- **Элементы теории вероятностей и математической статистики**

IV. 0. Элементы ТВ и МС

Статистический (стохастический) эксперимент* - *эксперимент, результат которого заранее неизвестен.

***Лежит в основе* теории вероятностей и математической статистики.**

Проводится *по определенному плану*

***Цель эксперимента* - выявить закономерности случайных явлений, процессов.**

***Результат эксперимента* - случайные величины (СВ), дискретные или непрерывные, определенные на всей числовой оси или на некотором интервале.**

Для определения СВ необходимо знать как СВ, так и ее *вероятность p*, значение которой лежит в пределах

$0 <= p <= 1$

IV. 0. Элементы ТВ и МС

Случайная величина и ее вероятность связаны функциями:

***1. З*авис**имость между значениями случайной величины и соответствующими вероятностями называется:**

- Для непрерывной случайной величины - *плотность распределения f(x)*

***- Для дискретной случайной величины- закон распределения* -**

2. Функция распределения F(x) – вероятность того, что случайная величина меньше некоторого значения X

Между функциями f(x) и F(x) существует связь

$$F(x) = \int_{-\infty}^{x} f(t).dt \qquad f(x) = \frac{dF}{dx}$$

***Вероятность попадания непрерывной случайной величины X в заданный интервал (a, b)* определяется по формуле**

$$P(a < X < b) = \int_{a}^{b} f(x)dx$$

IV. 0. Элементы ТВ и МС

Числовые характеристики случайной величины:

1. Меры положения

- *<u>математическое ожидание m</u>*- среднее центральное значение, вокруг которого распределены возможные значения СВ;

- **для дискретной СВ** **для непрерывной СВ**

$$m = \sum_i x_i . p_i \qquad m = \int_{-\infty}^{\infty} x . f(x) dx$$

- *<u>мода Мо</u>*- случайная величина, имеющая наибольшую вероятность;

- *<u>медиана Ме</u>* – случайная величина, расположенная в середине диапазона, в котором СВ определена.

Для ряда $\{x_i\} = \{2,3,5,6,7\}$ *Me* = 5;

Для ряда $\{x_i\} = \{3,5,6,7\}$ *Me* = (5 +6)/2 = 5.5

IV. 0. Элементы ТВ и МС

2. ***Меры разброса***

- ***дисперсия D*** **– математическое ожидание квадрата отклонения СВ от ее математического ожидания**

для дискретной СВ **для непрерывной СВ**

$$D = \sum_i (x_i - m)^2 . p_i \qquad D = \int_{-\infty}^{\infty} (x - m)^2 . f(x) dx$$

- ***среднеквадратическое отклонение*** $\sigma = \sqrt{D}$

- ***размах R*** **=** $x_{max} - x_{min}$

- ***коэффициент вариации (ковариация)*** **- отношение выборочного среднеквадратического отклонения к выборочной средней, выраженное в процентах.**

$$V = \frac{S}{\tilde{x}} \cdot 100(\%)$$

Ковариация характеризует однородность совокупности

При V < 33% - совокупность считается количественно однородной

IV. 0. Элементы ТВ и МС

3. **Меры формы. Функция плотности распределения f(x) может быть: - симметричной или ассиметричной;**

- крутовершинной или плосковершинной;

- одномодальной или полимодальной.

Для оценки формы графика функции f(x) вводятся меры формы : - коэффициент асимметрии A_s и эксцесс Е

Для нормального закона распределения (закона Гаусса) коэффициенты A_s и Е равны нулю

