Элементы
теории вероятностей и математической статистики

Цель: *повторить основные понятия теории вероятностей и математической статистики, которые используются в курсе эконометрики.*

ВОПРОСЫ

I. Случайные величины и их числовые характеристики.

II. Основные законы распределения случайных величин.

III. Точечные и интервальные оценки параметров.

IV. Статистические гипотезы и процедура их проверки.

I. Случайные величины и их числовые характеристики

Вероятность события –

это количественная мера, которая вводится для сравнения событий по степени возможности их появления.

Вероятность события A равна отношению числа случаев m, благоприятствующих ему, к общему числу случаев n, т.е.

$$P(A) = m / n$$

Случайной величиной называется переменная, которая под воздействием случайных факторов может с определенными вероятностями принимать те или иные значения из некоторого множества чисел.

Различают
дискретные и ***непрерывные***
случайные величины.

Дискретной называется случайная величина, которая принимает отдельные, изолированные друг от друга значения.

Число возможных значений дискретной случайной величины конечно или счётно.

Непрерывной называется случайная величина, множество значений которой непрерывно заполняет некоторый числовой промежуток.

Число возможных значений непрерывной случайной величины бесконечно и несчётно.

Полным описанием случайной величины является ее закон распределения.

Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Наиболее общей формой задания распределения случайной величины является функция распределения.

Она используется как для дискретных, так и для непрерывных случайных величин.

Функцией распределения случайной величины X называется функция $F(x)$, выражающая для каждого x вероятность того, что случайная величина X примет значение меньшее x:

$$F(x) = P(X < x)$$

Для *непрерывной* случайной величины вместо функции распределения обычно используют плотность вероятности.

Плотностью вероятности $f(x)$ непрерывной случайной величины X называется производная ее функции распределения:

$$f(x) = F'(x).$$

Для описания случайных величин часто используются ***числовые характеристики:***

- ✓математическое ожидание,
- ✓дисперсия,
- ✓среднее квадратическое отклонение.

Математическое ожидание $M(X)$ характеризует среднее ожидаемое значение случайной величины X, т. е. приближенно равно ее среднему значению.

Для дискретной случайной величины:

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

Для непрерывной случайной величины:

$$M(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Дисперсией случайной величины X называется математическое ожидание квадрата отклонения случайной величины от математического ожидания:

$$D(X) = M(X - M(X))^2$$

Дисперсия характеризует отклонение (разброс, рассеяние, вариацию) значений случайной величины относительно среднего значения.

Средним квадратическим отклонением случайной величины X называется квадратный корень из ее дисперсии.

$$\sigma(X) = \sqrt{D(X)}$$

II. **Основные законы распределения случайных величин**

Большинство случайных величин подчиняется определенному закону распределения.

В эконометрическом анализе часто используются нормальное распределение, распределения χ^2 ,Стьюдента, Фишера.

1. Нормальное распределение

Непрерывная случайная величина X имеет нормальный закон распределения с параметрами a и σ^2, если ее плотность вероятности имеет вид:

