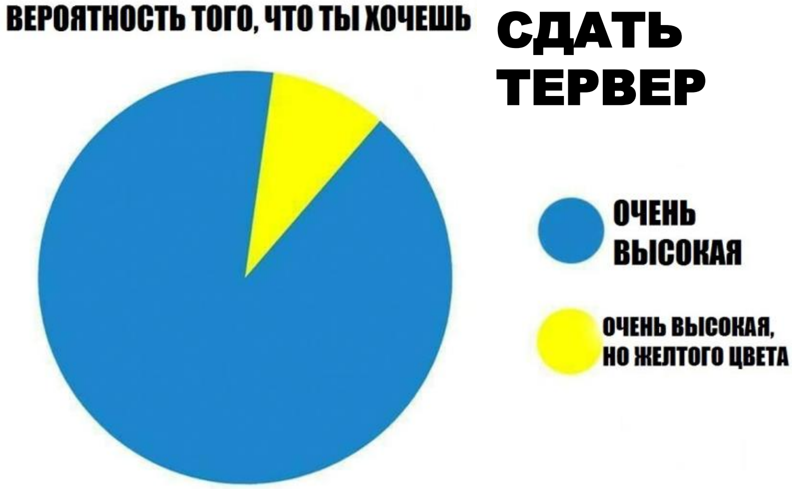
Математическая статистика

Математическая статистика, как и теория вероятности, изучает случайные явления и их закономерности

Теория вероятности исследует случайные явления, заданные полностью их вероятностной моделью и еще до проведения эксперимента позволяет определить закономерности.



Математическая статистика

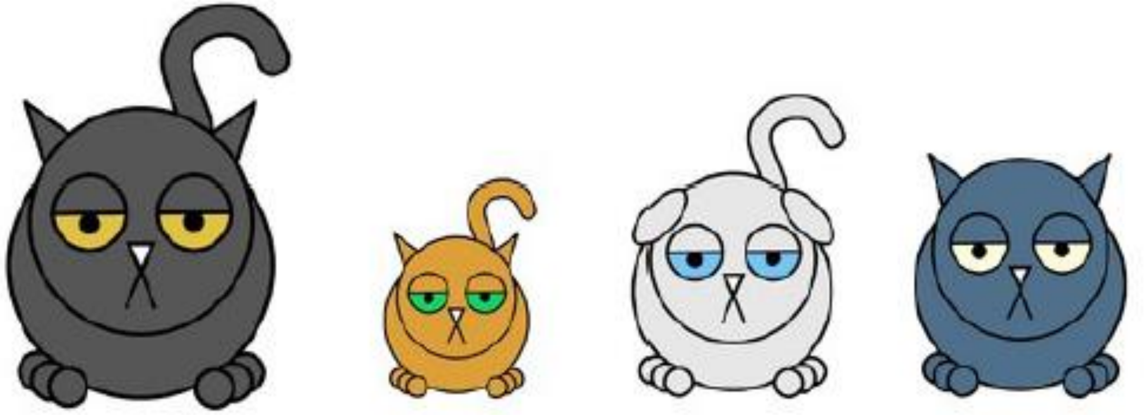
В ***мат. статистике*** вероятностная модель случайного явления (объекта) либо неизвестна, либо известна частично.

Недостаток информации компенсируется тем, что нам «позволено» проводить наблюдения над явлением (объектом) и на их основе *восстанавливать* «картину» явления.

Случайные признаки

Объекты исследования обладают теми или иными интересующими исследователя свойствами или **признаками,** способными изменять свое количественное или качественное состояние при переходе от одного объекта к другому.

«признак» = «показатель» = «переменная» = «фактор» = величины, которые могут принимать любые значения из заданного множества значений.



Случайные признаки

Основные признаки, фиксируемые при наблюдении, устанавливаются в зависимости от объекта и целей исследования. При наблюдении фиксируются **только некоторые** признаки.

Пример: исследование металла.
Признаки: предел прочности при сжатии, модуль упругости и т.д.

<u>Одновременное наблюдение</u> большого числа основных признаков *бесполезно*:

1. Довольно сложно
2. Статистические характеристики утрачивают ясность и смысл при большом числе признаков

Случайные признаки

Признаки

Качественные:

- Цвет глаз
- Профессия рабочего
- Отрасль промышленности

Количественные:

- Рост человека
- Длина детали
- Процент выполнения плана

Дискретные:

- Число членов семьи
- Количество бракованных изделий в партии

Непрерывные:

- Средний доход населения
- Величина ошибки измерения

Случайные величины

Случайные признаки

Необходимо знать все возможные значения признака, а также как часто те или иные значения появляются в одних и тех же условиях.

Закон распределения случайного признака: любое соотношение между возможными значениями и соответствующими им вероятностями.

