Вопросы по теме лекции: **Системный подход в научных исследованиях**

1) Дайте определение понятию системный подход.
2) Какие основные задачи системного подхода?
3) Дайте определение понятию системный анализ.
4) Назовите свойства кибернетических систем?
5) Назовите основные этапы выполнения системного анализа?

Лекция 4

Основы математической статистики

<u>Цель лекции:</u> изучить основы математической статистики и применение законов распределения параметров технологического процесса

<u>План лекции:</u>

1. Предмет теории математической статистики
2. Случайна величина и ее характеристики
3. Методы определения законов распределения
4. Последовательность построения законов распределения
5. Критерии согласия
6. Основные законы распределения случайных величин
7. Определение размера выборки

Рекомендуемая литература для изучения основ математической статистики

1. Елисеева И.И. Общая теория статистики: учебник для вузов / И.И. Елисеева, М.М. Юзбашев; под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2009. – 656 с.
2. Ефимова М.Р. Практикум по общей теории статистики: учебное пособие для вузов / М.Р. Ефимова и др. – М.: Финансы и статистика, 2007. – 368 с.
3. Мелкумов Я.С. Социально-экономическая статистика: учебно-методическое пособие. – М.: ИМПЭ-ПАБЛИШ, 2007. – 200 с.
4. Общая теория статистики: Статистическая методология в изучении коммерческой деятельности: учебник для вузов / О.Э. Башина и др.; под ред. О.Э. Башиной, А.А. Спирина. - М.: Финансы и статистика, 2008. – 440 с.
5. **Савченко А. Г., Пасічник О. В. Статистика. Макроекономіка: навч.-метод. посіб. К.: КНЕУ, 2006 – 221 с.**
6. **Громыко Г. Л. Теория статистики: учеб. для студентов вузов. М.: ИНФРА-М, 2000 – 360 с.**
7. **Шинкаренко В. Г. Теорія статистики: Навч. посіб. Х.: ХНАДУ 2005 – 150 с.**
8. **Галушко В.Г. Случайные процессы и их применение на автотранспорте. – К.: Высш. шк., 1980. – 272 с.**

1. Предмет теории математической статистики

Предмет прикладной науки – **математическая статистика** – разработка методов регистрации, описания и анализа статистических экспериментальных данных, получаемых в результате наблюдений <u>массовых</u> случайных величин.

ПРИКЛАДНОЕ ИССЛЕДОВАНИЕ

случайные явления

X_1 X_2 …… X_n

ЭКСПЕРИМЕНТ

Эксперимент – научно поставленный опыт или испытание, в процессе которого исследователь проверяет реально или искусственно вызванное им явление в точно учитываемых условиях.

получение опытных (статистических) данных

x_1 x_2 …… x_n

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

предмет

разработка методов регистрации, описания и анализа опытных данных, получаемых в результате наблюдений массовых случайных явлений

имеется конечный результат, но причины, обусловившие его появление, неизвестны

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

основные факты известны, но предсказать результат с абсолютной достоверностью невозможно

Основные задачи статистического анализа:

- статистическая проверка гипотез;
- определение числа наблюдений и получение выборки;
- определение характеристик генеральной совокупности на основе характеристик выборочной совокупности;
- построение уравнений корреляционной связи (уравнений регрессии);
- создание модели наблюдений (закон распределения);
- оценка параметров модели;
- изучение согласия между моделью и наблюдениями;
- реальное решение задач посредством оценки параметров и критериев значимости.

2. Случайна величина и ее характеристики

Случайной называется величина, которая в результате опыта может принять то или иное (но только одно) значение
(до опыта неизвестно какое именно).
Случайная величина характеризуется возможными значениями и вероятностями.

Дискретными случайными величинами
называются такие, которые принимают
только отдельные друг от друга значения и могут быть заранее перечислены.

Непрерывной случайной величиной
называется такая величина, возможные значения которой непрерывно заполняют некоторый промежуток (интервал числовой оси).

Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины (x_1, x_2, ….., x_n) и соответствующими им вероятностями.

