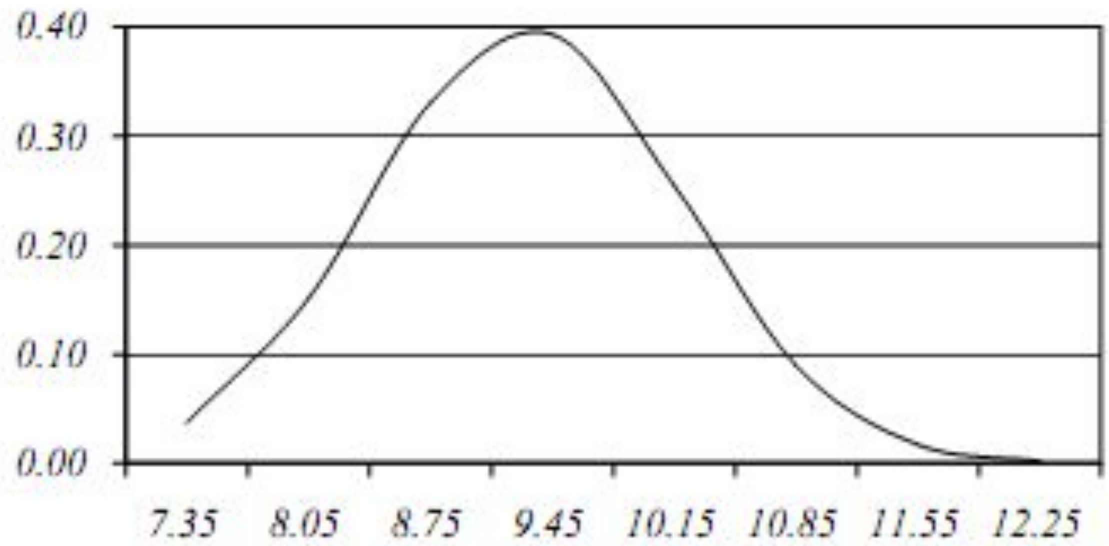
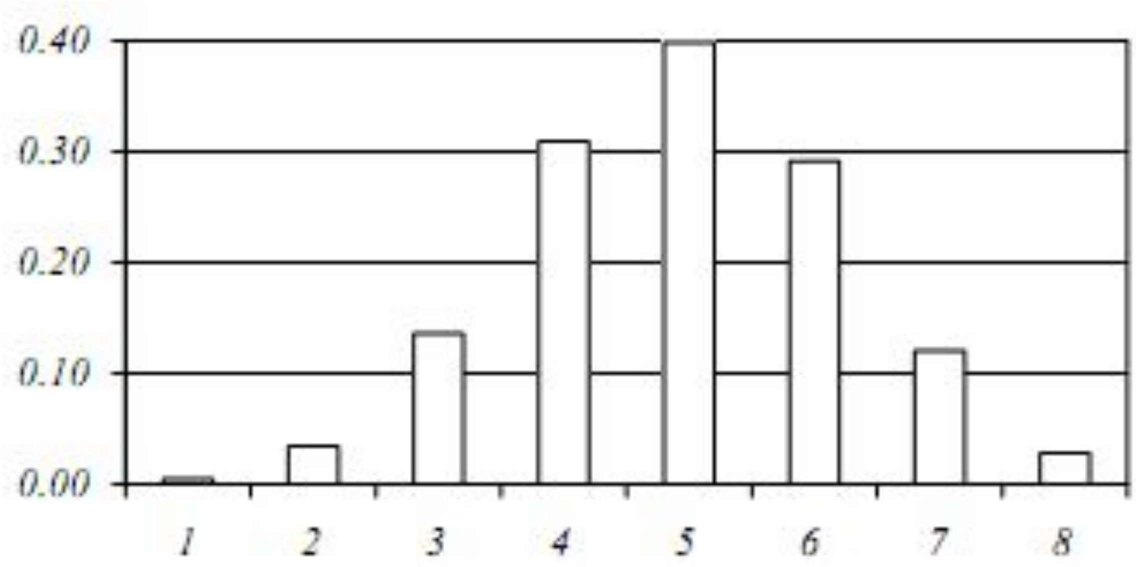
Типы распределения признаков

