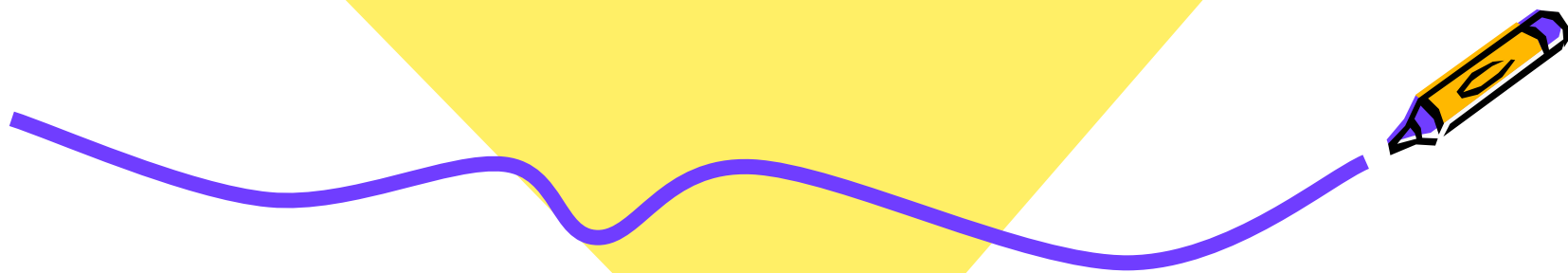


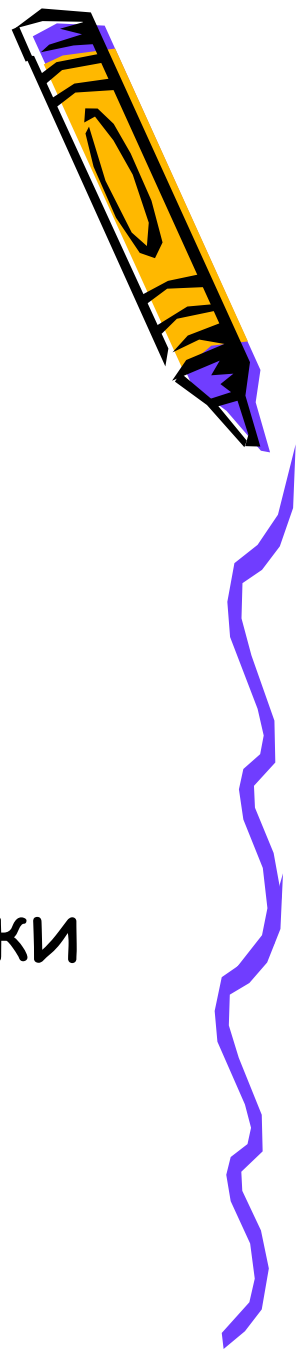


ЭЛЕМЕНТЫ
МАТЕМАТИЧЕСКОЙ
СТАТИСТИКИ



Статистика - дизайн информации





Цель:

- Дать понятие генеральной и выборочной совокупности, полигону и гистограмме частот
- Научиться строить полигон и гистограмму частот
- Познакомится с параметры оценки генеральной совокупности



Генеральная совокупность и выборка

- *Опр 1: Генеральной совокупностью* называется совокупность, из которой отбирают часть объектов.
- *Опр 2: Выборка (или выборочная совокупность)* - это множество объектов, случайно отобранных из генеральной совокупности.
- *Опр 3: Число объектов генеральной совокупности и выборки называют соответственно **объемом** генеральной совокупности и **объемом** выборки.*



- *Опр 4:* Если выборку отбирают по одному объекту, который обследуют и снова возвращают в генеральную совокупность, то выборка называется **повторной**. Если объекты выборки уже не возвращаются в генеральную совокупность, то выборка называется **бесповторной**.



Статистическое распределение выборки



- Пусть из генеральной совокупности извлечена выборка, причем x_1, x_2, \dots, x_k объёма N .
- *Опр 5:* Наблюдаемые значения x_1, x_2, \dots, x_k называют **вариантами**, а последовательность вариантов, записанная в возрастающем порядке, - **вариационным рядом**.
- *Опр 6:* Числа наблюдений n_1, n_2, \dots, n_k называют частотами, а их отношения к объёму

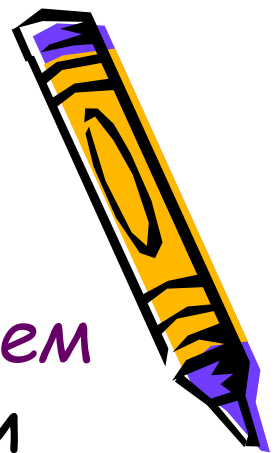
$$\frac{n_1}{n} = \omega_1, \quad \frac{n_2}{n} = \omega_2, \quad \dots, \quad \frac{n_k}{n} = \omega_k$$

- **относительными частотами**.

