Описательная статистика

Выборочный метод

<u>Задачи математической статистики</u>: 1) Сбор данных 2) Анализ 3) Интерпретация 4) Прогноз

ТВ — случайная величина → X ← генеральная совокупность — МС

$x_1, x_2, \ldots, x_n$ выборка из генеральной совокупности

<u>Идея выборочного метода</u>: по выборке получить представление о генеральной совокупности

Репрезентативность выборки

Способы отбора: 1) случайный повторный
2) случайный бесповторный

Описательная статистика

Способы представления выборочных данных:

1) первоначальные статистические данные
2) вариационный ряд
3) статистический ряд
4) сгруппированный статистический ряд

$x_1, x_2, \dots, x_n$ - числовые значения в порядке отбора

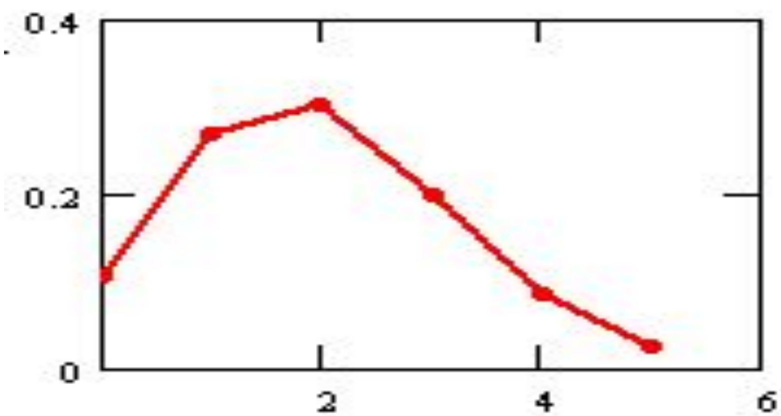
$x_{(1)}, x_{(2)}, \dots, x_{(n)}$ - упорядочены по возрастанию

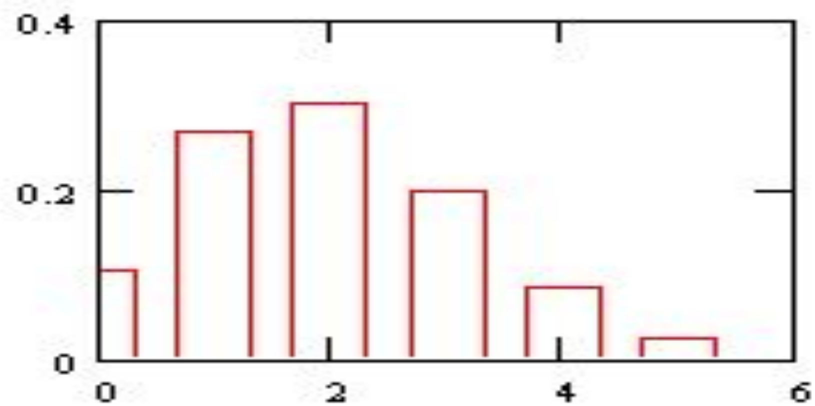
$\tilde{x}_j$	различные варианты
n_j	количество наблюдений каждой

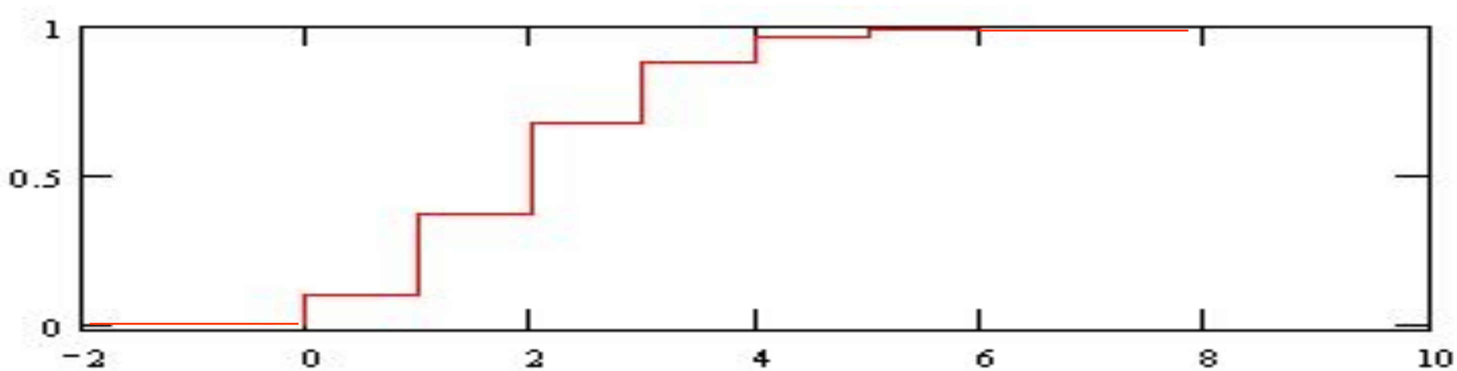
$[\tilde{x}_j; \tilde{x}_j)$	интервалы значений вариант
n_j	кол-во наблюдений в каждом