Нормальное распределение



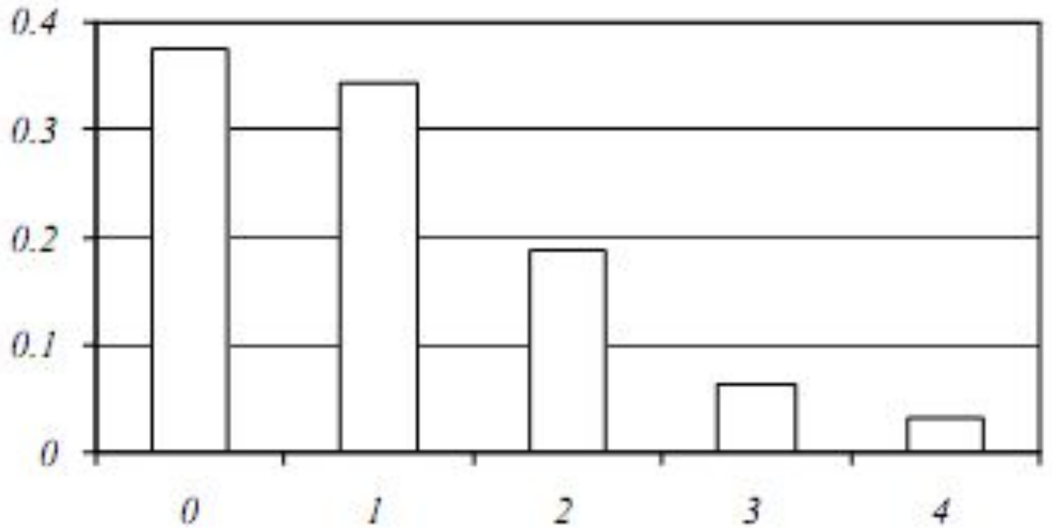
Распределение симметрично, причем крайние значения (наибольшие и наименьшие) появляются редко, но чем ближе значения признака к центру (к средней арифметической), тем оно чаще встречается

Биномиальное распределение



Характеризует поведение дискретных признаков, выраженных целыми числами. Как правило, для описания биологических признаков подходит симметричное биномиальное распределение, у которого дисперсия много меньше средней. Распределение организуется в процессе обора проб (объемом больше одного, m > 1). Число классов больше двух, k > 2.

Распределение Пуассона



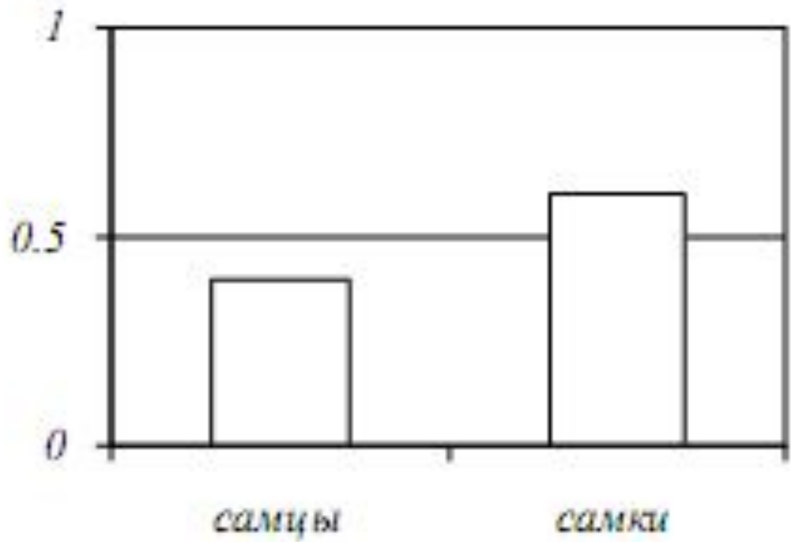
Это вариант описания стохастического поведения дискретных количественных признаков для случаев, когда вероятность элементарных альтернативных событий неодинакова, одно из них наблюдается заметно чаще другого ($p << q$). Закон Пуассона описывает редкие события, происходящие 1, 2, 3 и т. д. раз на сотни и тысячи обычных событий. Поведение биологических объектов, соответствующее закону Пуассона, наблюдается в том случае, когда по пробам случайно распределены редкие объекты.

