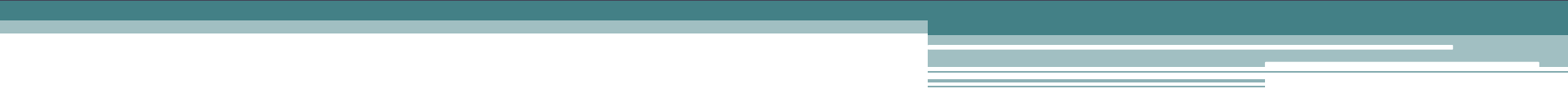
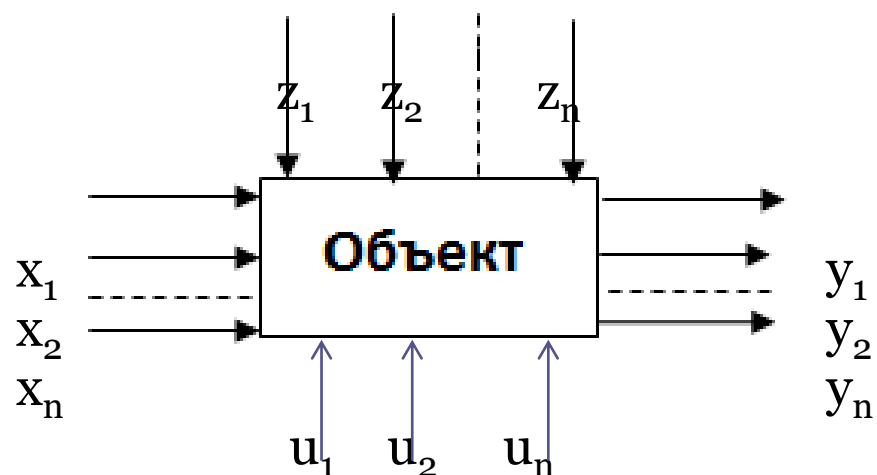


Составление математических моделей экспериментально- статистическими методами



- Основной и необходимый источник информации для построения статистической модели является эксперимент, а обработка экспериментальных данных осуществляется методами теории вероятности и математической статистики.
- Технологический объект представляется в виде «**черного ящика**»

Объект в виде «черного ящика»



- x_i – входные параметры;
- y_i – выходные параметры
- z_i – случайные воздействия
- u_i - управляющие воздействия

Рисунок 1- Схематическое изображение объекта

Статистические модели объекта

- аналитическое выражение функции отклика неизвестно, поэтому математическая модель представляется в этом случае в виде полинома:

$$y = \beta_0 + \sum_{i=1}^N \beta_i x_i + \sum_{\substack{i,j=1 \\ i \neq j}}^N \beta_{ij} x_i x_j + \sum_{i=1}^N \beta_{ii} \cdot x_i^2 + \dots$$

(1)

β_i , β_{ij} – теоретические коэффициенты, характеризующие соответственно линейные эффекты, эффекты взаимодействия и квадратичные эффекты. Они называются коэффициентами регрессии, а уравнение (1) – уравнением регрессии (общий вид математической модели)

- Результат эксперимента на сложном объекте обычно величина случайная.
- Поэтому при обработке экспериментальных данных можно определить, так называемые, выборочные коэффициенты регрессии: $b_0, b_i, b_{ij}, b_{ii} \dots$, которые являются лишь **оценками** для теоретических коэффициентов регрессии.
- Параметры генеральной совокупности обозначаются греческими буквами, а их выборочные оценки – латинскими.

- Уравнение регрессии, полученное на основании экспериментальных данных, имеет следующий

$$\hat{y} = b_0 + \sum_{i=1}^N b_i x_i + \sum_{i,j=1}^N b_{ij} x_i \cdot x_j + \sum_{i=1}^N b_{ii} x_i^2 + \dots$$

(2)

- y – выборочная оценка для y ; b_0 – свободный член уравнения регрессии; b_i – линейные эффекты; b_{ij} – эффекты взаимодействия; b_{ii} – квадратичные эффекты.
- С точки зрения исследования физико-химических свойств процесса, эта модель не имеет никакой информации. Справедлива такая модель только для объекта, на котором проводили эксперимент.

