Гипергеометрическое распределение

Гипергеометрическое распределение, как и биномиальное, позволяет оценить количество успехов в серии из n испытаний. Разница между ними заключается в способе получения исходных данных. В биномиальной модели данные выбираются либо из конечной генеральной совокупности с возвращением либо из бесконечной генеральной совокупности без возвращения. В гипергеометрической модели данные извлекаются только из конечной генеральной совокупности без возвращения.[1] Таким образом, в то время как в биномиальной модели вероятность успеха p остается постоянной, а испытания не зависят друг от друга, в гипергеометрической модели эти условия не выполняются. Наоборот, в гипергеометрической модели каждый исход зависит от предыдущих исходов.

Гипергеометрическое распределение, описывающее вероятность X успехов при заданных параметрах n, N и A:

$$(1)\quad P(X) = \frac{\binom{A}{X}\binom{N-A}{n-X}}{\binom{N}{n}}$$

где $P(X)$ — вероятность X успехов при заданных n, N и A, n — объем выборки, N — объем генеральной совокупности, A — количество успешных исходов в генеральной совокупности, $N – A$ — количество неудачных исходов в генеральной совокупности, X — количество успехов в выборке, $N – X$ — количество неудачных исходов в выборке.

Количество успехов X в выборке не может превосходить количество успехов A в генеральной совокупности либо объем выборки n. Таким образом, диапазон значений, которые может принимать случайная величина, подчиняющаяся гипергеометрическому распределению, ограничен либо объемом выборки (как и диапазон биномиальной переменной), либо объемом генеральной совокупности.

Математическое ожидание гипергеометрического распределения:

$$(2)\quad \mu = E(X) = nA/N$$

Стандартное отклонение гипергеометрического распределения:

$$(3)\quad \sigma = \sqrt{\frac{nA(N-A)}{N^2}}\sqrt{\frac{N-n}{N-1}}$$

Выражение $\sqrt{\frac{N-n}{N-1}}$ называется *поправочным коэффициентом конечной генеральной совокупности*. Он необходим, поскольку элементы выборки извлекаются из конечной генеральной совокупности.

Например, предположим, что некая организация пытается создать группу из 8 человек, обладающих определенными знаниями о производственном процессе. В организации работают 30 сотрудников, обладающих необходимыми знаниями, причем 10 из них работают в конструкторском бюро. Какова вероятность того, что в группу попадут два сотрудника из конструкторского бюро, если членов группы выбирают случайно? Объем генеральной совокупности в этой задаче $N = 30$, объем выборки $n = 8$, а количество успехов $A = 10$.

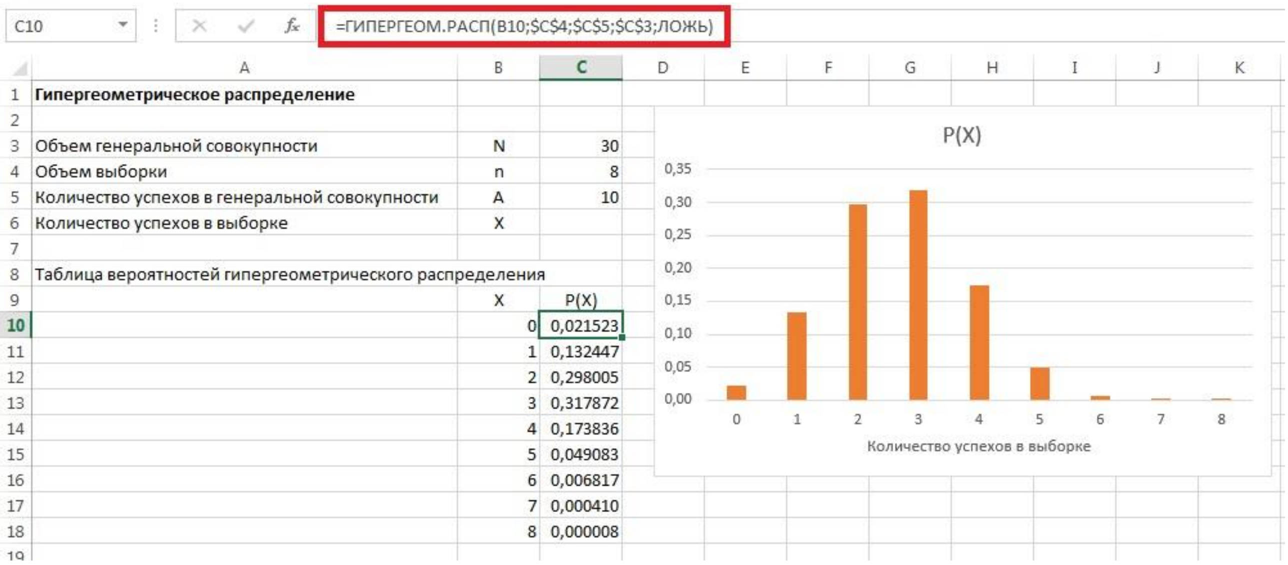
Используя формулу (1), получаем:

$$P(X=2) = \frac{\binom{10}{2}\binom{20}{6}}{\binom{30}{8}} = \frac{\frac{10!}{2!8!}\times\frac{20!}{6!14!}}{\frac{30!}{8!22!}} = 0{,}298$$

[1] Используются материалы книги Левин и др. Статистика для менеджеров. – М.: Вильямс, 2004. – с. 316–318

Таким образом, вероятность того, что в группу попадут два сотрудника из конструкторского бюро, равна 0,298 (или 29,8%).