Распределение Пуассона

Число повторных отловов, x	Число отловленных животных, a	Число случаев повторного отлова, $x \cdot a$
0	15	0
1	7	7
2	7	14
3	2	6
4	1	4
n	32	31

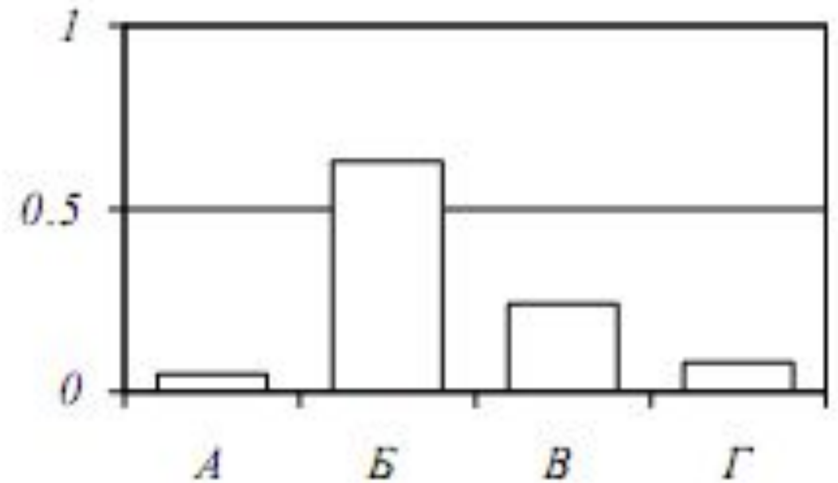
В последующие пять лет часть из них отлавливали повторно: 7 экз. по одному разу, 7 – по два, 2 – по три, 1 экз. – четыре раза, 15 экз. окольцованных птиц повторно не попадались. Число классов составляет k = 4, интервал dx = 1. Асимметрия в частотах встречаемости птиц позволяет предполагать распределение Пуассона.

Альтернативное распределение



Распределение дискретной случайной величины, имеющей лишь два противоположных (разнокачественных) значения (два класса, $k = 2$). В одной пробе (в одном наблюдении) содержится одна варианта ($m = 1$), одно из двух возможных значений. Вероятности каждого из них могут быть равны ($p = q$) либо не равны ($p < q$; $p > q$).

Полиномиальное распределение



Наблюдается для качественных признаков, имеющих не два альтернативных свойства, но несколько возможных проявлений качества. Примеры полиморфизма популяций – из этой области. В их числе варианты окраски покровов и волос, типы рисунков в определенных областях тела, способы жилкования листьев растений или крыльев насекомых, варианты расположения и формы щитков рептилий и др.

Оценка генеральных параметров

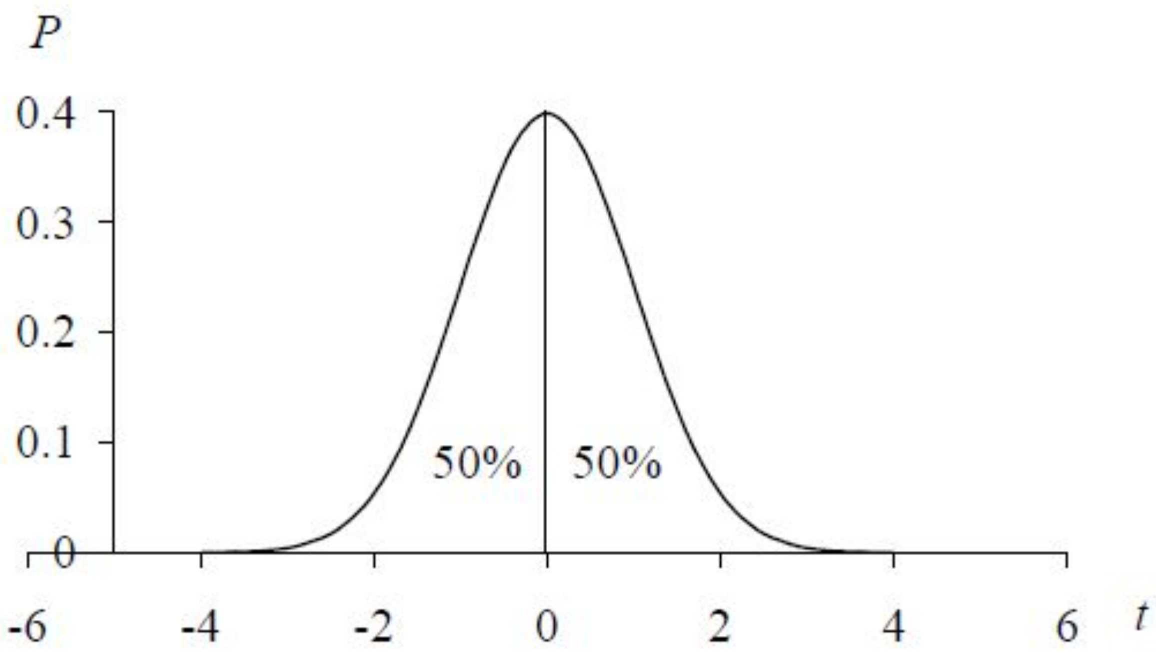
Распределение – это соотношение между значениями случайной величины и частотой их встречаемости.