***Асимметрией теоретического распределения* называют характеристику, которая определяется формулой**

$$A_s = \mu_3/\sigma^3.$$

***Эксцессом теоретического распределения* - характеристика, которая определяется формулой**

$$E_k = (\mu_4/\sigma^4) - 3.$$

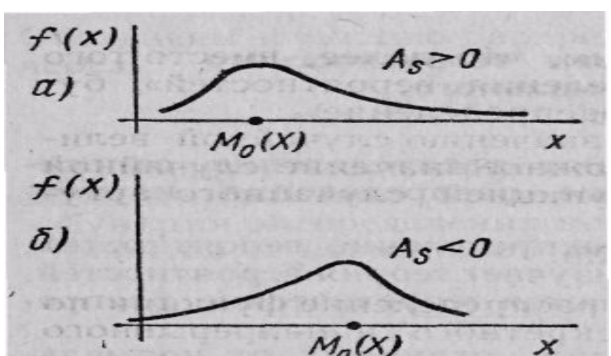


Рис. 10

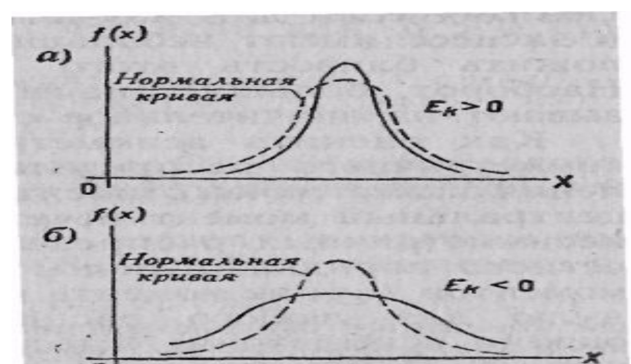


Рис. 11

1.1. Нормальный закон распределения (Гаусса)

Плотность распределения

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

Функция распределения

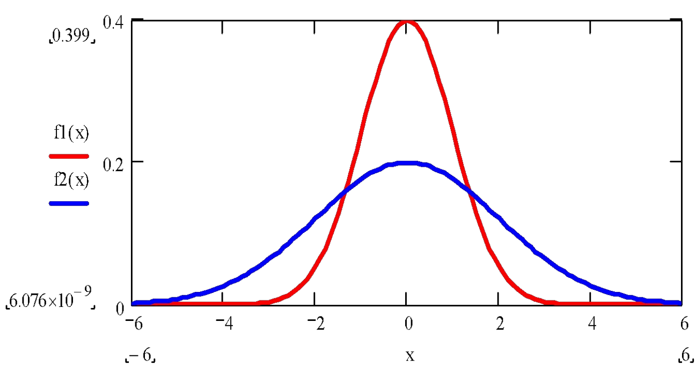
$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{\frac{-1}{2}\left(\frac{t-m}{\sigma}\right)^2} dt$$

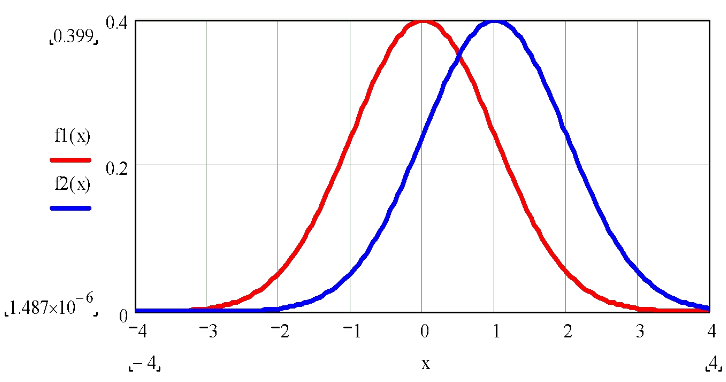
При нормировке (стандартное нормальное распределение)

$$z := \frac{x-m}{\sigma} \quad f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot z^2} \quad F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{1}{2}z^2} dz$$

На интервале [-3σ;+3σ] $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-3\sigma}^{+3\sigma} e^{-\frac{1}{2}z^2} dz = 0.9973 = 1-\alpha$