- Сумма относительных частот равна единице:

$$\omega_1 + \omega_2 + \dots + \omega_k = 1$$





- *Опр 7: Статистическим распределением выборки* называют перечень вариантов и соответствующих им частот или относительных частот.

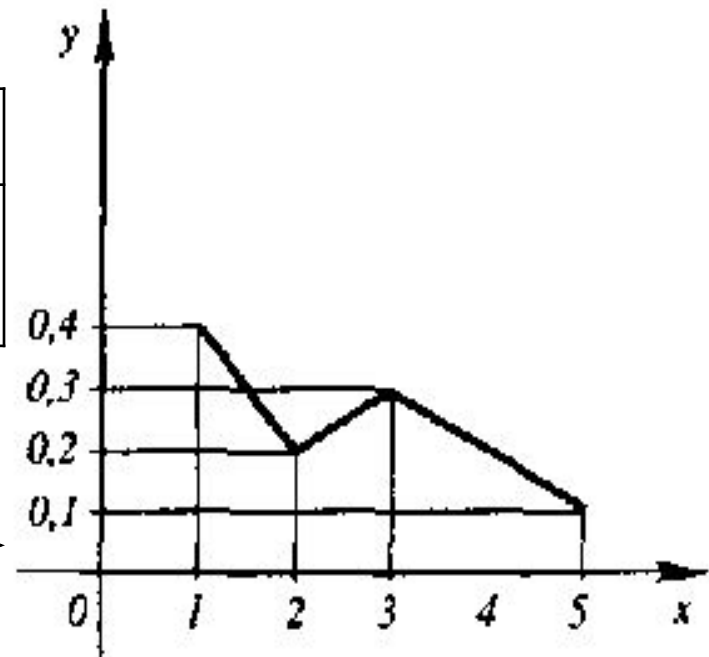


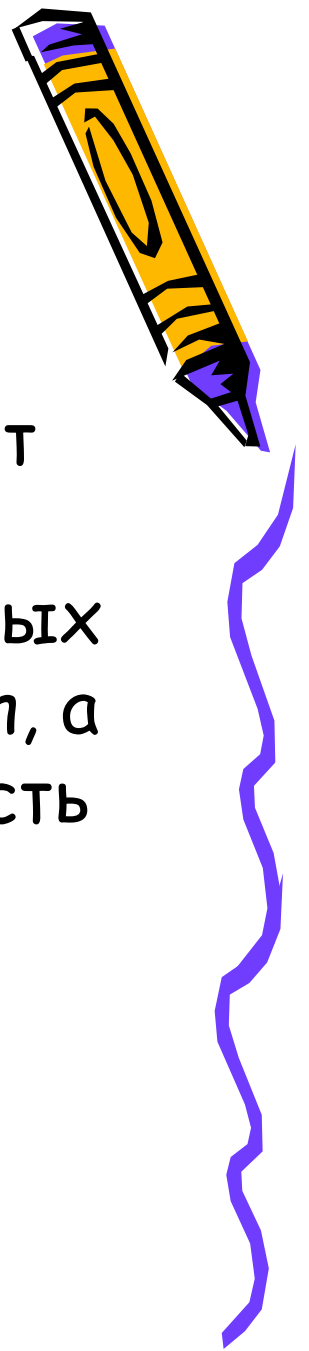


- *Опр 8: Полигоном частот называют ломанную отрезки которой соединяют точки .*
- Для построения полигона на оси Ox откладывают значения вариант x_i , на оси Oy - значения частот n_i (относительных частот w_i).

| | | | | |
|-----------------------------|-----|-----|-----|-----|
| Варианта x_i | 1 | 2 | 3 | 5 |
| Относительная частота p_i | 0,4 | 0,2 | 0,3 | 0,1 |