Случайные признаки

Дискретные случайные признаки задаются **рядом распределения**

$p_1 = P\{X=x_1\}$

$p_2 = P\{X=x_2\}$

…

$p_n = P\{X=x_n\}$

X	$\mathbf{x_1}$	$\mathbf{x_2}$	**…**	$\mathbf{x_n}$
P	p_1	p_2	…	p_n

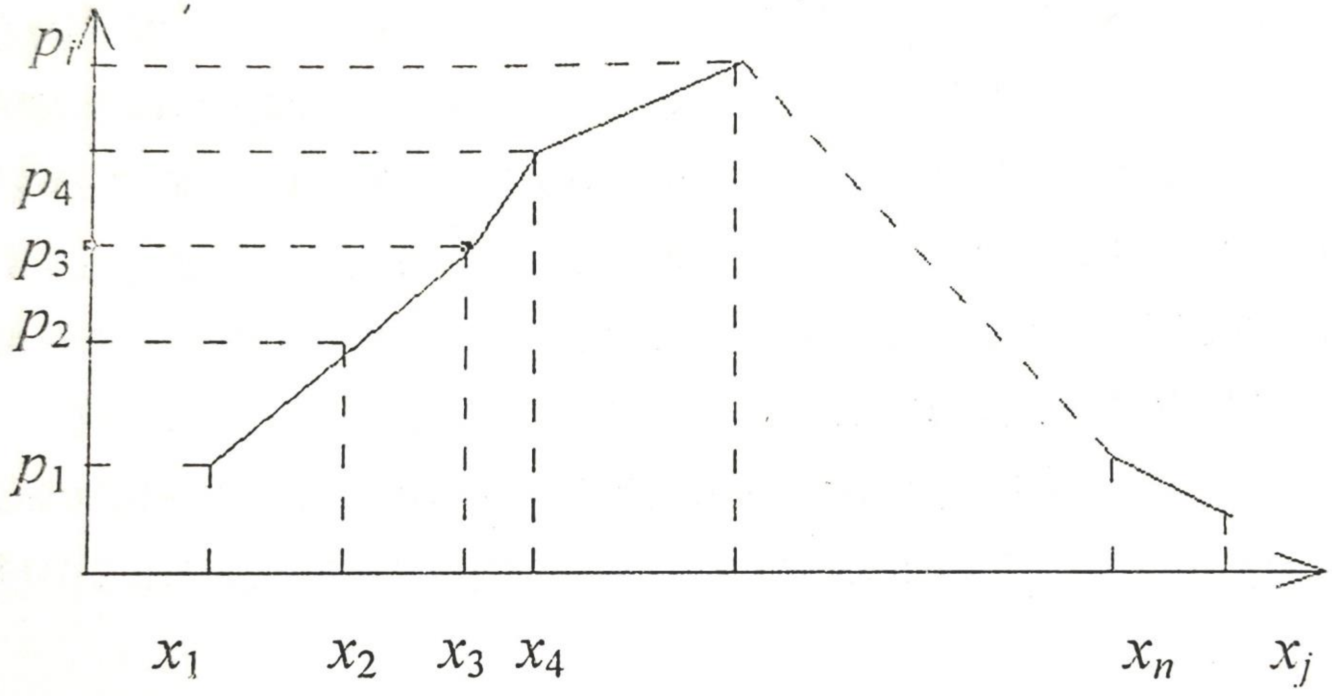


Рис. 1.1. Многоугольник распределения

Случайные признаки

Непрерывные случайные признаки задаются **плотностью распределения**, такой что

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

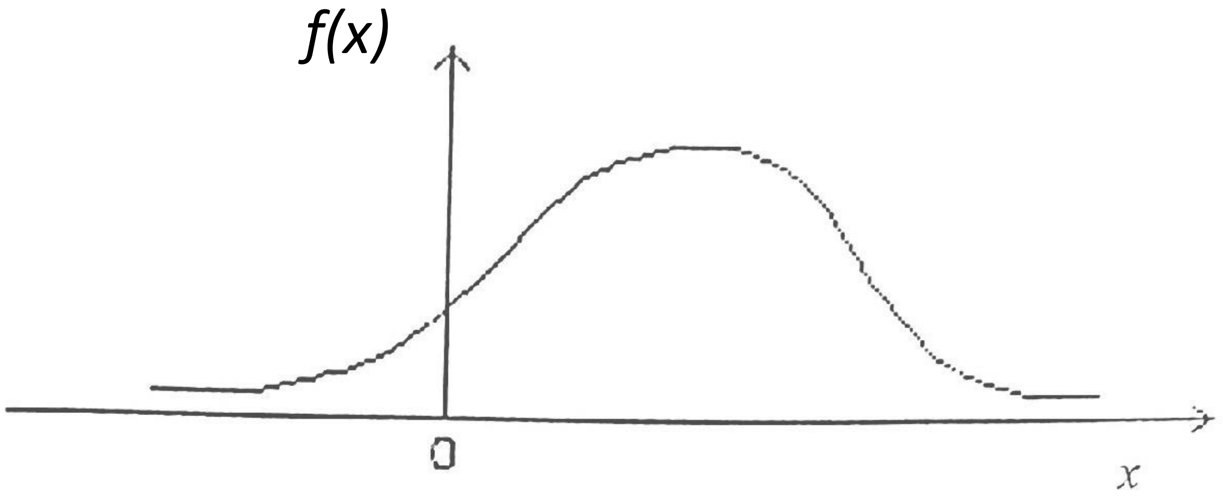


Рис. 1.2. Кривая распределения

Случайные признаки

Случайные признаки могут быть заданы **функцией распределения:**

$$F(x) = P\{X < x\}$$

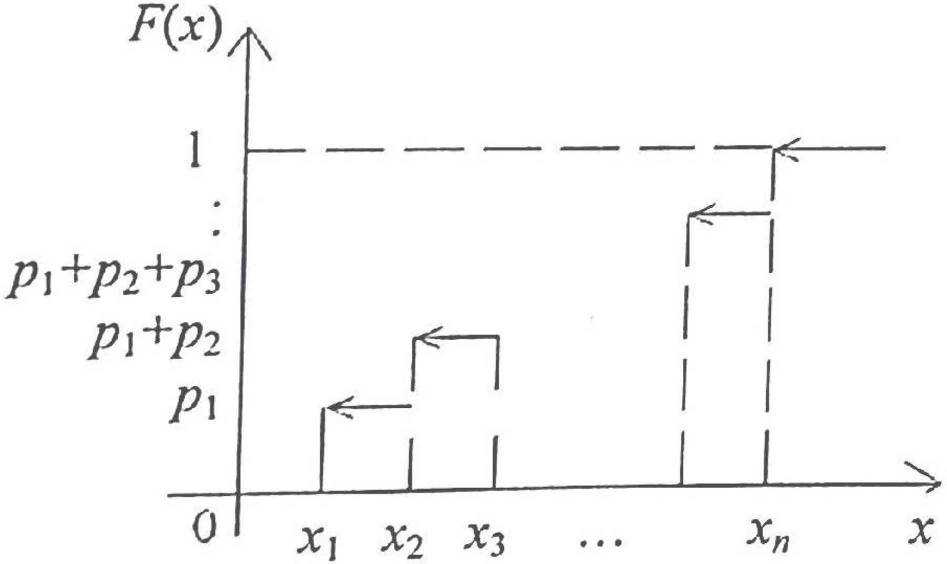


Рис. 1.3. Функция распределения дискретного признака

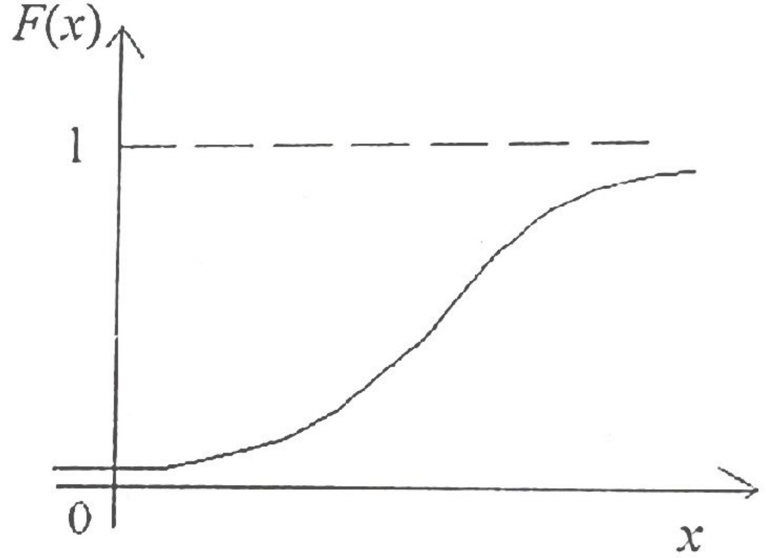


Рис. 1.4. Функция распределения непрерывного признака

Квантили распределения

При проведении исследований для получения тех или иных выводов с заданной надежностью используют понятие квантилей различных законов распределения.

Пусть X – непрерывный случайный признак с функцией распределения F(x) и плотностью f(x).

Квантилью порядка p или p-квантилью распределения F(x) называются величина x_p, являющаяся решением уравнения

$$F(x_p) = p, \ 0 < p < 1$$

Квантили распределения

Квантилью порядка p или p-квантилью распределения F(x) называются величина x_p, являющаяся решением уравнения

$$F(x_p) = p,\ 0 < p < 1$$

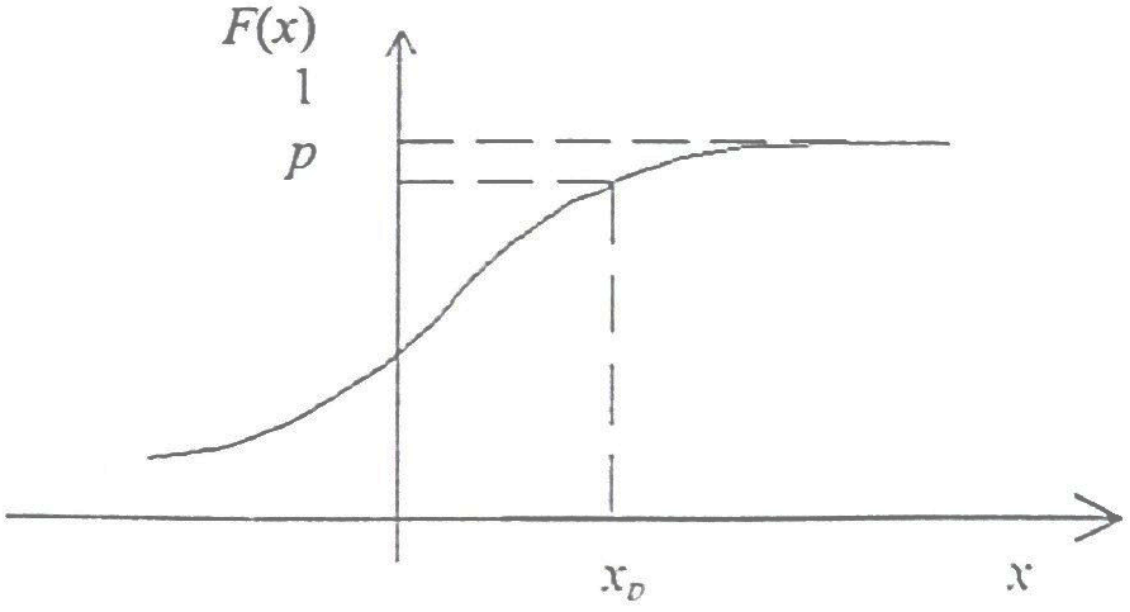


Рис. 2.1. Квантиль x_p распределения $F(x)$

Квантили распределения

Так как для непрерывного признака $F(x) = \int_{-\infty}^{x} f(x)dx$,
p-квантиль удовлетворяет соотношению

