24. Эмпирическая функция распределения. Гистограмма и полигон частот.



Эмпирическая функция распределения

Эмпирической функцией распределения называется

$$F_n^*(y) = rac{
u(y)}{n},$$

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

Гистограмма

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

Разбиваем на k интервалов:

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

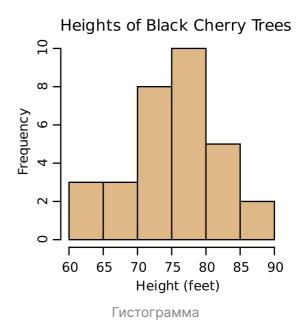
На равные интервалы, шириной h

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

 u_j - число наблюдений, попавших в j-ый интервал

$$\nu_1 + ... + \nu_k = n$$

Высота ступеньки $L_j=rac{\mu_j}{nh}$



Тогда сумма площадей всех ступенек

$$S_1 + ... + S_k = \sum_j L_j h = \frac{\nu_1 + ... + \nu_k}{n} = 1$$

Гистограмма аппроксимирует график функции плотности распределения:

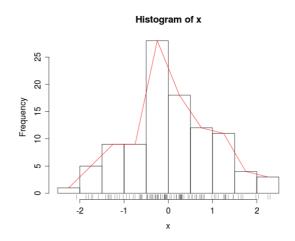
D

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

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

Полигон частот

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



Гистограмма и полигон частот