Полигон частот →

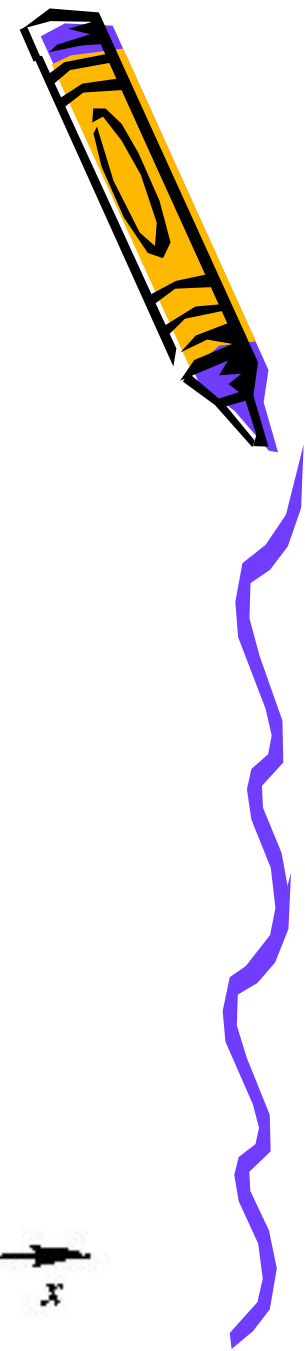




- *Опр 9: Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h , а высоты равны отношению $\frac{n_i}{h}$ (плотность частоты).*

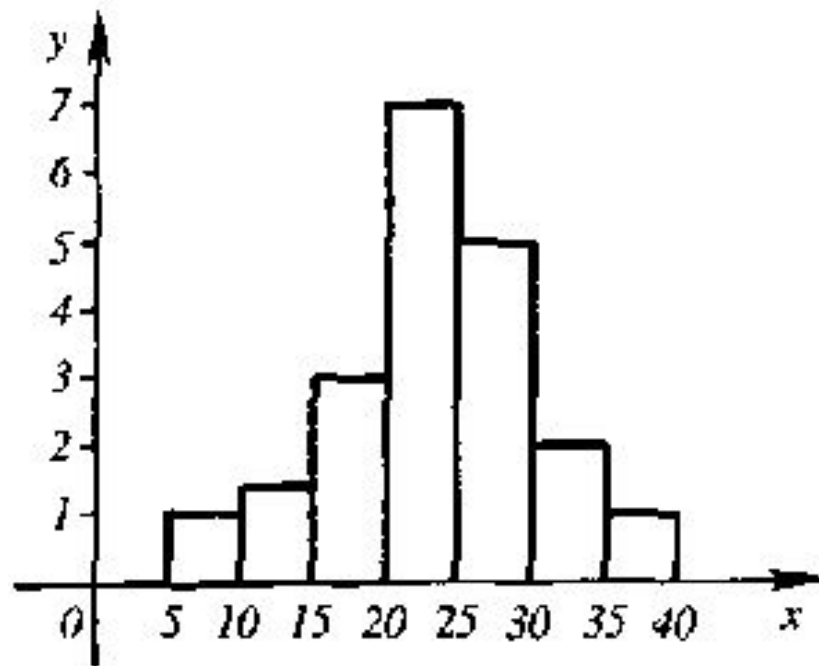


Непрерывное распределение объема $n = 100$



| Частичный интервал h | Сумма частот вариант частичного интервала n_j | $\frac{n_j}{h}$ |
|------------------------|---|-----------------|
| 5-10 | 4 | 0,8 |
| 10-15 | 6 | 1,2 |
| 15-20 | 16 | 3,2 |
| 20-25 | 36 | 7,2 |
| 25-30 | 24 | 4,8 |
| 30-35 | 10 | 2,0 |
| 35-40 | 4 | 0,8 |

