Тема 5. ОЦЕНКА СООТВЕТСТВИЯ МЕЖДУ ОЖИДАЕМЫМИ И НАБЛЮДАЕМЫМИ РАСПРЕДЕЛЕНИЯМИ ПО КРИТЕРИЮ χ^2 (хи-квадрат)

Объяснение. Если необходимо определить соответствие двух сравниваемых рядов распределения (эмпирического и теоретического или двух эмпирических) используют критерий χ^2.

Часто критерий χ^2 применяют:

– В генетическом анализе, когда необходимо убедиться в том, является ли обнаруженное отклонение от теоретически ожидаемого распределения (1 : 1, 1 : 3, 9 : 3, 9 : 7 и т.д.) закономерным или лежит в пределах возможных случайных колебаний.

– В защите растений, токсикологии, экологии при изучении качественных признаков для оценки соответствия эмпирических данных нулевой гипотезе (теоретической предпосылке H_0).

Нулевая гипотеза отвергается, если $\chi^2_{\text{факт.}} > \chi^2_{\text{теор.}}$ и не отвергается, если $\chi^2_{\text{факт.}} < \chi^2_{\text{теор.}}$. Теоретически ожидаемые частоты обозначают через M, а опытные, эмпирически полученные через m. Критерий χ^2 представляет собой сумму отношений квадратов разностей между частотами эмпирического и теоретического распределений к частотам теоретического распределения для данной группы.

$$\chi^2 = \frac{(m_1 - M_1)^2}{M_1} + \frac{(m_2 - M_2)^2}{M_2} + \cdots + \frac{(m_n - M_n)^2}{M_n} = \sum \frac{(m - M)^2}{M}$$

Величина $\chi^2_{\text{теор.}}$ зависит от числа степеней свободы, которое определяют по формуле $v = (c - 1) \times (\kappa - 1)$, где c – это число строк, а κ – число колонок в анализируемой таблице. Значение $\chi^2_{\text{теор.}}$, соответствующее определенному числу степеней свободы, находят в приложении 2.

Пример 1. При скрещивании двух сортов гороха Г. Мендель во втором поколении получил 355 желтых семян (m_1) и 123 зеленых семени (m_2). Общее количество полученных семян – 478 штук. Соответствуют ли результаты опыта теоретически ожидаемому отношению желтых семян к зеленым как 3:1? В качестве нулевой гипотезы (H_0) принимаем соотношение 3:1.

Решение. Расчеты удобно осуществлять в табл. 2. Сначала в соответствующие графы заносим ожидаемое расщепление и наблюдаемые частоты (m_1 и m_2) из условия задачи. Затем определим теоретически ожидаемые частоты (M_1 и M_2):

M_1 = ¾ × 478=358,5
M_2 = ¼ × 478=119,5

Таблица 2 – Результаты опыта и последовательность расчета критерия χ^2

Показатели	Семена		Сумма
	желтые	зеленые	
Ожидаемое расщепление H_0	3	1	4
Наблюдаемые частоты (m)	355	123	478
Ожидаемые частоты (M)	358,5	119,5	478
Разность ($m – M$)	-3,5	3,5	-
Квадрат разности $(m – M)^2$	12,25	12,25	-
Отношение $(m – M)^2 / M$	0,034	0,103	χ^2=0,137

Подставляя эмпирические и теоретически ожидаемые частоты в формулу для вычисления критерия χ^2, получим:

$$\chi^2_{\text{факт.}} = \sum \frac{(m-M)^2}{M} = \frac{(355-358{,}5)^2}{358{,}5} + \frac{(123-119{,}5)^2}{119{,}5} = 0{,}137$$

При числе степеней свободы равном 1 ($\nu = (c – 1) \times (\kappa – 1) = (2 – 1) \times (2 – 1) = 1$) теоретическое значение χ^2_{05} = 3,84 (приложение 2).

Вывод: так как $\chi^2_{\text{факт.}}$ (0,137) < χ^2_{05} (3,84), то нулевая гипотеза H_0 не отвергается и различия между фактическими и теоретически ожидаемыми частотами несущественны.

