

# Информационные технологии в управлении качеством и защита информации

## Построение гистограмм

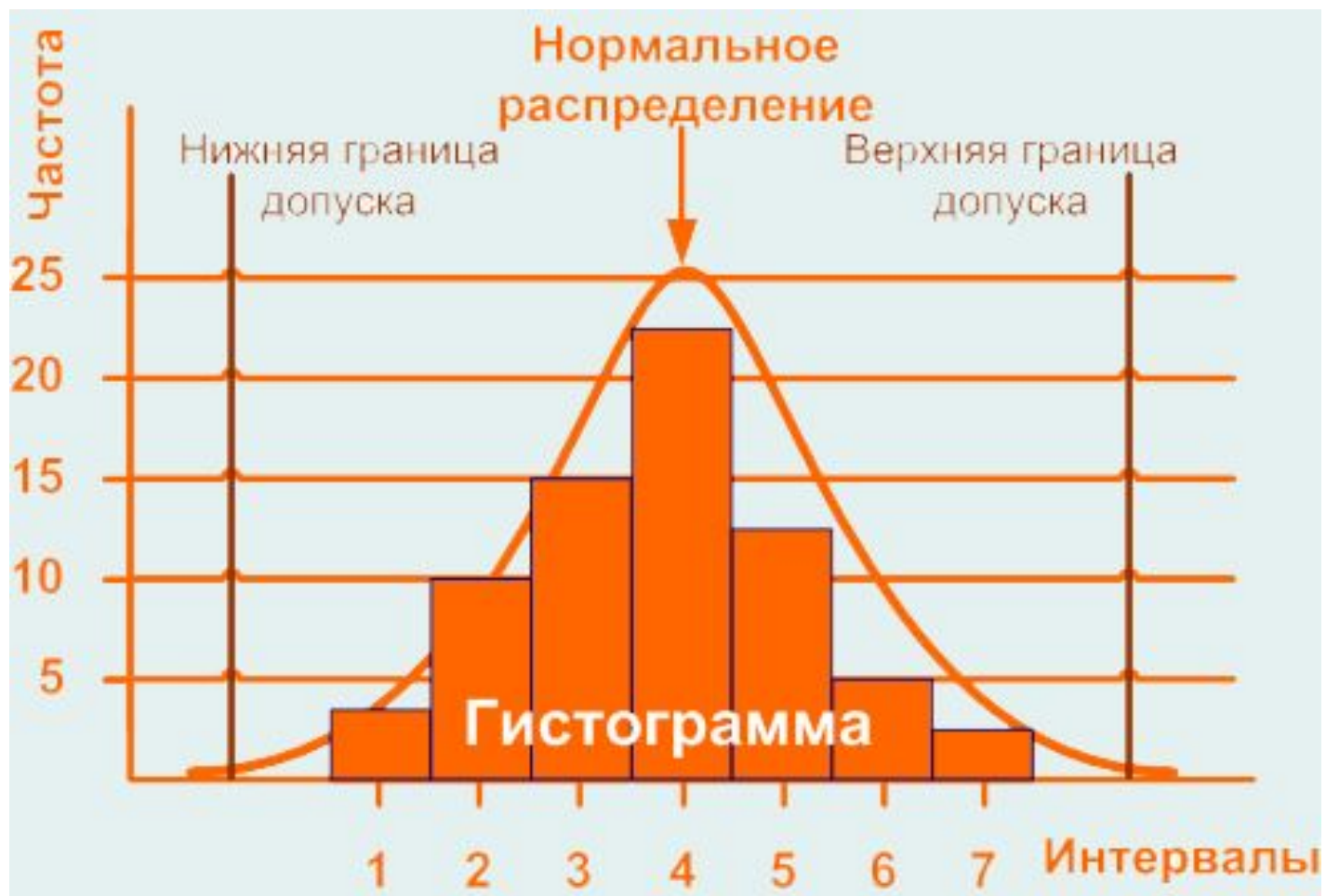
# Гистограмма

- способ представления статистических данных в виде столбчатой диаграммы. Она отображает распределение отдельных измерений параметров изделия или процесса. Иногда ее называют частотным распределением, так как гистограмма показывает частоту появления измеренных значений параметров объекта.

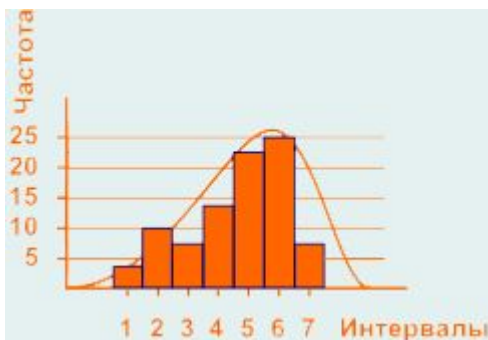
# Порядок построения гистограмм

1. Собираются статистические данные – результаты измерений параметра объекта (не менее 30 значений).
2. Выявляется наибольшее и наименьшее значение показателя среди полученных результатов измерений.
3. Определяется ширина диапазона значений показателя – из наибольшего значения показателя вычитается наименьшее значение.
4. Выбирается надлежащее число интервалов в пределах которых необходимо сгруппировать результаты измерений.
5. Устанавливаются границы интервалов.
6. Подсчитывается число попаданий значений результатов измерений в каждый из интервалов.
7. Строится гистограмма – на оси абсцисс (горизонтальной оси) отмечаются интервалы, а на оси ординат (вертикальной оси) отмечается частота попаданий результатов измерений в каждый интервал.