$$\int_{-\infty}^{x_p} f(x)dx = p$$

На рисунке заштрихованная площадь равна p

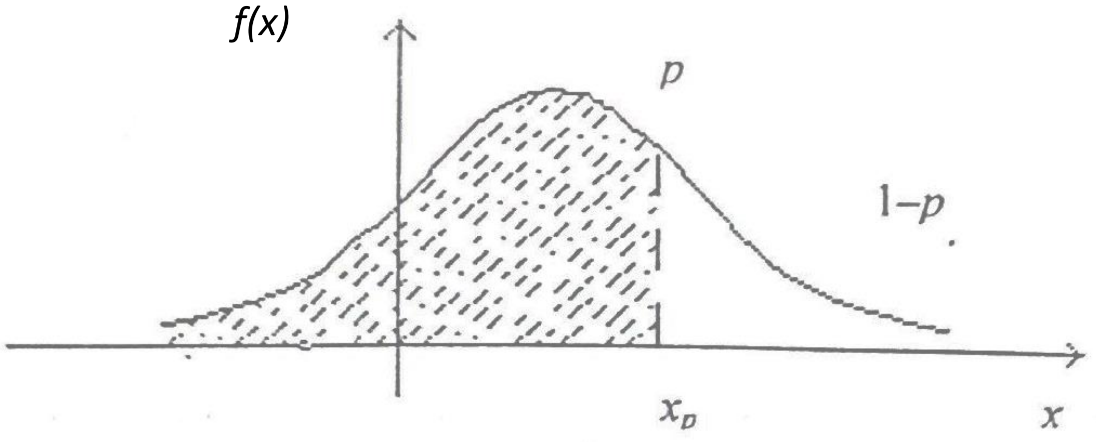
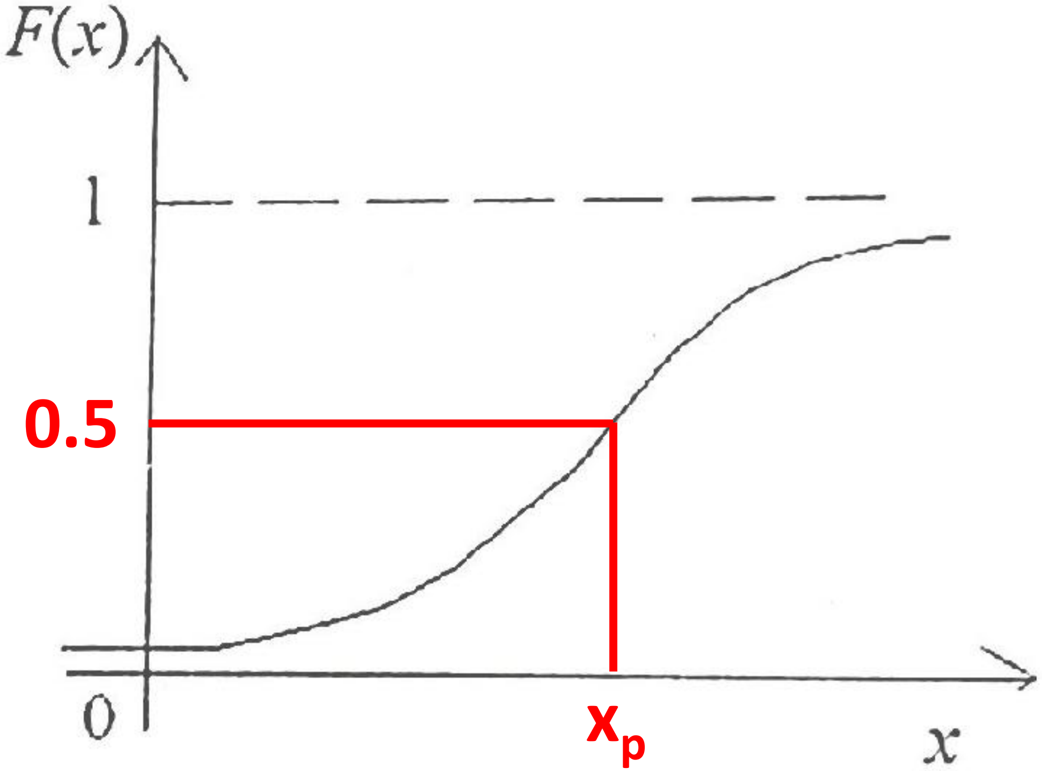
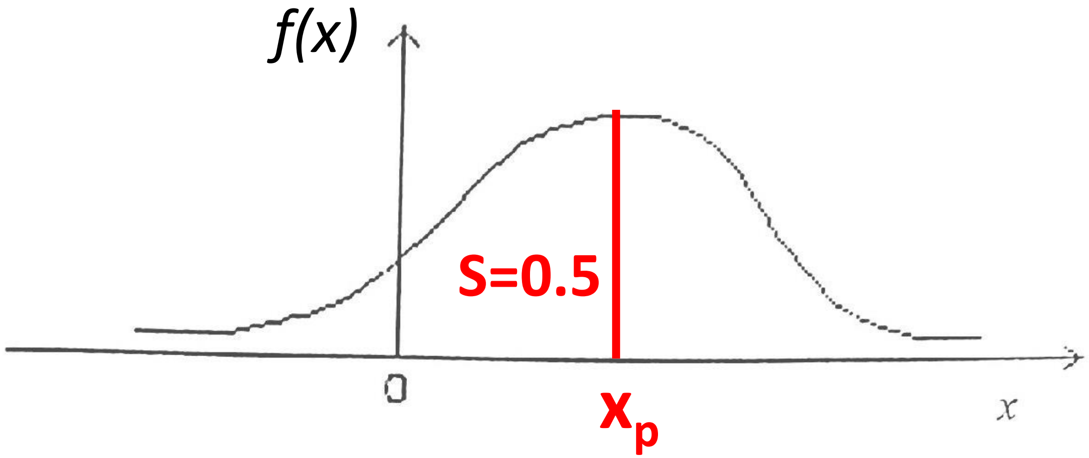


Рис. 2.2. Квантиль x_p распределения $p(x)$

Квантили распределения

Квантиль порядка p = 0.5 называется **медианой** распределения





Распределения

Наиболее часто используемые в математической статистике распределения:

- Нормальное распределение
- Распределение «хи-квадрат»
- Распределение Стьюдента

Рассмотрим распределения и способы нахождения их квантилей.

Нормальное распределение

Случайный признак *X* распределен по **нормальному закону** распределения с параметрами a и σ $(-\infty < a < +\infty, \sigma > 0)$, то есть $X \sim N(a, \sigma)$, если плотность распределения *X* задается формулой:

$$p(x) = p(x; a{,}\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Функция распределения:

$$F(x) = F(x; a{,}\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt = 0.5 + \Phi_0\left(\frac{x-a}{\sigma}\right),$$

где $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ – функция Лапласа

Нормальное распределение

Стандартное нормальное распределение: $a = 0$ и $\sigma = 1$

$$\varphi(x) = \varphi(x; a,\sigma) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\Phi(x) = F(x; 0,1) = 0.5 + \Phi_0(x)$$

Квантили $\boldsymbol{u_p(n)}$ порядка *p* для нормального распределения приведены в таблице. Величины $u_p(n)$ находятся как решение уравнения

$$\Phi(u_p) = p \Rightarrow \int_{-\infty}^{u_p} \varphi(x)dx = p$$

Нормальное распределение

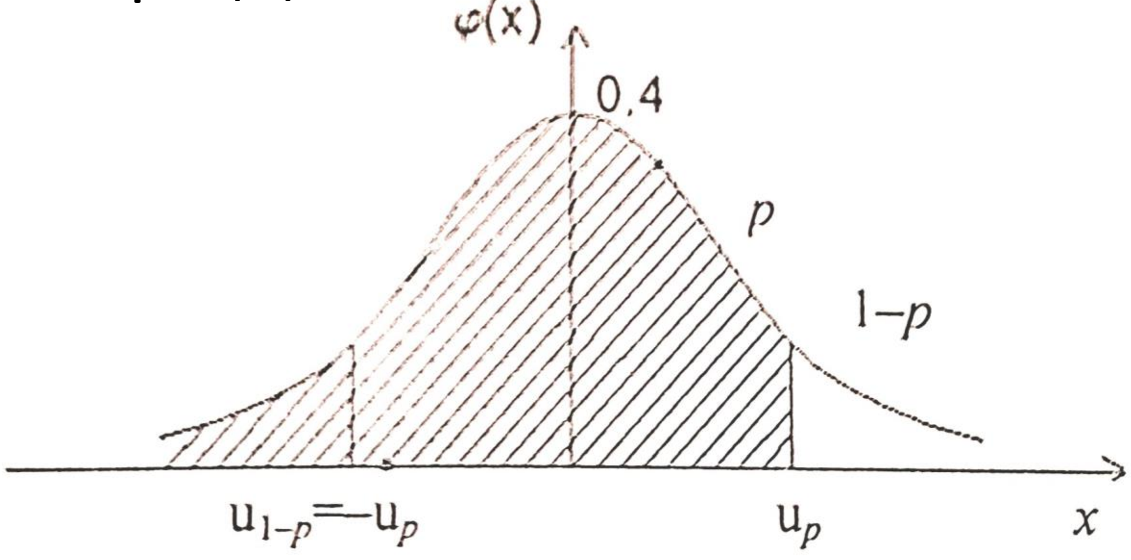


Рис. 2.3. Квантиль u_p стандартного нормального распределения

Квантили $u_p(n)$ порядка *p* для нормального распределения могут быть также найдены как решение уравнения

$$\Phi_0(u_p) = p - 0.5$$

Из симметричности распределения следует симметричность квантилей:

$$u_{1-p} = -u_p, 0 < p < 1$$

Распределение «хи-квадрат»

Пусть $X_1, X_2, \dots X_n$ – независимые случайные признаки, имеющие стандартное нормальное распределение.

Распределение случайной величины

$$\chi^2(n) = X_1^2 + X_2^2 + \cdots + X_n^2$$

называется **распределением (хи-квадрат) с n степенями свободы**.

