Тема лекции 12. **Сопоставление эмпирических и теоретических частот.**

Эмпирический вариационный ряд и его график – вариационная кривая – не позволяют с полной уверенностью судить о законе распределения совокупности, из которой взята выборка. На величине любого варьирующего признака сказывается влияние многочисленных, в том числе и случайных, факторов, искажающих четкую картину варьирования. Между тем знание закона распределения позволяет избежать возможных ошибок в оценке генеральных параметров по выборочным характеристикам.

Критерий соответствия. Критерий соответствия χ^2 (хи-квадрат) предложен К.Пирсоном в 1900 году и используется для проверки гипотез путем сравнения фактического распределения с теоретическим.

Сравнение фактических (эмпирических) и теоретических частот с помощью критерия хи-квадрат необходимо для выяснения достоверных различий между сравниваемыми выборками. При сопоставлении частот двух рядов исходят от *нулевой гипотезы* (H_o), сущность которой сводится к предположению, что разница между средними и другими выборочными показателями сравниваемых групп равно нулю и, что различия, наблюдаемые между выборочными характеристиками, носят не систематический, а исключительно случайный характер. Если же полученный эмпирический ряд достоверно отличается по распределению частот от распределения теоретического ряда, то полученные частоты подвергаются сомнению, а нулевая гипотеза отвергается и принимается противоположная ей *альтернативная гипотеза* (H_a).

Особенность и свойства критерия хи-квадрат:

1. Используют при сравнении частот двух эмпирических рядов или сравнении эмпирических рядов с теоретическими при проверке различных гипотез, при экологическом анализе, при оценке эффективности применения лекарственных препаратов и др.

2. Указывает на достоверность или недостоверность различий между частотами сопоставляемых рядов.

3. Формула для вычисления критерия соответствия:

$$\chi^2 = \sum \frac{(f_{\text{э}} - f_m)^2}{f_m} \qquad (63)$$

где $f_{\text{э}}$ – фактическое распределение частот (эмпирическая частота),

f_m – теоретически ожидаемое распределение частот (теоретическая частота).

4. Критерий хи-квадрат – показатель приближенный. Величина критерия хи-квадрат всегда положительна и может выражаться любой величиной от нуля и выше.

5. Ограничения в использовании метода хи-квадрат:

5.1. применим для выборок численностью 20 особей и более ($n<20-30$);

5.2. не может быть применен к значениям частот, выраженных в относительных величинах (в процентах или долях).

6. Хи-квадрат зависит от числа степеней свободы γ и уровней вероятности (P):

Формула	Пути установления γ	Формула	Пути установления γ
$\gamma = n - 3$	Нормальное распределение	$\gamma = l - 1$	Для вариационных рядов, оформленных в классы
$\gamma = n - 2$	Распределение Пуассона	$\gamma = (l_x - 1)\cdot(l_y - 1)$	Для таблиц, оформленных в виде корреляционных решеток

7. Нулевая гипотеза опровергается или остается в силе. Для проверки нулевой гипотезы нужно фактически полученную величину $\chi_э^2$ сравнить с её критическим значением χ_{st}^2. Если $\chi_э^2 \geq \chi_{st}^2$, то нулевая гипотеза отвергается на принятом уровне значимости с числом степеней свободы γ, в случае, если $\chi_э^2 \leq \chi_{st}^2$, то нулевая гипотеза применяется.

Для понятия сущности данной темы освоим методику вычисления хи-квадрат при изучении распределении пауков определенного вида по окрасу (качественный признак) на изучаемой площадке. В качестве примера рассмотрим задачу, в которой представлено расщепление по окрасу изучаемых пауков. Так, из 5400 обнаруженных на участке пауков 16 были серо-желтыми. Остальные имели темный окрас. С использованием метода хи-квадрат выясним, что наблюдаемое в опыте расщепление соответствует ожидаемому или нет.

