

Занятие 17

Тема. Предмет и основные задачи математической статистики. Первичная обработка данных наблюдений. Построение закона распределения по статистическим данным. Числовые характеристики выборки.

Математической статистикой называется наука, занимающаяся методами обработки экспериментальных данных, полученных в результате наблюдений над случайными явлениями. Первая задача математической статистики: указать способы сбора и группировки статистических данных, полученных в результате экспериментов. Вторая задача математической статистики: разработать методы анализа статистических данных.

Ко второй задаче относятся:

1. Оценка неизвестных параметров (вероятности события, функции распределения и её параметров и т.д.) с построением доверительных интервалов (методы оценивания).
2. Проверка статистических гипотез о виде неизвестного распределения и параметров распределения (методы проверки гипотез).

При этом решаются следующие в порядке сложности и важности задачи:

- Описание явлений, то есть, упорядочение поступившего статистического материала, представление его в наиболее удобном для обозрения и анализа виде (таблицы, графики).
- Анализ и прогноз, то есть приближённая оценка характеристик на основании статистических данных. Например, приближённая оценка математического ожидания и дисперсии наблюдаемой случайной величины и определение погрешностей этих оценок.
- Выработка оптимальных решений. Например, определение числа опытов n , достаточного для того, чтобы ошибка от замены теоретических числовых характеристик их экспериментальными оценками не превышала заданного значения. В связи с этим возникает задача проверки правдоподобия гипотез о параметрах распределения и о законах распределения случайной величины, решением которой является возможность сделать один из выводов:
 - отбросить гипотезу, как противоречащую опытным данным;
 - принять гипотезу, считать ее приемлемой.

Результатом проведенного статистического наблюдения является огромный массив статистических анкет, бланков, опросных листов и т.п., каждый из которых содержит сведения лишь об одной единице наблюдения.

Проанализировать весь этот разрозненный массив информации - задача не из легких... А вот если ее каким-либо образом сгруппировать, упорядочить, свести в единую информационную базу, то становится возможным и ее научный анализ. Таким образом, мы подошли к следующему этапу статистического исследования - сводке и группировке полученных статистических материалов.

Статистическая сводка - это сведение первичной статистической информации, полученной о единицах наблюдения, в упорядоченный по возрастанию или убыванию значений какого-либо признака (либо по определенному рангу) массив данных для выявления закономерностей в развитии исследуемого явления.

Результатом проведения статистической сводки является получение обобщающих статистических таблиц, которые содержат итоговые данные по показателям, характеризующим единицы наблюдения. Этими итоговыми данными могут быть суммарные значения показателей, рассчитанные как для всей совокупности в целом, так и для отдельных групп единиц, если производилась разбивка на группы; средние значения, расчетные относительные показатели. Чаще всего сводка осуществляется на основе проведения статистической группировки.

Под **статистической группировкой** понимается распределение единиц наблюдения на группы, однородные по одному или нескольким признакам. Эти признаки называются **группировочными**.

Сгруппированная информация позволяет лучше проанализировать типы экономических явлений, изучить их структуру, закономерности, связи между показателями, характеризующими единицы наблюдения. В зависимости от задач исследования строят типологические, структурные и аналитические группировки. Одной из важнейших задач математической статистики является *установление теоретического закона распределения случайной величины*, характеризующей изучаемый признак по опытному (эмпирическому) распределению, представляющему вариационный ряд.

Для решения этой задачи необходимо определить вид и параметры закона распределения.

Среднее арифметическое (M) – одна из основных характеристик выборки. Этот показатель характеризуется тем, что сумма отклонений от него выборочных значений (с учетом знака) равна нулю.

Для вычисления среднего арифметического сумму всех значений признака делим на объем выборки.

Пример: x_i : 20, 15, 15, 20, 30, среднее арифметическое равно 20. При этом сумма отклонений вариант от среднего арифметического равна нулю: сумма отклонений = $0 + (-5) + (-5) + 0 + 10 = 0$.

$$M = (20 + 15 + 15 + 20 + 30) / 5 = 20$$

Следует заметить, что среднее арифметическое измеряется в тех же единицах, что и признак. Например, если масса человека измеряется в кг, то и среднее арифметическое измеряется в кг.

Среднее арифметическое, вычисленное на основе выборочных данных, то есть данных, полученных на выборке, называется **выборочным средним арифметическим**. Оно обозначается как M . Среднее арифметическое генеральной совокупности называется генеральным средним. Оно обозначается буквой μ .

Мода (M_o) – характеристика положения. Представляет собой значение признака, встречающееся в выборке наиболее часто.

В качестве примера рассмотрим выборку: x_i : 3; 3; 3; 5; 5; 3; 4; 6; 7; 5; 3.

В выборке цифра «3» встречается 5 раз, поэтому $M_o = 3$.

Медиана (M_e) – характеристика положения, представляет собой такое значение признака, при котором одна половина значений меньше ее, а другая – больше.

В качестве примера рассмотрим выборку: x_i : 3; 3; 3; 5; 5; 3; 4; 6; 7; 5; 3.

Чтобы легко было определить медиану расположим варианты по возрастанию.

x_i : 3; 3; 3; 3; 3; **4**; 5; 5; 5; 6; 7. Варианта со значением «4» стоит в середине этой выборки. Это и есть медиана.

Средние значения не дают полной информации о вариации признака, поэтому наряду со средними значениями вычисляют характеристики вариативности.

К этим характеристикам относятся:

- размах вариации (R);
- дисперсия (S^2)
- стандартное отклонение (S)
- коэффициент вариации (V%)

Размах вариации

Размах вариации (R) вычисляется как разность между максимальным и минимальным значением признака:

$$R = X_{\max} - X_{\min}.$$

Размах вариации измеряется в тех же единицах, что и признак.

Информативность этого показателя невелика, так как эмпирические распределения результатов могут иметь одинаковый размах варьирования, а их форма будет очень отличаться.

Дисперсия

Дисперсия (S^2) – средний квадрат отклонений значений признака от среднего арифметического. Если признак измеряется в метрах, то дисперсия – в m^2 .

Это является недостатком, поэтому наиболее часто в публикациях приводится не дисперсия, а **стандартное отклонение (S)**. Этот показатель также называется среднеквадратическим отклонением или СКО. Стандартное отклонение представляет собой корень квадратный из дисперсии. *Чем больше стандартное отклонение, тем больше варьирует признак.*

Коэффициент вариации

Коэффициент вариации (V%). Чтобы сопоставить вариативность признаков, измеренных в различных единицах, используется относительный показатель (V%), который называется коэффициентом вариации.

Коэффициент вариации рассчитывается следующим образом. Стандартное отклонение делится на среднее арифметическое и умножается на 100%.

$$V\% = 100\% (S / M)$$

Например, если среднее арифметическое роста спортсменок равно $M=170$ см, а стандартное отклонение $S=5$ см, тогда коэффициент вариации равен: $V\% = 100\% (5/170)=2,94$.

Коэффициент вариации часто используют для оценки однородности выборки. Если $V < 10\%$ – выборка однородна, то есть, получена из одной генеральной совокупности.

Коэффициент асимметрии (As) характеризует “скошенность” эмпирического распределения. Если коэффициент асимметрии равен нулю – распределение симметричное. Если больше нуля – скошено влево, если меньше нуля – вправо.

Коэффициент эксцесса (Ex) определяет характер эмпирического распределения: остро- или плосковершинный.