$$p = \frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2},$$

где $t = \frac{(x-M)^2}{S}$ – нормированное отклонение;

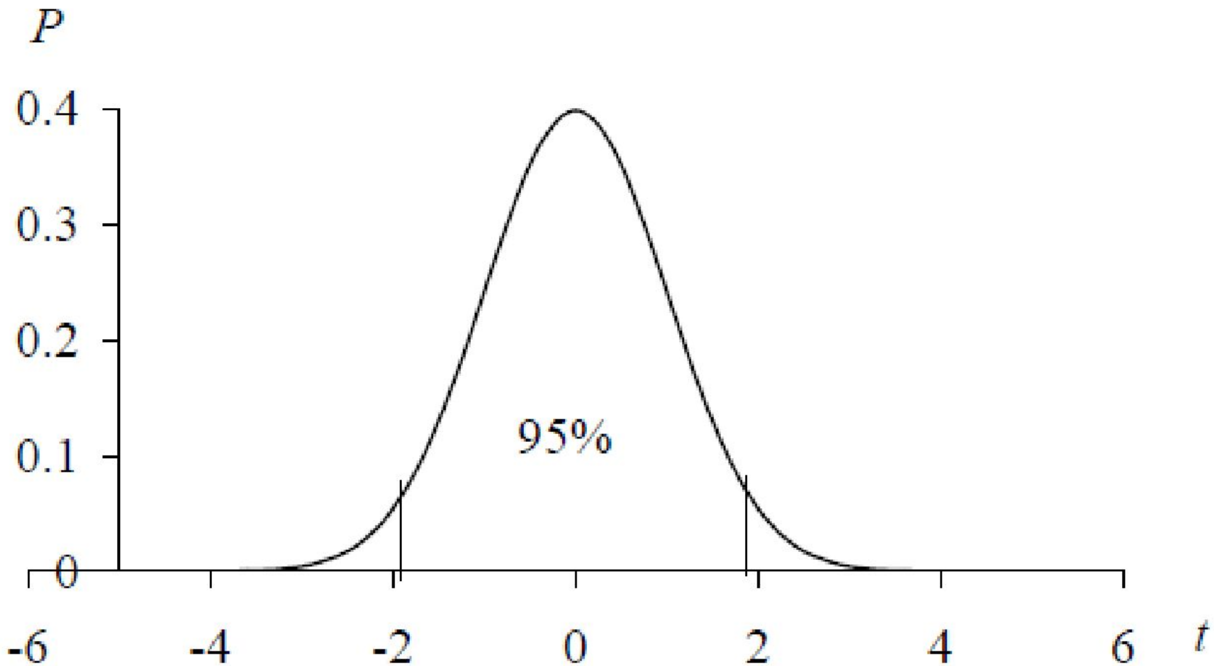
Вероятность (статистическая, или частость) – численная мера возможного, определяется как отношение числа вариант (исходов испытаний) определенного вида к общему числу вариант (опытов).