Интервал[-3σ;+3σ] является областью статистического допуска параметра качества





Нормальное распределение Гаусса

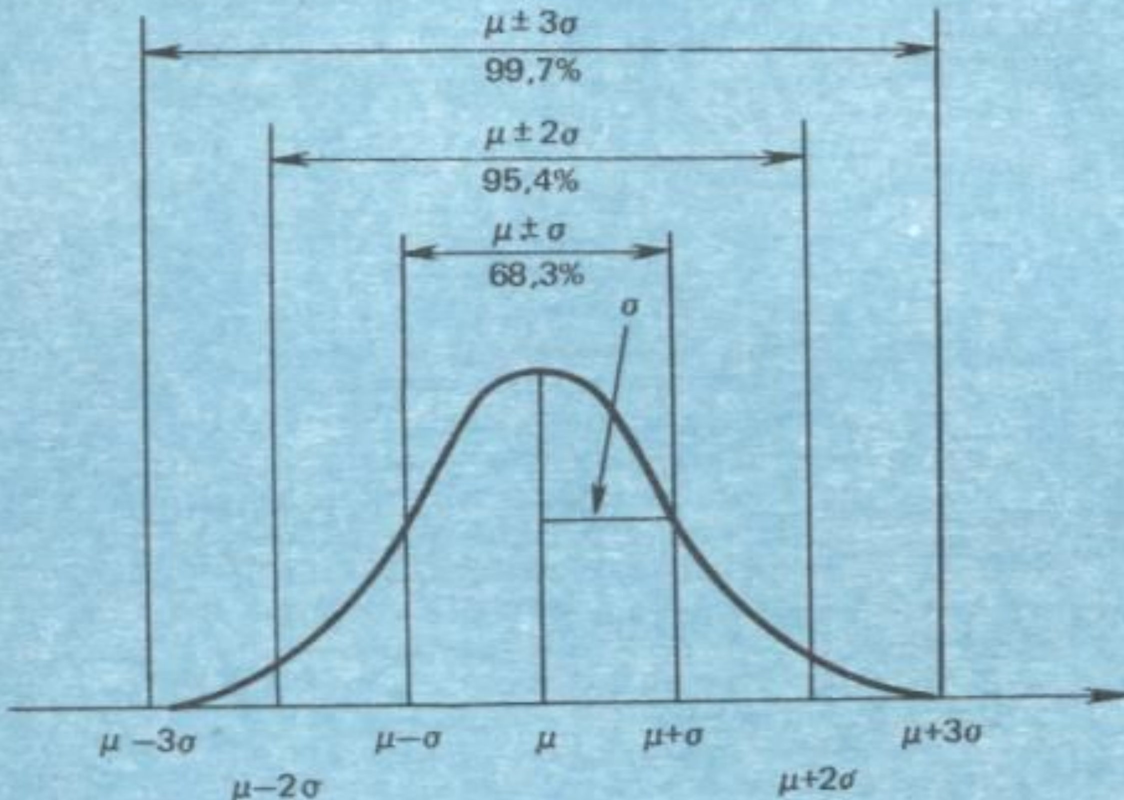


Рис. 5.7. Параметры μ, σ и вероятности для нормального распределения

IV. 0. Элементы математической статистики

Математическая статистика **– раздел математики, в котором изучаются методы сбора, систематизации, обработки и анализа результатов наблюдений массовых случайных явлений для выявления существующих закономерностей.**

Первая задача **математической статистики - определение способов сбора и систематизации статистической информации.**

Вторая задача **математической статистики – разработка методов обработки и анализа статистических данных**

Вторая задача **решается в несколько этапов:**

1. **Предварительный анализ данных – анализ и исключение грубых ошибок, вычисление параметров (статистик) выборочных данных**
2. **Точечные и интервальные оценки параметров модели**
3. **Выбор типа модели, описывающей данные эксперимента**
4. **Проверка модели о согласии модели и эмпирических данных**

IV.0. Элементы математической статистики

Все изучаемые объекты формируют генеральную совокупность (ГС) данных. Объем генеральной ГС обозначают N.

Сплошное обследование **– анализ *всех данных ГС.* Не всегда возможно – из за большого объема ГС или необходимости уничтожения объекта.**

Обычно из совокупности выбирают ограниченное число объектов (выборку) и их подвергают изучению, применяют *выборочный метод* обследования. *Объем выборки обозначают n*

Чтобы по данным выборки можно было уверенно судить об интересующем нас признаке ГС, необходимо, чтобы объекты выборки правильно представляли ГС.

Выборка должна быть *репрезентативной (представительной)*. Для этого каждый из объектов выборки должен быть отобран из ГС случайным образом.

Существуют специальные приёмы отбора, обеспечивающие репрезентативность выборки.