НЕКОТОРЫЕ ЭЛЕМЕНТЫ ТЕОРИИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ



Случайные величины

- Случайная величина – величина, значение которой принципиально нельзя предсказать исходя из условий опыта.
- Истинное значение переменной – это такое значение, которое получилось бы при некотором измерении, если бы отсутствовали элементы случайности, связанные с измерением.
- Различают непрерывные и дискретные случайные величины.

- **Дискретные** переменные могут принимать только отдельные значения, в некотором интервале, которые можно заранее перечислить.
- Значения **непрерывной** случайной величины заранее посчитать нельзя. Они непрерывно заполняют некоторый промежуток, т.е. могут принимать любое значение из заданного интервала.

Определение вероятности

- Пусть проведена серия опытов, которая включает N экспериментов. При этом непрерывное X (результат измерений) произошло M_i раз. Тогда отношение M_i/N называется частотой появления события X .
- Частота M_i/N – тоже величина случайная и изменяется в зависимости от количества проведенных опытов. При большом числе опытов она может стабилизироваться около некоторого значения (P).
- Предел, к которому стремится это отношение при неограниченном возрастании числа экспериментов называется вероятностью случайного события X .

$$P\{X\} = \lim_{n \rightarrow \infty} \frac{m_i}{n};$$

- *Вероятность – это отношение числа благоприятных исходов события к полному числу при большой продолжительности эксперимента.*

Характеристика случайных величин

Распределение случайной величины

- Для характеристики случайной величины необходимо задать распределение этой случайной величины.
- Соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими вероятностями называется законом распределения случайной величины.

Распределение дискретной случайной величины

- Дискретную случайную величину можно задать вероятностным рядом.

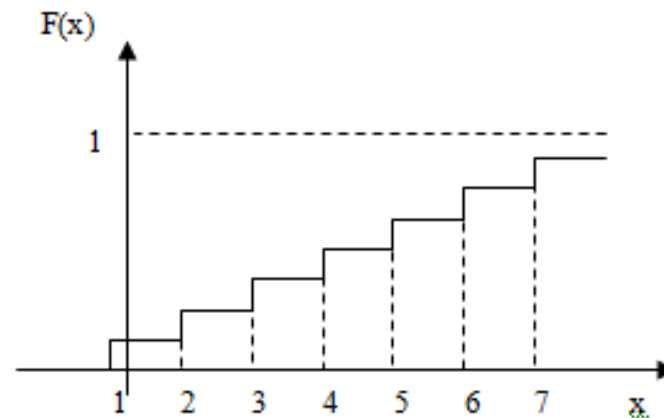
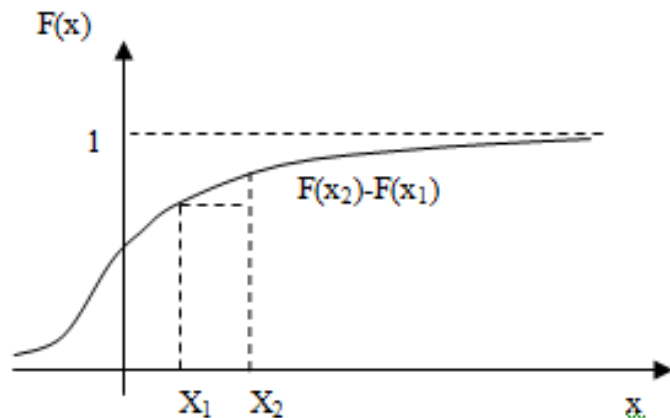
$$\frac{x_1 \quad x_2 \dots x_n}{P(x_1)P(x_2)\dots P(x_n)} \quad - \text{вероятностный ряд}$$

Распределение непрерывной случайной величины

- Для непрерывных случайных величин принимается вероятность того, что в результате опыта значение случайной величины попадет в некоторый интервал.
- Пользуются вероятностью события $P(X < x)$, где x – произвольное действительное число, а X – случайная величина: тогда
$$P(X < x) = F(x)$$
 – называется функцией распределения случайной величины.
- Это **интегральная** функция распределения случайной величины показывает вероятность того, что случайная величина X не превышает некоторого заданного или текущего значения x .

Функция распределения

- В виде функции распределения можно задать распределение как непрерывной, так и дискретной случайной величины



- Ордината кривой, соответствующая т. x_1 , есть вероятность того, что случайная величина X при испытании окажется меньше x_1 .
- Вероятность того, что значение случайной величины X заключено между x_1 и x_2 и равно разности функции распределения, вычисленных в этих точках.