# Гистограмма

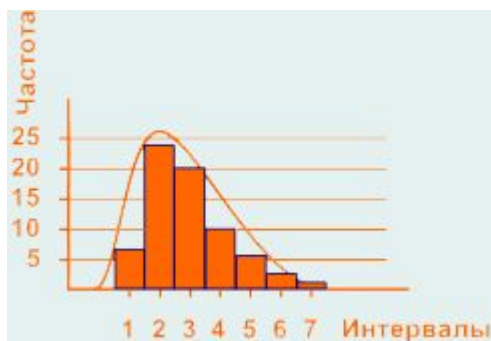


# Возможные отклонения от нормального распределения



## Гистограмма смещена влево (асимметрия влево):

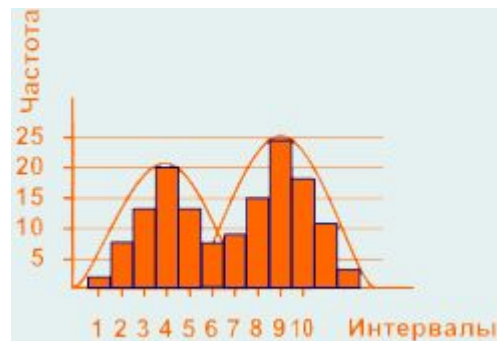
- смещение процесса к верхней границе допуска,
- из множества измерений отсортированы результаты, которые выпадают за пределы верхней границы допуска,
- природа процесса физически запрещает любые измерения больше чем максимальные значения допуска.



## Гистограмма смещена вправо (асимметрия вправо):

- Смещение процесса к нижней границе допуска,
- из множества измерений отсортированы результаты, которые выпадают за пределы нижней границы допуска,
- природа процесса физически запрещает любые измерения меньше чем минимальные значения допуска.

# Возможные отклонения от нормального распределения



## Бимодальность:

- Гистограмма отображает два совмещенных процесса. результаты измерений получены от двух разных устройств, двух операторов, контролеров, разных измерительных инструментов, или с разных точек измерения.

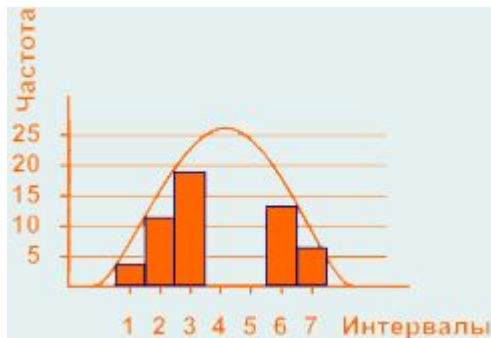


## Гистограмма усечена:

Распределение не является нормальным, т.к. нет постепенного снижения частоты результатов измерений от центра к границам допуска.

- процесс не способен удовлетворять спецификациям и часть измерений отсортирована с двух сторон при приближении к границам допуска,
- потеряны чересчур малые значения результатов измерений.

# Возможные отклонения от нормального распределения



## Гистограмма не имеет центра:

Центр распределения был отсортирован из набора данных результатов измерений.

- недостаточные требования в инженерной спецификации.



## Гистограмма содержит выступы на границах:

- часть измерений на удаленных от центра сторонах распределения была изменена, чтобы привести характеристики процесса в соответствие с установленным полем допуска
- измерения, выходящие за пределы поля допуска были записаны как входящие в поле допуска.