*Варианта - н*аблюдаемое значение количественного признака x_i,
Вариационный ряд - последовательность вариант, записанных в порядке возрастания
Частота n_i, - число наблюдений значения признака x_i,
Относительная частота v_i - отношение n_i к объёму выборки n
Справедливы соотношения

$$\sum_{i=1}^{n} n_i = n; v_i = \frac{n_i}{n}; \sum_{i=1}^{n} v_i = 1$$

Статистическое (эмпирическое) распределением выборки – соответствие между вариантами x_i *, записанными в порядке возрастания, и относительными частотами* v_i
Накопленные (относительные) *частоты* - сумма частот (относительных частот) со значением признака x меньше X.
*Эмпирическая функция распределения - с*оответствие между вариантами и накопленными частотами.
Интервальный вариационный ряд - варианты, объединенные в группы.
- В интервальном вариационном ряду К – количество групп, fi – частоты попадания варианта в i-ую группу, ∑ fi = n
- В интервальном ряду накопленные частоты (относительные частоты) показывают сумму частот (относительных частот) со значением признака x меньше X.

Элементы математической статистики и производственный процесс. Долгосрочная и краткосрочная вариации

Производственный процесс может отслеживаться в текущем режиме или течении длительного промежутка времени. В первом случае наблюдаем *краткосрочную вариацию*, во втором – *долгосрочную*.

<u>Краткосрочная вариация</u> абсолютно случайна,* зависит от большого количества общих причин, *<u>является вариацией по общим причинам.</u>

<u>Долгосрочная вариация</u>* содержит информацию о неслучайных причинах вариаций, это *<u>вариация по особым причинам</u>

Обнаружение особых причин приводит к попытке их устранения и улучшения характеристик процесса. Особыми причинами могут быть износ оборудования, различия в сырье, квалификация персонала и др. Чтобы обнаружить общие и особые причины, необходимо запланировать и провести статистический эксперимент, а затем проанализировать его результаты.

Существует аналогия между статистическим распределением выборки и законом распределения дискретной случайной величины.
В данном случае вместо возможных значений случайной величины фигурируют варианты, а вместо соответствующих вероятностей - относительные частоты.
В силу этой аналогии по известному эмпирическому распределению можно по тем же формулам, что и для дискретного распределения, найти выборочные аналоги математического ожидания и дисперсии.
Для оценки числовых параметров выборки и , в дальнейшем, генеральной совокупности, в математической статистике используют следующие числовые характеристики –

- ***статистики или меры процесса:***
- ***1. меры положения*** **– средние значения, медиана, мода;**
- ***2. меры разброса*** **– размах, выборочная дисперсия, выборочное среднеквадратическое отклонение;**
- ***3. меры формы*** **– коэффициент асимметрии, эксцесс**

Графическое представление экспериментальных данных

Экспериментальные данные по мере наблюдений заносятся в таблицы определенной формы.
Для наглядного представления экспериментальных данных используют графики и диаграммы.
К основным графическим формам отнесены:

1. Точечные диаграммы
2. Гистограммы
3. Диаграммы изменения процесса во времени
4. Диаграммы рассеяния

Наряду с перечисленными, применяют и другие наглядные средства: полигон, кумулятивная кривая, диаграммы рассеяния и др.
На следующих слайдах представлены некоторые графические формы

Корреляционный, регрессионный, дисперсионный, временной анализ

Основная задача *корреляционного анализа* – выявление связи между случайными переменными

Основная задача *регрессионного анализа* – *установление формы связи* между случайными переменными

Основная задача *дисперсионного анализа* - *оценка влияния различных факторов* на результат эксперимента.