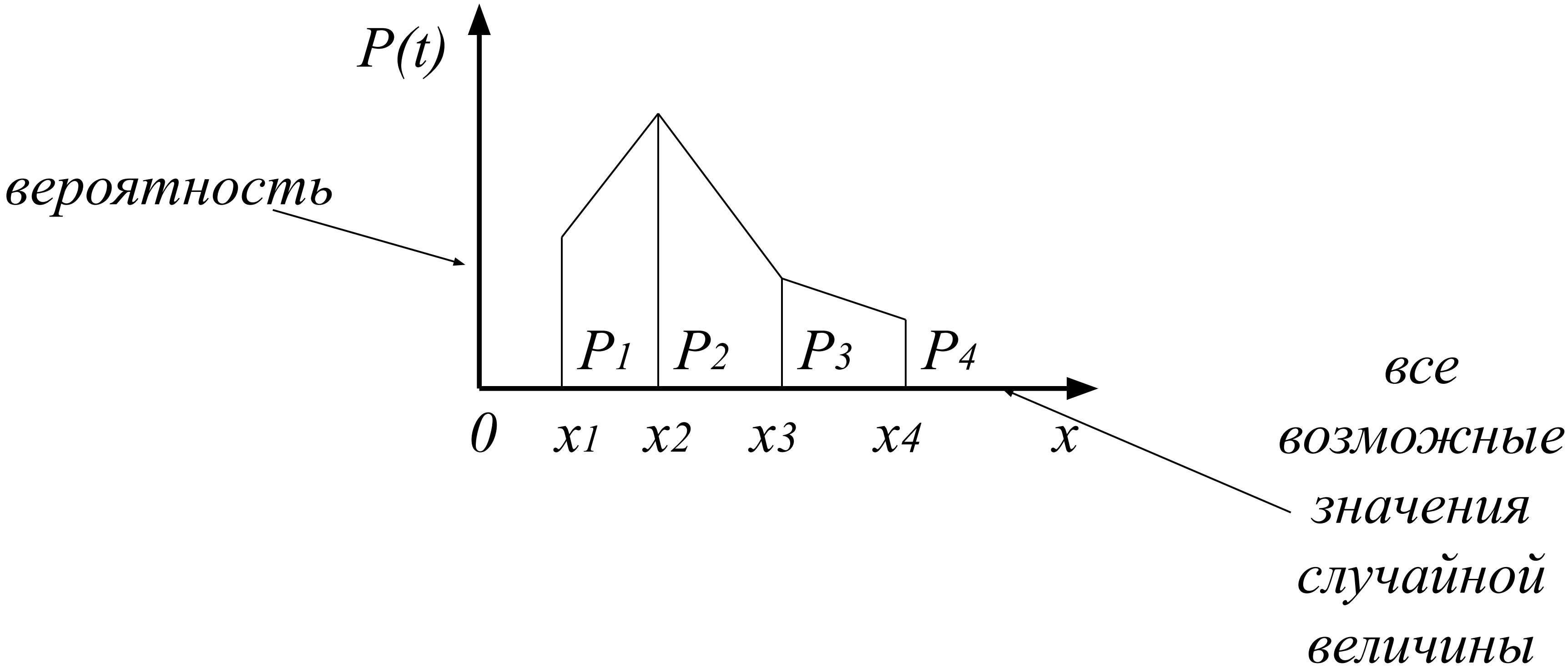
Простейшей формой задания закона распределения дискретной случайной величины ***X*** является **ряд распределения или таблица**

x_1	x_2	……	x_n
p_1	p_2	……	p_n

возможные значения случайной величины

вероятности

Многоугольник распределения дискретной случайной величины



Соединяются вершины только для наглядности, так как в промежутках между x1 и x2, x2 и x3 и т.д. случайная величина X значений принять не может, так как она дискретная, а ее ***вероятность в этих промежутках равна нулю.***

Случайная величина однозначно определяется следующими параметрами:

1) закон распределения (интегральная функция распределения или функция плотности распределения случайной величины);
2) параметр масштаба (параметр формы);
3) параметр расположения.

Числовые характеристики случайных величин:

1) ***Математическим ожиданием (средним значением) дискретной случайной величины*** называется сумма произведений всех возможных значений случайной величины на соответствующие им вероятности:

$$M(X) = \sum_{i=1}^{\infty} x_i p_i,$$

2) ***Модой случайной величины*** называют ее наиболее вероятное значение для дискретной случайной величины, и значение, которому соответствует максимум плотности вероятности, для непрерывной случайной величины.

3) ***Медианной*** случайно величины называется такое ее значение, относительно которого равновероятно получение большего или меньшего значения случайной величины

Числовые характеристики случайных величин:

4) Дисперсией называется математическое ожидание квадрата отклонения случайных величин от математического ожидания:

$$D[X] = M\left[(X - M[X])^2\right]$$

5) Среднеквадратическое отклонение - показатель рассеивания значений случайной величины относительно её математического ожидания.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Числовые характеристики случайных величин:

6) Моментом *k*-порядка называется математическое ожидание *k*-й степени отклонения случайной величины ***X*** от некоторой постоянной ***с***.

Если в качестве ***с*** берется нуль, моменты называются начальными

$$\mu_k = M(X)^k$$

Если **с = М(Х)**, то моменты называются центральными

$$\mu_k = M[X - M(X)]^k$$

3. Методы определения законов распределения

<u>Приемы определения законов распределения:</u>

1) Часто принципиальный характер кривой известен из теоретических соображений, связанных с существом задачи, или из аналитических задач, а из опыта (эксперимента) нужно определить лишь входящие в закон числовые параметры.
2) В некоторых случаях теоретическую кривую выбирают, учитывая внешний вид статистического распределения (гистограммы).
3) Иногда полезно использовать систему кривых Джонсона или Пирсона, каждая из которых, в общем случае, зависит от четырех параметров, а ее выбор можно осуществить с помощью специально разработанных графиков.
4) При использовании компьютерных программ при заданных статистических данных можно определить несколько законов распределения и выбрать наилучший. В качестве критерия:

   1 – наилучшее согласие теоретического и эмпирического распределений;

   2 – минимум параметров;

   3 – необходимость (и возможность) дальнейшего использования.

Методы определения параметров закона распределения:

1) метод моментов: параметры теоретической кривой должны быть равны соответствующим статистическим характеристикам (самый распространенный метод);

2) метод наименьших квадратов: сумма квадратов отклонений теоретической кривой от эмпирических данных должны быть минимальны;

3) метод наибольшего правдоподобия: пусть плотность вероятности *f(t)* случайной величины *T* зависит от параметра *a* (например – среднее), которое нужно определить на основании значений. Функция правдоподобия

$$L(t_1,...,t_2,...,t_n,a)=f(t_1,a)f(t_2,a)...f(t_n,a)\to max$$

4. Последовательность построения законов распределения

Закон распределения случайной величины X в общей форме задается функцией распределения (кумулятивной функцией распределения), которая определяет для каждого значения $\boldsymbol{x}$ вероятность наступления события $P(X \leq x)$

$$F(x) = P(X \leq x),$$

где $P(X \leq x)$ – вероятность того, что значение случайной величины не превышает $\boldsymbol{x}$.

