***Тема №7. Нормально распределенные данные. Основные характеристики.***

Эту тему мы начнем с примера двух групп данных, про которые мы будем говорить, что для них отличны ***показатели разнообразия признака в совокупности***.

**Пример.***Две разные числовые последовательности: -100; -20; 100; 20 и 0,1; -0,2; 0,1. Для обеих из них средние арифметические значения абсолютно одинаковы и равны 0. Означает ли, что последовательности обладают одинаковыми характеристиками? Нет, так как диапазоны разброса данных этих последовательностей относительного среднего значения сильно различны.*

Для характеристики любой группы данных наряду со средним значением мы будем использовать показатели разнообразия. Они могут быть различны. Основными критериями разнообразия признака в статистической совокупности являются: лимит, амплитуда, среднее квадратическое отклонение, дисперсия, коэффициент осцилляции, коэффициент вариации, перцентильная оценка. На предыдущем занятии обсуждалось, что средние величины дают лишь обобщающую характеристику изучаемого признака в совокупности и не учитывают значения отдельных его вариант: минимальное и максимальное значения, выше среднего, ниже среднего и т.д.

Показатели измерения вариации признака бывают **абсолютные** и **относительные**. К абсолютным показателям вариации относят: размах вариации, лимит, среднее квадратическое отклонение, дисперсию, перцентильную оценку. Коэффициент вариации и коэффициент осцилляции относятся к относительным показателям вариации.

**Лимит (lim)–** это критерий, который определяется крайними значениями вариант в вариационном ряду. Другими словами, данный критерий ограничивается минимальной и максимальной величинами признака:

$$lim=v\_{min}÷v\_{max}$$

**Амплитуда (Am)**или **размах вариации –** это разность крайних вариант. Расчет данного критерия осуществляется путем вычитания из максимального значения признака его минимального значения, что позволяет оценить степень разброса вариант:

$$Am=v\_{max}-v\_{min}$$

Недостатком лимита и амплитуды как критериев вариабельности является то, что они полностью зависят от крайних значений признака в вариационном ряду. При этом не учитываются колебания значений признака внутри ряда.

Наиболее полную характеристику разнообразия признака в статистической совокупности дает **среднее квадратическое отклонение **$σ$ (сигма), которое является общей мерой отклонения вариант от своей средней величины. Среднее квадратическое отклонение часто называют также **стандартным отклонением**.

В основе среднего квадратического отклонения лежит сопоставление каждой варианты $v\_{i}$ со средней арифметической $M$ данной совокупности. Так как в совокупности всегда будут варианты как меньше, так и больше, чем она, то сумма отклонений $\left(v\_{i}-M\right)$, имеющих знак "$-$", будет погашаться суммой отклонений, имеющих знак "$+$", т.е. сумма всех отклонений $\sum\_{i=1}^{n}\left(v\_{i}-M\right)$ равна нулю. Для того, чтобы избежать влияния знаков разностей берут отклонения вариант от среднего арифметического в квадрате, т.е. $\left(v\_{i}-M\right)^{2}$. Сумма квадратов отклонений не равняется нулю. Чтобы получить коэффициент, способный измерить изменчивость, берут среднее от суммы квадратов – это величина носит название **дисперсии:**

$$D=\sum\_{i=1}^{n}\frac{(v\_{i}-M)^{2}}{n}$$

По смыслу, дисперсия – это средний квадрат отклонений индивидуальных значений признака от его средней величины. Дисперсия **–** квадрат среднего квадратического отклонения $σ^{2}$.

Дисперсия является размерной величиной (именованной). Так, если варианты числового ряда выражены в метрах, то дисперсия дает квадратные метры; если варианты выражены в килограммах, то дисперсия дает квадрат этой меры (кг2), и т.д.

**Среднее квадратическое отклонение** – квадратный корень из дисперсии:

$$σ=\sqrt{\sum\_{i=1}^{n}\frac{(v\_{i}-M)^{2}}{n}}=\sqrt{\frac{(v\_{1}-M)^{2}+(v\_{2}-M)^{2}+\cdots +(v\_{n}-M)^{2}}{n}}$$

**В том случае, если число элементов совокупности** $n<30$**, то при расчете дисперсии и среднего квадратического отклонения в знаменателе дроби вместо** $n$ **необходимо ставить** $(n-1)$**.**

Расчет среднего квадратического отклонения можно разбить на шесть этапов, которые необходимо осуществить в определенной последовательности:

1. определить среднюю арифметическую *M* имеющейся совокупности
2. рассчитать отклонение каждой варианты от средней величины: $d\_{i}=v\_{i}-M$
3. каждое отклонение возвести в квадрат: $d\_{i}^{2}$ (Для получения обобщающей характеристики числового ряда использовать сумму отклонений от среднего нельзя. Это связано с тем, что сумма всех отрицательных и положительных отклонений от среднего всегда равна нулю.)
4. посчитать сумму всех $d\_{i}^{2}$
5. разделить получившуюся сумму на число элементов совокупности *n*
6. из полученного результата извлечь квадратный корень

***Применение среднеквадратического отклонения:***

а) для суждения о колеблемости вариационных рядов и сравнительной оценки типичности (представительности) средних арифметических величин. Это необходимо в дифференциальной диагностике при определении устойчивости признаков.

