

24. Эмпирическая функция распределения. Гистограмма и полигон частот.

Status	Completed
--------	-----------

Эмпирическая функция распределения

Эмпирической функцией распределения называется

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

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

Гистограмма

$$X_i \in [a, b] \forall i$$

Разбиваем на k интервалов:

$$a = t_0 < \dots < t_k = b$$

На равные интервалы, шириной h

$$t_j = t_{j-1} + h; h = \frac{b-a}{k}$$

ν_j - число наблюдений, попавших в j -ый интервал

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

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



Гистограмма

Тогда сумма площадей всех ступенек

$$S_1 + \dots + S_k = \sum L_j h = \frac{\nu_1 + \dots + \nu_k}{n} = 1$$

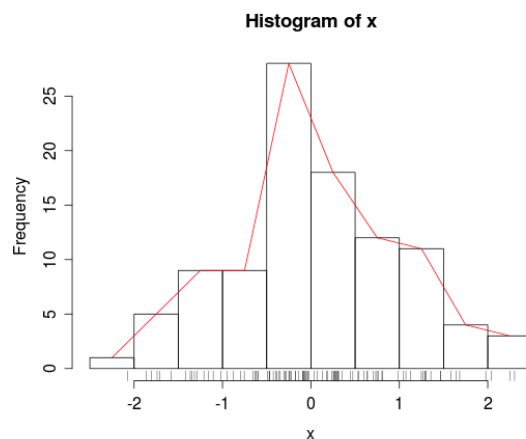
Гистограмма аппроксимирует график функции плотности распределения:

Выбирать число интервалов k можно по разному, можно пользоваться **формулой Стерджесса**:

$$k = \lceil \log_2 n \rceil + 1$$

Полигон частот

Соединим отрезками середины горизонтальных линий, ограничивающих ступени. Данная непрерывная кривая называется **полигоном частот**.



Гистограмма и полигон частот