



Эмпирическая функция распределения. Гистограмма и полигон частот.

Def. Эмпирической функцией распределения называется

$$F_n^*(y) = \frac{\nu(y)}{n}$$

где $\nu(y)$ — число наблюдений X_i таких, что $X_i < y$

Гистограмма

$$\forall i \in [1, n] \quad X_i \in [a, b]$$

Разбиваем на k интервалов: $a = t_0 < \dots < t_k = b$

На равные интервалы, шириной h ($t_j - t_{j-1} = h$)

$$h = \frac{b-a}{k}, \quad \nu_j - \text{число наблюдений попавших в } j\text{-ый интервал}$$

miro

$$\nu_1 + \dots + \nu_k = n$$

$$\text{Высота ступеньки } L_j = \frac{\nu_j}{nh}$$

$$\text{Тогда сумма площадей всех ступенек: } S_1 + \dots + S_k = \sum_j L_j h = \frac{\nu_1 + \dots + \nu_k}{n} = 1$$

$$\text{Ф-ла Стерджесса для кол-ва интервалов } k = \lceil \log_2 n \rceil + 1$$

miro

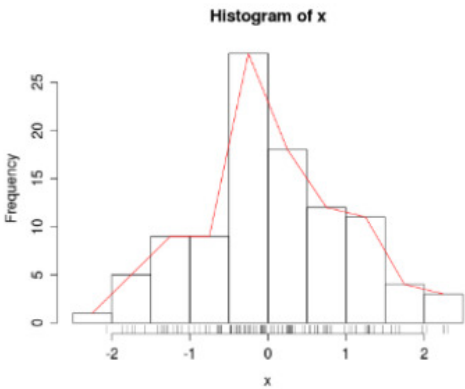
Полигон частот

Соединим отрезками середины верха ступенек. Дадная непрерывная кривая наз. полигоном частот.

Пример:

Пример 3. Имеется вариационный ряд из примера 2:
(0; 1; 1; 2; 2,6; 2,6; 2,6; 3,1; 4,6; 4,6; 6; 6; 7; 9; 9).
Разобьём отрезок $[0, 10]$ на четыре равных отрезка. Отрезку $[0, 2,5)$ принадлежат четыре элемента выборки, отрезку $[2,5, 5)$ — шесть, отрезку $[5, 7,5)$ — три, и отрезку $[7,5, 10]$ — два элемента выборки. Строим гистограмму (рис. 2). На рис. 3 — гистограмма для той же выборки, но при разбиении области на пять равных отрезков.

$n = 15, h = \frac{10}{4} = 2,5$
 $L_1 = \frac{4}{15 \cdot 2,5} = \frac{8}{75}$



Гистограмма и полигон частот

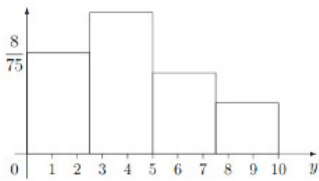


Рис. 2. Гистограмма при $k = 4$

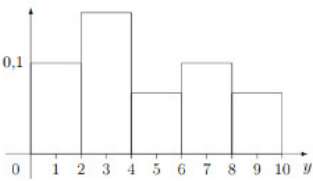


Рис. 3. Гистограмма при $k = 5$

miro