$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$$

Плотность распределения вероятности

- Для непрерывной случайной величины часто применяется производная функции распределения (**дифференциальная** функция распределения вероятностей) – это **плотность** распределения вероятности:

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

$$f(x) = \frac{dF(x)}{dx} = \frac{\lim_{\Delta x \rightarrow 0} (x < X \leq x + \Delta x)}{\Delta x}$$

- т.е. $f(x)$ равно отношению вероятности попадания случайной величины в интервал $(x, x + \Delta x)$ к длине этого интервала (Δx) , когда Δx – бесконечно малая величина.

Числовые характеристики случайных величин

Кроме функции и плотности распределения, существуют другие числовые характеристики случайных величин. Некоторые основные свойства случайных величин могут быть описаны с помощью определенных числовых характеристик.

Наиболее часто применяются на практике два параметра:

- **Математическое ожидание** случайной величины
- **Дисперсия случайной** величины

Математическое ожидание

- **Математическое** ожидание случайной величины – это параметр, который характеризует **центр рассеяния** (центр распределения) случайной величины. Принято обозначать $M[x]$, m_x , m
- Иногда $M[x]$ называют «средневзвешенным» значением случайной величины или «генеральным» средним значением.
- Для случайной дискретной величины:

$$m_x = M[x] = \sum_{i=1}^n x_i \cdot p_i \quad (1)$$

- Для случайной непрерывной величины:

$$m_x = M[x] = \int_{-\infty}^{\infty} x_i \cdot f(x) \cdot dx \quad (2)$$

$$D[x] = M\left(\int_{-\infty}^{\infty} (x_i - m_x)^2 \cdot f(x) \cdot dx\right)$$

Дисперсия

- *Дисперсия* – параметр, характеризующий **степень отклонения (рассеяния)** случайной величины от ее **среднего** значения. Т.е. это параметр, характеризующий разброс значений этой величин.
- Дисперсией случайной величины X называется математическое ожидание квадрата разности случайной величины от ее математического ожидания.

$$D[x] = M(X - m_x)^2$$

- Для случайной дискретной величины:

$$D[x] = \sum_{i=1}^n (x_i - M[x])^2 \cdot P_i$$

- Для случайной непрерывной величины:

$$D[x] = \int_{-\infty}^{\infty} (x_i - m_x)^2 \cdot f(x) \cdot dx$$

- Дисперсия обозначается: $D[x]$, σ_x^2 , σ^2

Пример:

В результате испытаний двух приборов (расходомеров) установлена вероятность наблюдения помех, оцениваемая по 2-х балльной системе

Первый расходомер		Второй расходомер		
Балл (x_j)	1	2	1	2
Вероятность наблюдения помех, P_i	0.2	0.065	0.03	0.15

$$m_{x_1} = 0.2 \cdot 1 + 0.065 \cdot 2 = 0.33$$

$$m_{x_2} = 0.03 \cdot 1 + 0.15 \cdot 2 = 0.33$$

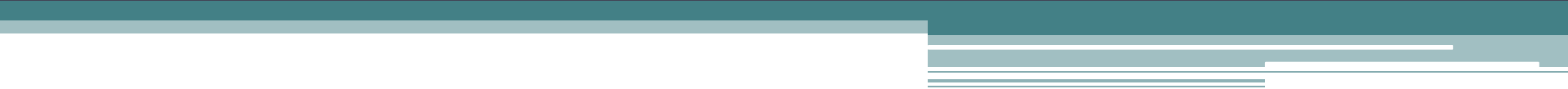
Т.о. средний уровень помех у расходомеров одинаков, и по этому показателю нельзя выбрать лучший прибор. Определим устойчивость показаний (разброс вокруг среднего) D_x . Посчитаем дисперсию уровня помех:

$$D_1[x] = (1 - 0.33)^2 \cdot 0.2 + (2 - 0.33)^2 \cdot 0.065 = 0.11$$

$$D_2[x] = (1 - 0.33)^2 \cdot 0.03 + (2 - 0.33)^2 \cdot 0.15 = 0.43$$

Следовательно, лучшим является первый расходомер.

Законы распределения случайной величины



Законы распределения

При аппроксимации результатов эксперимента используются различные законы, например:

- Биномиальное распределение
- Равномерный закон (равномерное распределение).
- Нормальный закон распределения (закон Гаусса)
- Экспоненциальное распределение

Биномиальное распределение

- Рассмотрим простейший случай повторения одного и того же испытания при постоянных условиях.
- Пусть проведено n независимых опытов, в каждом из которых может появиться или не появиться некоторое событие X . Обозначим: появление события X^+ , не появление — X^- .