Гистограмма частот



Оценка параметров генеральной совокупности

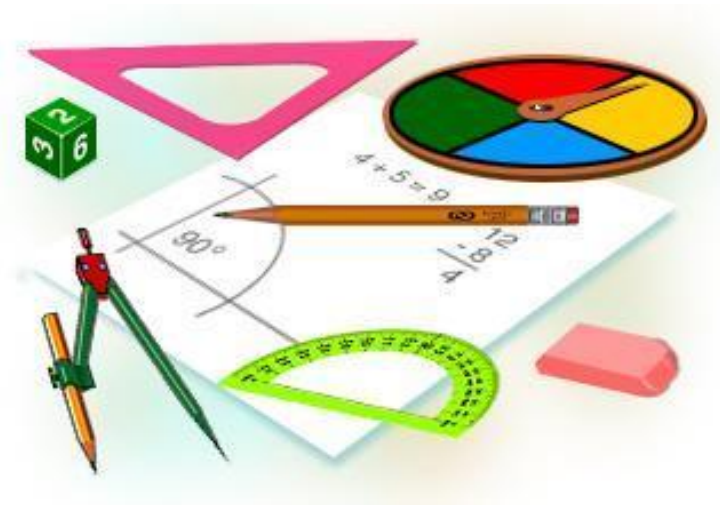


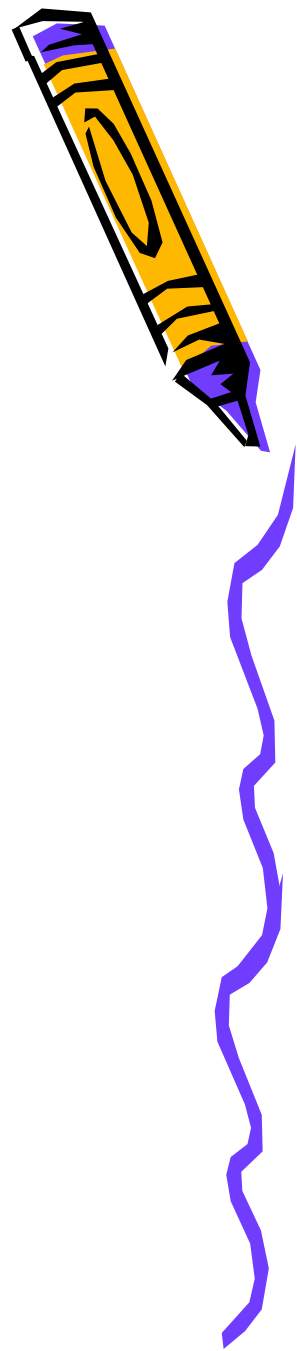
- *Опр 10: Статистической оценкой Θ^** неизвестного параметра Θ теоретического распределения называют функцию $f(X_1, X_2, \dots, X_n)$ от наблюдаемых случайных величин.
- *Опр 11: Точечной* называют статистическую оценку, которая определяется одним числом $\Theta^* = f(x_1, x_2, \dots, x_n)$, где x_1, x_2, \dots, x_n - результаты n наблюдений над количественным признаком X (выборка).





- Опр 12: *Несмещенной* называют точечную оценку, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки.
- Опр 13: *Смещенной* называют точечную оценку, математическое ожидание которой не равно оцениваемому параметру.





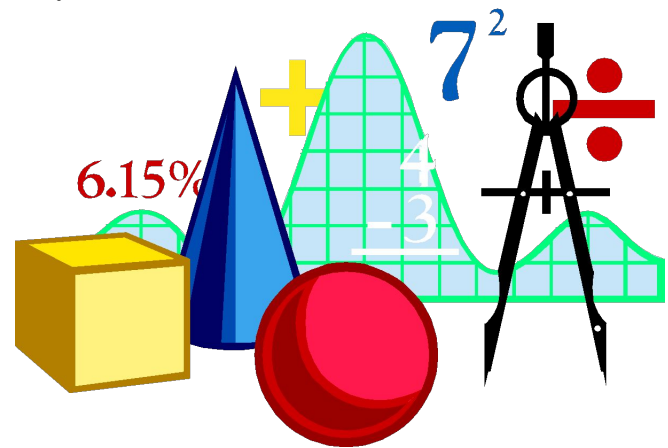
- *Опр 14: Выборочной средней $\bar{x}_{\hat{a}}$* называют среднее арифметическое значений признака выборочной совокупности.
- *Опр 15: Выборочной дисперсией D_B* называется среднее арифметическое квадратов отклонений наблюдаемых значений признака X от выборочного среднего $\bar{x}_{\hat{a}}$.

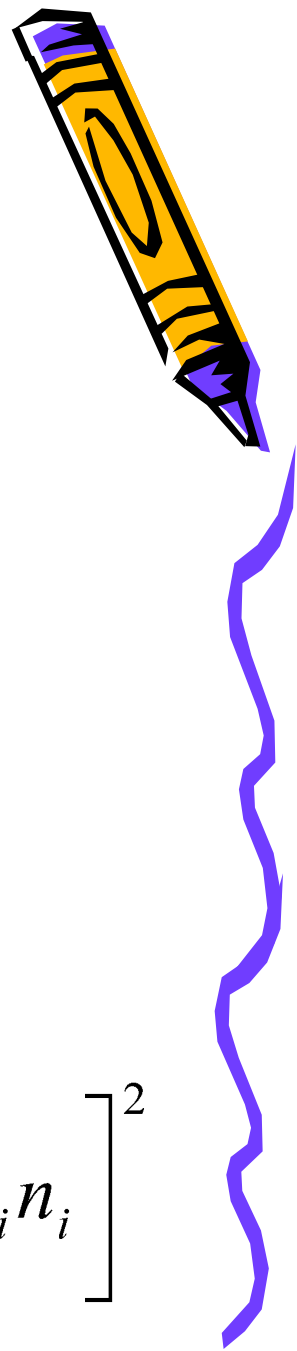




- Несмещенной оценкой генеральной средней (математического ожидания) служит выборочная средняя $\bar{x}_a = \frac{1}{N} \cdot \sum_{i=1}^k n_i x_i$ где x_i - варианты выборки, n_i - частота варианты x_i ,

$$N = \sum_{i=1}^k n_i \quad - \text{объем выборки.}$$




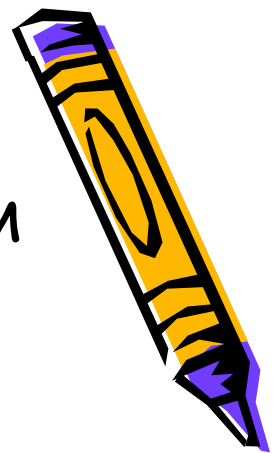


- Несмещенной оценкой генеральной дисперсии служит исправленная выборочная дисперсия $s^2 = \frac{N}{N-1} \cdot D_{\hat{a}}$

