s_{\text{в}}^2}.$

Оценка параметров генеральной совокупности по её выборке

- Характеристики нормального закона распределения M(X), D(X), δ(X) для генеральной совокупности представляют собой постоянные величины или параметры. По отношению к ним соответствующие выборочные характеристики $\overline{X}_{В}$, $S_{В}^{2}$, $S_{В}$ - являются **оценками** генеральных параметров.
- **<u>Опр.</u>** Оценкой параметра генеральной совокупности называют всякую однозначно определённую функцию результатов наблюдений, с помощью которой судят о значении параметра.

- **Оценкой математического ожидания** служит **выборочная средняя**, если данные представлены в виде вариационного ряда:
- $\bar{\mathrm{x}} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{x}_\mathrm{i} \cdot m_i$, где x_i - варианта выборки, m_i - частота встречаемости варианты, k - число классов.
- Для получения несмещённой точечной **оценки дисперсии** генеральной совокупности служит **исправленная выборочная дисперсия**, если данные представлены в виде вариационного ряда:
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x})^2 \cdot m_i.$

- **Оценкой среднего квадратического отклонения** выборочной средней, или ошибкой выборочной средней, будет:

- $m_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$

Пример 5.1.

Имеется выборка: 2,4,5,3,6,4. Найти выборочную среднюю, выборочную дисперсию и ошибку выборочной средней.

- $\bar{\mathrm{x}} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i} = \frac{2+4+5+3+6+4}{6} = 4;$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x})^2 =$$

$$\frac{(2-4)^2+(4-4)^2+(5-4)^2+(3-4)^2+(6-4)^2+(4-4)^2}{6-1} = 2;$$

$$m_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{2}{6}} = 0{,}58.$$