2.2. Нахождение критических точек распределения χ^2

Найдем число степеней свободы k но формуле:

$$k = s - r - 1,$$

где s - число интервалов после объединения, r - число параметров, оцениваемых по выборке.

Итак

$$k = 7 - 2 - 1 = 4.$$

По таблице критических точек распределения χ^2 (приложение 3), по заданному уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = 4$ находим критическую точку $\chi^2_{табл}(\alpha, k)$:

$$\chi^2_{табл}(0{,}05;4) = 9{,}5.$$

Так как $\chi^2_{набл} < \chi^2_{табл}$ - гипотезу о нормальном распределении генеральной совокупности принимаем. Расхождение между эмпирическими и теоретическими частотами незначимо (случайно).

Требуется: по результатам практических работ №1и №2 проверить гипотезу о нормальном распределении случайной величины по критерию Пирсона.

Практическая работа № 4

Однофакторный дисперсионный анализ

1. Одинаковое число испытаний на всех уровнях

Пусть на количественный нормально распределенный признак X воздействует фактор F, который имеет р постоянных уровней $F_1, F_2, \ldots, F_p$. На каждом уровне произведено по q испытаний. Результаты наблюдений -числа (x_{ij}, где $i = 1, 2, \ldots, q$- номер испытания, $j = 1,2, \ldots, p$- номер уровня фактора), записывают в виде таблицы.

Таблица 4.1

Номер испытания	*Уровни фактора*			
i	F_1	F_2	…	F_p
1	x_{11}	x_{12}	…	x_{1p}
2	x_{21}	x_{22}	…	x_{2p}
…	…	…	…	..
q	x_{q1}	x_{q2}	…	x_{qp}
Групповая средняя $\overline{x}_{гр}$	$\overline{x}_{гр1}$	$\overline{x}_{гр2}$	…	$\overline{x}_{грp}$

Ставится задача: на уровне значимости α проверить нулевую гипотезу о равенстве групповых средних при допущении, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы. Для решения этой задачи вводятся: *общая сумма* квадратов отклонений наблюдаемых значений признака от общей средней

$$S_{общ} = \sum_{j=1}^{p}\sum_{i=1}^{q}\left(x_{ij} - \bar{x}\right)^2;$$

факторная сумма квадратов отклонений групповых средних от общей средней (характеризует рассеяние «между группами»)

$$S_{факт} = q\sum_{j=1}^{p}\left(\bar{x}_{грj} - \bar{x}\right)^2;$$

остаточная сумма квадратов отклонений наблюдаемых значений группы от своей групповой средней (характеризует рассеяние «внутри групп»)

$$S_{ост} = \sum_{j=1}^{p}\sum_{i=1}^{q}\left(x_{ij} - \bar{x}_{грj}\right)^2;$$

Практически остаточную сумму находят по формуле

$$S_{ост} = S_{общ} - S_{факт}$$

Разделив уже вычисленные факторную и остаточную суммы на соответствующее число степеней свободы, находят факторную и остаточную дисперсии:

$$S^2_{факт} = \frac{S_{факт}}{p-1},$$

$$S^2_{ост} = \frac{S_{ост}}{p(q-1)}.$$

Сравнивают факторную и остаточную дисперсии по критерию Фишера-Снедекора. Для этого находят *наблюдаемое* и *критическое* значения критерия.

$$F_{набл} = \frac{S^2_{факт}}{S^2_{ост}}$$

Если факторная дисперсия окажется меньше остаточной, то уже отсюда следует справедливость нулевой гипотезы о равенстве групповых средних, поэтому дальнейшие вычисления (сравнение дисперсий с помощью критерия F) излишни. В противном случае, если факторная дисперсия окажется больше остаточной, используют таблицу критических точек распределения Фишера-Снедекора и находят критическое значение критерия.

$$F_{крит} = F_{k_1;k_2;\alpha}$$

где $k_1 = p - 1$ и $k_2 = p(q-1)$ - числа степеней свободы факторной и остаточной дисперсий, α - уровень значимости.

Если $F_{набл} < F_{крит}$ - различие групповых средних незначимое.

