Презентация по статистике. **Методы распределения.**

Для подробного описания особенностей распределения используются дополнительные характеристики, в частности, определяются *мо менты распределения.*

Моментом k-го порядка **называется средняя из *k-х* степеней от клонений** вариантов *х* от **некоторой постоянной величины *А:***

$$M_k = \overline{(x - A)^k}.$$

При использовании в качестве весов частот или частостей моменты называются *эмпирически ми,* а при использовании вероятностей — *теоретическими.*

Эмпирический момент *k-го* порядка*:*

$$M_k = \frac{\sum_i (x_i - A)^k f_i}{\sum_i f_i}$$

- 1. *Начальные моменты (М^)* получаются, если постоянная величина *А* равна нулю (Л = О):

$$M_k = \overline{(x_i - 0)^k} = \frac{\sum_i x_i^k f_i}{\sum_i f_i}.$$

- 2. *Условные и начальные относительно X0 моменты (тк)* получаются при *А* равном не нулю, а некоторой производной величине *X0* **(начало** отсчета):

$$m_k = \overline{(x_i - x_0)^k} = \frac{\sum_i (x_i - x_0)^k f_i}{\sum_i f_i}.$$

- С помощью условных моментов упрощается расчет основных характеристик ряда распределения. При подстановке различных значений *k* получаем начальные моменты относительно *Xo.* Так, например, если *k* = 1, то:

$$m_1 = \frac{\sum_i (x_i - x_0) f_i}{\sum_i f_i} = \frac{\sum_i x_i f_i}{\sum_i f_i} - \frac{x_0 \sum_i f_i}{\sum_i f_i} = \bar{x} - x_0.$$

- Из этой формулы вытекает, что *x = x0+m1* т.е. средняя арифметическая равна началу отсчета плюс начальный момент первого порядка. Если отклонения *(xi- x0)* имеют общий множитель *C,* то на него можно разделить отклонения, а по окончании вычислить полученный момент, умножив на этот множитель в соответствующей степени, т. е.:

$$m_k = \frac{\sum\limits_{i=1} \left(\frac{x_i - x_0}{C} \right)^k \cdot f_i}{\sum\limits_{i=1} f_i} \cdot C^k.$$

- Отсюда следует, что при *k* = 1 x=x0+m1*C.

- 3. *Центральные моменты* (μ k) получаются, если за постоянную величину *A* взять среднюю арифметическую *(A=x):*

$$\mu_k = \overline{(x_i - \bar{x}_k)^k} = \frac{\sum_i (x_i - \bar{x})^2 f_i}{\sum_i f_i}.$$

Закономерности распределения

Каждому ряду распределения свойственна определенная закономерность, выражением которой является кривая распределения, представляющая собой функцию распределения. Можно выделить определенную зависимость между изменением частот и изменением значений признаков: частоты изменяются закономерно с изменением варьирующего признака, т. е. с увеличением значения варьирующего признака частоты первоначально увеличиваются, затем, достигнув какой-то максимальной величины в середине ряда, уменьшаются. Такие закономерности изменения частот в вариационных рядах называются закономерностями распределения.

Эмпирическим распределением называют распределение частот (относительных частот), соответствующих отдельным значениям признака, функционально связанных с изменением вариант.

Если в качестве весов при расчете центрального момента взять не частоты (f), а вероятности (p), то получим теоретические моменты распределения. Отсюда –теоретическим называют распределение вероятностей.

Если имеется эмпирический ряд распределения, то необходимо найти функцию распределения, т. е. подобрать такую теоретическую кривую распределения, которая бы наиболее полно отражала закономерность распределения.

Под кривой распределения понимается графическое изображение в виде непрерывной линии изменения частот (вероятностей), функционально связанных с изменением вариант.

Закон распределения случайной величины может быть задан в виде таблицы, функции распределения либо плотности распределения.

В статистике широко используются различные виды теоретических распределений: распределение Стьюдента, Пуассона, нормальное распределение, хи-квадрат распределение, распределение Фишера, биномиальное (распределение Бернулли), равномерное распределение. Каждое из теоретических распределений имеет специфику и свою область применения в различных отраслях знаний.