б) для реконструкции вариационного ряда, т.е. восстановления его частотной характеристики на основе **правила «трех сигм»**. **В интервале (*М±3σ)*находится 99,7% всех вариант ряда, в интервале (*М±2σ)*— 95,5% и в интервале (*М±1σ)*— 68,3% вариант ряда** (рис.1).

в) для выявления «выскакивающих» вариант

г) для определения параметров нормы и патологии с помощью сигмальных оценок

д) для расчета коэффициента вариации

е) для расчета средней ошибки средней арифметической величины.

**Для характеристики любой генеральной совокупности, имеющей нормальный тип распределения, достаточно знать два параметра: среднюю арифметическую и среднее квадратическое отклонение.**



Рисунок 1. Правило «трех сигм»

**Пример.**

*В педиатрии среднеквадратическое отклонение используется для оценки физического развития детей путем сравнения данных конкретного ребенка с соответствующими стандартными показателями. За стандарт принимаются средние арифметические показатели физического развития здоровых детей. Сравнение показателей со стандартами проводят по специальным таблицам, в которых стандарты приводятся вместе с соответствующими им сигмальными шкалами. Считается, что если показатель физического развития ребенка находится в пределах стандарт (среднее арифметическое) ±σ, то физическое развитие ребенка (по этому показателю) соответствует норме. Если показатель находится в пределах стандарт ±2σ, то имеется незначительное отклонение от нормы. Если показатель выходит за эти границы, то физическое развитие ребенка резко отличается от нормы (возможна патология).*

Кроме показателей вариации, выраженных в абсолютных величинах, в статистическом исследовании используются показатели вариации, выраженные в относительных величинах. **Коэффициент осцилляции -** это отношение размаха вариации к средней величине признака. **Коэффициент вариации -** это отношение среднего квадратического отклонения к средней величине признака. Как правило, эти величины выражаются в процентах.

Формулы расчета относительных показателей вариации:

$$V\_{Am}=\frac{Am}{M}∙100\% V\_{Am}=\frac{σ}{M}∙100\%$$

Из приведенных формул видно, что чем больше коэффициент *V* приближен к нулю, тем меньше вариация значений признака. Чем больше *V*, тем более изменчив признак.

В статистической практике наиболее часто применяется коэффициент вариации. Он используется не только для сравнительной оценки вариации, но и для характеристики однородности совокупности. Совокупность считается однородной, если коэффициент вариации не превышает 33% (для распределений, близких к нормальному). Арифметически отношение σ и средней арифметической нивелирует влияние абсолютной величины этих характеристик, а процентное соотношение делает коэффициент вариации величиной безразмерной (неименованной).

Полученное значение коэффициента вариации оценивается в соответствии с ориентировочными градациями степени разнообразия признака:

- слабое — до 10 %

- среднее — 10 - 20 %

- сильное — более 20 %

Использование коэффициента вариации целесообразно в случаях, когда приходится сравнивать признаки разные по своей величине и размерности.

Отличие коэффициента вариации от других критериев разброса наглядно демонстрирует **пример**.

*Таблица 1*

Состав работников промышленного предприятия

|  |  |  |  |
| --- | --- | --- | --- |
| Учетный признак | Среднее арифметическое | Среднее квадратическое отклонение σ | Коэффициент вариации, % |
| Стаж работы (лет)  | 8,7 | 2,8 | 32,1 |
| Возраст (лет)  | 37,2 | 4,1 | 11,0 |
| Образование (классов)  | 9,2 | 1,1 | 11,9 |

На основании приведенных в примере статистических характеристик можно сделать вывод об относительной однородности возрастного состава и образовательного уровня работников предприятия при низкой профессиональной устойчивости обследованного контингента. Нетрудно заметить, что попытка судить об этих социальных тенденциях по среднему квадратическому отклонению привела бы к ошибочному заключению, а попытка сравнения учетных признаков «стаж работы» и «возраст» с учетным признаком «образование» вообще была бы некорректной из-за разнородности этих признаков.