Числовые характеристики:

$$M\chi^2(n) = n, \qquad D\chi^2(n) = 2n$$

Распределение «хи-квадрат»

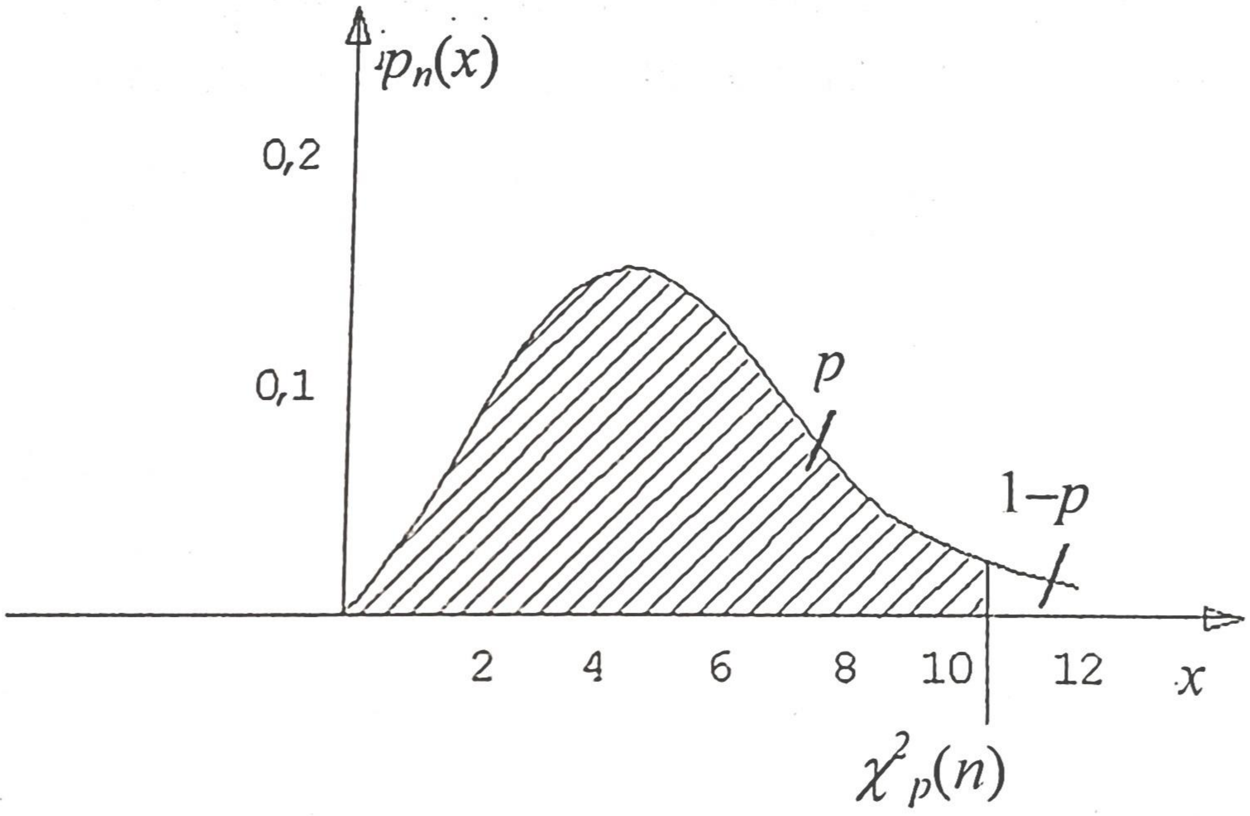


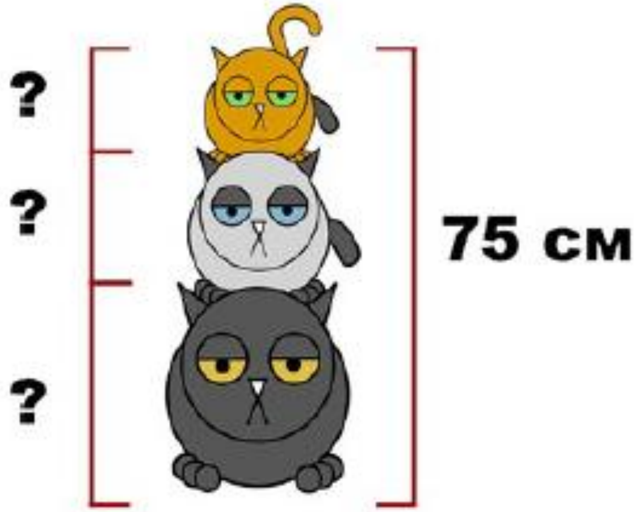
Рис.2.4. Квантиль $\chi^2_p(n)$ распределения χ^2 $(n = 6; p = 0{,}90)$

Квантили $\chi^2_p(n)$ порядка *p* для распределения χ^2 с *n* степенями свободы приведены в таблице.

Распределение «хи-квадрат»

Понятие «**степень свободы**»:

Пусть известна сумма размеров трех котиков – 75 см, но неизвестна величину каждого конкретного котика. Эти величины будут неизвестны ровно до тех пор, пока вы не начнете их измерять.

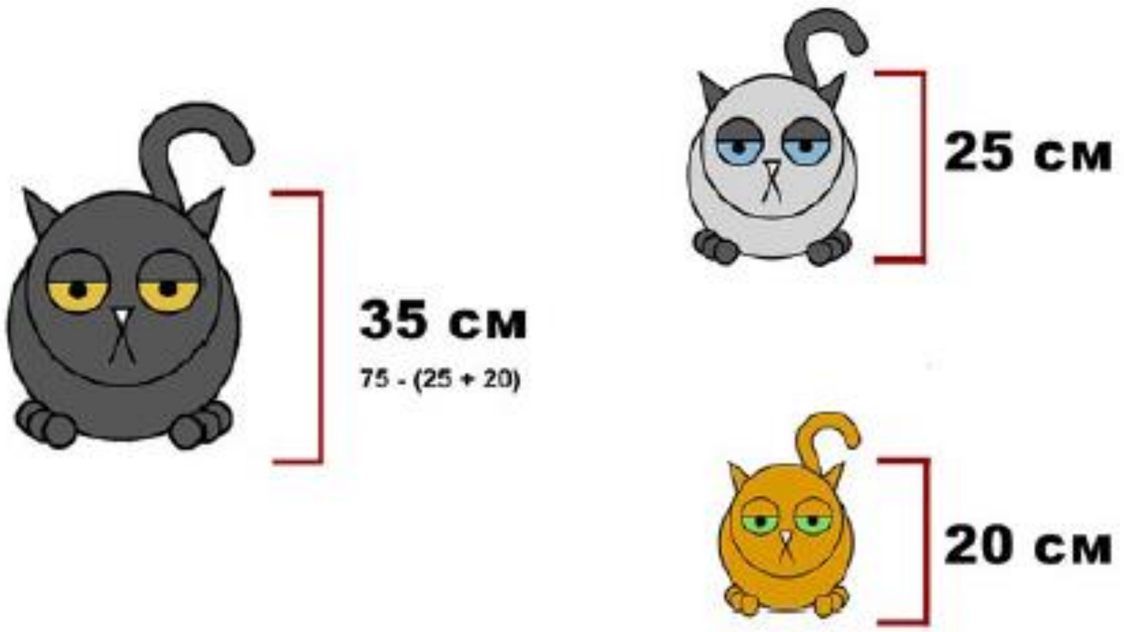


Представим, что мы узнали размер первого – 20 см. А значит, сумма размеров оставшихся котиков будет 55 см. При этом их конкретные размеры до сих пор неизвестны.

Распределение «хи-квадрат»

Измерим второго котика. Он оказался равен 25 см. Размер третьего кота перестал быть неизвестным — теперь мы можем его вычислить.

Число степеней свободы — это то количество котиков, которое мы должны измерить, чтобы однозначно узнать размер всех котиков при известном среднем или дисперсии. Если у вас только одна котиковая выборка, то это количество котиков минус единица.



Распределение «хи-квадрат»

При достаточно большом *n* распределение χ^2 имеет приближенно нормальное распределение с параметрами *n* и *2n.* Следовательно, при $n \to \infty$ для квантилей распределения «хи-квадрат» имеет место *асимптотическое соотношение*

$$\chi_p^2(n) \approx n + u_p\sqrt{2n},$$

где u_p – квантиль порядка p стандартного нормального распределения.

Пример: для *n* = 100 и *p* = 0.975
по таблице $\chi_{0.95}^2(100) = \mathbf{129.561}$,
а по приближенной формуле **127.72**

Распределение Стьюдента

Пусть $Y, X_1, X_2, \ldots X_n$ – независимые случайные признаки, имеющие стандартное нормальное распределение.
Распределение

$$t(n) = \frac{Y}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}} = \frac{Y}{\sqrt{\chi^2(n)\ /\ n}}$$

называется **распределение Стьюдента (t-распределением) с *n* степенями свободы**.