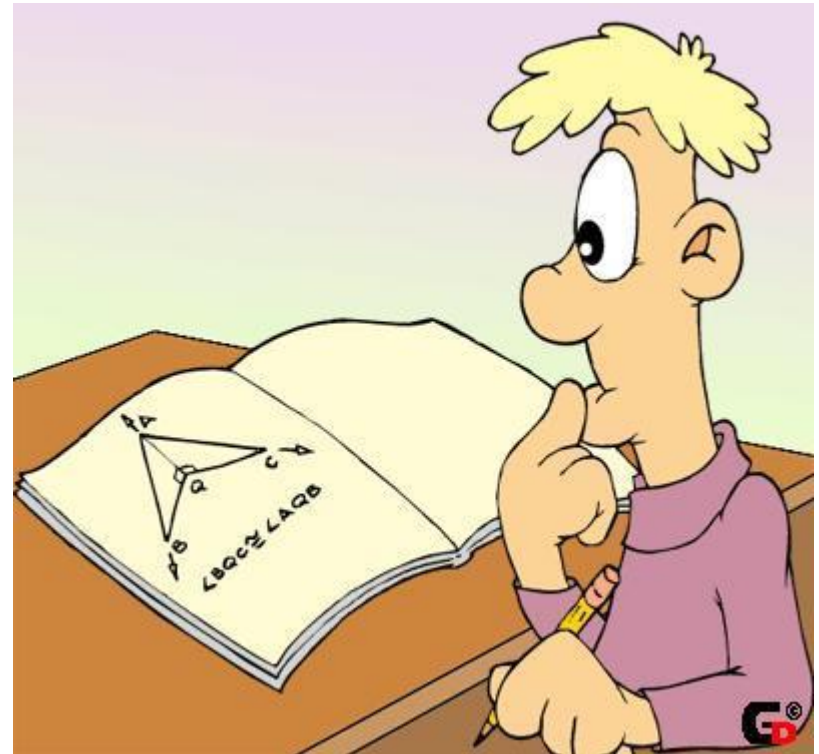
или $s^2 = \frac{1}{N-1} \cdot \sum_{i=1}^k x_i^2 n_i - \frac{N}{N-1} \bar{x}^2.$

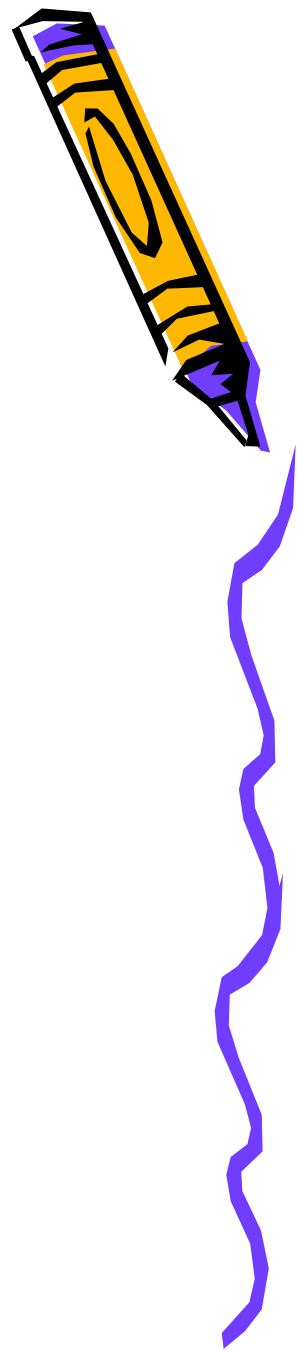
- Смещенной оценкой генеральной дисперсии служит выборочная дисперсия


$$D_{\hat{a}} = \bar{x}^2 - [\bar{x}]^2 = \frac{1}{N} \sum_{i=1}^k x_i^2 n_i - \left[\frac{1}{N} \sum_{i=1}^k x_i n_i \right]^2$$



- Выборочным средним квадратическим отклонением (стандартом) называют квадратный корень из выборочной дисперсии $\sigma_{\hat{a}} = \sqrt{D_{\hat{a}}}$.