Решение. Для решения необходимо заполнить таблицу 12.1, где по форме изложена методика вычисления. В первую графу записываем классы; в нашем случае их два – доминантный и рецессивный. Во вторую графу вводим сведения из условия задачи: из 5400 пауков 16 имеют окрас рецессивный и 5384 – доминантный. В третью графу записываем теоретически возможное расщепление, которое исходит из второго закона Г.Менделя. Поэтому из 5400 пауков 75% особей введем в доминантный, а 25% - в рецессивный класс.

Далее производим вычисления по представленным формулам таблицы (графы 4,5,6). В итоге, при сложении показателей, записанных в шестой графе, мы получаем хи-квадрат, который равен 0,059.

Таблица 12.1

Проверка генетической гипотезы о моногибридном наследовании окраса тела у пауков с использованием критерия хи-квадрат

Класс	$f_э$	f_m	$f_э - f_m$	$(f_э - f_m)^2$	$\frac{(f_э - f_m)^2}{f_m}$
Доминанта	5384	5383	1	1	0,0002
Рецессива	16	17	-1	1	0,0588
Итого	5400	5400	-	-	0,059

Таблица 12.2

Стандартные значения χи –квадрат при разных уровнях вероятности

Число степеней свободы γ (n-1)	Вероятность (p)			Число степеней свободы γ (n-1)	Вероятность (p)		
	0,95	0,99	0,999		0,95	0,99	0,999
1	3,8	6,6	10,8	17	27,6	33,4	40,8
2	6,0	9,2	13,8	18	28,9	34,8	42,3
3	7,8	11,3	16,3	19	30,1	36,2	43,8
4	9,5	13,3	18,5	20	31,4	37,6	45,3
5	11,1	15,1	20,5	21	32,7	38,9	46,8
6	12,6	16,8	22,5	22	33,9	40,3	48,3
7	14,1	18,5	24,3	23	35,2	41,6	49,7

Сравним при γ =1 фактически полученную величину $\chi_ф{}^2$ с её критическим значением $\chi_{st}{}^2$. С этой целью используем из таблицы 12.2 уровни «вероятности» и строку γ = 1. В этой строке стоят три значения χ^2: 3,8 (Р=0,05); 6,6 (Р=0,01); 10,8 (Р=0,001). Вычисленное значение χ^2 значительно меньше табличных. Следовательно, наблюдаемое в опыте расщепление соответствует ожидаемому, а поэтому нулевая гипотеза, т.е. расщепление по фенотипу в соотношении 3 : 1, остается в силе. Поэтому, отклонения от данного расщепления, т.е. 17 (теоретическая частота) против 16 (эмпирическая частота) рецессив носит случайный, а не систематический характер.

Ответ. Наблюдаемое в опыте расщепление соответствует ожидаемому, а поэтому нулевая гипотеза, т.е. расщепление по фенотипу в соотношении 3 : 1, остается в силе.

Критерий λ. Данный критерий предложен А.Н.Колмогоровым и Н.И.Смирновым и может применяться для определения достоверности расхождения между фактическими и теоретическими распределениями, а также различий между любыми двумя распределениями частот одного и того же признака даже в том случае, когда число классов и число дат у этих распределений неодинаково. Для применения критерия лямбда не требуется определять число степеней свободы и не нужны таблицы для определения трех предельных значений критерия, так как для любого числа классов эти предельные значения одинаковы: 1,36; 1,63; 1,95 и соответствуют обычным трем степеням вероятности достоверного различия – P = 0,95; 0,99 и 0,999. Единственным условием применения критерия лямбда является достаточная численность сравниваемых распределений – не менее нескольких десятков дат.

Литература:

Основная – 1 [84-128]; 2 [т.1-117-137]; 3 [419-492].

Дополнительная – 2 [46-102]; 3 [64-90]; 5 [241-322].

Контрольные вопросы:

1. Метод хи-квадрат: применение, основные формулы вычисления.
2. В чем суть нулевой гипотезы?
3. Какие условия следует соблюдать при применении метода хи-квадрат?
4. Критерий Ястремского.
5. λ критерий.