3.4 Тема. Критерий X^2 («хи-квадрат») К.Пирсона

Цель. Освоение метода X^2, использование его при решении задач

Как бы точно ни вычислялись теоретические частоты, они, как правило, не совпадают с эмпирическими частотами ряда. Отсюда возникает необходимость сопоставления эмпирических частот с вычисленными, или ожидаемыми, частотами, с тем, чтобы установить достоверность или случайность наблюдаемого между ними расхождения.

Для сопоставления эмпирических и теоретических частот количественных и качественных признаков К.Пирсон (1900) предложил использовать критерий «хи-квадрат» (X^2), или критерий соответствия. Формула критерия «хи-квадрат» включает сумму дробей, полученную от деления квадрата разности между эмпирическими ($P_{эмп}$) и теоретическими частотами ($P_{теор}$) на частоты теоретические ($P_{теор}$):

$$X^2 = \sum \frac{(Pэмп - Pтеор)^2}{Pтеор} \quad (35)$$

Символ X^2 – не квадрат какого-то числа, он выражает лишь исходную величину, определяемую данной формулой. Так как отклонения эмпирических частот возводятся в квадрат, величина критерия X^2 всегда положительная.

При полном совпадении эмпирических частот с частотами вычисленными или ожидаемыми Σ(Рэмп-Ртеор)=0 и критерий X^2 тоже будет равен нулю. Если же Σ(Рэмп-Ртеор)$\neq$0, это укажет на несоответствие вычисленных частот эмпирическим частотам ряда. В таких случаях необходимо оценить значимость критерия X^2, который теоретически может изменяться от 0 до ∞. Это производится путем сравнения фактически полученной величины $X^2_ф$ с его критическим значением (X^2_{st}). Нулевая гипотеза, т.е. предположение, что расхождение между эмпирическими и теоретическими или ожидаемыми частотами носит случайный характер, опровергается, если $X^2_ф \geq X^2_{st}$ для принятого уровня значимости и числа степеней свободы. Критические значения X^2 для разных уровней значимости и чисел степеней свободы содержатся в таблице 2 приложений.

Распределение вероятных значений случайной величины X^2 непрерывно и асимметрично. Оно зависит от числа степеней свободы и приближается к нормальному распределению по мере увеличения числа наблюдений. Поэтому применение критерия X^2 к оценке дискретных распределений сопряжено с некоторыми погрешностями, которые сказываются на его величине, особенно на малочисленных выборках. Для получения более точных оценок выборка, распределяемая в вариационный ряд, должна иметь не менее 50 вариант. Правильное применение критерия хи-квадрат требует также, чтобы частоты вариант в крайних классах не были бы меньше 5; если их меньше 5, то они объединяются с частотами соседних классов, чтобы в сумме составляли величину, большую или равную 5. Соответственно объединению частот уменьшается и число классов. Число степеней свободы устанавливается по вторичному числу классов с учетом числа ограничений свободы вариации.

Пример. Требуется оценить результат испытания нового препарата для предупреждения инфекционного заболевания кроликов. Из 50 кроликов 20 получали профилактический препарат (опытная группа), а 30 – не получали (контроль). В опытной группе заболело 7 особей.

Доказывают ли результаты опыта профилактическое действие препарата или случайность причин?

Чтобы сделать определенное заключение, следует все данные опыта внести в таблицу и провести соответствующую их обработку.

Таблица 3.4.1 Расчет критерия соответствия при определении достоверности различий между кроликами двух групп

Группа животных	Число заболевших животных		Число здоровых животных		Всего животных в группе
	Наблюдаемые (Ф)	Теоретически ожидаемые (Т)	Наблюдаемые (Ф)	Теоретически ожидаемые (Т)	
Опытная	7	8,4 (T_1)	13	11,6 (T_2)	20
Контрольная	14	12,6 (T_3)	16	17,4 (T_4)	30
Итого	21	21	29	29	50

Следует подсчитать теоретически ожидаемые частоты – Т для заболевших и здоровых животных в опытной и контрольной группах; $T_1=\frac{20.21}{50}=8.4$; T_2 можно найти вычитанием из числа животных в опытной группе ожидаемой величины – T_1, т.е. T_2=20-8,4=11,6; $T_3=\frac{30.14}{50}=12.6$; T_4=30-12,6=17,4.

Подставив все величины в формулу (35) получим:

$$x^2=\frac{(7-8.4)-1(2)^2}{8.4}+\frac{(13-11.6)-1(2)^2}{116}+\frac{(14-12.6)-1(2)^2}{12.6}+\frac{(16-17.4)^2}{17.4}=0.26$$

Для того, чтобы провести сравнение вычисленного значения хи-квадрат с табличным, необходимо знать число степеней свободы, которое на единицу меньше числа классов. При расчетах по четырехпольным таблицам число степеней свободы равно единице. Сравнивая полученное в нашем примере значение хи-квадрат со стандартным, находим, что вычисленная нами величина (0,26) меньше всех стандартных ее значений в строке таблицы, соответствующей одной степени свободы. Следовательно, профилактическое действие препарата не может считаться доказанным.

Задание 1. При скрещивании черноколосой персидской пшеницы с пшеницей красноколосой все растения первого поколения оказались черноколосыми, а во втором поколении, т.е. от посева гибридных семян, получилось расщепление на 154 черноколосых, 40 красноколосых и 15 белоколосых растений.

Проверьте с помощью предположение, что в данном случае расщепление соответствует ожидаемому отношению 12:3:1.

Задание 2. При изучении защитного действия индло-3-пропиогидросомовой кислоты и экспериментальном заражении кроликов болезнью Ауески из 20 особей выжило 8, пало 12, а при изучении терапевтического эффекта – из 17 кроликов выжило 6, пало 11. Проверьте гипотезу об эффективности терапевтического и защитного действия этого препарата.

Контрольные вопросы.

1. Что такое критерий соответствия (хи-квадрат) и как он используется в генетических исследованиях?
2. Как применяют хи-квадрат при изучении наследования качественных признаков?
3. Как применяют хи-квадрат при определении достоверности различий между двумя группами животных?