- Вероятность появления события X в каждом опыте

- $P(X) = P = 1,$

вероятность не появления его

$$q = 1-p,$$

тогда

$$p + q = 1.$$

Требуется найти вероятность того, что событие X в этих n опытах появится m раз. Результат каждого испытания будем отмечать, ставя X^+ или X^- на соответствующем месте.

Например, при трех испытаниях возможны следующие $2^3=8$ исходов:

- $X^- X^- X^-$, $X^+ X^+ X^+$, $X^- X^+ X^+$, $X^+ X^- X^+$, $X^+ X^+ X^-$, $X^- X^- X^+$, $X^- X^+ X^-$, $X^+ X^- X^-$.
- (Событие X не появилось 3 раза; событие X появилось 3 раза, событие X не появилось в первом испытании, появилось во втором и в третьем; событие X появилось в первом и третьем, не появилось во втором испытаниях и т.д.).

- Так как испытания независимы, то вероятность каждого результата может быть найдена путем перемножения вероятностей событий X^+ и X^- в соответствующих испытаниях, например:

- $P^0_3 = q^3;$

- $P^1_3 = qqr + qqr + qqr = 3q^2p;$

- $P^2_3 = prr + prr + prr = 3p^2q;$

- $P^3_3 = ppp.$

- $q^3 + 3q^2p + 3p^2q + p^3 = (p+q)^3=1$

В общем случае для $X^+X^-X^-X^+ \dots X^-$ можно записать

- $P(X^+)P(X^-)P(X^-)P(X^+) \dots P(X^-)=pqqr\dots q.$

- Если в написанной последовательности X^+ встречается **m** раз, а, соответственно X^- встречается **$n-m$** раз, то вероятность такого результата будет **p^mq^{n-m}** .

- Если мы рассмотрим вероятность $P_n(m)$ ровно m раз наблюдать событие X в течение n испытаний, то
- получим

$$P_n(m) = C_n^m p^m q^{n-m}, \quad (1)$$

- где C_n^m - число сочетаний из n элементов по m , определяемое по формуле

$$C_n^m = n! / (m!(n - m)!) \quad (2)$$

- n - общее число точек сосредоточения массы;
- m - число точек сосредоточения массы, в которой определяется плотность вероятности;
- P - параметр распределения, вероятность выпадения данной случайной величины.
- Совокупность вероятностей $P_n(m)$ при $m=0,1,2,\dots,n$, т.е. $P_n(0), P_n(1) \dots P_n(m)$ называется биномиальным распределением вероятностей или параметром распределения (вероятностью выпадения данной случайной величины.)

