Случайные события

Событие можно назвать достоверным ($p(A) = 1$), если в результате испытания оно обязательно произойдет. Невозможное событие не произойдет никогда ($p(A) = 0$).

Операции над случайными событиями:

- A+B - наступило либо A либо B ($p(A \vee B)$)
- A*B - наступило и A и B
- $\overline{A}$ - событие A не наступило

Для случайного события есть понятие относительной частоты — это отношение количества состоявшихся событий к общему числу испытаний:

$$\mathbf{W(A) = m/n}$$

- W(A) — это относительная частота события A;
- m — число состоявшихся событий A;
- n — общее число испытаний.

При достаточно большом количестве испытаний n величина относительной частоты W будет стремиться к конкретному числу. Оно называется статистической вероятностью и обозначается как P(A) :

$$\mathbf{P(A) = m/n}$$
$$p(A) \in [0;1]$$

Совместные события - могут произойти вместе, **несовместные** - не могут.
Совместные и несовместные события

Вероятности несовместных событий можно складывать:

$$\mathbf{P(A+B) = P(A) + P(B)}$$

В случае совместных событий формула суммы событий другая: из суммы вероятностей отдельных событий вычитается вероятность их совместного появления.

$$\mathbf{P(A+B) = P(A) + P(B) - P(AB)}$$

Независимыми события называют, когда появление одного из них не влияет на появление другого. С зависимыми - наоборот.

Вероятность одновременного появления двух независимых событий вычисляется по формуле:

$$\mathbf{P(A*B)=P(A)*P(B)}$$

Вероятность появления двух зависимых событий:

$$P(A*B)=P(A)*P(B|A)=P(B)*P(A|B)$$

Если событие A может наступить только при появлении событий B1,B2,…,Bn, образующих полную группу несовместных событий, то вероятность A вычисляется по формуле:

$$P(A)=P(B1)\cdot P(A|B1)+P(B2)\cdot P(A|B2)+...+P(Bn)\cdot P(A|Bn)$$

Формула Байеса: чтобы определить вероятность события B при условии, что событие A уже произошло, используют формулу Байеса:

$$P(B|A)=P(B)\cdot P(A|B) / P(A)$$

Комбинаторика

Число сочетаний из n элементов по k элементов в каждом (в сочетаниях порядок не важен):

$$C(kn)=n! / (k!(n-k)!)$$

Определим **число размещений** из n элементов по k элементов в каждом. При размещениях порядок важен, поэтому вариантов размещения может быть больше, чем сочетаний при заданных k и n.

$$A(kn)=n! / (n-k)!$$

Число перестановок из n элементов — при перестановках важен порядок, но отличие от размещений в том, что применяются все имеющиеся n элементов:

$$P(n)=n!$$

Случайная величина

Случайная величина — та, что в результате испытания принимает только одно возможное значение. Дискретные и непрерывные.

Закон распределения вероятностей — соответствие между возможными значениями и их вероятностями.

Биномиальное распределение

$$P_n(X = k) = C_n^k p^k q^{n-k}$$

где **p** — это вероятность наступления события **A** в **n** независимых испытаниях, а **q = 1 – p.**

Матожидание:

$$M(X) = np$$

Дисперсия:

$$D(X) = npq$$

Распределение Пуассона

Вероятность того, что событие произойдет **m** раз в **n** испытаниях:

$$P_m \approx \frac{\lambda^m}{m!} e^{-\lambda}$$

где $\lambda = np.$

Описательная статистика

Генеральная совокупность — это множество, которое содержит данные обо всех объектах, соответствующих определённым характеристикам. **Выборка** — это случайным образом выбранная часть генеральной совокупности.

Мат ожидание

Это - среднее значение случайной величины (распределение вероятностей стационарной случайной величины), при стремлении количества выборок/измерений к бесконечности.

Оценка мат ожидания

$$M(X) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Квадратичное отклонение

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}$$

Дисперсия

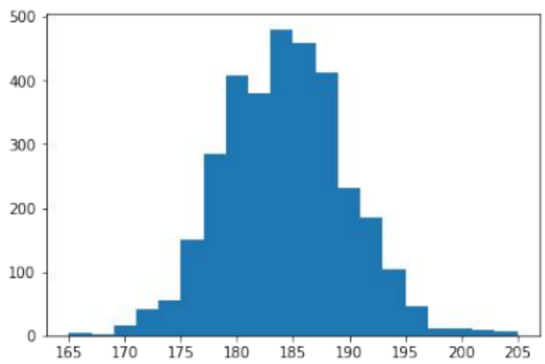
$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$$

Несмещенная оценка дисперсии

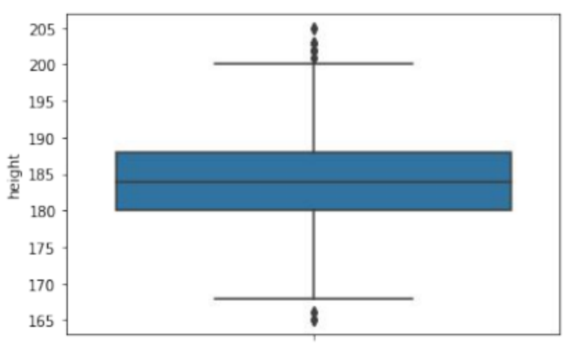
$$\sigma^2_{\text{несмещ.}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n - 1}$$