- Доверительный интервал - это интервал, который с заданной вероятностью покрывает неизвестную характеристику.

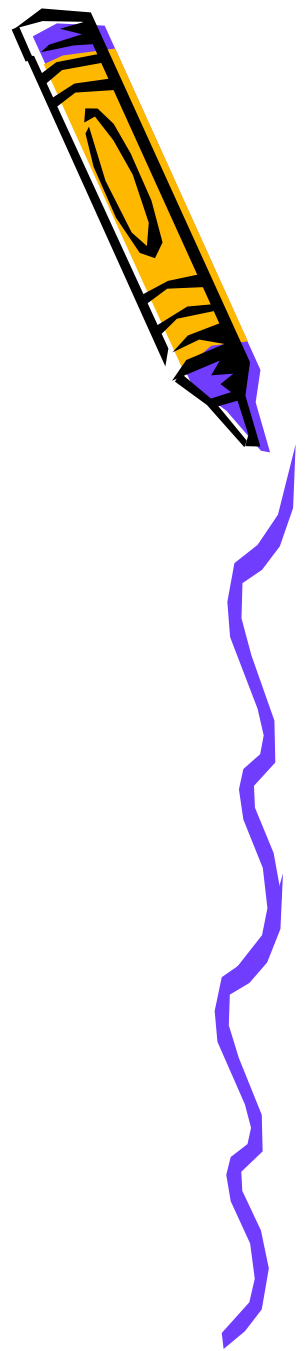




Доверительный интервал для математического ожидания

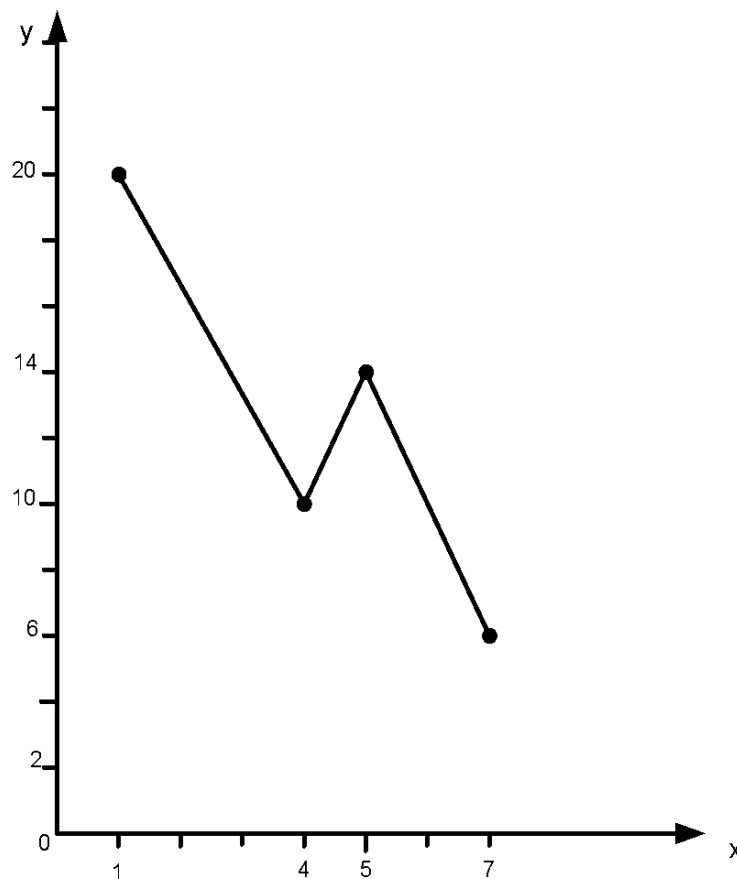
$$\bar{x} - t_{\gamma} \cdot \frac{s}{\sqrt{N}} < M(X) < \bar{x} + t_{\gamma} \cdot \frac{s}{\sqrt{N}}$$

где t_{γ} - аргумент распределения
Стьюдента, соответствующей
доверительной вероятности γ и $(N-1)$
степени свободы.



Пример 1: Построить полигон частот по данному распределению

| | | | | |
|-------|----|----|----|---|
| x_i | 1 | 4 | 5 | 7 |
| n_i | 20 | 10 | 14 | 6 |