Числовые характеристики:

- $Mt(n) = 0$, если $n > 1$
- $Mt(1)$ не существует
- $Dt(n) = \frac{n}{n-2}$, если $n > 2$
- $Dt(n)$ не существует при $n \leq 2$

Распределение Стьюдента

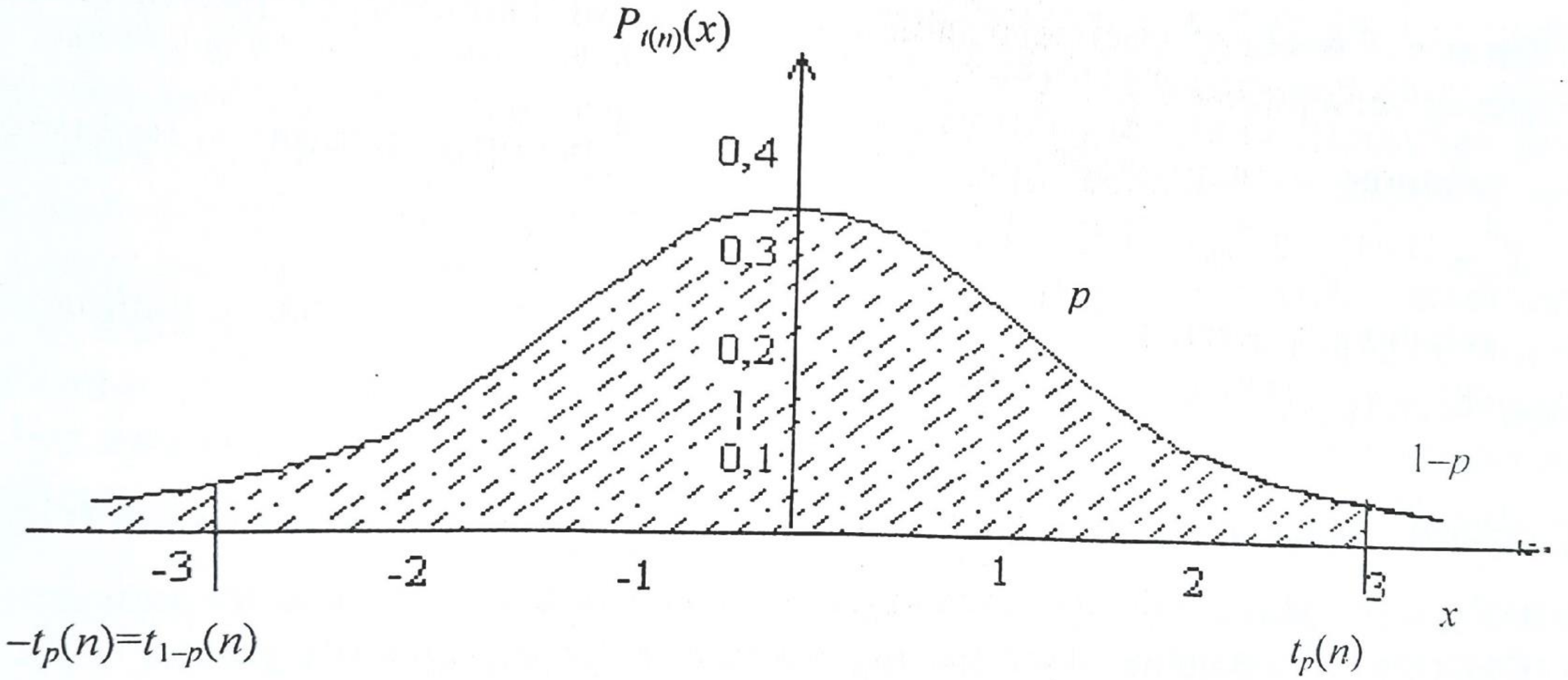


Рис. 2.5. Квантиль $t_p(n)$ распределения Стьюдента ($n = 4$; $p = 0{,}975$)

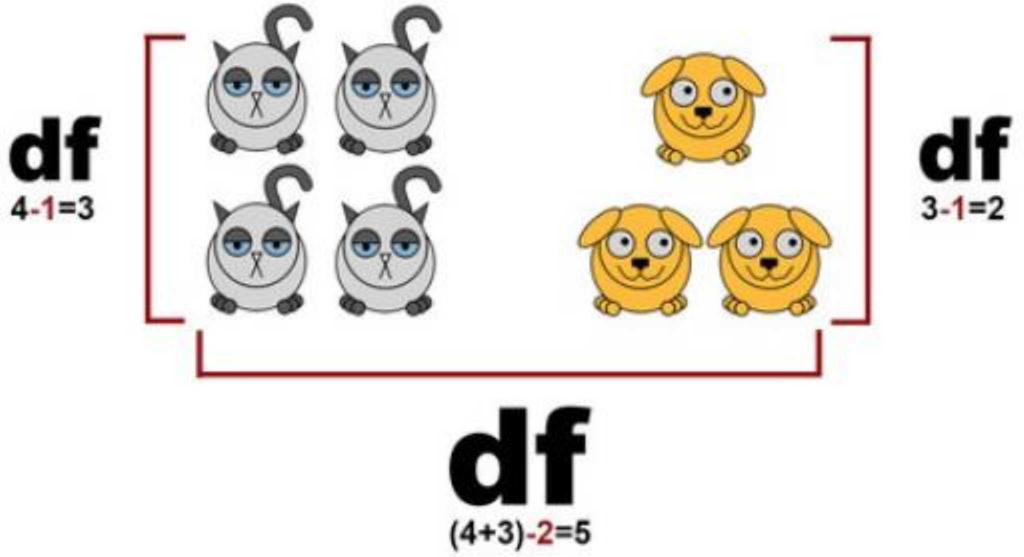
Квантили $t_p(n)$ порядка *p* для распределения Стьюдента с *n* степенями свободы приведены в таблице.

В силу симметричности t-распределения $t_{1-p}(n) = -t_p(n)$

Распределение Стьюдента

«Степень свободы»:

Если к выборке котиков добавляются еще и выборка пёсиков, то общее количество степеней свободы — это просто сумма степеней свободы котиков и пёсиков. Или по-другому — общее количество животных вычесть двойку.



Распределение Стьюдента

Если $n \to \infty$, то распределение Стьюдента сходится к стандартному нормальному распределению. Поэтому для соответствующих квантилей имеет место *асимптотическое соотношение* для больших значений n

$$t_p(n) \approx u_p,$$

где u_p – квантиль порядка p стандартного нормального распределения.

Пример: для $n = 100$ и $p = 0.975$
по таблице $t_{0.95}(100) = \mathbf{1.984}$,
а по приближенной формуле **1.960**.

Выборочный метод

Выборочная совокупность (выборка) –
совокупность случайно отобранных объектов исследования

Генеральная совокупность – совокупность объектов,
из которой производится выборка

Объем совокупности
(генеральной или выборочной) –
число объектов в ней.

Выборочный метод

Выборочная совокупность

Повторная выборка – отобранный объект возвращается в генеральную совокупность перед отбором следующего

Бесповторная выборка – отобранный объект не возвращается в генеральную совокупность

Если объем генеральной совокупности велик,
а выборка составляет лишь небольшую ее часть,
то различие между повторной и бесповторной выборками мало.

Выборочный метод

Выборочный метод – исследование выборок из генеральной совокупности

Основная цель метода – по выборочной совокупности вынести суждения о свойствах всей генеральной совокупности

Выборка

Совокупность определенных числовых данных

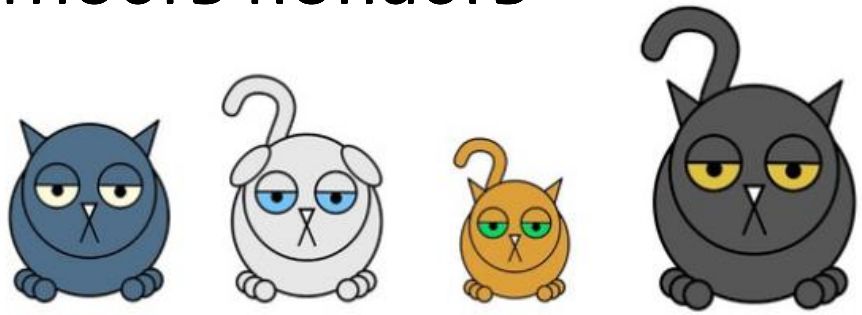
Совокупность случайных величин – результатов наблюдений

$x_1, x_2, \ldots, x_n$ – *случайная выборка* объема n

Выборочный метод

Репрезентативность (представительность) выборки – полнота и адекватность представления ею интересующих нас свойств исследуемой генеральной совокупности

Случайность выбора: каждый объект генеральной совокупности имеет одну и ту же вероятность попасть в выборочную совокупность.



Достижение случайности выбора часто сложно на практике. Сочетание регулярного (направленного) и случайного выбора иногда оказывается более эффективным. Способ сбора исходных данных должен тщательно планироваться.

Статистические данные

Установление неких закономерностей основано на изучении статистических данных. **Статистические данные** – сведения относительно некоторой совокупности объектов.

Допустимо заменить каждый объект связанным с ним *признаком* (или признаками) и говорить не о совокупности объектов, а о совокупности значений изучаемого признака (признаков). Мы отождествляем объекты исследования и множество характеризующих признаков.

