Тема 3. ВЫЧИСЛЕНИЕ И АНАЛИЗ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК КОЛИЧЕСТВЕННОЙ ИЗМЕНЧИВОСТИ

Объяснение. Различают два типа изменчивости: *количественную*, которая может быть измерена, и *качественную*, которая не поддается измерению.

Под количественной изменчивостью понимают изменчивость количественных признаков − это такие свойства изучаемых объектов, разные состояния которых можно выразить при помощи чисел, например масса, высота, концентрация и др.

Различают два вида количественной изменчивости: прерывистую, или *дискретную*, и *непрерывную*. В первом случае различия между вариантами выражаются целыми числами, между которыми нет и не может быть переходов, например, число растений на квадратном метре, количество детенышей у животных и т. д. Во втором случае значения вариант выражаются мерами объема, длины, массы и т. д., между которыми возможны любые переходы с неограниченным числом возможных значений.

Выделяют также порядковые (ранговые) переменные, когда признаки невозможно измерить количественно, однако при этом они поддаются ранжированию в порядке возрастания степени проявления признака. Например, 10-балльная система оценивания в отечественной системе образования; оценка интенсивности запаха природных вод; степени агрессивности лабораторных мышей и т. д.

Качественной изменчивостью называется такое варьирование, когда различия между вариантами выражаются качественными показателями, которые одни варианты имеют, а другие нет (цвет, вкус, форма изучаемого предмета). Если признак принимает только два взаимоисключающих значения (больной – здоровый, безостый – остистый), то изменчивость называется *альтернативной*.

Основными статистическими характеристиками количественной изменчивости являются: средняя арифметическая ($\bar{x}$), дисперсия (S^2), среднее квадратическое отклонение (S), коэффициент вариации (V), ошибка средней арифметической ($S_{\bar{x}}$) и относительная ошибка ($S_{x\%}$).

Средняя арифметическая является абстрактной величиной, характеризующей всю совокупность в целом. В простейшем случае она находится путем сложения всех вариант выборки и деления на их число.

$$\bar{x} = \frac{\sum X}{n}$$

Как обобщающая характеристика средняя арифметическая ($\bar{x}$) обладает рядом свойств:

1. С увеличением или уменьшением индивидуальных значений выборки на или во сколько-то раз, средняя арифметическая изменяется таким же образом.
2. На числовой оси средняя арифметическая является центром колебания индивидуальных значений выборочной совокупности.
3. Сумма всех положительных и всех отрицательных отклонений от средней арифметической равна нулю.

$$\sum(X - \bar{x}) = 0$$

4. Сумма квадратов отклонений от средней арифметической при нормальном распределении меньше суммы квадратов отклонений от любого произвольного числа A, не равного средней:

$$\sum(X - \bar{x})^2 < \sum(X - A)^2$$

Средняя арифметическая не показывает индивидуальных колебаний изучаемых признаков. Основными мерами вариации, рассеяния изучаемого признака служат дисперсия и стандартное отклонение.

Выборочная несмещенная *дисперсия* представляет собой частное от деления суммы квадратов отклонений на число всех измерений без единицы (n – 1). Если количество элементов в выборке меньше 30, то знаменатель дроби принимает значение n.

$$S^2 = \frac{\sum(X - \bar{x})^2}{n - 1}$$

Из математической статистики известно, что при определении любых средних величин сумму всех показателей необходимо делить на число независимых друг от друга величин. В связи с этим в формулах сумму квадратов отклонений $\sum(X - \bar{x})^2$ делят не на общее число наблюдений, а на число без единицы, так как одно любое отклонение зависимое и может быть найдено из равенства $\sum(X - \bar{x})^2 = 0$. Остальные отклонения могут свободно варьировать, принимать любые значения. Число свободно варьирующих величин называется *числом степеней свободы.* Оно обозначается v и в простейшем случае равно n – 1.

Свойства дисперсии:

1. Дисперсия показывает возможную площадь рассеяния индивидуальных признаков от среднего значения.
2. Будучи показателем четной степени, она позволяет избавиться от отрицательных значений отклонений от средней, приводящих результаты сложения к нулю.

3. Размерность дисперсии равна квадрату размерности изучаемого признака.

Дисперсия представляет собой средний квадрат отклонений. То есть вначале рассчитывается среднее значение, затем берется разница между каждым исходным и средним значением, возводится в квадрат, складывается и затем делится на количество значений в данной совокупности. Разница между отдельным значением и средним арифметическим отражает меру отклонения. В квадрат возводится для того, чтобы все отклонения стали исключительно положительными числами и во избежание взаимоуничтожения положительных и отрицательных отклонений при их суммировании.