Функция плотности распределения случайной величины

$$f(x) = \frac{dF(x)}{dx}.$$

Вычислительная схема определения числовых характеристик закона распределения случайных величин

(предположение - закон известен)

- представление экспериментальных (статистических) данных в форме **статистического ряда** или графически в виде **гистограммы** для непрерывных случайных величин, или **полигона** – для дискретных.
- определение **параметров** закона распределения;
- проверка **согласия** теоретического и статистического распределения по критериям согласия Пирсона или Колмогорова;
- построение **графика** теоретической кривой распределения (при необходимости).

Пример определения закона распределения непрерывных случайных величин

Имеются статистические данные случайной величины *T*: $t_1, ..., t_2, ..., t_n$ Для наглядности и компактности данные преобразуют в статистический ряд. В случае непрерывных случайных величин определяют размах $R = t_{max} - t_{min}$ Затем делят *R* на интервалы или «разряды» с шириной, равной *h*. При этом обычно *h* определяют из соотношения:

$$h \approx \frac{R'}{1 + 3{,}2\, lg\, N}$$

где *N* – размер выборки (количество наблюдений или данных).

Пример определения закона распределения непрерывных случайных величин (продолжение)

Далее делится *R* на интервалы:

$$K = \frac{R}{h}$$

и подсчитывают количество значений попавших в интервал m_N (частота). Затем определяют частость - r_N, соответствующую данному интервалу

$$r_N = \frac{m_N}{N}$$

Пример определения закона распределения непрерывных случайных величин (продолжение)

Поделив r_i на ширину интервала h_i, получают эмпирическую плотность:

$$P_i' = \frac{r_i}{h_i}$$

Для наглядности статистические данные оформляют в виде гистограммы по частотам или частостям (предпочтительней), пользуясь данными статистического ряда, можно <u>приближенно</u> построить функцию (интеграл) *F(t)* распределения случайной величины *T*. Обычно достаточно построить ее по граничным точкам или серединному интервалу (лучше), используя значения r_i или $p`_i h_i$

$$F^*(t_N) = \sum_{i=1}^{N} r_i = \sum_{i=1}^{N} p_i^* h_i = 1$$

Пример определения закона распределения непрерывных случайных величин (продолжение)

Таблица 1 – Статистическая обработка данных (пример)

№	Граница интервала	Частота	Частость	Ширина интер-вала	Плотность	Функция
1	0; 50	2	0,035	50	0,0007	0,0035
2	51; 150	29	0,51	100	0,0051	0,545
…						
N						

5. Критерии согласия

Для проверки согласованности теоретического и эмпирического распределения наиболее широко применяется критерий Пирсона (χ^2) и Колмогорова (λ).

Критерий согласия Пирсона, или *критерий согласия χ^2 (Хи-квадрат)* — наиболее часто употребляемый критерий для проверки гипотезы о принадлежности наблюдаемой выборки $x_1, x_2, \ldots, x_n$ объёмом n некоторому теоретическому закону распределения $F(x, \theta)$. Свойства критерия были впервые исследованы Карлом Пирсоном в 1900 году

Критерий может использоваться при проверке простых гипотез вида

$$H_0 : F_n(x) = F(x, \theta),$$

где θ — **известный** вектор параметров теоретического закона, и при проверке сложных гипотез вида

$$H_0 : F_n(x) \in \{F(x, \theta), \theta \in \Theta\},$$

когда оценка $\hat{\theta}$ скалярного или векторного параметра распределения $F(x, \theta)$ вычисляется по той же самой выборке.

1. По Пирсону.

1.1 Определяют метод расхождения:

$$\chi^2 = \sum_{i=1}^{K} \frac{\left(m_i - N \cdot P_i\right)^2}{N \cdot P_i}$$

где K – количество интервалов;

m_i – частота в i-ом интервале;

N – общее количество наблюдений;

P_i – теоретическое значение вероятности в i-ом интервале.

Для удобства применяют такую формулу (для непрерывной случайной величины):

$$\chi^2 = N\sum_{i=1}^{K}\frac{\left[P_i^* - f(t_i)\right]^2 h_i}{f(t_i)}$$

где P_i^* - эмпирическая плотность (вероятность);

$f(t_i)$ - теоретическая вероятность в i-ом интервале;

h_i - ширина i-го интервала.

1.2. Определяют число степеней свободы (f_1 или r) как разность между числом интервалов и положенных связей (условий) S^*:

$$f_1 = K - S^*$$

Поскольку при построении законов распределения всегда требуется

$\sum P_i^* = 1$, т.е. всегда имеется как *min* одна связь, то $S = S^* - 1$ и тогда

$$f_1 = K - S - 1$$

По f_1 и X^2 определяют вероятность согласия p_a теоретического и эмпирического (статистического) распределения. Если вероятность больше 0,05 ($p_\alpha \geq 0{,}05$), то эмпирический согласуется с теоретическим, если меньше, то отвергается.