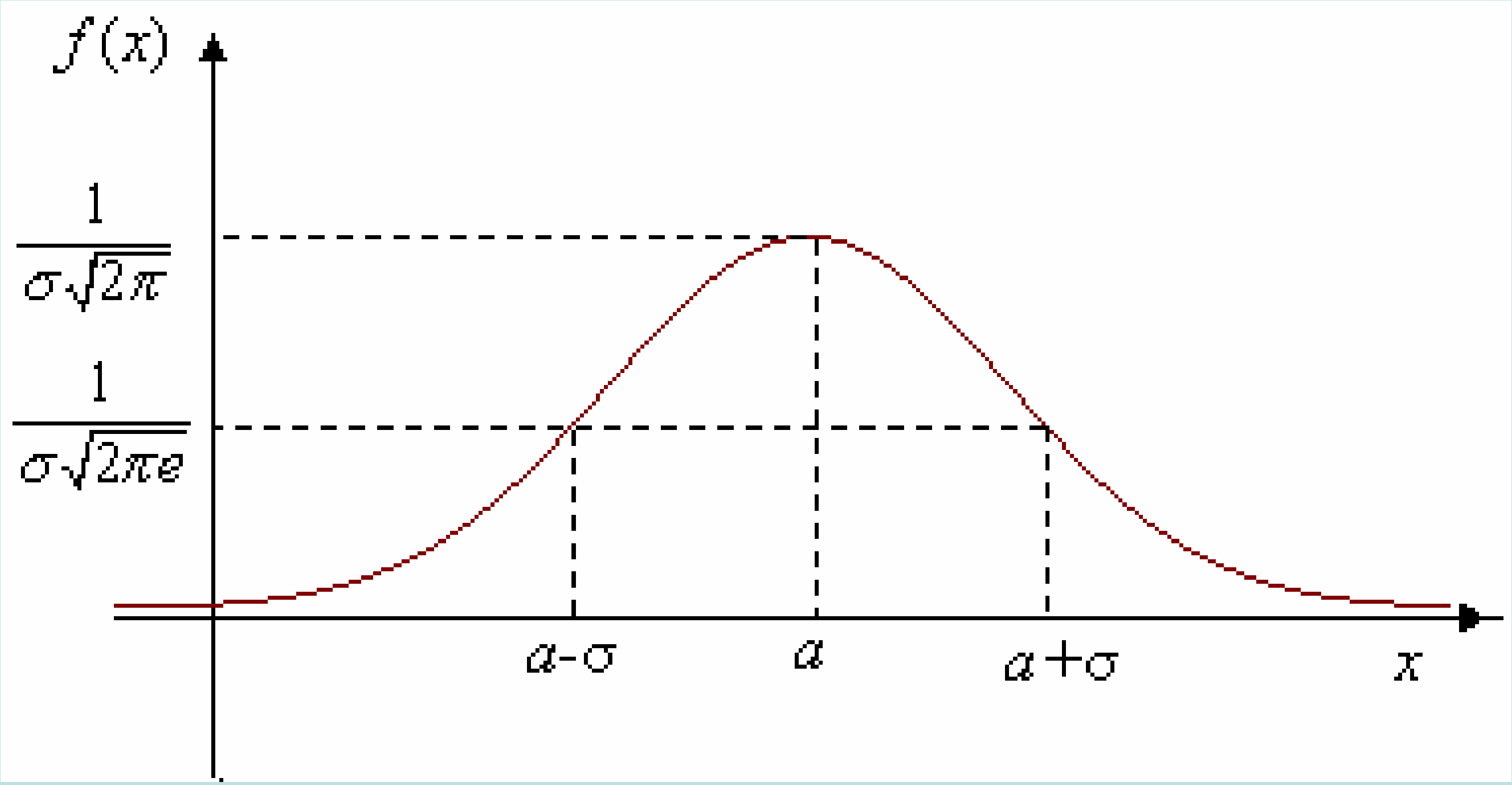
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

где a - математическое ожидание, σ^2 – дисперсия,

e = 2,7, π = 3,14

График функции плотности нормального распределения ***f (x)*** имеет колоколообразный вид, он симметричен относительно прямой ***x=a***, имеет максимум в точке ***x=a*** и две точки перегиба

$x_1 = a - \sigma$; $x_2 = a + \sigma$.



Нормальный закон распределения с параметрами a=0 и σ^2 =1 называется ***стандартным*** или ***нормированным***.

2. Распределение χ^2 (хи-квадрат)

Распределение χ^2 с k степенями свободы называется распределение суммы квадратов k независимых случайных величин, распределенных по стандартному нормальному закону, т.е.

$$\chi^2 = Z_1^2 + Z_2^2 + .. + Z_k^2 = \sum Z_i^2,$$

где случайная величина Z_i $(i=1,2,\ldots k)$ имеет нормальное стандартное распределение.

Число степеней свободы исследуемой случайной величины определяется числом случайных величин, ее составляющих, уменьшенных на число линейных связей между ними.

3. Распределение Стьюдента
(t - распределение) –

это распределение случайной величины

$$t = \frac{Z}{\sqrt{\frac{1}{k}\chi^2}}$$

где случайные величины $\boldsymbol{Z}$ и χ^2 независимы,

$\boldsymbol{Z}$ имеет стандартное нормальное распределение, а χ^2 – распределение хи – квадрат с $\boldsymbol{k}$ степенями свободы.

При $\boldsymbol{k} \to \infty$ $\boldsymbol{t}$ – распределение приближается к нормальному.

4. ***Распределение Фишера***

(F- распределение)– это распределение случайной величины

$$F = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}$$

где случайные величины $\chi^2(\boldsymbol{k}_1)$ и $\chi^2(\boldsymbol{k}_2)$ независимы и имеют распределения хи – квадрат с числом степеней свободы $\boldsymbol{k}_1$ и $\boldsymbol{k}_2$ соответственно.

III. Точечные и интервальные оценки параметров

Генеральная совокупность - это множество всех значений случайной величины, которые она может принять в процессе наблюдения.

Выборочная совокупность (***выборка***) - часть генеральной совокупности, отобранная для изучения.

Изучение генеральной совокупности во многих случаях невозможно или нецелесообразно в силу больших материальных затрат, уничтожения или порчи исследуемых объектов.

Задача математической статистики состоит в исследовании свойств выборки и обобщении этих свойств на генеральную совокупность.