Оценка закона распределения по выборочным данным

1) полигон - оценка многоугольника распределения
2) гистограмма – оценка плотности распределения
3) эмпирическая функция распределения







Распределения статистик

Статистические копии

X - генеральная совокупность

θ - параметр генеральной совокупности

$\tilde{\theta}\ ;\ \hat{\theta}$ - оценка параметра по выборке

$x_1, x_2, \ldots, x_n$ - числовые значения

$X_1, X_2, \ldots, X_n$ - случайные величины

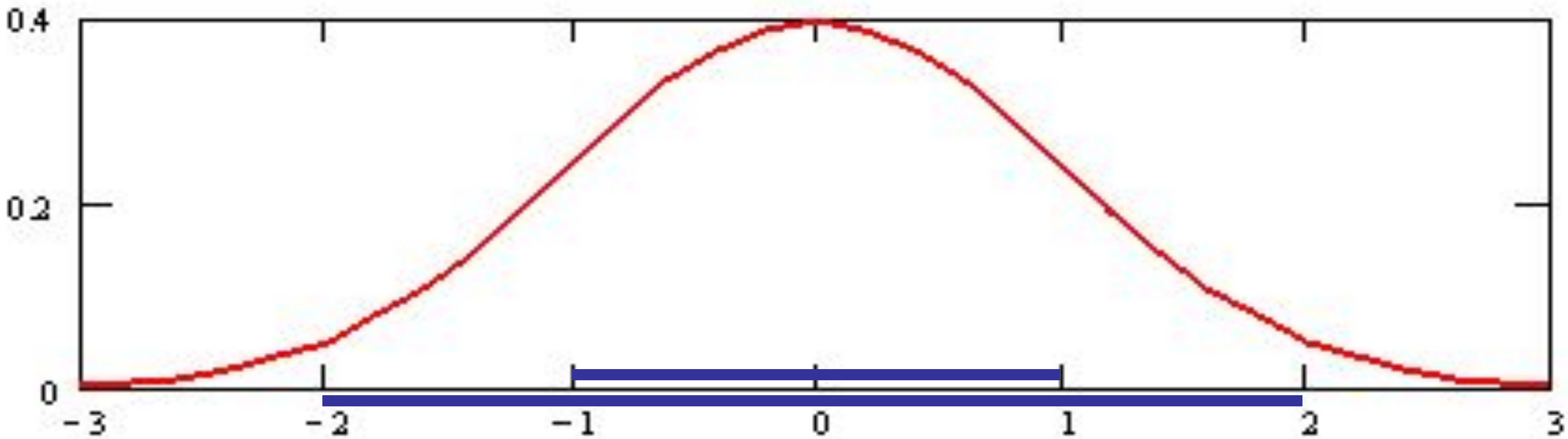
Статистические копии: 1) независимые 2) одинаково распределенные

$K(X_1, X_2, \ldots, X_n)$ - статистика (выборочная функция)

Распределения статистик

Стандартное нормальное распределение

$$U \sim N(0,1) \qquad M(U) = 0 \qquad D(U) = 1$$



68% 95% 99,7%

$$P\left(\left|X - m_X\right| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)$$