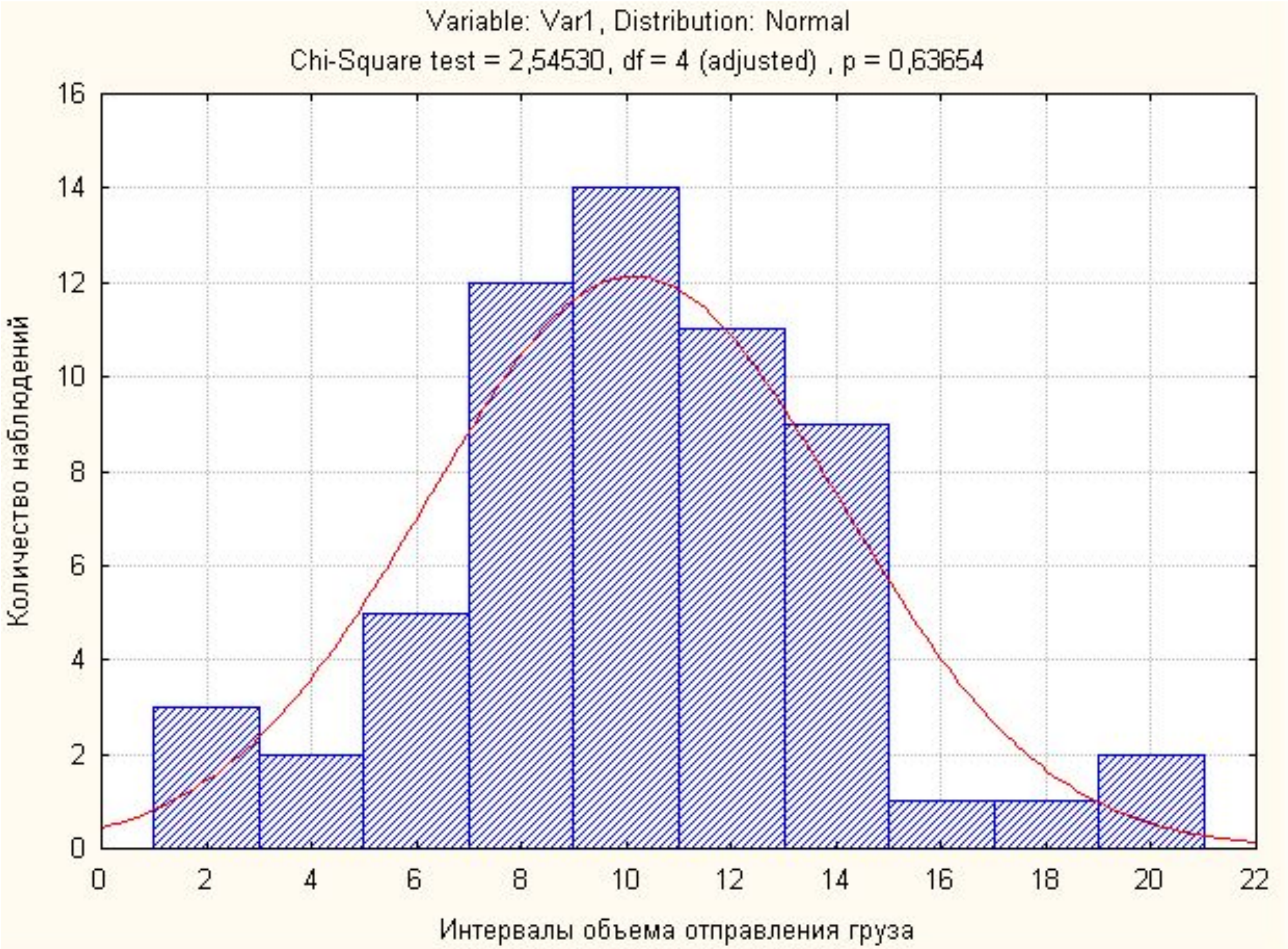
Чем больше f_1, тем больше «допустимое» X^2, чем меньше X^2, тем больше p_a

$$f_1 \uparrow \sim p_\alpha \uparrow \qquad X^2 \uparrow \sim p_\alpha \downarrow$$

Значения χ^2 в зависимости от вероятности и числа степеней свободы (фрагмент)

ρ f1	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,71	3,84	5,41	6,64	10,83
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	0,115	0,185	0,352	0,584	1,005	1,424	2,37	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	0,297	0,429	0,711	1,064	1,649	2,20	3,36	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	0,554	0,752	1,145	1,610	2,34	3,00	4,35	6,06	7,29	9,24	11,07	13,39	15,09	20,5
6	0,872	1,134	1,635	2,20	3,07	3,83	5,35	7,23	8,56	10,64	12,59	15,03	16,81	22,5
7	1,239	1,564	2,17	2,83	3,82	4,67	6,35	838	9,80	12,02	14,07	16,62	18,48	24,3
8	1,646	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	15,51	18,17	20,1	26,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,66	12,24	14,68	16,92	19,68	21,7	27,9

Пример определения закона распределения с помощью программы Statistica



2. По Колмогорову

2.1 Определяется эмпирическое и теоретическое значения функции распределения $F^*(t)$ и $F(t)$.

2.2 Вычисляются абсолютные значения разности между теоретической и эмпирической функциями распределения при одинаковых значениях аргумента, а затем выбирается наибольшая

$$k_0 = \max\left|F(t_N) - F^*(t_N)\right|.$$

2.3 Вычисляется

$$\lambda = k_o\sqrt{n}.$$

2.4 Определяется вероятность согласия $P(\lambda)$ теоретического и эмпирического распределений по табличным данным для вычисленного λ. Если $P(\lambda) > 0,05$, то согласие будет удовлетворительным.

Примечание. Применяется когда закон распределения известен.