При увеличении генеральной совокупности и объема выборки вычисления гипергеометрического распределения становятся все более утомительными. Однако гипергеометрическое распределение можно вычислить с помощью функции Excel =ГИПЕРГЕОМ.РАСП() (рис. 1).



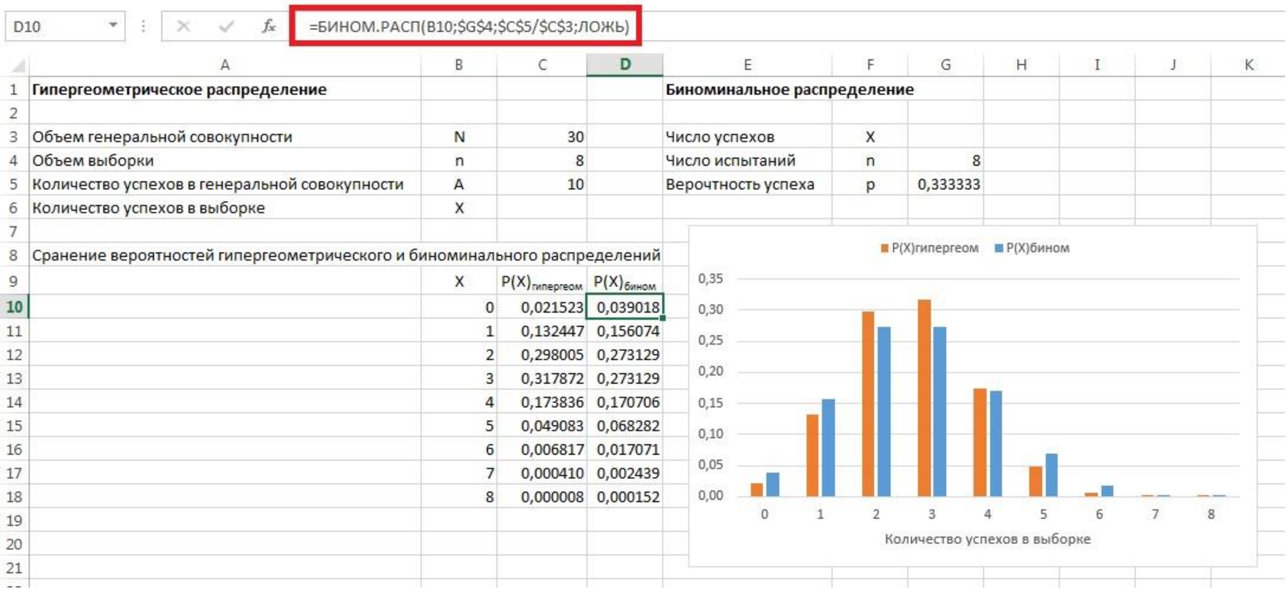
C10 =ГИПЕРГЕОМ.РАСП(B10;$C$4;$C$5;$C$3;ЛОЖЬ)

	A	B	C
1	**Гипергеометрическое распределение**		
2			
3	Объем генеральной совокупности	N	30
4	Объем выборки	n	8
5	Количество успехов в генеральной совокупности	A	10
6	Количество успехов в выборке	X	
7			
8	Таблица вероятностей гипергеометрического распределения		
9		X	P(X)
10		0	0,021523
11		1	0,132447
12		2	0,298005
13		3	0,317872
14		4	0,173836
15		5	0,049083
16		6	0,006817
17		7	0,000410
18		8	0,000008

Рис. 1. Вычисление в Excel гипергеометрического распределение при *N* = 30, *A* = 10 и *n* = 8

Таким образом, в рамках рассмотренного выше примера, наиболее вероятно, что в группе из 8 сотрудников трое будут из конструкторского бюро.

Видно (рис. 6), что гипергеометрическое и биноминальное распределения довольно похожи.



D10 =БИНОМ.РАСП(B10;$G$4;$C$5/$C$3;ЛОЖЬ)

	A	B	C	D	E	F	G
1	**Гипергеометрическое распределение**				**Биноминальное распределение**		
2							
3	Объем генеральной совокупности	N	30		Число успехов	X	
4	Объем выборки	n	8		Число испытаний	n	8
5	Количество успехов в генеральной совокупности	A	10		Верочтность успеха	p	0,333333
6	Количество успехов в выборке	X					
7							
8	Сранение вероятностей гипергеометрического и биноминального распределений						
9		X	$P(X)_{гипергеом}$	$P(X)_{бином}$			
10		0	0,021523	0,039018			
11		1	0,132447	0,156074			
12		2	0,298005	0,273129			
13		3	0,317872	0,273129			
14		4	0,173836	0,170706			
15		5	0,049083	0,068282			
16		6	0,006817	0,017071			
17		7	0,000410	0,002439			
18		8	0,000008	0,000152			

6. Сравнение гипергеометрического и биноминального распределений

Предыдущая заметка [Биноминальное распределение]

Следующая заметка

К оглавлению [Статистика для менеджеров с использованием Microsoft Excel]