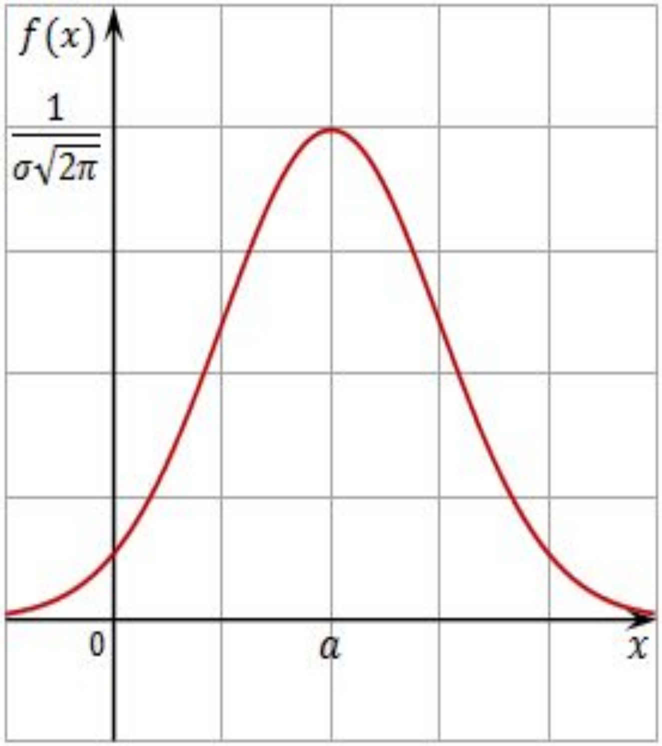
Первым фундаментальным по значимости является нормальный закон распределения (ЗНР).

Подчиненность закону нормального распределения тем точнее, чем больше факторов действует вместе. Часто возникают распределения, хотя и не отвечающие строго нормальному распределению, но имеющие с ним сходство, а именно: вероятность min и max значений тем меньше, чем больше отклонение отдельных вариант от общей средней. Иными словами: минимальные и максимальные варианты встречаются много реже, чем серединные.

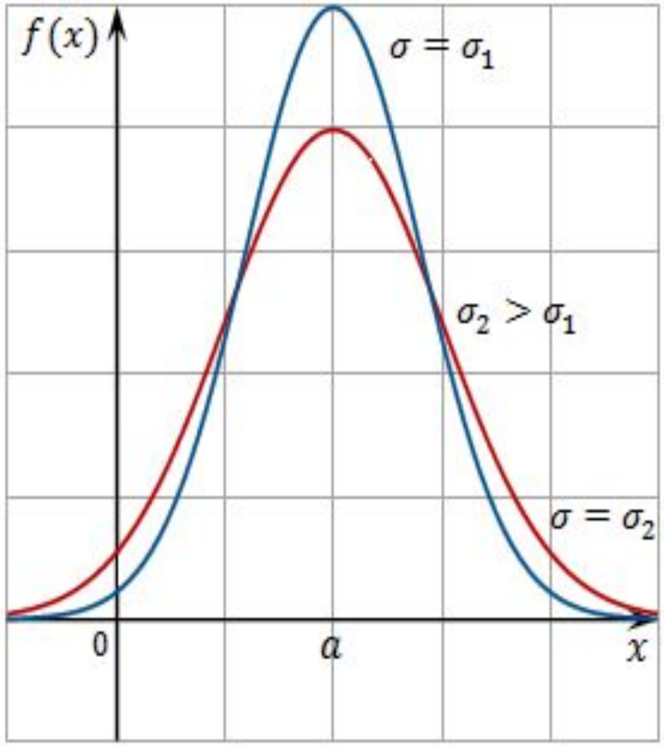
Нормальное распределение полностью определяется двумя входными параметрами: средней арифметической и среднеквадратическим отклонением (б).

Кривая распределения симметрична относительно точки максимума $x=a(\mu)$.