***Дисперсионный анализ* применяется и для последующего планирования экспериментов**

Важнейшей задачей *исследования временных рядов* – *выявление и статистическая оценка основной тенденции* развития изучаемого процесса

Элементы корреляционного и регрессионного анализа

На практике параметры регрессии определяются на базе данных выборочного эксперимента. Графическим представлением результатов этого эксперимента является *корреляционная диаграмма, поле корреляции, диаграмма рассеяния*

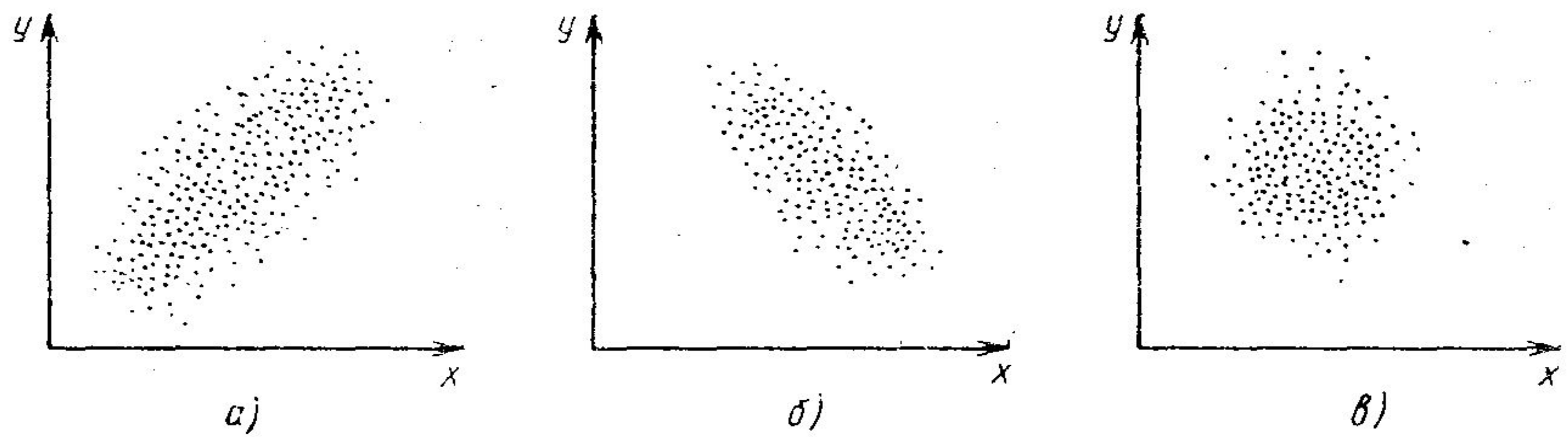


Рис. 2.10. Диаграммы рассеяния:
а — положительная корреляция; *б* — отрицательная корреляция; *в* — отсутствие корреляции

Корреляция. Поле корреляции

По виду диаграммы рассеяния можно судить о характере и силе корреляционной связи. **Если удается провести *прямую* (линию тренда) через группу точек поля корреляции, то между факторами имеется линейная связь Плотная группировка точек вокруг прямой говорит о сильной связи, угол наклона прямой – о направлении корреляции. При наклоне до 90^0 увеличение фактора X вызывает рост фактора Y. Сила корреляционного эффекта зависит от угла наклона прямой – чем круче линия тренда, тем сильнее фактор X влияет на фактор Y**

Для определения количественной силы связи вычисляют **коэффициент корреляции r . Коэффициент корреляции изменяется в пределах -1<= r <=1.**

При r =0 две случайные величины независимы.

При r >0 наблюдается положительная связь между факторами (признаками), рост значений одного фактора ведет к росту значений другого.

При r <0 – связь обратная, рост значений одного фактора приводит к уменьшению другого.