Таблица - Значения критерия Колмогорова (фрагмент)

λ	α
λ ≤0,29	1,00000
0,30	0,99999
0,31	0,99998
0,32	0,99995
0,33	0,99991
0,34	0,99993
0,35	0,9997
0,36	0,9995
0,37	0,9992
0,38	0,9987
0,39	0,9981
0,40	0,9972
0,41	0,9960
0,42	0,9945
0,43	0,9926
0,44	0,9903

λ	α
0,76	0,6104
0,77	0,5936
0,78	0,5770
0,79	0,5605
0,80	0,5441
0,81	0,5280
0,82	0,5120
0,83	0,4962
0,84	0,4806
0,85	0,4653
0,86	0,4503
0,87	0,4355
0,88	0,4209
0,89	0,4067
0,90	0,3927
0,91	0,3791

λ	α
1,23	0,0970
1,24	0,0924
1,25	0,0879
1,26	0,0836
1,27	0,0794
1,28	0,0755
1,29	0,0717
1,30	0,0681
1,31	0,0646
1,32	0,0613
1,33	0,0582
1,34	0,0551
1,35	0,0522
1,36	0,0495
1,37	0,0469
1,38	0,0444

λ	α
1,70	0,0062
1,71	0,0058
1,72	0,0054
1,73	0,0050
1,74	0,0047
1,75	0,0044
1,76	0,0041
1,77	0,0038
1,78	0,0035
1,79	0,0033
1,80	0,0031
1,81	0,0029
1,82	0,0027
1,83	0,0025
1,84	0,0023
1,85	0,0021

λ	α
2,17	0,0002
2,18	0,0001
2,19	0,0001
2,20	0,0001
2,21	0,0001
2,22	0,0001
2,23	0,0001
2,24	0,0001
2,25	0,0001
2,26	0,0001
2,27	0,0001
2,28	0,0001
2,29	0,0001
2,30	0,0001
2,31	0,000046
2,32	0,000042

6. Основные законы распределения случайных величин

РАСПРЕДЕЛЕНИЯ

НЕПРЕРЫВНЫЕ

- **равномерное;**
- **нормальное;**
- **гамма-распределение (Эрланга);**
- **экспоненциальное (показательное);**
- **логистическое;**
- Грама-Шарлье;
- Стьюдента;
- Вейбулла;
- Максвелла и др.

ДИСКРЕТНЫЕ

- **Пуассона;**
- **биноминальное;**
- **геометрическое.**

Равномерное распределение

Равномерное распределение - это распределение случайной величины, принимающей значения, принадлежащие интервалу [a, b], характеризующееся тем, что плотность вероятности на этом интервале постоянна. Определяется параметром расположения ***a*** – нижняя граница области значений, и параметром масштаба ***b*** – размер области значений.

$$a = \bar{x} - \sqrt{3 \cdot s^2},$$

$$b = \sqrt{12 \cdot s^2},$$

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

где $\bar{x}$ – среднеарифметическое значение;
s^2 – выборочная дисперсия без поправки;
n – объём выборки.

Обозначение	$\mathcal{U}(a, b)$, $Rav(a, b)$
Параметры	$a, b \in (-\infty, \infty)$, a — коэффициент сдвига, $b - a$ — коэффициент масштаба
Носитель	$a \leqslant x \leqslant b$
Плотность вероятности	$\frac{1}{b-a}$ $\quad a \leqslant x \leqslant b$ 0 $\quad x < a$, $x > b$
Функция распределения	0 $\quad x < a$ $\frac{x-a}{b-a}$ $\quad a \leqslant x < b$ 1 $\quad x \geqslant b$
Математическое ожидание	$\frac{a+b}{2}$
Медиана	$\frac{a+b}{2}$
Мода	любое число из отрезка $[a, b]$
Дисперсия	$\frac{(b-a)^2}{12}$

Нормальное (Гауссово) распределение

Норма́льное распределе́ние, также называемое распределением Гаусса, также называемое распределением Гаусса — распределение вероятностей, которое в одномерном случае

Определяется параметром расположения

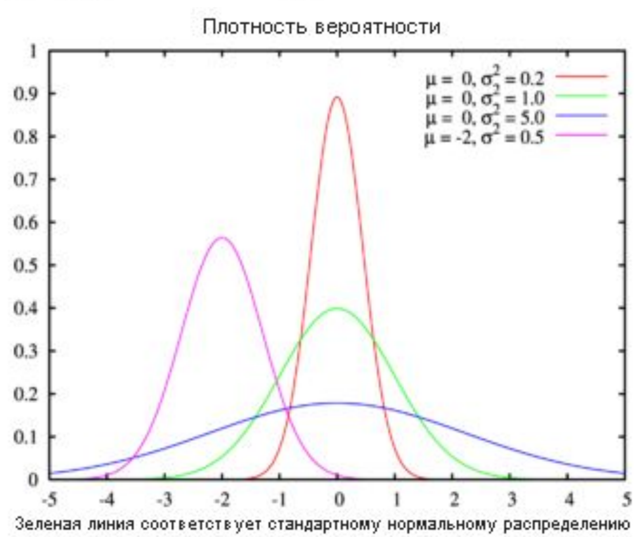
μ – математическое ожидание, и параметром масштаба

σ – среднеквадратическое отклонение.

$$\mu = \bar{x}.$$

Функция плотности вероятности

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp(-\frac{(x-\mu)^2}{2\sigma^2}).$$