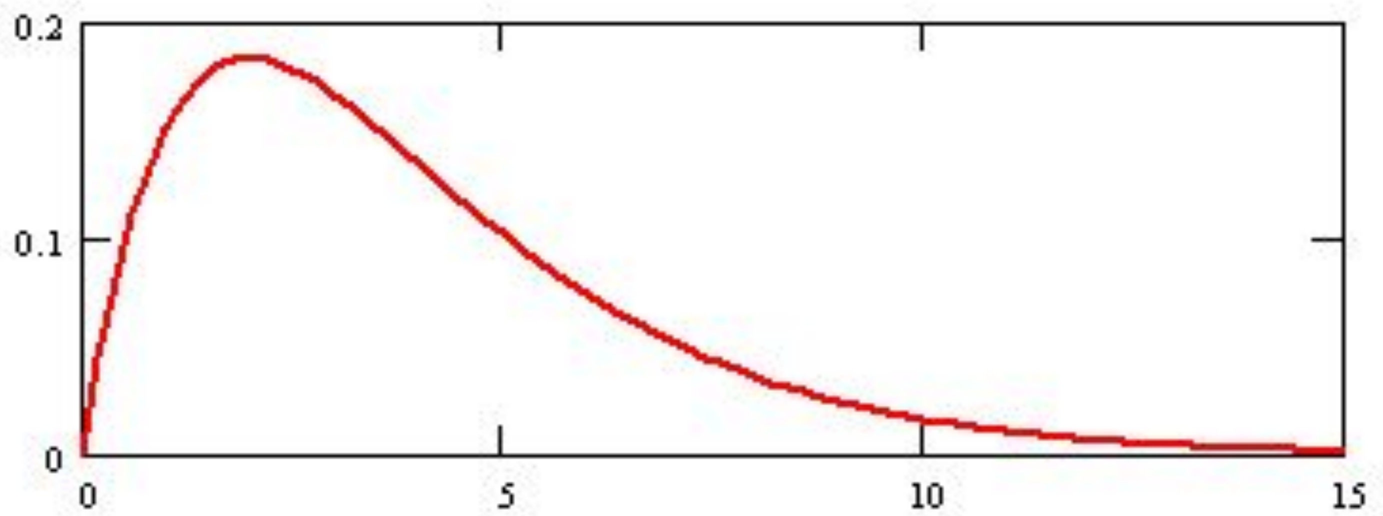
$\Phi(x)$ - функция Лапласа

Распределение «хи-квадрат»