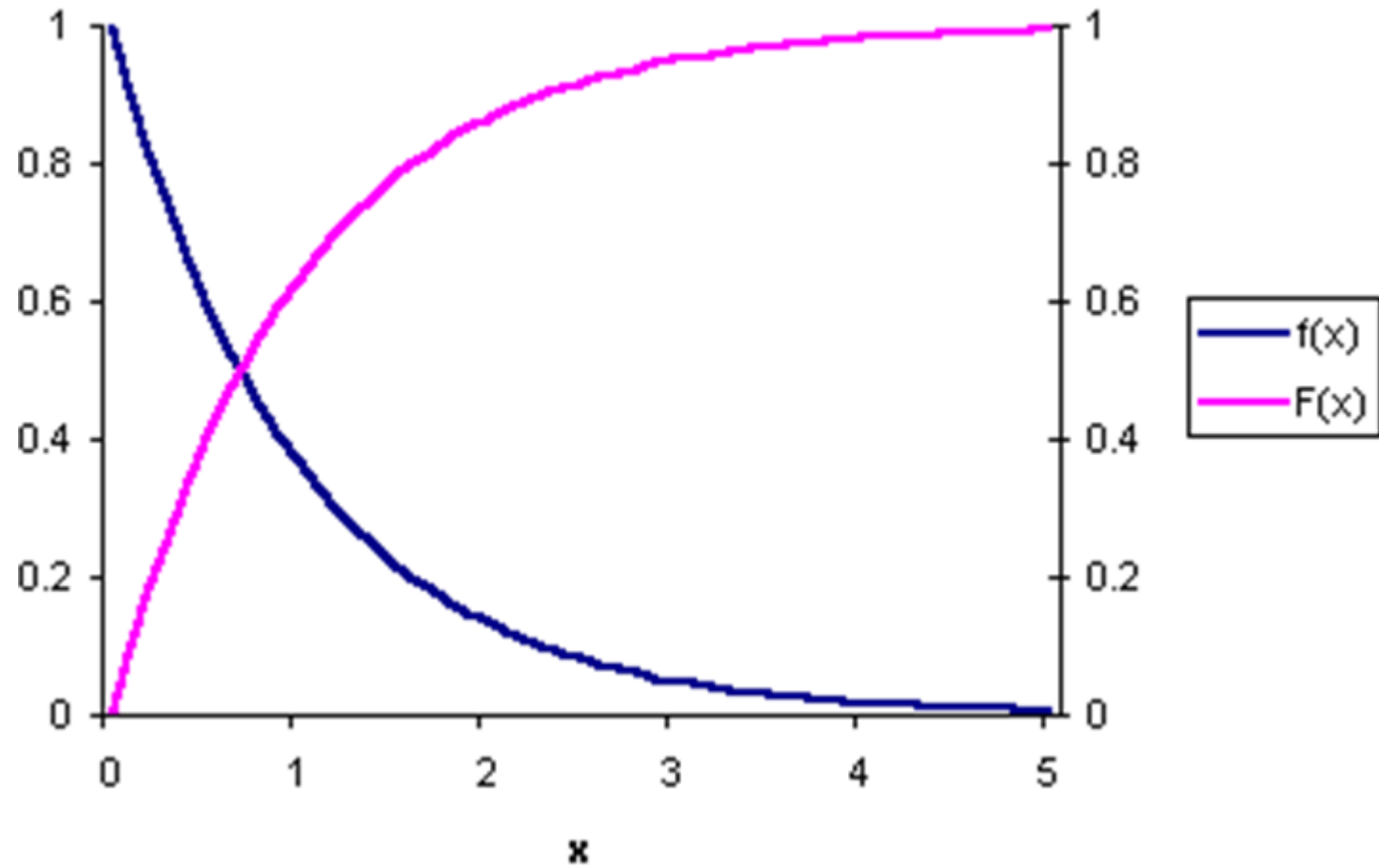
Экспоненциальный закон распределения

Говорят, что случайная величина X имеет **экспоненциальное** распределение с параметром $\theta > 0$, если она непрерывна, принимает только положительные значения, и имеет плотность распределения и функцию распределения:

$$f(x) = \theta \cdot e^{-\theta x}, \quad x > 0;$$

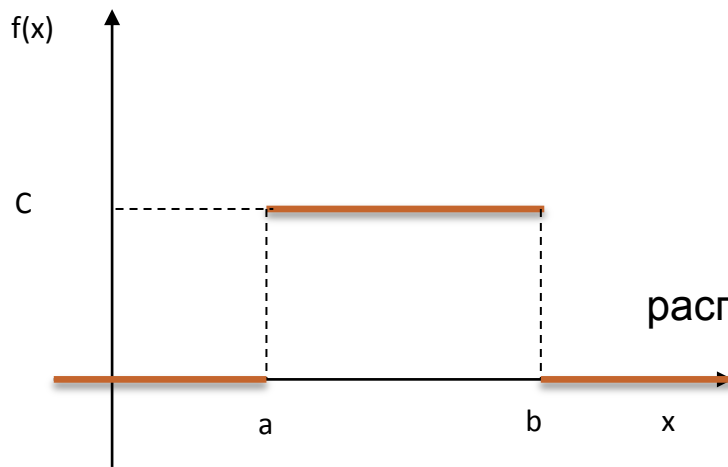
$$F(x) = 1 - e^{-\theta x}, \quad x > 0;$$

Для $Q=1$



Равномерное распределение

- Распределение, плотность вероятности которого постоянна на некотором участке, а вне его равна **0**, называется непрерывным равномерным распределением.