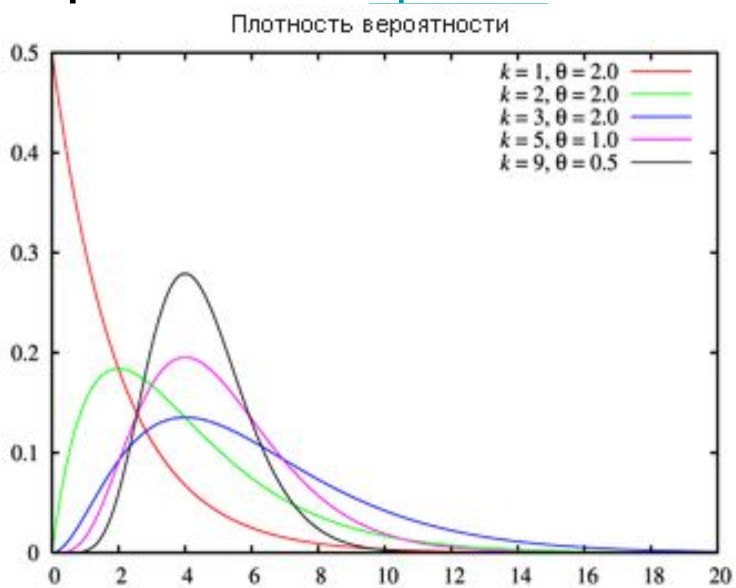


Обозначение	$N(\mu, \sigma^2)$
Параметры	μ — коэффициент сдвига (вещественное число) $\sigma > 0$ — коэффициент масштаба (вещественный, строго положительный)
Носитель	$x \in (-\infty; +\infty)$
Плотность вероятности	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
Функция распределения	$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right]$
Математическое ожидание	μ
Медиана	μ
Мода	μ
Дисперсия	σ^2

Распределение Эрланга

Га́мма-распределе́ние — это двухпараметрическое семейство абсолютно непрерывных распределений — это двухпараметрическое семейство абсолютно непрерывных распределений. Если параметр принимает целое значение, распреде
называет
распреде

Обозначение	$\Gamma(k,\theta)$, $\mathrm{Gamma}(k,\theta)$
Параметры	$k>0$, $\theta>0$ - коэффициент масштаба
Носитель	$x\in[0;\infty)$
Плотность вероятности	$x^{k-1}\dfrac{e^{-\frac{x}{\theta}}}{\Gamma(k)\theta^k}$
Функция распределения	$\dfrac{\gamma(x/\theta,k)}{\Gamma(k)}$
Математическое ожидание	$k\theta$
Медиана	
Мода	$(k-1)\theta$, когда $k\geq 1$
Дисперсия	$k\theta^2$



Экспоненциальное (показательное) распределение

Экспоненциальное или **показательное распределение** — абсолютно непрерывное распределение, моделирующее время между двумя последовательными свершениями одного и того же события.

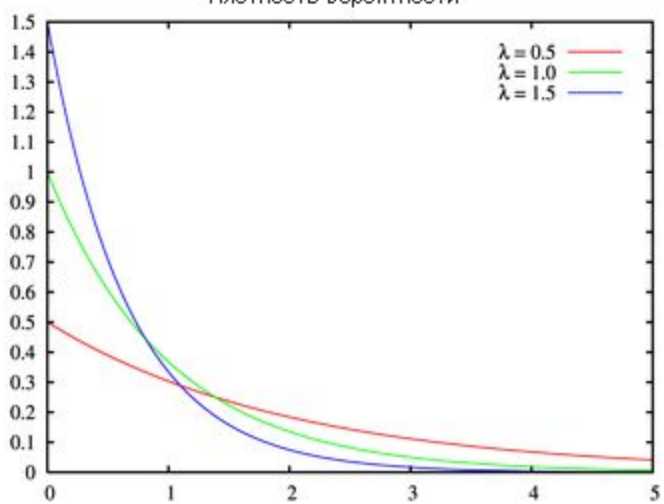
Определяется параметром масштаба b – среднее, часто употребляется параметр $\lambda = 1/b$ – отношение риска.

$$b = \bar{x}.$$

Функция плотности вероятности имеет вид

$$f(x) = \frac{1}{b} \cdot \exp(-\frac{x}{b}).$$

Плотность вероятности



Обозначение	$\mathrm{Exp}(\lambda)$
Параметры	$\lambda > 0$ - интенсивность или обратный коэффициент масштаба
Носитель	$x \in [0; \infty)$
Плотность вероятности	$\lambda e^{-\lambda x}$
Функция распределения	$1 - e^{-\lambda x}$
Математическое ожидание	λ^{-1}
Медиана	$\ln(2)/\lambda$
Мода	0
Дисперсия	λ^{-2}

Логистическое распределение

Логисти́ческое распределе́ние — один из видов [абсолютно непрерывных распределений]() — один из видов абсолютно непрерывных распределений. Формой напоминает [нормальное распределение](), но имеет

Обозначение	$L(\mu, s)$
Параметры	μ $s > 0$
Носитель	$x \in (-\infty; +\infty)$
Плотность вероятности	$\dfrac{e^{-(x-\mu)/s}}{s\left(1+e^{-(x-\mu)/s}\right)^2}$
Функция распределения	$\dfrac{1}{1+e^{-(x-\mu)/s}}$
Математическое ожидание	μ
Медиана	μ
Мода	μ
Дисперсия	$\dfrac{\pi^2}{3}s^2$