Для перевода меры рассеяния в величину с размерностью изучаемой величины применяют *стандартное или среднеквадратичное отклонение*. Его получают извлечением квадратного корня из дисперсии: $S = \sqrt{S^2}$.

При вычислении дисперсии и стандартного отклонения по основным формулам нередко возникают технические неудобства. Средняя арифметическая обычно получается в виде дробного числа, поэтому отклонения $(X - \bar{x})$ и особенно их квадраты $(X - \bar{x})^2$ получаются многозначными, что затрудняет расчеты и ведет к ошибкам. Поэтому разработаны способы вычислений, основанные на том, что для получения суммы квадратов центральных отклонений $(X - \bar{x})^2$ достаточно взять отклонения от любого произвольного числа A (удобно брать целое число, близкое к предполагаемой средней) и произвести расчеты по формуле:

$$\sum (X - \bar{x})^2 = \sum (X - A)^2 - \frac{[\sum (X - A)]^2}{n} = \sum X_1^2 - \frac{(\sum X_1)^2}{n}$$

Стандартное отклонение служит показателем, который дает представление о наиболее вероятной средней ошибке отдельного, единичного измерения, взятого из данной совокупности. В интервал плюс-минус стандартное отклонение относительно средней ($\bar{x} \pm 1S$) укладывается примерно ⅔ всех наблюдений (около 68% всех вариант), то есть основное ядро изучаемого ряда величин. Возможны отклонения от средней арифметической, превосходящие ($\bar{x} \pm 1S$), но вероятность их по мере удаления от $\pm 1S$ все время уменьшается. Так, вероятность встретить варианту, отклоняющуюся от средней на величину больше $\pm 2S$ составляет менее 5%, а больше $\pm 3S$ – менее 0,3%. Поэтому утроенное значение стандартного отклонения принято считать предельной ошибкой отдельного наблюдения.

Стандартное отклонение, выраженное в процентах к средней арифметической данной совокупности, называется *коэффициентом вариации* (V):

$$V = \frac{S}{\bar{x}} \cdot 100$$

Коэффициент вариации является относительным показателем изменчивости. Изменчивость принято считать незначительной, если коэффициент вариации не превышает 10%, средней, если V выше 10%, но менее 20%, и значительной, если коэффициент вариации более 20%.

Для характеристики степени выравненности материала иногда целесообразно использовать величину, дополняющую значение коэффициента вариации до 100. Этот показатель называют *коэффициентом выравненности* и определяют по равенству: $B = 100 - V$.

Вычисление основных статистических характеристик выборочным методом всегда связано с неизбежными ошибками. *Ошибка средней арифметической для выборочной совокупности* является мерой отклонения выборочной средней от средней генеральной совокупности. Она прямо пропорциональна стандартному отклонению и обратно пропорциональна корню квадратному из объема выборки:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

Ошибка средней арифметической обладает рядом свойств:

1. Ошибка средней арифметической подчиняется закону нормального распределения.

2. Ошибку средней арифметической можно регулировать, изменяя объем выборки. Для каждой величины ошибки объем выборки определяют по соотношению:

$$n = \left(\frac{S}{S_{\bar{x}}}\right)^2$$

Ошибка средней арифметической выражается в тех же единицах измерения, что и сама средняя. Поэтому сравнить между собой ошибочность определения средних разных признаков одной и той же совокупности бывает затруднительно или даже невозможно. Такое сравнение можно провести, если ошибку выразить в процентах. Ошибка выборки, выраженная в процентах от соответствующей средней, называется *относительной ошибкой выборочной средней*:

$$S_{x\%} = \frac{S_{\bar{x}}}{\bar{x}} \cdot 100$$

Относительная ошибка средней прямо пропорциональна абсолютной ошибке и обратно пропорциональна средней арифметической.

Приводя числовые данные в тексте или в таблицах вначале указывают среднее значение параметра, затем знак ±, а после него – значение стандартного отклонения или ошибки. Например, средняя продолжительность пребывания птицы у гнезда при кормлении птенцов составляет 8 ± 2 сек. При описании данных обязательно указать объем анализируемых выборок и показатель разброса, приведенный вместе со средним значением. Далее следует выполнить индивидуальное задание (приложения 4, 5).