Если $F_{набл} > F_{крит}$ - различие групповых средних значимое.

Пример 4.1. Опыт по определению влияния доз удобрений на рост некоторой культуры дали следующие результаты.

Таблица 4.2

Номер наблюдения	Урожайность при различных дозах удобрения		
	I	II	III
1	70	86	46
2	80	70	68
3	64	79	55
4	78	85	71

Методом дисперсионного анализа определить, существенны ли различия между средними. Надежность оценки принять равной 0,95.

Решение:

Определим число уровней фактора - p =3.

Определим число опытов на каждом уровне - q = 4.

Найдем общую среднюю:

$$\bar{x} = \frac{70+80+64+78+86+70+79+85+46+68+55+71}{12} = 71$$

Найдем групповые средние:

$$\bar{x}_{гр1} = \frac{70+80+64+78}{4} = 73;$$

$$\bar{x}_{гр2} = \frac{86+70+79+85}{4} = 80;$$

$$\bar{x}_{гр3} = \frac{46+68+55+71}{4} = 60.$$

Найдем общую сумму квадратов отклонений наблюдаемых значений признака от общей средней:

$$S_{общ} = (70-71)^2 + (80-71)^2 + (64-71)^2 + (78-71)^2 + (86-71)^2 + (70-71)^2 + (79-71)^2 + (85-71)^2 + (46-71)^2 + (68-71)^2 + (55-71)^2 + (71-71)^2 = 1556.$$

Найдем факторную сумму квадратов отклонений:

$$S_{факт} = 4 \cdot \left((73-71)^2 + (80-71)^2 + (60-71)^2\right) = 824.$$

Найдем остаточную сумму квадратов отклонений:

$$S_{ост} = S_{общ} - S_{факт} = 1556 - 824 = 732.$$

Найдем дисперсии факторную и остаточную:

$$S^2_{факт} = \frac{824}{3-1} = 412,$$

$$S^2_{ост} = \frac{732}{3\cdot(4-1)} = 81{,}3.$$

Найдем наблюдаемое значение критерия:

$$F_{набл} = \frac{412}{81{,}3} = 5{,}07.$$

Найдем критическое значение критерия:

$\gamma = 0{,}95$, следовательно $\alpha = 1 - \gamma = 0{,}05$, $k_1 = 2$, $k_2 = 9$, $F_{крит} = 4{,}26$.

$F_{набл} > F_{крит}$ - нулевую гипотезу отвергаем. Средние между собой различаются существенно, т.е. влияние доз удобрений на рост культуры значимо.

2. Неодинаковое число испытаний на разных уровнях

Если число испытаний на первом уровне равно q_1, на втором уровне - q_2 и тд., а на уровне p - соответственно q_p, то вычисления проводят следующим образом. Общую сумму квадратов отклонений вычисляют, как и в случае одинакового числа испытаний на всех уровнях. Факторную сумму квадратов отклонений находят по формуле:

$$S_{факт} = q_1\cdot\left(\overline{x}_{гр1} - \overline{x}\right)^2 + q_2\cdot\left(\overline{x}_{гр2} - \overline{x}\right)^2 + \ldots + q_p\cdot\left(\overline{x}_{гр} - \overline{x}\right)^2;$$

где общее число испытаний n равно:

$$n = q_1 + q_2 + \ldots + q_p.$$

Затем находят остаточную сумму:

$$S_{ост} = S_{общ} - S_{факт}.$$

Число степеней свободы факторной суммы соответственно равно:

$$k_1 = p - 1$$

Число степеней свободы остаточной суммы находят по формуле:

$$k_2 = k - k_1$$

где $k = n - 1$ - число степеней свободы общей суммы.

Остальные вычисления проводят таким же образом, как и в случае одинакового числа испытаний на разных уровнях.

Пример 4.2. На Узбекской опытной рисовой станции проводились опыты по изучению влияния удобрений на урожай риса. В 1-ой опытной группе применяли удобрение $P_2O_5 + N$; во 2-ой - то же удобрение, а также предпосевную обработку почвы. Получены следующие данные (в ц/га):

Таблица 4.3

Группы	Урожаи на делянках						
Контрольная	35	33	31	37	42	35	40
Опытная 1-я	43	48	54				
Опытная 2-я	36	31	42	36			

Проанализируйте полученные результаты.