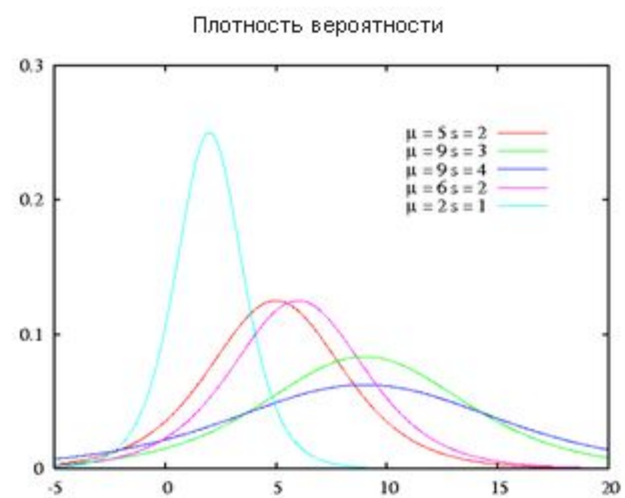
Определяется параметром расположения a – среднее, и параметром масштаба b – стандартное отклонение.

Часто в качестве параметра масштаба используется

$$k = \frac{b \cdot \sqrt{3}}{\pi}.$$

Функция плотности вероятности имеет вид

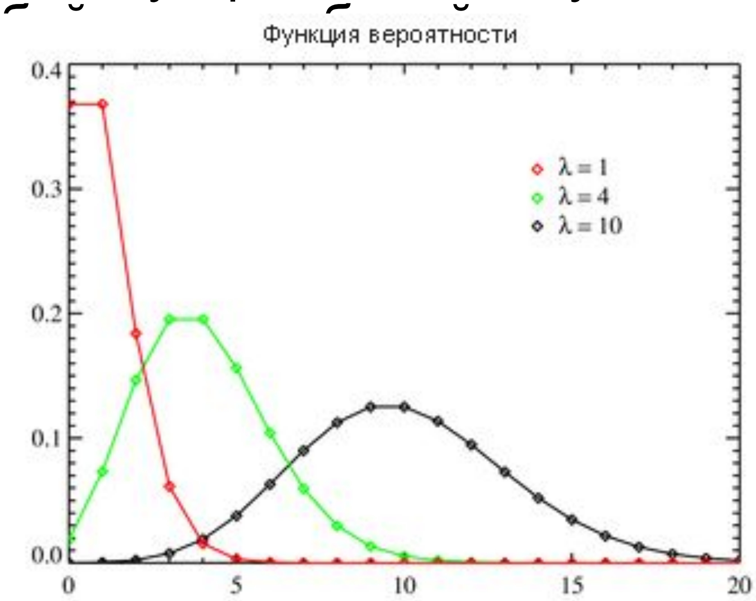
$$f(x) = \frac{\exp(\frac{x-a}{k})}{k \cdot \left[1 + \exp(\frac{x-a}{k})\right]^2}.$$



Распределение Пуассона

Распределение Пуассона — вероятностное распределение дискретного типа, моделирует случайную величину — вероятностное распределение дискретного типа, моделирует случайную величину, представляющую собой число событий — вероятностное распределение дискретного типа, моделирует случайную величину, представляющую сс
пр
ф
ус
пр
ф
ин
и

Обозначение	$\mathrm{P}(\lambda)$
Параметры	$\lambda \in (0, \infty)$
Носитель	$k \in \{0, 1, 2, \ldots\}$
Функция вероятности	$\frac{e^{-\lambda}\lambda^k}{k!}$
Функция распределения	$\frac{\Gamma(k+1, \lambda)}{k!}$
Математическое ожидание	λ
Медиана	$\approx \lfloor \lambda + 1/3 - 0.02/\lambda \rfloor$
Мода	$\lfloor \lambda \rfloor$
Дисперсия	λ

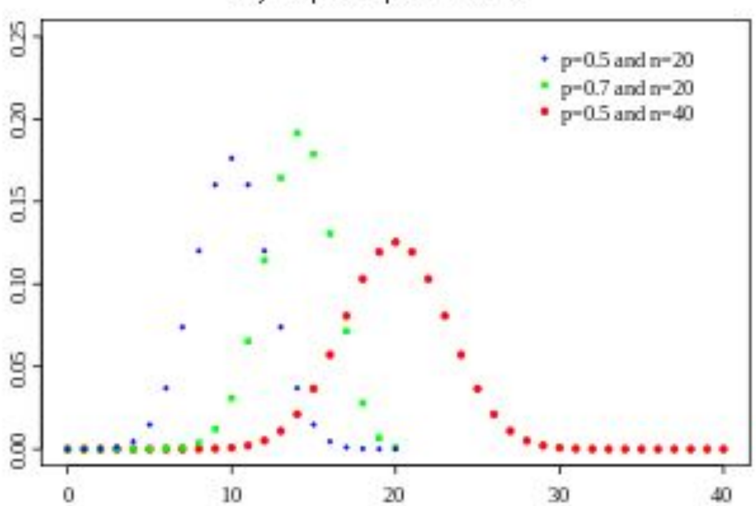


Биноминальное распределение

Биномиа́льное распределе́ние — распределение — распределение количества «успехов» в последовательности из независимых — распределение количества «успехов» в последовательности из независимых случайных экспериментов — распредел