Если учесть величину среднеквадратического отклонения б, то окажется, что при больших значениях б значение плотности вероятности $f(x)$ мало и наоборот – при малых значениях б плотность вероятности (ордината точки максимума) неограниченно возрастает. Отсюда: среднеквадратическое отклонение нормально распределенной СВ существенно влияет на форму нормальной кривой. Максимальная ордината кривой обратно пропорциональна среднеквадратическому отклонению б. Вся площадь, ограниченная кривой распределения и осью абсцисс, равна 1.



f(x)
1/σ√2π
0
a
x



f(x)
σ = σ1
σ2 > σ1
σ = σ2
0
a
x

Плотность вероятности нормального распределения выражается следующей формулой:

$$f\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

или

$$f\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

t – нормированное отклонение:

$$t = \frac{x - \bar{x}}{\sigma}$$

В это выражение входит две константы:

$$\bar{x}\ (a;\ \mu);\ \ \sigma$$

$$M(X) = a\ (\mu) = \bar{x}; \quad D(X) = \sigma^2$$

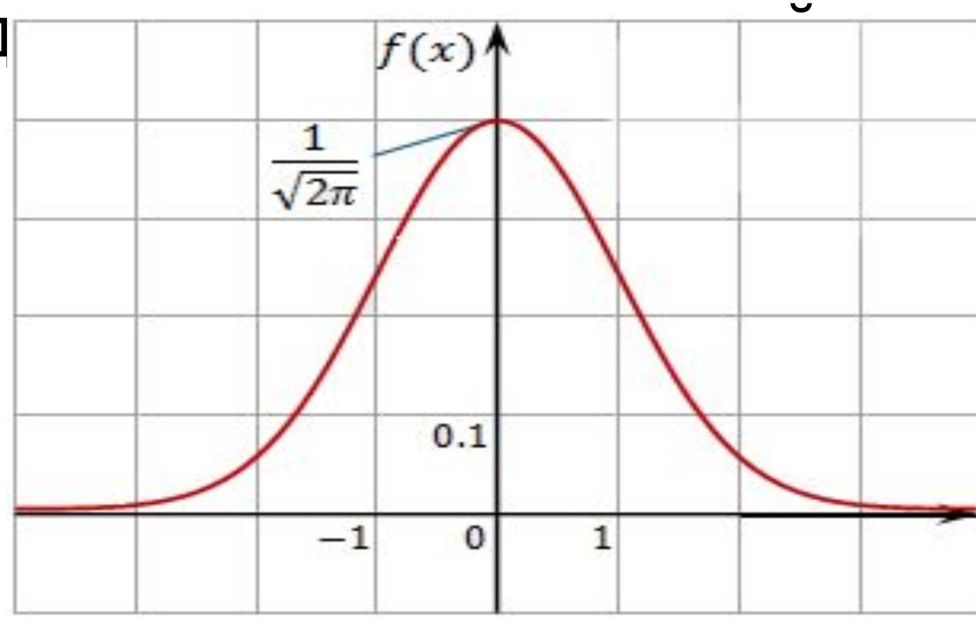
Это распределение характерно тем, что в соответствующих пределах заключено соответствующее количество всех частот:

$$\bar{x} \pm 1\sigma \rightarrow 0.683\ (68.3\%)$$

$$\bar{x} \pm 2\sigma \rightarrow 0.954\ (95.4\%)$$

$$\bar{x} \pm 3\sigma \rightarrow 0.997\ (99.7\%)$$

Последний результат означает, что с вероятностью, близкой к единице (0,9973), случайная величина, подчиняющаяся закону нормального распределения, не выйдет за пределы заданного интервала. Это утверждение называют правилом трёх сигм. Вероятность того, что СВ примет значение за пределами зад ... ала:
(1- 0,9973=0,0027)



Характеристика асимметрии и эксцесса

Выяснение общего характера распределения предполагает оценку его однородности, а также расчет показателей асимметрии и эксцесса.

При сравнительном изучении асимметрии нескольких распределений с разными единицами измерениявычисляется **относительный показатель асимметрии**:

$$A_S = \frac{\bar{x} - Mo}{\sigma}, \quad \textit{или} \quad A_S = \frac{\bar{x} - Me}{\sigma}$$

Его величина может быть положительной (для правосторонней асимметрии) и отрицательной (для левосторонней асимметрии).