$U_i \sim N(0,1)$
и независимы $\Rightarrow$

$$\boxed{K = \sum_{i=1}^{n} U_i^2 \sim \chi^2_{(n)}}$$

$$M(K) = n \qquad D(K) = 2n$$



$n = 4$

$M_0=2$

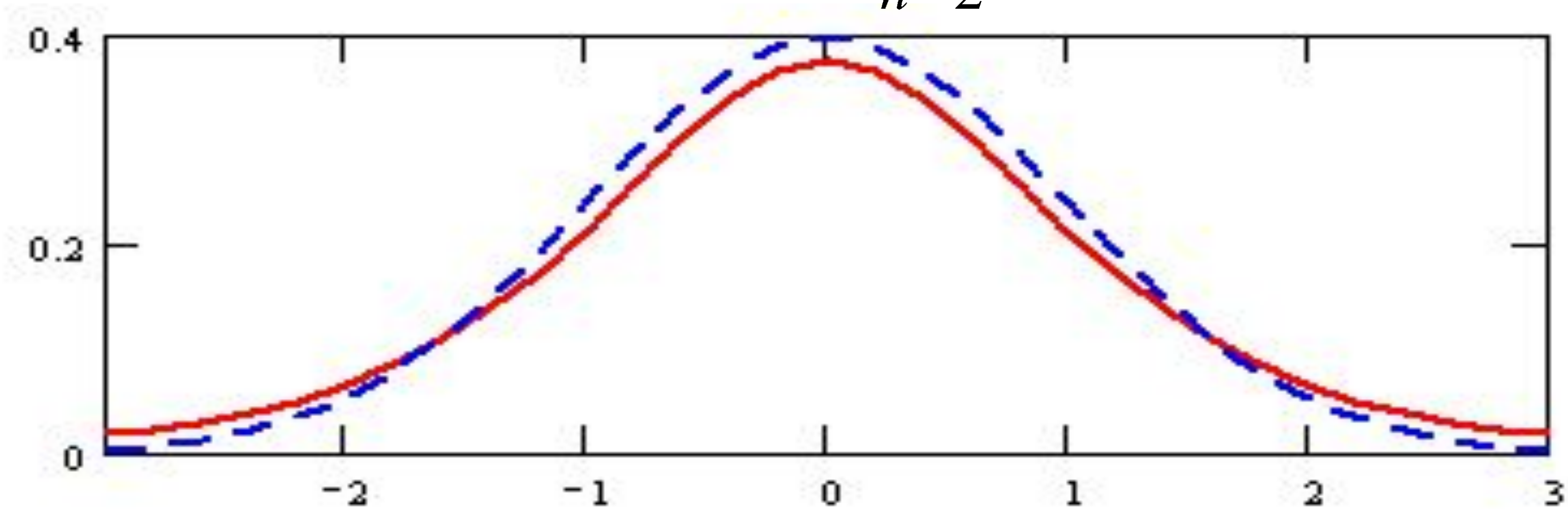
Распределения статистик

Распределение Стьюдента

$$U \sim N(0,1),\ K \sim \chi^2_{(n)}$$
и независимы $\Rightarrow$

$$\boxed{\frac{U}{\sqrt{K/n}} \sim T(n)}$$

$$M(T)=0 \quad D(T)=1+\frac{2}{n-2}\ (n>2)$$



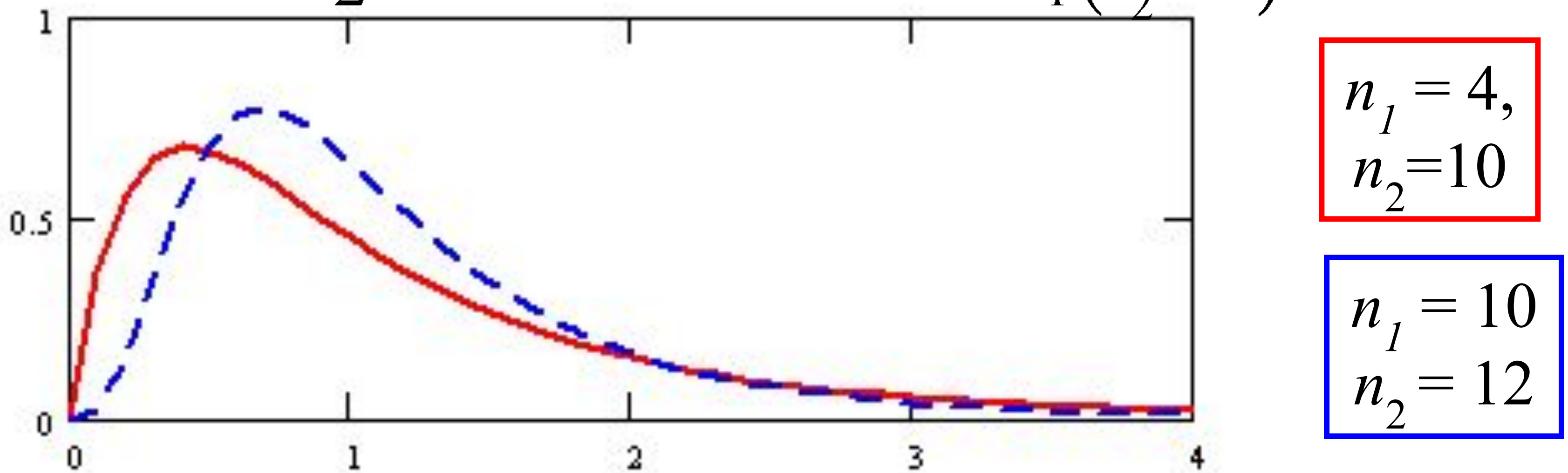
$n = 4$ D=2

Распределения статистик

Распределение Фишера

$$K_1 \sim \chi^2_{(n_1)},\ K_2 \sim \chi^2_{(n_2)}$$
и независимы $\Rightarrow$ $\boxed{\dfrac{K_1/n_1}{K_2/n_2} \sim F_{(n_1, n_2)}}$

$$M(F) = \frac{n_2}{n_2 - 2} \qquad D(F) = \frac{2n_2^2\left(n_1 + n_2 - 2\right)}{n_1\left(n_2 - 2\right)} \ (n_2 > 4)$$



Оценки параметров

Точечные оценки и их свойства

Оценка $\tilde{\theta}$ называется **несмещенной** оценкой параметра θ ,если

$$\boxed{M\left(\tilde{\theta}\right)=\theta}$$

Примеры

1) $\theta = m$ - параметр; его оценка: $\tilde{\theta} = \overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$

$$M\left(\tilde{\theta}\right) = M\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\left(\sum_{k=1}^{n} M\left(X_k\right)\right) =$$

$$= \frac{1}{n}\cdot n \cdot m = m \Rightarrow$$

среднее арифметическое - несмещенная оценка математического ожидания

Точечные оценки и их свойства

Примеры $\quad 2)\ \theta = D$

$$\tilde{D} = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$$

$$\tilde{D} = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - m + m - \overline{X}\right)^2 =$$

смещенная оценка

$$= \frac{1}{n}\sum_{k=1}^{n}\left(X_k - m\right)^2 - \left(\overline{X} - m\right)^2$$

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$$

$$M\left(\tilde{D}\right) = \frac{1}{n}\sum_{k=1}^{n} M\left(X_k - m\right)^2 - M\left(\overline{X} - m\right)^2 =$$