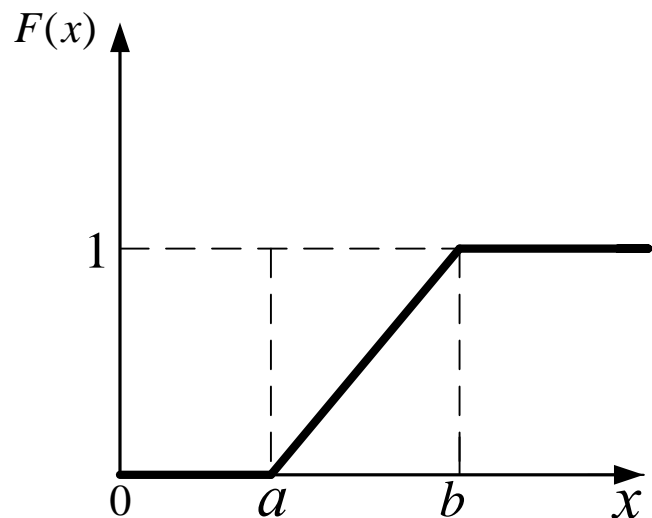
$$f(x) = \begin{cases} c & a \leq x \leq b \\ 0 & x < a \text{ или } x > b \end{cases}$$

Так как площадь, ограниченная кривой распределения, равна 1, то **$C(b-a)=1$** , тогда

$$c = \frac{1}{b-a};$$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & x < a \text{ или } x > b \end{cases}$$

Равномерное распределение



$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & x < a \text{ или } x > b \end{cases} \rightarrow$$

$$F(x) = \begin{cases} 0 & \text{при } x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Нормальное распределение (закон Гаусса)

- Случайная непрерывная величина X называется распределенной по нормальному закону, если ее плотность распределения имеет вид:

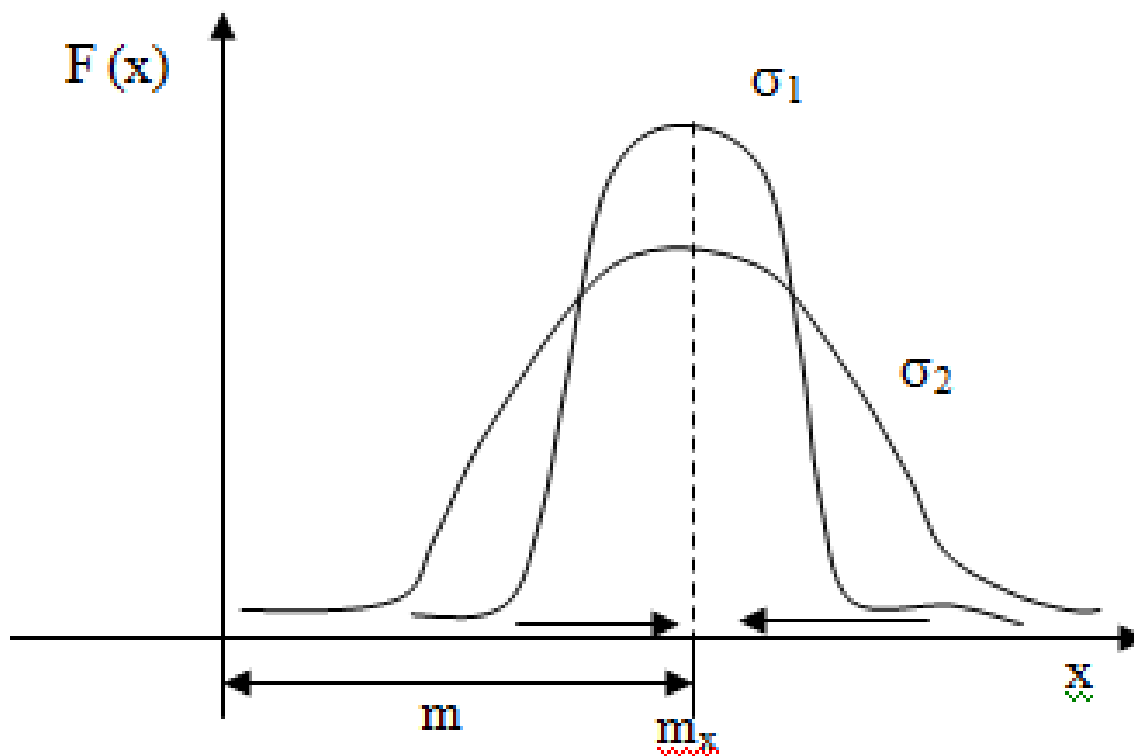
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}; \text{ или } f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}; \quad (1)$$

$-\infty < x < \infty$

- m_x – математическое ожидание; σ^2 – дисперсия случайной величины X – стандартное отклонение от математического ожидания, симметричное m ; σ – среднее квадратичное отклонение (стандартное);
- σ и σ^2 – характеризуют разброс данных, т.е. ошибку измерений.
- Чем больше σ , тем больше ошибка.

