Распределение дискретной случайной величины

В одной из предыдущих заметок указывалось, что исход испытания может представлять собой числовую переменную. В свою очередь, числовые переменные разделяются на дискретные и непрерывные. Дискретные переменные характерны для перечислений и подсчета, а непрерывные — для измерений. В этой и нескольких последующих заметках будут рассмотрены общие положения и наиболее распространенные распределения, описывающие дискретные случайные величины.[1]

Распределение дискретной случайной величины — это исчерпывающий список всех возможных значений случайной переменной, где каждому исходу поставлена в соответствие его вероятность. Например, на рис. 1 приведено распределение количества ипотечных займов, выданных в течение недели местным филиалом банка. Поскольку в таблице приведены все возможные исходы, сумма их вероятностей равна 1.

Количество ипотечных займов, выданных за неделю	Вероятность
0	0,10
1	0,10
2	0,20
3	0,30
4	0,15
5	0,10
6	0,05

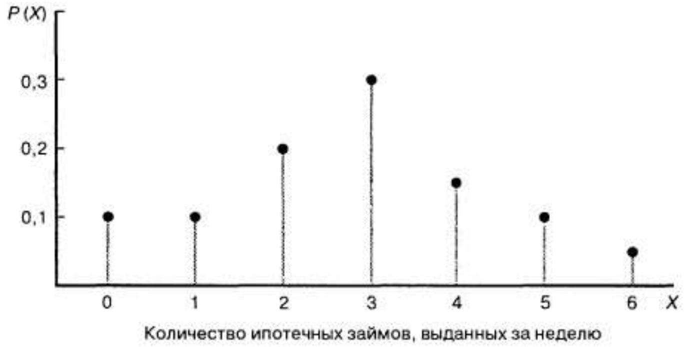

Рис. 1. Распределение количества ипотечных займов, выданных за неделю

Математическим ожиданием *μ* дискретной случайной величины *X* называется среднее значение ее распределения. Эта величина равна сумме произведений всех значений случайной величины *X* на соответствующие вероятности *P(X)*. Другими словами, математическое ожидание дискретной случайной величины *X* — это взвешенное среднее всех возможных исходов, где в качестве весов служат вероятности каждого исхода.

$$(1)\quad \mu = E(X) = \sum_{i=1}^{N} X_i P(X_i)$$

где X_i — i-е значение дискретной случайной величины X, $P(X_i)$ — вероятность i-го значения дискретной случайной величины X.

Математическое ожидание количества ипотечных займов, выданных за неделю:
р = 0х0,01+1х0,1+2х0,2+3х0,3+3х0,15+5х0,1+6х0,05 = 0 + 0,1 + 0,4 + 0,9 + 0,6 + 0,5 + 0,3 = 2,8

Обратите внимание: математическое ожидание количества ипотечных займов, выданных за неделю, выражается числом, которое не имеет буквального смысла, поскольку количество займов может измеряться только целыми числами.

Дисперсия σ^2 дискретной случайной величины *X* представляет собой взвешенное среднее квадратов разностей между всеми ее возможными значениями и математическим ожиданием. В качестве весов служат вероятности соответствующих исходов:

---

[1] Используются материалы книги Левин и др. Статистика для менеджеров. – М.: Вильямс, 2004. – с. 294–297

$$(2)\quad \sigma^2=\sum_{i=1}^{N}\left(X_i-\mathrm{E(X)}\right)^2\mathrm{P}(X_i)$$

где X_i — i-е значение дискретной случайной величины X, $P(X_i)$ — вероятность i-го значения дискретной случайной величины X.

Стандартное отклонение σ дискретной случайной величины:

$$(3)\quad \sigma=\sqrt{\sum_{i=1}^{N}\left(X_i-\mathrm{E(X)}\right)^2\mathrm{P}(X_i)}$$

В Excel для расчета описательных статистик дискретной случайно величины нет стандартных функций, поэтому, как правило, просто используют дополнительные столбцы для промежуточных вычислений по формулам (1), (2) и (3), см. рис. 2. Единственное исключение – математическое ожидание – его можно определить сразу (без промежуточных вычислений) с помощью функции =СУММПРОИЗВ().

а

	A	B	C	D
1	X	P(X)	[X - E(X)]^2	[X - E(X)]^2*P(X)
2	0	0,10	7,84	0,78
3	1	0,10	3,24	0,32
4	2	0,20	0,64	0,13
5	3	0,30	0,04	0,01
6	4	0,15	1,44	0,22
7	5	0,10	4,84	0,48
8	6	0,05	10,24	0,51
9	Итого	1,00		

б

	F	G	H	I
1	Статистика	Обозначение	Значение	Формула для расчета
2	математическое ожидание	E(X)	2,80	=СУММПРОИЗВ(A2:A8;B2:B8)
3	дисперсия	σ^2	2,46	=СУММ(D2:D8)
4	стандартное отклонение	σ	1,57	=КОРЕНЬ(H3)

Рис. 2. Последовательное вычисление описательных статистик дискретной случайно величины: (а) исходные данные и промежуточные вычисления; (б) финальные расчеты

Существует возможность обойтись и без промежуточных вычислений. Для этого следует воспользоваться формулами массива (рис. 3, см. также соответствующий лист приложенного Excel-файла). Если вы не применяли такие формулы ранее, рекомендую для начала прочитать Excel. Введение в формулы массива. Любопытно, что в Excel некоторые стандартные функции уже являются формулами массива, хотя их и немного. В частности, использованная выше =СУММПРОИЗВ().

	A	B	C	D	E	F	G
1	X	P(X)		Статистика	Обозначение	Значение	Формула для расчета
2	0	0,10		математическое ожидание	E(X)	2,80	{=СУММ(A2:A8*B2:B8)}
3	1	0,10		дисперсия	σ^2	2,46	{=СУММ((A2:A8-F2)^2*B2:B8)}
4	2	0,20		стандартное отклонение	σ	1,57	{=КОРЕНЬ(СУММ((A2:A8-F2)^2*B2:B8))}
5	3	0,30					
6	4	0,15					
7	5	0,10					
8	6	0,05					
9	Итого	1,00					

Рис. 3. Вычисление описательных статистик дискретной случайно величины с помощью формул массива

Заметим, что для чистоты эксперимента, можно вообще обойтись без ссылок на промежуточное значение E(X) (такие ссылки на ячейку F2 используются в формулах расчета σ^2 и σ, см. ячейки G3 и G4 на рис. 3). В этом случае, например, для расчета σ^2 получится чуть более громоздкая формула: {=СУММ((A2:A8-СУММ(A2:A8*B2:B8))^2*B2:B8)}.

Предыдущая заметка Условная вероятность. Теорема Байеса

Следующая заметка

К оглавлению Статистика для менеджеров с использованием Microsoft Excel