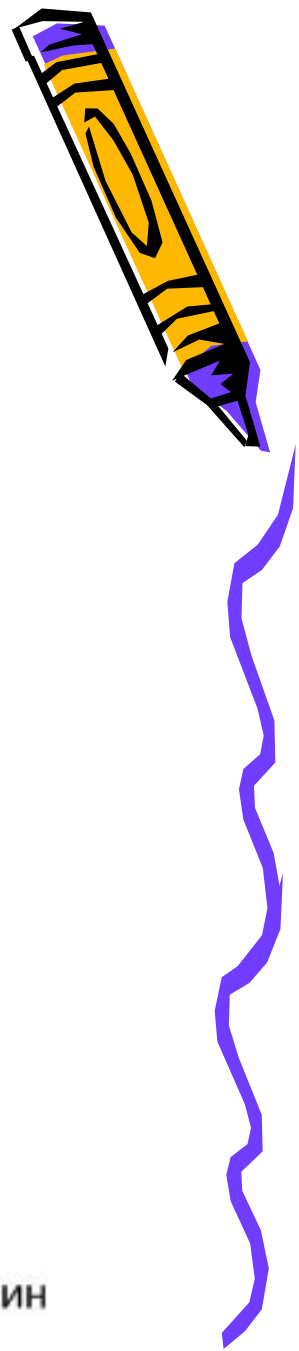
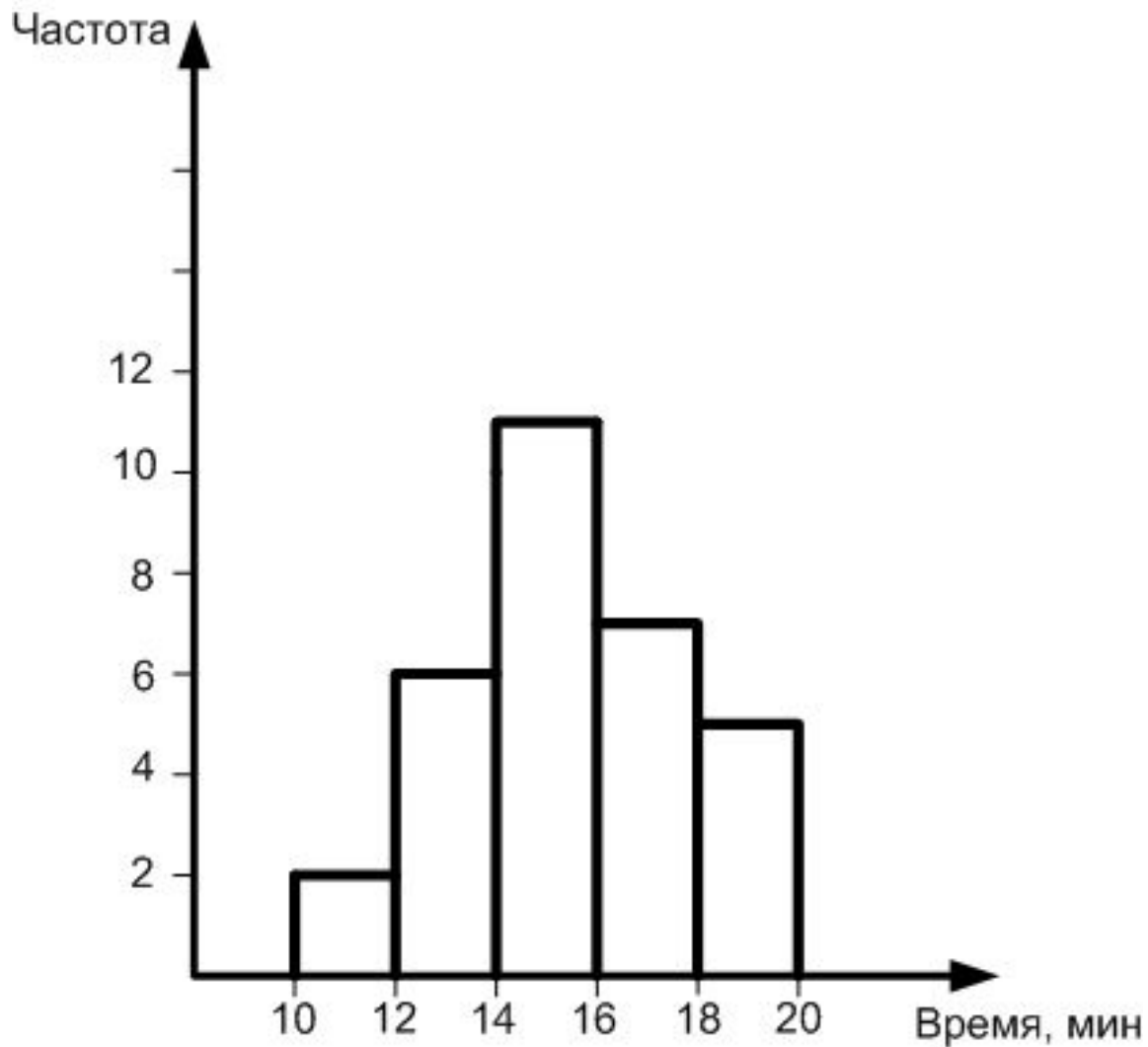
Пример 2: Наблюдая за работой бригады токарей, установили, сколько времени тратили они на обработку одной детали. Обобщая полученные данные составили таблицу.