Применение данного показателя дает возможность определить не только величину асимметрии, но и проверить ее наличие в генеральной совокупности. Принято считать, что асимметрия выше 0,5 (независимо от знака) считается значительной. Если асимметрия меньше 0,25, она считается незначительной.

Наличие асимметрии в генеральной совокупности проверяется с помощью определения оценки существенности на основе средней квадратической ошибки:

$$\sigma_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}$$

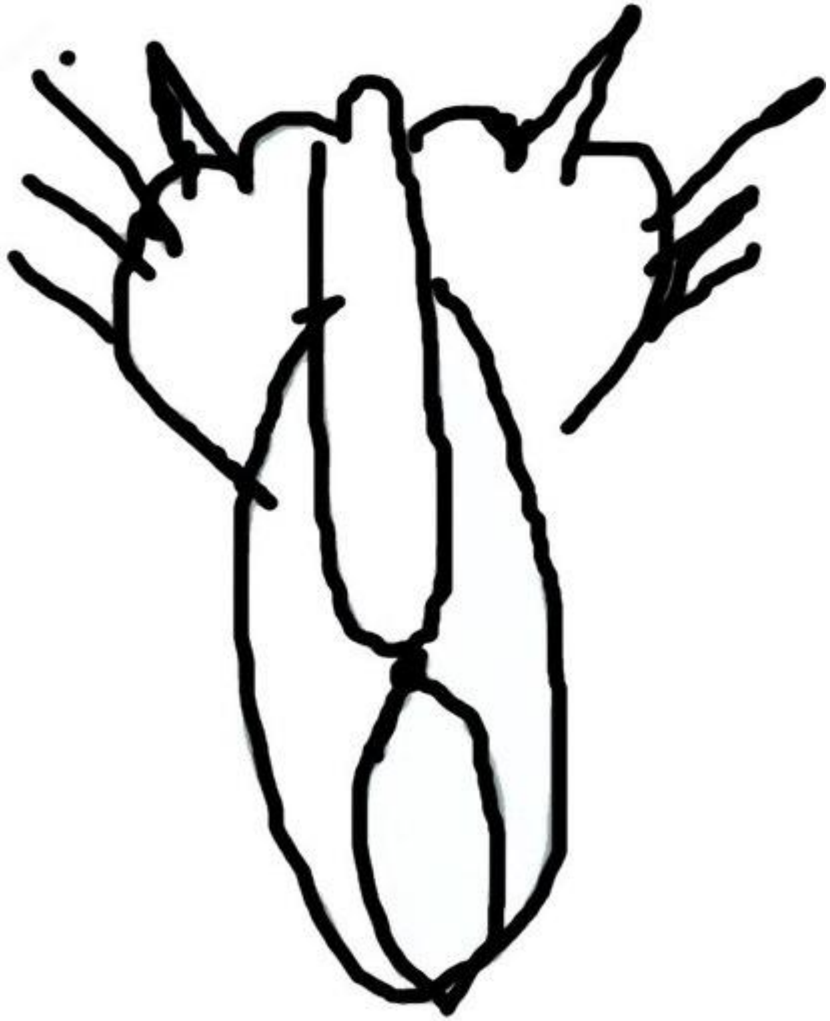
В случае, если $\frac{|A_s|}{\sigma_{A_s}} > 3$, асимметрия считается существенной
и
распределение признака в генеральной совокупности несимметрично и неслучайно, а закономерно.

Для симметричных распределений может быть рассчитан показатель эксцесса, который показывает, насколько резкий скачок имеет изучаемое явление. Показатель эксцесса определяется на основе центрального момента четвертого порядка по формуле:

$$E_k = \frac{\mu_4}{\sigma^4} - 3 = \frac{\frac{\sum (x_i - \bar{x})^4 f_i}{\sum f_i}}{\sigma^4} - 3$$

Если показатель эксцесса больше нуля, то распределение островершинное и скачок считается значительным, если коэффициент эксцесса меньше нуля, то распределение считается плосковершинным и скачок считается незначительным. Среднеквадратическая ошибка эксцесса показывает, насколько существенен скачок в явлении и рассчитывается по формуле:

$$\sigma_{E_k} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}$$



Спасибо за внимание!