Следствия



Все варианты лежат в интервале плюс-минус бесконечность. Слева и справа от средней арифметической лежит по 50% вариант (свойство симметрии нормального распределения), т. е. с вероятностью Р = 0.5 (50%) можно предсказать появление новой варианты в интервалах $M - \infty$ и $M + \infty$

Следствия



Между M – 1.96S и M + 1.96S лежит 95% вариант. Это позволяет с 95%-ой вероятностью предполагать, что новая варианта окажется в интервале M ± 1.96S (округленно M ± 2S – так называемое правило двух стандартных отклонений).

Следствия

С вероятностью Р = 0.99 значение новой варианты будет заключено в пределах М ± 2.58S и с вероятностью Р = 0.999 – в интервале М ± 3.3S.

Термин «доверительная вероятность Р = 0.95» означает, что, согласно принятому допущению, 95% вариант достаточно полно характеризуют изучаемое явление.

Уровень значимости – это тот теоретический процент значений нормального распределения, который можно отбросить, не учитывать, дабы с меньшими усилиями получить основную информацию об изучаемом явлении.

Для практического понимания достаточно знать, что уровень значимости – это вероятность ожидаемой ошибки наших выводов, вероятность того, что данный статистический вывод не верен.

Генеральная совокупность

Генеральная совокупность – все варианты одного типа.

Выборочная совокупность, **выборка** – это множество вариант одного типа, ограниченное способом отбора из генеральной совокупности.

Чем меньше объем выборок, тем менее точным будут выборочные оценки генеральных параметров, и, напротив, чем больше выборка, тем ближе выборочные средние и дисперсии лежат к генеральным значениям (**закон больших чисел**).

Ошибка репрезентативности

Отличия значений выборочных параметров от генеральных называются ошибкой репрезентативности данного параметра.

Величина ошибки тем больше, чем больше варьирование признака (S) и чем меньше выборка (n).

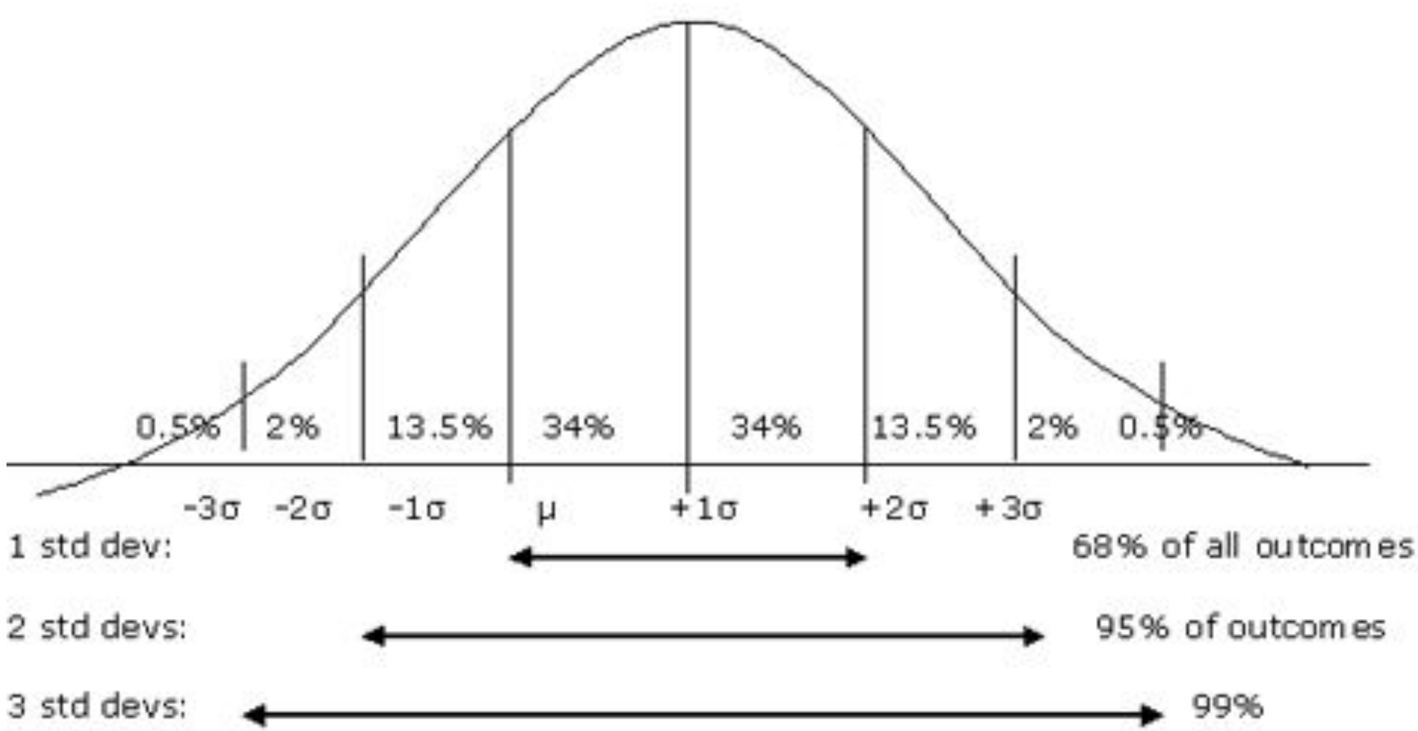
Ошибка средней: $m_M = \dfrac{S}{\sqrt{n}}$,