- Пользуясь таблицей, постройте гистограмму частот, характеризующую распределение токарей бригады по времени, затрачиваемому на обработку одной детали.

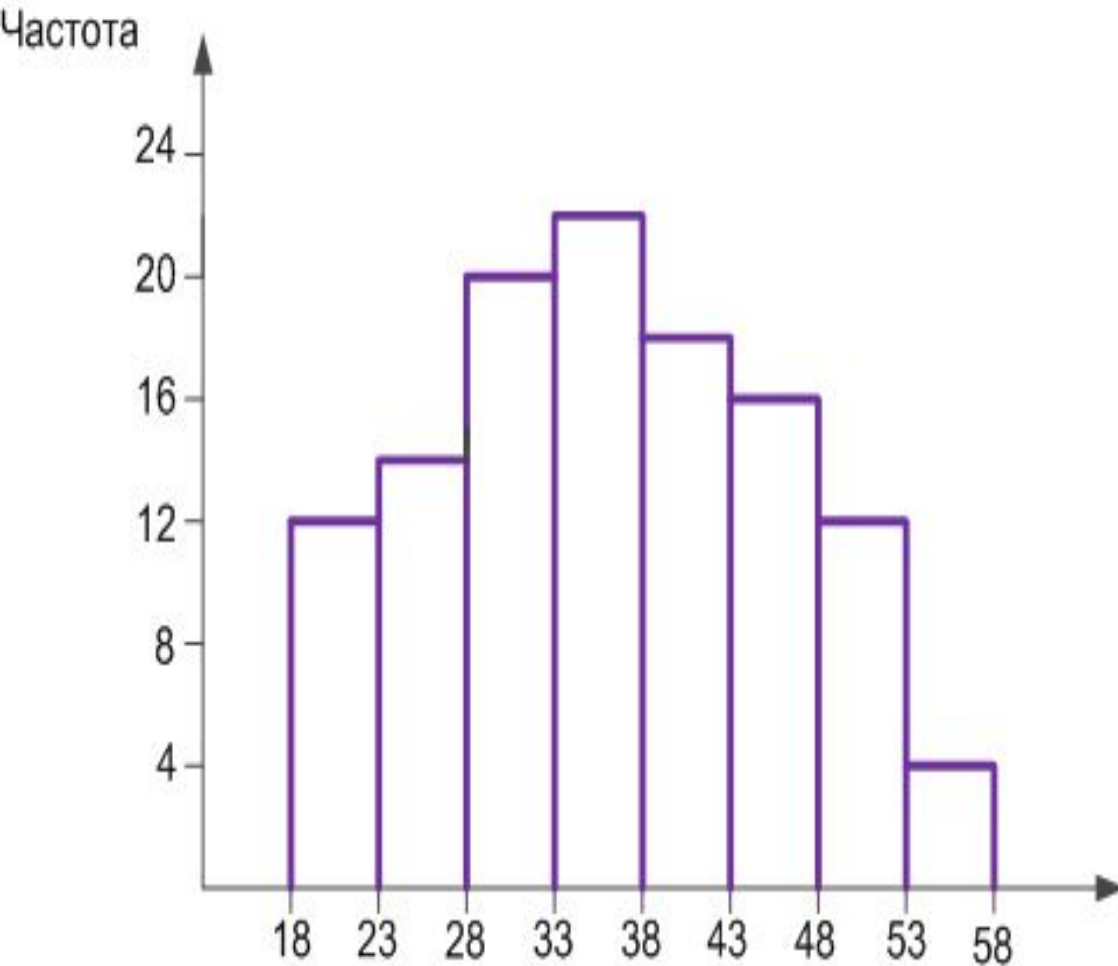
| Время, мин | Число токарей |
|---------------|------------------|
| 10-12 | 2 |
| 12-14 | 6 |
| 14-16 | 11 |
| 16-18 | 7 |
| 18-20 | 5 |



Решение:



Пример 3: На гистограмме представлены данные о распределения рабочих строительной организации по возрастным группам:



Пользуясь гистограммой, найдите:

- число рабочих строительной организации в возрасте от 18 до 23 лет;
- возрастную группу, к которой относится наибольшее число рабочих;
- общее число рабочих строительной организации.

- Независимо от того, в какой отрасли знания получены числовые данные, они обладают определёнными свойствами, для выявления которых может потребоваться особого рода научный метод обработки. Последний известен как статистический метод или, короче, статистика.



Дж.Юл.М.Кендалл,
«Теория статистики»