При исследовании различных параметров генеральной совокупности на основе выборки возможно лишь получение ***оценок*** этих параметров.

Оценки строятся на основе ограниченного набора данных, что влечет за собой вероятность погрешности.

Процесс нахождения оценок по определенному правилу (формуле) называют ***оцениванием***.

Цель любого оценивания - получение наиболее точного значения оцениваемой характеристики.

Различают два вида оценок -
точечные и ***интервальные.***

Точечной оценкой $\boldsymbol{\theta}^*$ параметра $\boldsymbol{\theta}$ называется числовое значение этого параметра, полученное по выборке объема $\boldsymbol{n}$.

Так как выборка носит случайный характер, то оценка θ^* является случайной величиной, принимающей различные значения для различных выборок.

В силу случайности точечной оценки $\boldsymbol{\theta}^*$ она может рассматриваться как случайная величина со своими числовыми характеристиками –

математическим ожиданием и дисперсией.

Качество оценок характеризуется *основными свойствами:*

- ✓ *несмещённость,*
- ✓ *эффективность,*
- ✓ *состоятельность.*

1. Несмещенность оценок

Оценка $\boldsymbol{\theta}^*$ называется *несмещённой оценкой* параметра $\boldsymbol{\theta}$, если ее математическое ожидание равно оцениваемому параметру генеральной совокупности:

$$M(\theta^*) = \theta$$

Разность $M\ (\theta^*)\ -\ \theta$ называется *смещением*.

2. Эффективность оценок

Оценка $\boldsymbol{\theta}^*$ называется *эффективной оценкой* параметра $\boldsymbol{\theta}$, если ее дисперсия $D(\boldsymbol{\theta}^*)$ меньше дисперсии любой другой альтернативной оценки при фиксированном объеме выборки $\boldsymbol{n}$.

3. Состоятельность оценок

Оценка $\boldsymbol{\theta}^*$ называется *состоятельной оценкой* параметра $\boldsymbol{\theta}$, если при $n \to \infty$, она стремится по вероятности к оцениваемому параметру.

(иначе, состоятельная оценка дает точное значение при достаточно большом объеме выборки независимо от входящих в нее конкретных наблюдений).

Точечная оценка может быть дополнена ***интервальной оценкой,*** т.е. интервалом ($\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$), внутри которого с наперед заданной вероятностью γ находится точное значение оцениваемого параметра $\boldsymbol{\theta}$.

Интервал называют ***доверительным интервалом***,

γ - надежностью, с которой оцениваемый параметр $\boldsymbol{\theta}$ попадает в интервал ($\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$).

Для определения доверительного интервала заранее выбирают число α =1- γ, (0< α <1), называемое ***уровнем значимости***, и находят два числа θ_1 и θ_2, зависящих от точечной оценки θ^*, такие что

$$\boldsymbol{P}\,(\theta_1 < \theta < \theta_2) = 1 - \alpha = \gamma.$$

Обычно используется α = 0,1; 0,05; 0,01, что соответствует 90, 95, 99%-м доверительным интервалам.

IV. Статистические гипотезы и процедура их проверки

Статистической гипотезой называется предположение относительно параметров или вида распределения случайной величины.

Нулевой (основной) называют выдвинутую гипотезу $\boldsymbol{H}_0$,

альтернативной - гипотезу $\boldsymbol{H}_1$, которая противоречит нулевой.

Проверку статистической гипотезы осуществляют на основании данных выборки. Поскольку выборка имеет ограниченный объем, появляется возможность принятия ошибочного решения.

Вероятность α того, что будет отвергнута правильная нулевая гипотеза, называется ***уровнем значимости***.

Случайная величина, которая служит для проверки нулевой гипотезы, называется ***статистическим критерием***.

В качестве статистического критерия выбирается такая случайная величина t, точное или приближенное распределение которой известно.

После выбора критерия множество всех его возможных значений разбивают на две непересекающиеся области:

критическую и **область принятия гипотезы.**

Критической областью называется совокупность значений критерия, при которых гипотеза H_0 отвергается.

Областью принятия гипотезы называется совокупность значений критерия, при которых гипотеза H_0 не отклоняется.

Точки, разделяющие критическую область и область принятия гипотезы, называются **критическими**.

Критические точки $\boldsymbol{t_{кр.}}$ определяются по таблицам известного распределения выбранного критерия $\boldsymbol{t}$ при заданном уровне значимости и числе степеней свободы.

Основной принцип проверки статистической гипотезы:

- ✓ если наблюдаемое значение критерия (вычисленное по выборке) принадлежит критической области, то нулевую гипотезу отклоняют;
- ✓ если наблюдаемое значение критерия принадлежит области принятия гипотезы, то нулевую гипотезу принимают.