При Abs (r) = 1 связь функциональная

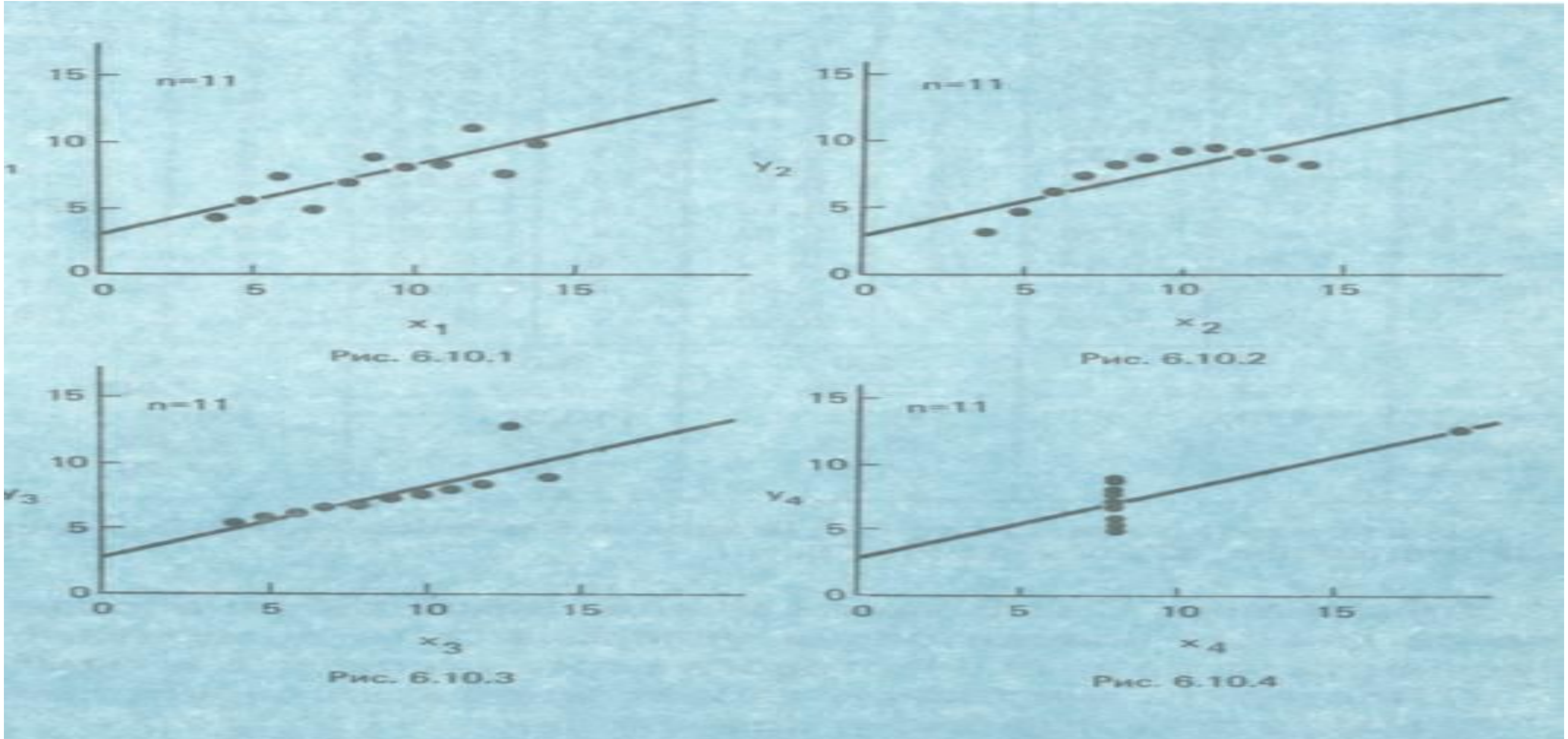


Рис. 6.10. Разные диаграммы рассеивания, имеющие одинаковые линии регрессии

Статистическая связь между одним фактором и средним значением другого может быть сильной, слабой, отсутствовать. Уравнение регрессии (линия регрессии) показывает силу влияния одной характеристики на другую. Чем круче линия регрессии, тем сильнее влияние одного параметра на другой.

значением другого может быть сильной, слабой, отсутствовать. Уравнение регрессии (линия регрессии) показывает силу влияния одной характеристики на другую. Чем круче линия регрессии, тем сильнее влияние одного параметра на другой.

Элементы анализа временных рядов

Графическим представлением временных рядов являются временные диаграммы. Данные наносятся по мере поступления.

Временные диаграммы позволяют

1. **Обнаружить выбросы. Выбросы являются отклонением от нормы**
2. **Обнаружить тренд. Тренд – устойчивое изменение во времени среднего процесса**
3. **Обнаружить серию. Серии возникают чаще всего из-за дефектов оборудования, проблем калибровки, некоторой совокупности дефектов**
4. **Обнаружить сдвиги, скачки. Характеризуют безвозвратно наступившие изменения в системе**

Временные диаграммы в виде контрольных карт применяются для анализа производственного процесса в течении смены, месяца или более длительного периода.