ошибка стандартного отклонения: $m_S = \dfrac{S}{\sqrt{2 \cdot n}}$,

ошибка коэффициента вариации: $m_{CV} = \dfrac{CV}{\sqrt{2 \cdot n}}$.

Вычисленные значения ошибок подставляют к соответствующим параметрам со знаками плюс-минус (параметр ± ошибка) и в такой форме представляют в научных отчетах и публикациях.

Доверительный интервал



Генеральная средняя находится в диапазоне Мвыбор. ± 1.96m, т. е. предсказывать ширину интервала, в котором заключен генеральный параметр, давать интервальную оценку генеральному параметру.

Доверительный интервал

В соответствии с законом нормального распределения можно ожидать, что генеральный параметр (истинное значение) окажется в интервале от M − tm до M + tm, где m – ошибка средней арифметической,
t – квантиль распределения Стьюдента при данном числе
степеней свободы (df) и уровне значимости (обычно α = 0.05).

Доверительный интервал – это интервал значений изучаемого признака, в котором с той или иной вероятностью находится значение генерального параметра.

Число степеней свободы, df	Доверительная вероятность (P) Уровень значимости (α)		
	$P = 0.095$ $\alpha = 0.05$	$P = 0.099$ $\alpha = 0.01$	$P = 0.0999$ $\alpha = 0.001$
2	4.303	9.925	31.598
3	3.182	5.841	12.941
4	2.776	4.604	8.610
5	2.571	4.032	6.859
6	2.447	3.707	5.959
7	2.365	3.499	5.405
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587
11	2.201	3.106	4.437
12	2.179	3.055	4.318
13	2.160	3.012	4.221
14	2.145	2.977	4.140
15	2.131	2.947	4.073
16	2.120	2.921	4.015
17	2.110	2.898	3.965
18	2.101	2.878	3.922
19	2.093	2.861	3.883
20	2.086	2.845	3.850
22	2.074	2.819	3.792
25	2.060	2.787	3.725
30	2.042	2.750	3.646
35	2.030	2.724	3.591
40	2.021	2.704	3.551
45	2.014	2.690	3.520
50	2.008	2.678	3.496
55	2.004	2.669	3.476
60	2.000	2.660	3.460
70	1.994	2.648	3.435
80	1.989	2.638	3.416
90	1.986	2.631	3.402
100	1.982	2.625	3.390
120	1.980	2.617	3.373
>120	1.960	2.5758	3.2905

Определение точности опыта

Это отношение ошибки средней к самой средней арифметической, выраженное в процентах

$$\varepsilon = \frac{m}{M} \cdot 100\%$$

Точность считается хорошей, если ε меньше 3%, и удовлетворительной при $3 < \varepsilon < 5\%$

$$n = \left(\frac{t \cdot CV}{\varepsilon} \right)^2$$

где n – объем выборки,
t – граничное значение из таблицы распределения Стьюдента, соответствующее принятому уровню значимости при планируемом объеме выборки,
CV – приблизительное значение коэффициента вариации (%),
ε – планируемая точность оценки (погрешности) (%).