несмещенная оценка

$$= D - D\left(\overline{X}\right) = D - \frac{D}{n}$$

Точечные оценки и их свойства

Теорема 1

Эмпирические начальные моменты являются несмещенными оценками теоретических начальных моментов (если последние существуют)

$$m_r(X) = M\left(X^r\right) \qquad \widetilde{m}_r(X) = \frac{1}{n}\sum_{k=1}^{n} X_k^r$$

$$M(\widetilde{m}_r) = \frac{1}{n}\sum_{k=1}^{n} M\left(X_k^r\right) = \frac{1}{n}\sum_{k=1}^{n} m_r(X) = m_r$$

Точечные оценки и их свойства

Оценка $\widetilde{\theta}$ называется **состоятельной** оценкой параметра θ ,если

$$\boxed{\widetilde{\theta}\left(X_1,\ldots,X_n\right)\xrightarrow[n\to\infty]{P}\theta}$$

<u>Примеры</u> состоятельных оценок

$1)\ \theta = m$

(по теореме Чебышева)

$$\widetilde{\theta} = \overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

$2)\ \theta = p$

(по теореме Бернулли)

$$\widetilde{p} = p^* = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Точечные оценки и их свойства

Несмещенная оценка $\tilde{\theta}$ параметра θ является состоятельной, $\Leftrightarrow D\left(\tilde{\theta}\right) \xrightarrow[n\to\infty]{} 0$ (*)

<u>Теорема 2</u>

Эмпирические начальные моменты являются состоятельными оценками теоретических начальных моментов, если последние существуют

По теореме 1 эти оценки несмещенные

$$D\left(\tilde{m}_r\right) = \frac{1}{n^2}\sum_{k=1}^{n} D\left(X_k^r\right) = \frac{1}{n^2}\sum_{k=1}^{n}\left(M\left(X_k^{2r}\right) - \left(M\left(X_k^r\right)\right)^2\right) =$$

$$= \frac{1}{n}\left(m_{2r}(X) - \left(m_r(X)\right)^2\right) \xrightarrow[n\to\infty]{} 0$$

По свойству (*) оценки состоятельны

Точечные оценки и их свойства

Несмещенная оценка $\tilde{\theta}_1$ параметра θ называется **более эффективной** оценкой , чем оценка $\tilde{\theta}_2$, если ее дисперсия меньше дисперсии $\tilde{\theta}_2$

<u>Пример</u> $X \sim N(a, \sigma_0)$ $\theta = a$

$\tilde{\theta}_1 = \overline{X}$ несмещенные оценки $\tilde{\theta}_2 = \tilde{x}_{0.5}$

$$D(\tilde{\theta}_1) = \frac{\sigma_0^2}{n} \qquad D(\tilde{\theta}_2) \approx 1.6 \cdot \frac{\sigma_0^2}{n}$$

$\tilde{\theta}_1$ более эффективна, чем $\tilde{\theta}_2$

$$\frac{D(\tilde{\theta}_2)}{D(\tilde{\theta}_1)} \approx 1.6$$

для достижения той же точности требуется в 1.6 раз больше опытов

Точечные оценки и их свойства

Несмещенная оценка $\widetilde{\theta}$ называется **эффективной** оценкой параметра θ , если ее дисперсия достигает наименьшего для данного распределения значения

Свойства оценок:

несмещенность; состоятельность;
эффективность .

Методы получения оценок:

метод моментов;
метод максимального правдоподобия.

Методы получения точечных оценок

Пусть $f(x,\theta)$ - п.р. генеральной совокупности
Найти оценку параметра $\tilde{\theta}$ по выборке $x_1, x_2, \ldots, x_n$

Метод моментов

$$\boxed{m_k(\theta) = \tilde{m}_k}$$

Оценки: несмещенность (±);
состоятельность (+);
эффективность (-)

Пример $X \sim R(0,b)$

$$m_1(X) = \frac{b}{2} \qquad \tilde{m}_1 = \bar{x} \qquad \tilde{b} = 2\bar{x}$$

Методы получения точечных оценок

Метод максимального правдоподобия

$$L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)$$

$x_1, x_2, \ldots, x_n$

- функция правдоподобия

$$\tilde{\theta} = \operatorname{argmax} L(\theta)$$

Оценки: несмещенность (±);
состоятельность (+);
эффективность (±)

Пример

$$X \sim N(a,\sigma)$$

$$\theta_1 = a$$

$$\theta_2 = \sigma^2$$

$$L(\theta)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} exp\left(\frac{(x_i-a)^2}{2\sigma^2}\right)$$

$$\begin{cases} \tilde{a}=\bar{x} \\ \tilde{\sigma}^2=\frac{1}{n}\sum(x_i-\bar{x})^2 \end{cases}$$