в

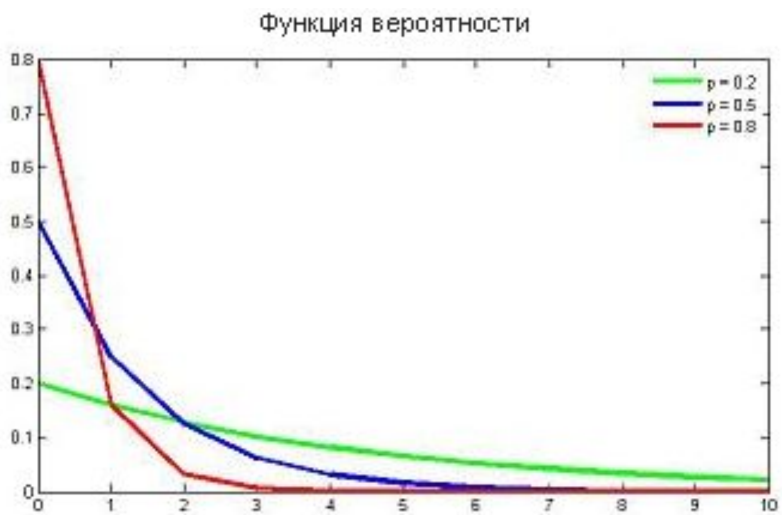
Функция вероятности



Обозначение	$B(n, p)$
Параметры	$n \geqslant 0$ — число «испытаний» $0 \leqslant p \leqslant 1$ — вероятность «успеха»
Носитель	$k \in \{0, \dots, n\}$
Функция вероятности	$\binom{n}{k} p^k q^{n-k}$
Функция распределения	$I_{1-p}(n - \lfloor k \rfloor, 1 + \lfloor k \rfloor)$
Математическое ожидание	np
Медиана	одно из $\{\lfloor np \rfloor - 1, \lfloor np \rfloor, \lfloor np \rfloor + 1\}$
Мода	$\lfloor (n+1)\, p \rfloor$
Дисперсия	npq

Геометрическое распределение

Геометри́ческое распределе́ние — распределение дискретной случайной величины равной количеству испытаний случайного эксперимента до наблюдения первого «успеха».



Обозначение	$\mathrm{Geom}(p)$	
Параметры	$n \ge 0$ — число «неудач» до первого «успеха» $0 < p \le 1$ — вероятность «успеха» $q \equiv 1 - p$ — вероятность «неудачи»	$n \ge 1$ — номер первого «успеха» $0 < p \le 1$ — вероятность «успеха» $q \equiv 1 - p$ — вероятность «неудачи»
Носитель	$n \in \{0, 1, 2, 3, \dots\}$	$n \in \{1, 2, 3, \dots\}$
Функция вероятности	$q^n p$	$q^{n-1} p$
Функция распределения	$1 - q^{n+1}$	$1 - q^n$
Математическое ожидание	$\frac{q}{p}$	$\frac{1}{p}$
Медиана	N/A	N/A
Мода	0	1
Дисперсия	$\frac{q}{p^2}$	$\frac{q}{p^2}$

7. Определение размера выборки

<u>Совокупность</u> – группа объектов, предметов или явлений, объединенных каким-либо общим признаком или свойством качественной или количественной характеристики (генеральная или выборочная совокупность).

Выборка или **выборочная совокупность** — часть генеральной совокупности элементов, которая охватывается экспериментом (наблюдением, опросом).

Характеристики выборки:

- Качественная характеристика выборки — что именно мы выбираем и какие способы построения выборки мы для этого используем.
- Количественная характеристика выборки — сколько случаев выбираем, другими словами объём выборки.

Необходимость выборки:

- Объект исследования очень обширный. Например, потребители продукции глобальной компании — огромное количество территориально разбросанных рынков.
- Существует необходимость в сборе первичной информации.

Для большинства практических задач, в которых законы распределения случайных величин описываются нормальным законом (или близким – Релея, Коши), объем выборки определяется по формуле:

$$n = \frac{t^2 \sigma^2}{\Delta^2},$$

где t – численное значение стандартного отклонения… - зависит от доверительной вероятности вывода (β) $t_{0,95} = 1,96$;

σ - среднеквадратическое отклонение совокупности;

Δ - дозволенная ошибка (абсолютная):

$$\Delta = \varepsilon a,$$

где ε - дозволенная относительная ошибка (0,05-0,1);

a – среднее значение совокупности (пробные выборки, известные a. «рекомендованные»).

При проведении выборочного наблюдения необходимо соблюдать следующие требования:

- единицы совокупности должны быть: легко различимы; на перекрывать друг друга; образовывать всю совокупность;
- выбор единиц совокупности должен соответствовать целям наблюдения;
- они должны быть удобны для работы;
- должна существовать возможность их перечисления (составление перечня);
- выборочная совокупность должна быть репрезентативной (представительской), т.е. давать представление обо всей совокупности для этого используется метод случайного отбора.

Пример определения объема выборки.

Пусть генеральная совокупность представляет значение средней эксплуатационной скорости для *N*=215 междугородних маршрутов Украины.

Необходимо определить размер выборки при следующих исходных предпосылках. Закон распределения скорости предполагается нормальным. Доверительная вероятность равна 0,95, точность вычисления скорости 1 км/ч.

Для решения данной задачи формируется совокупность 215 значений скорости и из них выбираются, например, 15 значений: 39; 42; 40; 29; 39; 43; 44; 50; 38; 32; 37; 49; 33; 40; 26.

Для определения среднеквадратического отклонения для нормального закона распределения можно воспользоваться правилом «трех сигм» (рассеивание случайной величины в основном укладывается на участке $a \pm 3\sigma$).

Таким образом, среднеквадратическое отклонение определяется делением разницы между максимальным и минимальным значением (размах) на 6.