**Критерий «хи-квадрат» для дисперсий**

При анализе числовых данных иногда бывает важно оценить не только их среднее значение, но и изменчивость. Для примера вернемся к задаче о расфасовке кукурузных хлопьев, описанной [ранее](http://baguzin.ru/wp/?p=5790). Будучи управляющим компании Oxford Cereal Company, вы отвечаете за процесс расфасовки кукурузных хлопьев по коробкам. Необходимо убедиться, что конвейер работает нормально, и каждая коробка содержит в среднем 368 г зерна. Предполагалось, что стандартное отклонение σ равно 15 г. Предположим теперь, что нам необходимо определить, отличается ли стандартное отклонение от заданного уровня, равного 15 г.[[1]](#footnote-1)

Пытаясь оценить изменчивость генеральной совокупности, мы должны сначала определить, какой статистический критерий можно использовать, чтобы представить распределение изменчивости выборочных данных. Если случайная переменная имеет нормальное распределение, для ответа на вопрос, равны ли дисперсия или стандартное отклонение заданной величине, применяется тестовая χ2-статистика:

*(1) χ2 =* $\frac{(n –1)S^{2}}{σ^{2}}$

где *n* — объем выборки, *S2* — выборочная дисперсия, *σ2* — гипотетическая дисперсия генеральной совокупности. Тестовая χ2-статистика имеет распределение *χ2* с *n – 1* степенями свободы.

Распределение χ2 имеет положительную асимметрию, форма которой зависит исключительно от количества степеней свободы (рис. 1).



Рис. 1. χ2-статистика при степенях свободы *n* = 5, 7, 10; графики построены с помощью функции Excel =ХИ2.РАСП(х;n;ЛОЖЬ), где *х* – координата оси абсцисс, *n* – количество степеней свободы

Математическое ожидание распределения χ2 равно количеству степеней свободы, а дисперсия — удвоенному количеству степеней свободы. Чтобы определить критические значения χ2-распределения можно воспользоваться функцией Excel =ХИ2.ОБР(0,025;24) для нижней границы χL2 , где 0,025 – половина уровня значимости α, 24 – число степеней свободы; =ХИ2.ОБР(0,975;24) для верхней границы χU2. Таким образом, для χ2- распределения с 24 степенями свободы, доля площади под всей кривой равная 0,025, отсекается χU2 = 39,36. В то же время доля площади, равная 0,975 отсекается χL2 = 12,40. (рис. 2).



Рис. 2. Определение нижнего и верхнего критических значений распределения χ2 с 24 степенями свободы, соответствующего уровню значимости, равному 0,05

Следовательно, для 24 степеней свободы вероятность того, что тестовая χ2-статистика будет равна критическому значению 12,401 или превысит его, равна 0,975. В то же время, вероятность того, что тестовая χ2-статистика будет равна критическому значению 39,36 или превысит его, равна 0,025. Таким образом, вероятность того, что тестовая χ2-статистика лежит между критическими значениями 12,40 и 39,36, равна 0,95. Следовательно, задав уровень значимости и определив количество степеней свободы, мы можем найти критическое значение тестовой χ2-статистики для любого конкретного распределения χ2.

Вернемся к задаче о расфасовке кукурузных хлопьев. Нас интересует, отличается ли стандартное отклонение от заданной величины, равной 15 г. Следовательно, можно применить двусторонний критерий, а нулевую и альтернативную гипотезы сформулировать следующим образом: *Н0: σ = 15 г; Н1: σ ≠ 15 г.*

Если для анализа создана выборка, состоящая из 25 коробок, то нулевая гипотеза отклоняется, если тестовая χ2-статистика попадает в область отклонения гипотезы, ограниченную левым или правым хвостами кривой распределения χ2 с (25 — 1) = 24 степенями свободы. Следовательно, решающее правило формулируется следующим образом. Гипотеза *Н0* отклоняется, если χ2 > χU2 = 39,364 или χ2 < χL2 = 12,401, в противном случае гипотеза не отклоняется.

Предположим, что выборочное стандартное отклонение *S*, вычисленное для выборки, состоящей из 25 коробок, равно 17,7 г. Для того чтобы проверить нулевую гипотезу при уровне значимости, равном 0,05, воспользуемся формулой (1):

*χ2 =* $\frac{(n –1)S^{2}}{σ^{2}}= \frac{(25 –1 )(17,7)^{2}}{15^{2}}$ *= 33,42*

Обратите внимание на то, что 33,42 — вычисленное значение тестовой χ2-статистики — лежит в интервале, ограниченном нижним и верхним критическими значениями, т.е. 12,40 и 39,36. То есть, гипотезу *Н0* отклонять нельзя. Следовательно, нет оснований утверждать, что стандартное отклонение генеральной совокупности отличается от 15 г.

Критерий χ2 для проверки гипотезы от дисперсии или стандартном отклонении считается классической параметрической процедурой. При проверке гипотезы о дисперсии генеральной совокупности или стандартном отклонении предполагается, что исходные данные имеют нормальное распределение. К сожалению, χ2-критерий довольно чувствителен к нарушению этих предположений (т.е. этот критерий не является устойчивым). Следовательно, если генеральная совокупность не имеет нормального распределения, особенно, когда объем выборки невелик, точность критерия может значительно снизиться.

Предыдущая заметка [Ранговый критерий Крускала-Уоллиса. Непараметрический метод для полностью рандомизированного эксперимента](http://baguzin.ru/wp/?p=6033)

Следующая заметка

К оглавлению [Статистика для менеджеров с использованием Microsoft Excel](http://baguzin.ru/wp/?p=5285)

1. Используются материалы книги Левин и др. Статистика для менеджеров. – М.: Вильямс, 2004. – с. 755–757 [↑](#footnote-ref-1)