Котики:

Барсик, Снежок, Рыжик, Черныш

$\Rightarrow$

Размер котиков:
50 см, 40 см, 20 см, 60 см

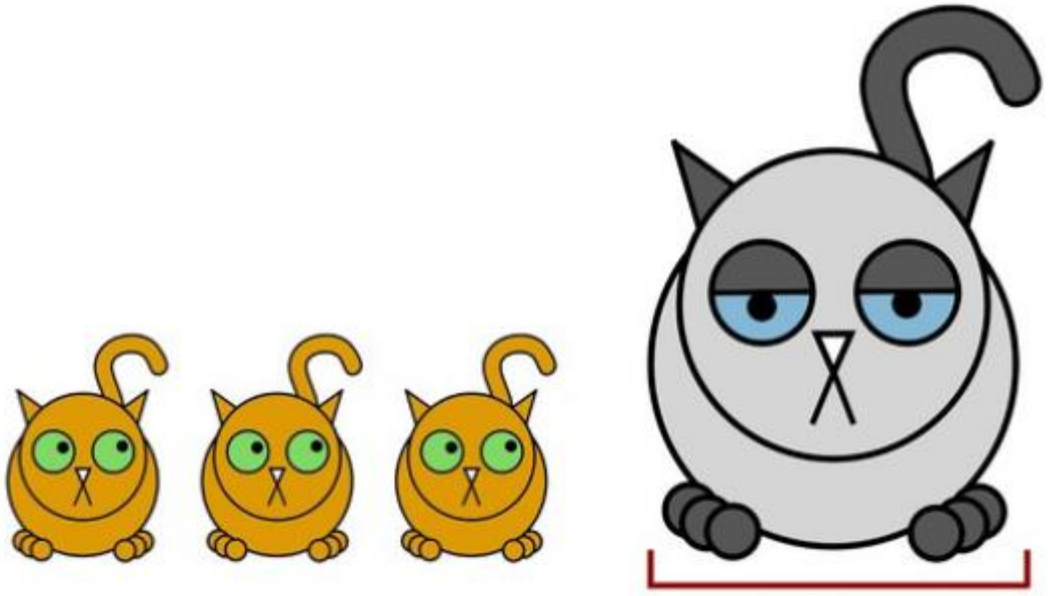
Статистические данные

Генеральная совокупность –
все возможные значения признака(признаков)

Выборка – множество значений исследуемого признака (признаков), полученное в результате наблюдения

Возможные размеры котиков:
От 20 до 80 см

Размер котиков из выборки:
50 см, 40 см, 20 см, 60 см



Статистические данные

Пусть X – некоторый признак (показатель, переменная, фактор), характеризующий свойство объекта или явления.

Результаты n наблюдений над признаком X
могут быть представлены различными способами.

В некоторых случаях анализа явлений и процессов важен **порядок получения статистических данных**.

В таком случае значения $x_1, x_2, \ldots, x_n$ записываются в порядке их регистрации и образуют простую выборку объема n.

Пример:
Динамика температуры в апреле, градусы: 0, 10, -5, 15, 2, 3, 7…

Статистические данные

Во многих случаях порядок получения статистических данных не играет роли. Кроме того, простая выборка труднообозрима и неудобна.

Поэтому применяют различные формы упорядочения данных в выборке: по возрастанию, по совпадающим значениям, по интервалам и т.д.

Часто удобно пользоваться не самой выборкой,
а **ее модификацией**.

Вариационный ряд

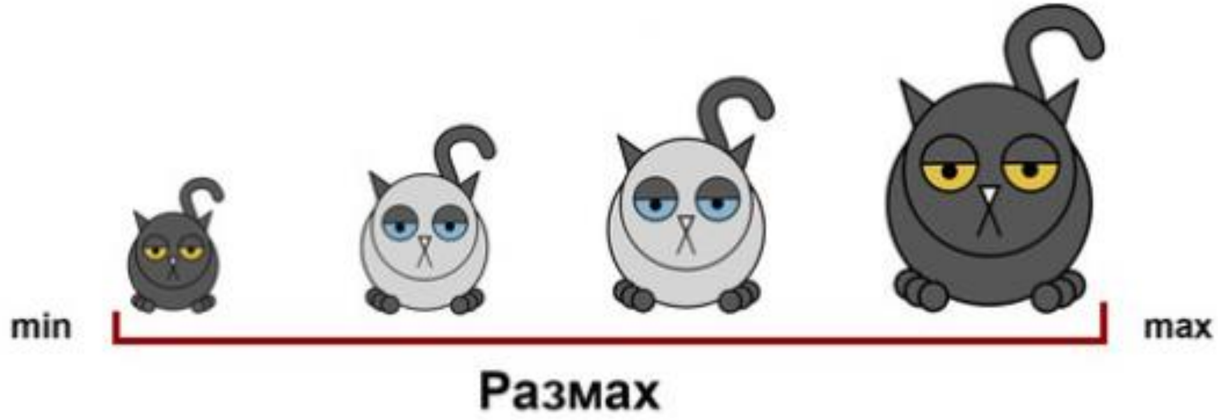
Если выборочные наблюдения $x_1, x_2, \dots, x_n$ упорядочить по возрастанию элементов, то получится **вариационный ряд**:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

Разность между крайними членами вариационного ряда

$x_{(1)} = x_{min} = min\{x_1, \dots, x_n\}$ и $x_{(n)} = x_{max} = max\{x_1, \dots, x_n\}$ называется **размахом выборки** и обозначается R:

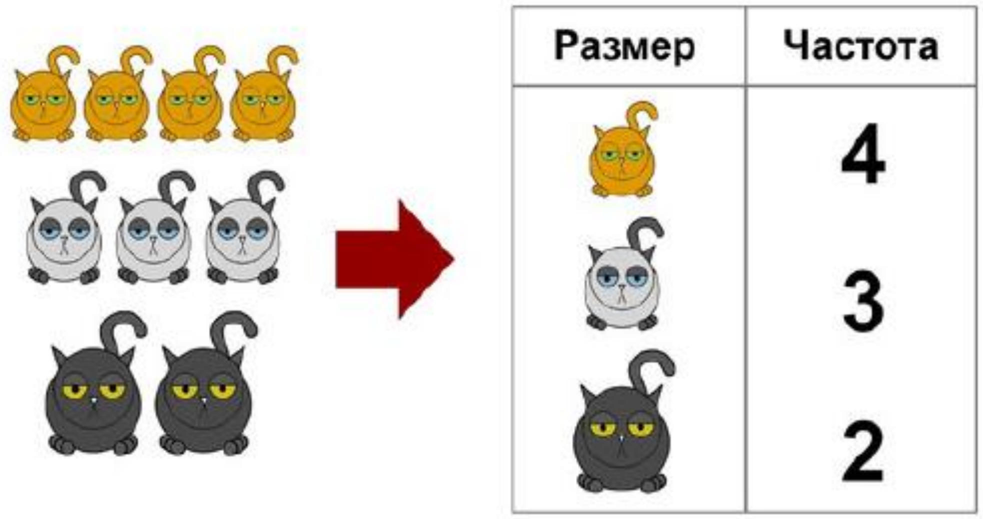
$$R = x_{(n)} - x_{(1)}$$



Группирование статистических данных

При наблюдении дискретных признаков в выборке могут часто встречаться одинаковые значения.

- Дискретные данные удобно группировать в **дискретный статистический ряд**.



- При большом объеме выборки ее элементы могут быть сгруппированы в **интервальный статистический ряд**.

Дискретный статистический ряд

Пусть количество различных данных в выборке равно k $(k \leq n)$ и значения упорядочены $x_1 \leq x_2 \leq \cdots \leq x_n$

Если значение x_i встретилось в выборке ν_i раз, то ν_i – ***частота значения*** x_i.

Тогда выборку можно сгруппировать и представить в виде дискретного статистического ряда:

Значение признака	x_1	x_2	…	x_k	$\sum\limits_{\boldsymbol{i=1}}^{\boldsymbol{k}} \nu_i$
Частота	ν_1	ν_2	…	ν_k	n

Интервальный статистический ряд

Все n значений выборки разбивают на k непересекающихся интервалов $[c_i, c_{i+1}), i = 1, \dots, k$, обычно одинаковой длины h

Число интервалов может быть выбрано по формуле Стерджеса

$$k = [\log_2 n] + 1$$

ν_i – ***число выборочных данных***, попавших в i-тый интервал группировки значения x_i.

Интервал группировки	$[c_1, c_2)$	$[c_2, c_3)$	…	$[c_k, c_{k+1})$	$\sum_{i=1}^{k} \nu_i$
Частота	ν_1	ν_2	…	ν_k	n

Дискретный статистический ряд

Пример: Исследователь, интересующийся тарифным разрядом рабочих механического цеха, в результате опроса 100 рабочих получил следующие сведения: 5, 1, 4, 5, 4, 3, 5, 5, 2, 5, 5, 6, 4, 3, 1, 5, 2, 5, 5, 5, 3, 3, 3, 6, 5, 6, 5, 3, 4, 5, 4, 6, 6, 5, 2, 1, 5, 5, 4, 5, 5, 3, 9, 4, 5, 5, 4, 3, 5, 5, 4, 5, 6, 1, 5, 2, 6, 4, 4, 3, 5, 6, 3, 5, 6, 2, 5, 4, 5, 5, 4, 6, 2, 5, 3, 4, 5, 6, 5, 5, 3, 5, 4, 6, 6, 5, 5, 4, 5, 5, 6, 5, 6, 6, 5, 5.

Исследуемый признак *X* – тарифный разряд, выборочная совокупность объема *n* = 100. Признак *дискретный*, поэтому представим данные в виде **дискретного статистического ряда**:

Тарифный разряд	**1**	**2**	**3**	**4**	**5**	**6**
Кол-во рабочих	4	6	12	16	44	18

Интервальный статистический ряд

Пример: Исследователь, изучающий выработку на одного рабочего-станочника механического цеха в отчетном году (в процентах к предыдущему) получил данные по 117 рабочим, представленные в таблице:

Выработка в процентах	**80÷90**	**90÷100**	**100÷110**	**110÷120**	**120÷130**	**130÷140**	**140÷150**
Кол-во рабочих	8	15	46	29	13	3	3

Исследуемый признак *X* – выработка в отчетном году (в процентах к предыдущему году). Признак *непрерывный*, поэтому данные представлены **интервальным статистическим рядом**.