- График плотности распределения вероятности называется нормальной кривой или кривой Гаусса. Это кривая колоколообразного вида, симметричная m_x .

При $x=m_x$, кривая имеет максимум равный $\frac{1}{\sigma\sqrt{2\pi}}$



- Площадь под всей кривой выразится **интегралом:**

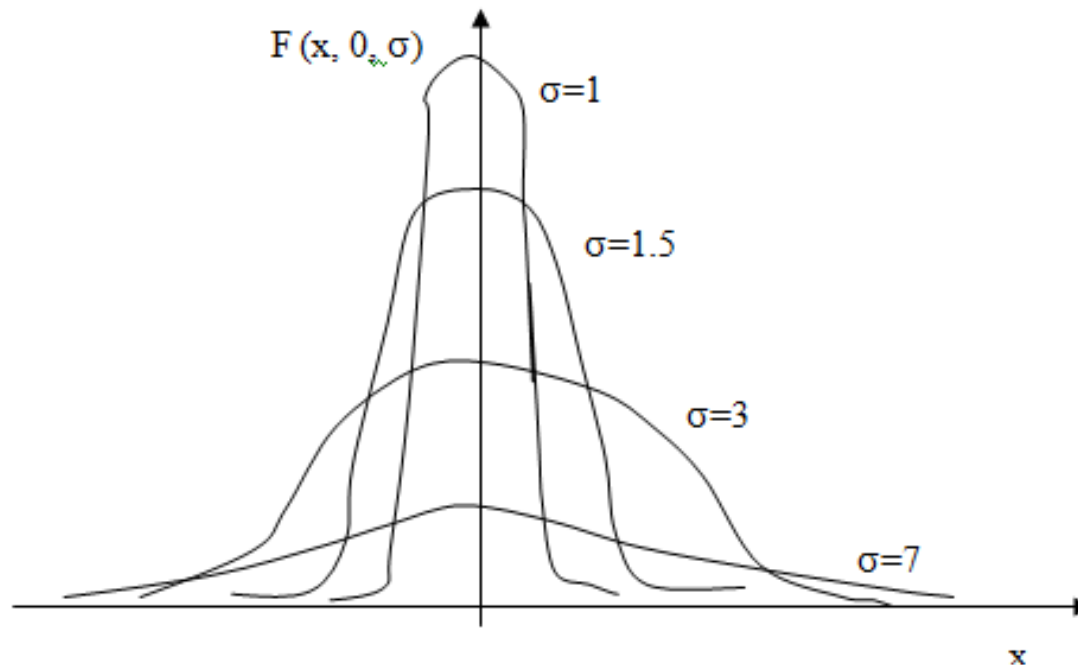
$$I = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} \cdot dx \quad (2)$$

- Нормальное распределение при $m=0$ и $\sigma=1$ называется **стандартным**.
- Если изменять m , то кривая будет перемещаться вдоль оси x , сохраняя свою форму.

- При $m_x=0$ будем иметь семейство нормальных кривых с центром в начале координат, зависящих только от σ .

$$f(x,0,\sigma) = y = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{x^2}{2\sigma^2}};$$

При уменьшении σ ордината кривой растет, одновременно происходит резкий спад ее к оси x , поэтому S (площадь) всегда равна 1. При очень малых σ кривая похожа на иглу вдоль оси y .



Стохастическая связь

- **Стохастической** называется связь, при которой с изменением одной величины изменяется другая.
- Для оценки тесноты стохастической связи пользуются показателями:
 - **Ковариация**
 - **Коэффициент корреляции**

- Зависимость между случайными величинами существует, если

$$\text{cov}_{xy} = M[(X - m_x)(Y - m_y)] \neq 0$$

- Эта величина называется **ковариацией** cov_{xy} .
- Коэффициент корреляции определяется:

$$\rho_{xy} = \frac{\text{cov}_{xy}}{\sigma_x \cdot \sigma_y} = \frac{M[(X - m_x)(Y - m_y)]}{\sigma_x \cdot \sigma_y}$$

- Коэффициент показывает меру линейной зависимости между x и y , где σ_x и σ_y – среднеквадратичные отклонения величин x и y .