# Достоинства и недостатки гистограмм

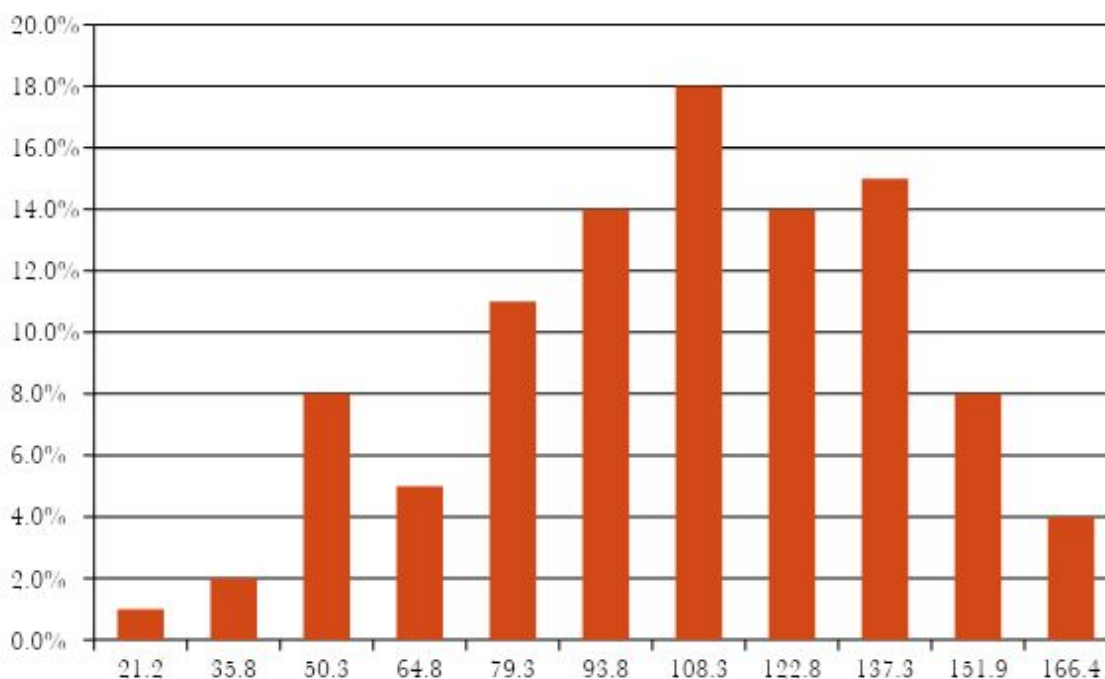
- наглядность,
  - простота,
  - возможность быстро представить вид распределения большого числа данных,
  - показывает взаимосвязь изменения контролируемых параметров по отношению к инженерным спецификациям.
- отсутствие возможности количественно оценить стабильность процесса,
  - отсутствие привязки ко времени,
  - необходимость большого числа данных для точной оценки структуры распределения,
  - возможность различного толкования результатов,
  - субъективность в представлении формы распределения.



# Пример

104,1	124,1	48,1	95,3	107,8	41,4	66,0	117,3	123,6	100,6	101,2	83,7	78,6	112,6	123,4	148,7	70,1	151,8	132,9	75,7
141,1	108,7	83,9	106,5	84,4	106,7	79,1	90,4	99,8	126,6	120,8	52,2	129,1	113,8	141,3	65,2	105,3	137,1	121,4	97,7
93,8	33,5	100,8	92,5	112,4	166,4	130,4	76,8	117,2	21,2	38,5	93,7	94,7	130,5	116,9	143,7	77,1	80,1	127,5	103,5
63,7	115,3	127,4	158,5	141,3	135,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0
115,0	76,2	141,2	65,3	64,3	127,4	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0	115,0

макс 166,4  
 мин 21,2  
 n 10,0  
 интервал 14,5



151,5 15,0 15,0%  
 151,9 8,0 8,0%  
 166,4 4,0 4,0%

# Стратификация

- один из инструментов качества, предназначенный для выявления какой-либо закономерности в массиве данных за счет их разделения.
- применяется в том случае, когда данные из различных источников сосредоточены вместе и это мешает определить структуру или их системность.
- используют совместно с другими инструментами анализа данных.

# Порядок выполнения стратификации

1. Определяются факторы, по которым будет проводиться стратификация (время, операторы, оборудование, условия производственных, материалы и средства измерения).
2. Определяется число страт (слоев). Количество страт берется соответственно количеству факторов, выявленных на предыдущем шаге.
3. Выбирается необходимый инструмент качества для графического представления статистических данных (диаграмма разброса, контрольная карта или гистограмма).
4. Определяется количество статистических данных, попадающих в каждую страту.
  - различия между значениями случайной величины внутри страты должны быть как можно меньше по сравнению с различием ее значений в исходной совокупности данных.
  - различия между стратами должны быть как можно больше.
5. На выбранный графический инструмент качества «наносятся» данные с указанием принадлежности этих данных к каждой из страт.
6. Проводится анализ подмножества данных. Анализ данных проводится для каждой страты отдельно.

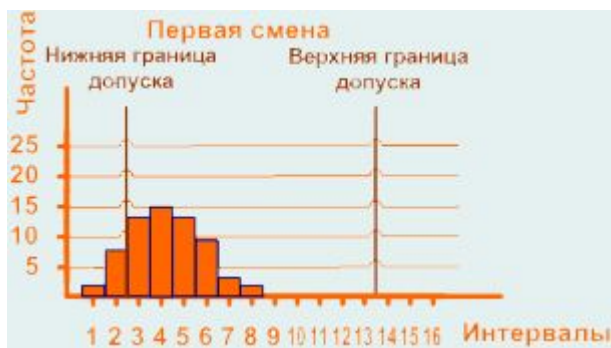
# Пример



**Распределение случайной величины до того, как применена стратификация данных:**

Форма гистограммы показывает распределение близкое к нормальному с выходом части значений случайной величины за пределы поля допуска

## Факторы стратификации: утренняя, дневная и вечерняя смены



распределение случайной величины близкое к нормальному, однако распределение смещено к нижней границе допуска



распределение случайной величины близкое к нормальному и находится в пределах поля допуска, но существует асимметрия влево



распределение случайной величины имеет асимметрию вправо и смещено к верхней границе допуска. Кроме того, разброс случайной величины для третьей смены больше чем для первой и второй смены