Пример 2. Рассмотрим пример с распределением кормящихся птиц по частям крон деревьев, где они собирают корм (или распределением гнезд по имеющимся искусственным дуплянкам в разных местообитаниях). Предположим, мы получили следующие данные распределения вида по 5 градациям среды (частям крон или местообитаниям) (второй столбец таблицы):

Таблица 3 – Результаты опыта и последовательность расчета критерия χ^2

Номер стации	Фактическое распределение m	Теоретическое распределение M	$m – M$	$(m – M)^2$	$(m – M)^2 / M$
1	35	40	-5	25	0,625
2	28	30	-2	4	0,133
3	18	20	-2	4	0,20
4	12	8	4	16	2,0
5	7	2	5	25	12,5
сумма	100	100			χ^2=15,46

Данные фактического распределения могут быть представлены как в виде частоты встречаемости (в процентах), так и в абсолютных значениях, однако с точки зрения расчетов теоретической частоты удобнее использовать проценты. Теоретической (ожидаемой нами) частотой распределения (помещена в третий столбец таблицы) в наших примерах могут быть, например, найденные нами данные о распределении биомассы корма по частям крон или распределение по исследуемым местообитаниям развешанных искусственных гнездовий.

Подставляем данные в формулу (см. таблицу 3) и получаем значение $\chi^2 = 15{,}46$. Наш результат оказался существенно выше критического (найденного по таблице) значения (при числе степеней свободы 4 на уровне значимости $P = 0{,}05$ критическое значение $\chi^2 = 9{,}49$), что свидетельствует о неслучайном (неравномерном) распределении нашего вида по частям крон или местообитаниям. Если бы полученное нами значение χ^2 было равно или меньше 9,49 мы бы доказали, что распределение птиц полностью соответствует распределению их ресурса (в данном случае корма или дуплянок).

Пример 3. Две группы насекомых (долгоносиков) содержали при разной температуре 25°C и 5°C. При первом температурном режиме из 84 насекомых погибли 16, при втором – из 50 погибли 25. Доказывают ли результаты опыта губительное действие низкой температуры на выживаемость долгоносиков?

Если между группами насекомых нет отличий по устойчивости к температуре, то доля погибших долгоносиков в обеих группах будет одинаковой, то есть ожидаемое число погибших особей должно быть пропорционально общему числу (H_0).

Решение. Условие задачи и решение оформляем в таблице 4.

Таблица 4 – Результаты опыта и расчеты теоретически ожидаемых частот

Вариант опыта (температура)	погибшие		выжившие		всего
	m	M	m	M	
1 (25°C)	16	25,7 (M_1)	68	58,3 (M_3)	84
2 (5°C)	25	15,3 (M_2)	25	37,4 (M_4)	50
Всего	41		93		134
%	30,6%	30,6%	69,4%	69,4%	100%

Теоретически ожидаемые частоты вычисляем согласно нулевой гипотезе. Если из общего числа насекомых (134 шт.) погибло 30,6% (41 шт.), то теоретически ожидаемая доля погибших в каждой группе будет такой же. Составим и решим соответствующие пропорции:

84 – 100%

X – 30,6%, отсюда X = (84 × 30,6) / 100 = 25,7шт. (M_1)
(50 × 30,6) / 100 = 15,3 шт.(M_2)
(84 × 69,4) / 100 = 58,3 шт. (M_3)
(50 × 69,4) / 100 = 37,4 шт. (M_4)

Подставляя эмпирические и теоретически ожидаемые частоты в формулу для вычисления критерия χ^2, получим:

$$\chi^2 = \frac{(16-25{,}7)^2}{25{,}7} + \frac{(25-15{,}3)^2}{15{,}3} + \frac{(68-58{,}3)^2}{58{,}3} + \frac{(25-34{,}7)^2}{34{,}7} = 14{,}137$$

При числе степеней свободы равном 3 ($v = (c-1) \times (\kappa-1) = (2-1) \times (4-1) = 3$) теоретическое значение $\chi^2_{05} = 7{,}81$.

Вывод: так как $\chi^2_{\text{факт.}}$ (14,137) > χ^2_{05} (7,81), то нулевая гипотеза H_0 отвергается и низкая температура действительно является причиной гибели долгоносиков. Далее следует выполнить индивидуальное **задание** (приложение 7).