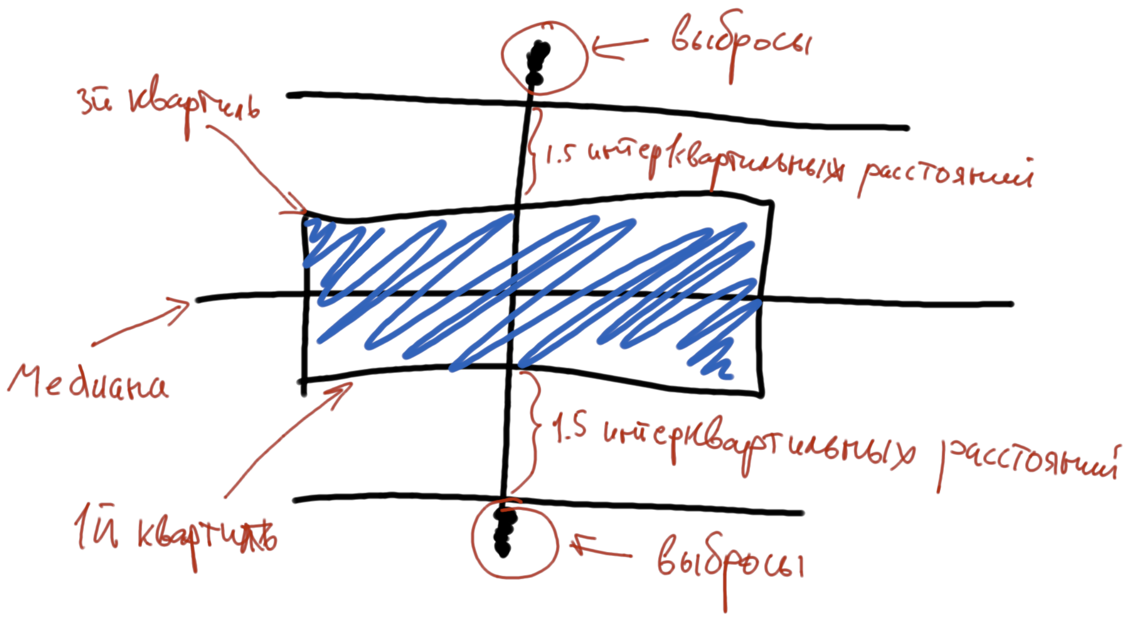
Интерквартильное расстояние — отрезок, равный разности третьей и первой квартили.

Гистограмма



Boxplot





Непрерывная случайная величина

Непрерывная случайная величина - принимает все возможные значения, содержащиеся на промежутке, который может быть как конечным (ограниченным), так и бесконечным.

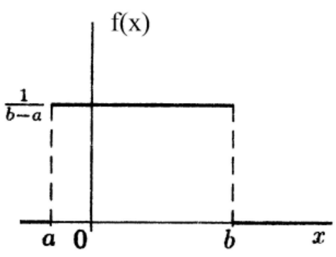
Функция распределения вероятностей - для каждого значения x показывает, какова вероятность того, что случайная величина меньше X. **Плотность распределения вероятностей** - равна производной функции распределения вероятностей:

$$f(x) = F'(x)$$

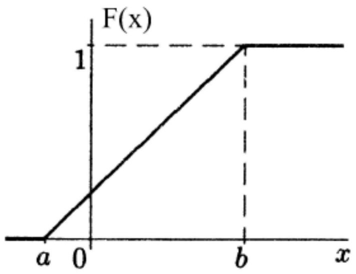
Равномерное распределение

$$f(x) = \begin{cases} 0, \text{если } x \le a; \\ \frac{1}{b-a}, \text{если } a < x \le b; \\ 0, \ \text{если } x > b. \end{cases}$$

Плотность равномерного распределения:

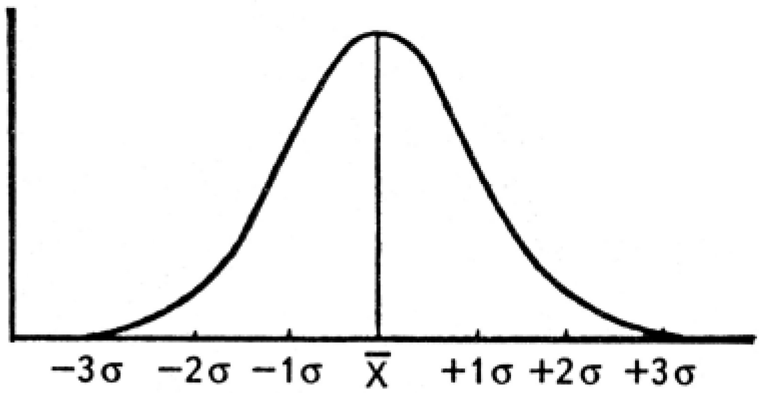


Функция равномерного распределения:



Нормальное распределение

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad \text{где } a = M(X),\ \sigma^2 = D(X).$$



На отрезке от −σ до +σ расположено около 68% наблюдений, от −2σ до +2σ — 95.4%, и от −3σ до +3σ — 99.72%. Одним из свойств нормального распределения считается то, что **значения среднего, медианы и моды совпадают**.

Центральная предельная теорема

Сумма достаточно большого числа слабо зависимых случайных величин, у которых примерно одинаковые масштабы, имеет распределение, близкое к нормальному.

Если у нас есть несколько выборок из генеральной совокупности, то среднее по этим выборкам также будет иметь нормальное распределение. Среднее достаточно большого числа независимых и нормально распределенных случайных величин также считается приблизительно нормально распределенным.