Выборочные статистические характеристики



Генеральная совокупность и случайная выборка

- **Генеральная** совокупность - все допустимые значения случайной величины.
- **Выборка** – это конечный набор значений случайной величины, полученный в результате наблюдений (экспериментов). Число элементов выборки N называется ее **объемом**. Это N объектов, которые извлекаются из генеральной совокупности.
- Смысл статистических методов заключается в том, чтобы по ограниченной выборке (N) обоснованно судить о свойствах генеральной совокупности.
- Подобное суждение может быть получено путем оценивания параметров генеральной совокупности при помощи, так называемых, **оценок**.

Требования к оценке

Обозначим:

α – параметр генеральной совокупности

α_n – выборочный параметр, т.е. параметр, который является оценкой параметра α .

1. **Состоятельность.**

Оценка (α_n) параметра α называется состоятельной, если при неограниченном увеличении объема выборки N значение α_n стремится к своему теоретическому значению α . $P\{\lim|\alpha_n - \alpha| \rightarrow 0\} = 1$

2. **Несмещенность.**

Оценка α_n называется **несмещенной**, если ее математическое ожидание при любом объеме выборки равно оцениваемому параметру

$$M[\alpha_n] = \alpha$$

3. **Эффективность**

Оценка α_n называется эффективной, если среди всех оценок параметра α она обладает наименьшей дисперсией.

$$D[\alpha_n] = \min.$$

Оценка математического ожидания

- Оценка математического ожидания – это *среднее арифметическое* измеряемой величины.

$$m_x \rightarrow \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- \bar{x} является очень хорошей оценкой m_x , оно удовлетворяет всем трем требованиям.

Оценка дисперсии

- Оценкой дисперсии является *выборочная* дисперсия:

$$\sigma^2_x \rightarrow S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)^2}{n-1} \text{ или}$$
$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}_i)^2}{n-1}}$$

- Каждый параметр, который входит в формулу для выборочной характеристики и определяется объемом выборки называется **связью (l)**.
- Разность между объемом выборки и числом связей называется **числом степеней свободы**.
- **f=n-l**
- Число связей определяет число выборочных характеристик.

Выборочный коэффициент корреляции

- Выборочный коэффициент корреляции r_{xy} представляет оценку коэффициента корреляции ρ_{xy} .

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1) \cdot S_x \cdot S_y}; \text{ где}$$

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}; S_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

- Если $r=0$, x и y – независимы, корреляции нет.
- Если $r=+1$ или $r=-1$ зависимость строго функциональная $y=f(x)$ (**линейная**).
- Если $r>0$, то с ростом x увеличивается y .
- Если $r<0$, то с ростом x уменьшается y .
- Коэффициент корреляции показывает, существует ли связь между x и y , но самого вида функции не дает.

Статистические критерии и их применение.

- Одной из основных задач статистического анализа результатов измерений является проверка статистических гипотез.
- **Статистическая гипотеза** – это некоторое предположение относительно свойств генеральной совокупности, из которой извлекается выборка.
- **Критерий** статистической гипотезы – это правило, позволяющее принять или отвергнуть данную гипотезу на основании выборки.
- Чтобы принять или отвергнуть гипотезу, еще до получения выборки (результатов измерений), задаются **уровнем значимости** критерия q , т.е. мы задаемся величиной, допустимой в данной ситуации погрешности (ошибки). Вероятность совершения ошибки – q . Обычно уровень значимости выбирают: 0.01, 0.05, 0.1, 0.02. наиболее часто – 0.05 или 5%. Если $q=5$, это значит, что в 95% случаев подтверждается данная гипотеза.
- Рассмотрим **критерии**, которые будем применять в дальнейшем

- Критерий Фишера (F – критерий) – применяется при проверке однородности двух дисперсий (или нескольких). При этом проверяется гипотеза, можно ли считать сравниваемые дисперсии оценками одной и той же генеральной совокупности.
- Критерий Кохрена (G-критерий)
- Используется для сравнения нескольких дисперсий, если выборочная дисперсия получена по одинаковым выборкам $m_1 = m_2 = \dots = m_n$, т. е. если во всех сериях экспериментов число параллельных опытов было одинаково.
- Критерий Стьюдента (t – критерий) применяется при определении необходимого числа экспериментов для достижения заданной степени точности, для определения грубых ошибок, для сравнения между собой двух средних полученных по выборкам.