Эмпирическая функция распределения

Наиболее полную характеристику признака дает функция распределения $F(x) = P\{ X < x \}$.

Если признак X распределен по некоторому закону F(x), то будем говорить, что и генеральная совокупность распределена по закону F(x).

Введем *выборочный (или статистический) аналог* функции F(x).

Эмпирическая функция распределения

Эмпирической (или выборочной) функцией распределения называется функция

$$F^*(x) = \frac{\nu_n(x)}{n},$$

где $\nu_n(x)$ – число элементов выборки $x_1, x_2, \dots, x_n$, значения которых меньше x.

В отличие от эмпирической (т.е. построенной по эмпирическим или выборочным данным) функции $F^*(x)$ функцию $F(x) = P(X < x)$ называют **теоретической (или генеральной) функцией распределения**.

Эмпирическая функция распределения

Построение эмпирической функции распределения удобно производить с помощью группированных данных, представленных статистическим рядом (дискретным или интервальным).

$$F^*(x) = \begin{cases} 0 & \text{при } x \leq x_1 \\ \sum_{j=1}^{i} p_j^*, & \text{при } x_i < x \leq x_{i+1}, i = 1, \dots, k-1 \\ 1 & \text{при } x > x_{k+1} \end{cases}$$

Функция $F^*(x)$ постоянна на промежутках $[x_i, x_{i+1})$, а в концевых точках x_i, x_{i+1} увеличивается на $p^* = \nu_n/n\,, i = 1, \dots k-1$

Эмпирическая функция распределения

$$F^*(x) = \begin{cases} 0 & \text{при } x \le x_1 \\ \sum_{j=1}^{i} p_j^*, & \text{при } x_i < x \le x_{i+1}, i = 1, \dots, k-1 \\ 1 & \text{при } x > x_{k+1} \end{cases}$$

Эмпирическая функция распределения $F^*(x)$ определена однозначно для всех x $(-\infty < x < +\infty)$ и обладает всеми свойствами функции распределения: изменяется от 0 до 1, не убывает и непрерывна слева.

При увеличении числа наблюдений *n* над признаком *X* происходит **сближение** $F^*(x)$ с теоретической функцией распределения $F(x)$.

Эмпирическая функция распределения

Пример: По данным из таблицы построить эмпирическую функцию распределения и сравнить ее с функцией распределения нормального закона.

Выработка в процентах	**80÷90**	**90÷100**	**100÷110**	**110÷120**	**120÷130**	**130÷140**	**140÷150**
Кол-во рабочих	8	15	46	29	13	3	3

Решение: Количество выборочных данных $n = 117$, размах выборки $R = 50$, число интервалов группировки $k = 7$, длина интервалов $h = 10$.

Эмпирическая функция распределения

№ п/п	**Граница интервала**		**Частота, ν_i**	**Относительная частота, p_i^***	**Функция $F^*(x)$**
	Левая	Правая			
1	80.0	90.0	8	0.07	0.07
2	90.0	100.0	15	0.13	0.20
3	100.0	110.0	46	0.39	0.59
4	110.0	120.0	29	0.25	0.84
5	120.0	130.0	13	0.11	0.95
6	130.0	140.0	3	0.03	0.97
7	140.0	150.0	3	0.03	1.00

Эмпирическая функция распределения

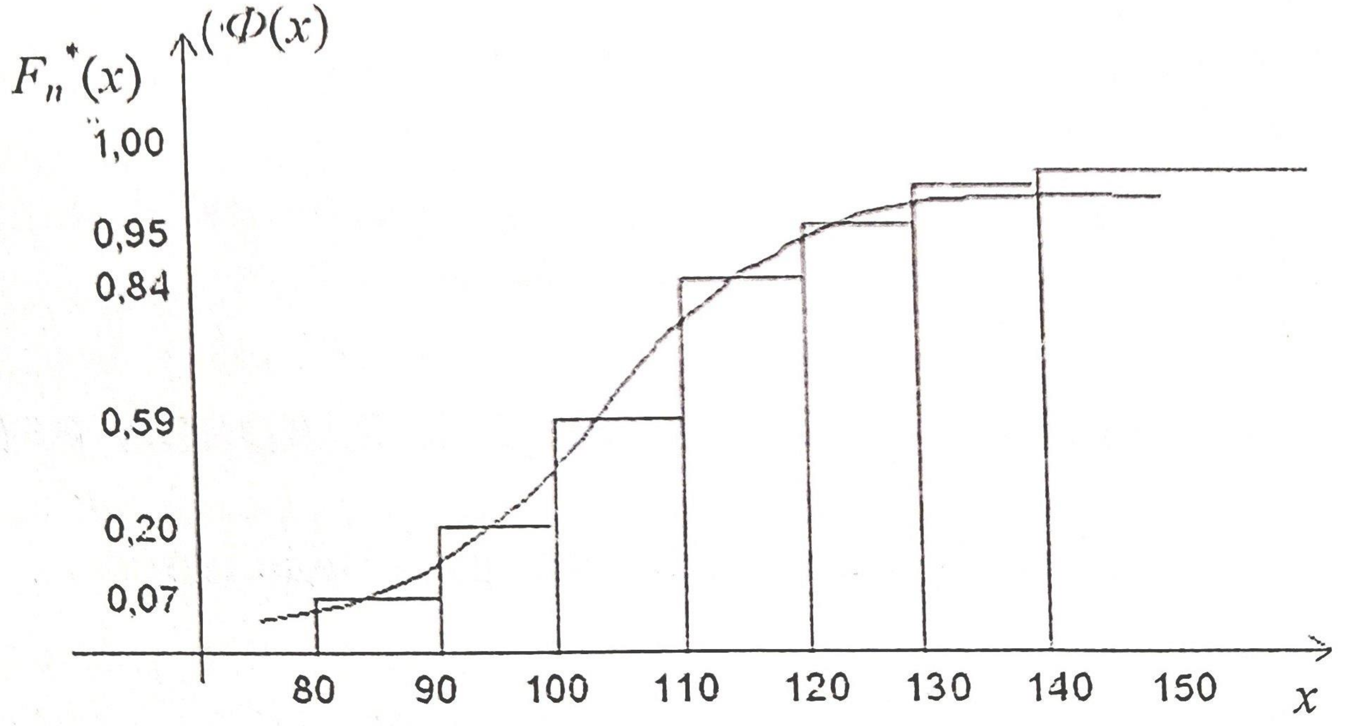


Рис. 5.1. Эмпирическая ф.р. и теоретическая ф.р. нормального закона

По рисунку можно сделать вывод, что предположение о близости распределения наблюдаемого признака (выработки станочников за год в процентах к предыдущему году) к нормальному закону вполне разумно.

Полигон частот

Существуют другие способы наглядного представления выборочных данных: **полигон и гистограмма**.

Полигон обычно используют для изображения *дискретного* статистического ряда.

На оси Ох откладываются все различные выборочные данные $x_1 < x_2 < \cdots < x_n$, а по оси Оу либо частоты $\nu_1, \nu_2, \dots \nu_k$, либо относительные частоты $p_1^* = \nu_1/n, p_2^* = \nu_2/n, \dots p_k^* = \nu_k/n$. Точки (x_i, ν_i) или (x_i, p_i^*) отмечаются и соединяются отрезками.

Полученная ломаная – **полигон частот** или полигон относительных частот.

Полигон частот

Полигон (относительных) частот – статистический аналог многоугольника распределения.

Позволяет визуально сравнивать и делать предположения о близости распределения признака к тому или иному закону распределения.

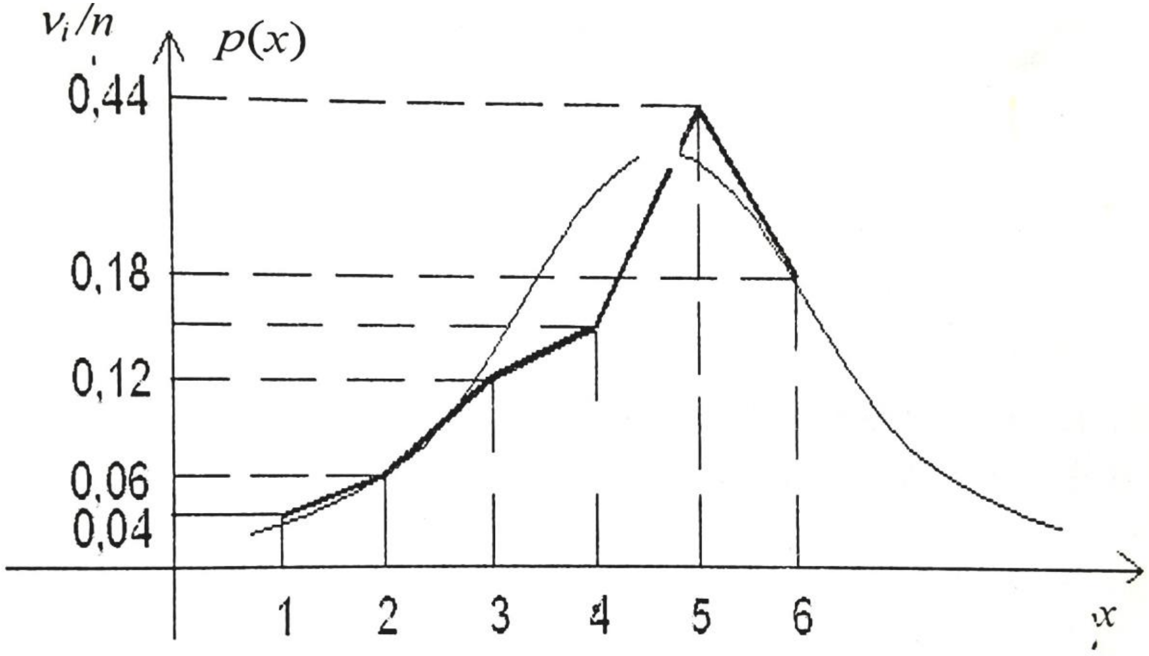


Рис. 5.2. Полигон и плотность распределения нормального закона

Гистограмма относительных частот

Гистограмма относительных частот – это графическое изображение *интервального* статистического ряда в виде прямоугольников разной высоты.

Основания прямоугольников – интервалы группировки, а высоты – относительные частоты этих интервалов

$$p^*(x) = \begin{cases} 0 & \text{при } x \le c_1 \\ \nu_i \,/\, (nh), & \text{при } c_i < x \le c_{i+1}, i = 1, \ldots, k \\ 1 & \text{при } x > c_{k+1} \end{cases}$$

При большом объеме выборки *n* и малых значениях длин интервалов группировки *h* гистограмма относительных частот близка к плотности распределения исследуемого признака *X*.

Гистограмма относительных частот

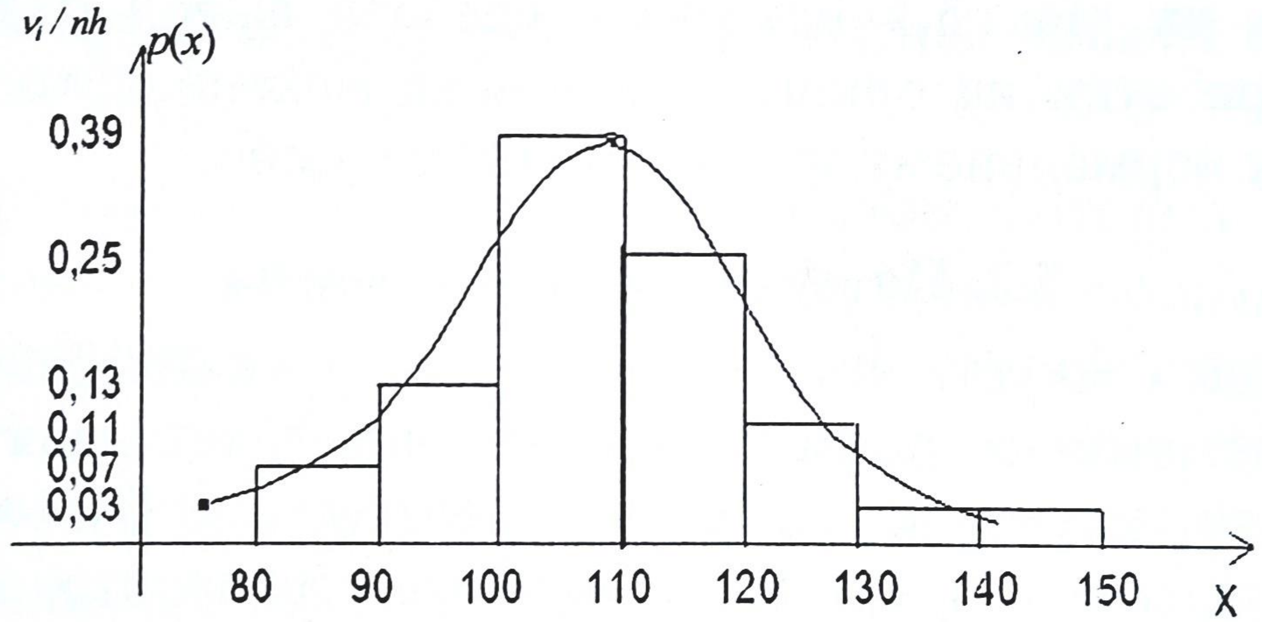


Рис. 5.3. Гистограмма и плотность распределения нормального закона

Сравнение графиков показывает, что предположение о близости распределения исследуемого признака *X* к нормальному вполне естественно

Точечные оценки

- Одна из основных целей статистической обработки данных – заменить исходное множество данных небольшим числом.
- Задача состоит в представлении множества обрабатываемых данных в виде сравнительно небольшого числа сводных **характеристик**, построенных на основании этих данных.
- При этом важно, характеристики сохраняло по возможности всю существенную информацию из исходных данных.

Такие сводные характеристики являются *функциями от исходных результатов* наблюдения и называются **статистиками**.

Точечные оценки

- Распределение исследуемого признака X характеризуется целым рядом числовых теоретических характеристик, или параметров: $\Theta_1, \Theta_2, \ldots, \Theta_n$.
- Сюда относятся среднее значение (мат. ожидание), дисперсия, среднее кв. отклонение, моменты и т.д.
- На практике распределение изучаемого признака, как правило, неизвестно или же известно частично.

Поэтому необходимо указать некоторую функцию от исходных данных (**статистику**), для использования *вместо неизвестного теоретического параметра* в качестве его приближения.

Это приближенное значение параметра Θ называется **точечной оценкой** и обозначается Θ^*.

Точечные оценки

При изучении случайных признаков часто используются числовые характеристики, позволяющие оценивать такие его свойства, как центр группирования значений исследуемого признака, меру их рассеивания относительно центра группирования, характеристики формы распределения.

Для оценки теоретических характеристик распределения применяются различные **выборочные характеристики или статистики**.

Опишем введение таких статистик для исходных данных в виде простой выборки $x_1, x_2, \dots, x_n$ объема n.

Точечные оценки

Признак: размер кота
Параметры признака: среднее, дисперсия

Генеральная совокупность

Выборка

Точечные оценки:

- Выборочное среднее
- Выборочная дисперсия

Являются **статистиками** – функциями от исходных данных

Точечные оценки

Выборочное среднее

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Выборочная дисперсия

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

Выборочное среднее квадратическое (стандартное) отклонение

$$S = +\sqrt{S^2}$$

Точечные оценки

Пусть исходные данные **сгруппированы** и представлены таблицей частот (дискретной или интервальной).

Пусть x_i - значение признака для дискретного случая (центр интервала группировки для интервального),

ν_i - *частота*, соответствующая i-му значению (интервалу),

$$i = 1, \dots k,$$

k – число различных выборочных *данных* или *интервалов* группировки.

Логично, что $\nu_1 + \nu_2 + \cdots + \nu_k = \sum_{i=1}^{k} \nu_i = n$ – объем выборки

Точечные оценки

Для сгруппированных данных формулы примут вид:

Выборочное среднее

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \nu_i$$

Выборочная дисперсия

$$S^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 \nu_i = \frac{1}{n}\sum_{i=1}^{k} x_i^2 \nu_i - (\bar{x})^2$$

Точечные оценки

Пример: Вычислить выборочное среднее значение, дисперсию и среднеквадратичное отклонение тарифного разряда рабочих по данным таблицы.

Тарифный разряд	**1**	**2**	**3**	**4**	**5**	**6**
Кол-во рабочих	4	6	12	16	44	18

Точечные оценки

Тарифный разряд	**1**	**2**	**3**	**4**	**5**	**6**
Кол-во рабочих	4	6	12	16	44	18

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i v_i = \frac{1*4+2*6+3*12+4*16+5*44+6*18}{100} \approx 4.44$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 v_i - (\bar{x})^2 = \frac{1^2*4+2^2*6+\cdots+6^2*18}{100} - (4.44)^2 \approx 1.70$$

$$S = \sqrt{1.70} \approx 1.30$$

Точечные оценки

Пример: Вычислить выборочное среднее значение, дисперсию и среднеквадратичное отклонение выработки одного станочника в отчетном году (в процентах к предыдущему году)

Выработка, %	**80÷90**	**90÷100**	**100÷110**	**110÷120**	**120÷130**	**130÷140**	**140÷150**
Кол-во раб.	8	15	46	29	13	3	3

Точечные оценки

Выработка, %	**80÷90**	**90÷100**	**100÷110**	**110÷120**	**120÷130**	**130÷140**	**140÷150**
Кол-во раб.	8	15	46	29	13	3	3

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i v_i = \frac{1}{117} * (85 * 8 + 95 * 15 + 105 * 46 +$$

$$+115 * 29 + 125 * 13 + 135 * 3 + 145 * 3) \approx 108.85$$

$$S^2 = \frac{1}{117}\sum_{i=1}^{k} x_i^2 v_i - (108.85)^2 \approx 160.08 \qquad S = \sqrt{160.08} \approx 12.65$$