Решение:

Определим число уровней фактора - $p = 3$.

Определим число наблюдений на каждом уровне фактора: $q_1 = 7;\ q_2 = 3;\ q_3 = 4.$;

Найдем общую среднюю:

$$\bar{x} = \frac{35+33+31+37+42+35+40+43+48+54+36+31+42+36}{14} = 38{,}79$$

Найдем групповые средние:

$$\bar{x}_{гр1} = \frac{35+33+31+37+42+35+40}{7} = 36{,}14;$$

$$\bar{x}_{гр2} = \frac{43+48+54}{3} = 48{,}33;$$

$$\bar{x}_{гр3} = \frac{36+31+42+36}{4} = 36{,}25.$$

Найдем общую сумму квадратов отклонений наблюдаемых значений признака от общей средней:

$$S_{общ} = (35-38{,}79)^2 + (33-38{,}79)^2 + (31-38{,}79)^2 + (37-38{,}79)^2 + \ldots = 558{,}36.$$

Найдем факторную сумму:

$$S_{факт} = 7\cdot(36{,}14-38{,}79)^2 + 3\cdot(48{,}33-38{,}79)^2 + 4\cdot(36{,}25-38{,}79)^2 = 348{,}08.$$

Найдем остаточную сумму:

$$S_{ост} = S_{общ} - S_{факт} = 558{,}36 - 348{,}08 = 210{,}27.$$

Найдем число степеней свободы $S_{факт}$: $k_1 = p - 1 = 3 - 1 = 2.$

Найдем общее число степеней свободы. Так как общее число испытаний $n = 7+3+4$, следовательно $k = n - 1 = 14 - 1 = 13$.

Найдем число степеней свободы $S_{ост}$: $k_2 = k - k_1 = 13 - 2 = 11$.

Найдем дисперсии факторную и остаточную:

$$S^2_{факт} = \frac{348{,}08}{3-1} = 174{,}04;$$

$$S^2_{ост} = \frac{210{,}27}{11} = 19{,}12.$$

Найдем наблюдаемое значение критерия:

$$F_{набл} = \frac{174{,}04}{19{,}12} = 9{,}10.$$

Найдем критическое значение критерия, приняв 5% уровень значимости: $F_{крит} = 3{,}98$.

$F_{набл} > F_{крит}$ нулевую гипотезу отвергаем. Средние между собой различаются существенно. Внесение удобрений влияет на урожай.

Самостоятельные задания.

Задачи 1—6 приведены в таблице 4.4. Показаны результаты конкурсного сортоиспытания культур (урожайность в ц/га). Каждый сорт испытывался на четырех участках. Методом дисперсионного анализа изучите влияние сорта на урожайность на 5-% уровне значимости.

Таблица 4.4

№ задачи	Сорт	Урожайность по повторностям (ц/га)			
		1	2	3	4
1	Озимая пшеница по черному пару				
	Гибрид 481	32,2	32,7	30,7	33,3
	Новоукраинка 84	35,2	35,2	32,2	33,8
	Безостая 4	45,7	44,2	43,7	44,0
	Скороспелка 3	42,5	54,5	35,7	53,6
	Приазовская	36,8	37,0	38,0	37,0
2	Яровая пшеница по кукурузе				
	Мильтурум 321	29,0	27,1	27,7	27,5
	Альбидиум 3700	34,2	31,3	33,3	32,8
	Цезиум 111	31,3	30,4	30,1	30,2
3	Яровая пшеница по пласту трав				
	Мильтурум 553	29,7	33,1	32,8	29,7
	Альбидиум 3700	25,8	26,3	25,4	28,1
	Смена	25,2	25,8	24,3	26,3
	Лютесценс	28,5	27,9	27,9	27,6
4	Озимая пшеница по подсолнечнику				
	Новоукраинка 84	42,4	37,4	40,7	38,2
	Безостая 4	52,5	50,1	53,8	50,7
	Скороспелка 3	52,3	53,0	51,4	53,6
5	Озимая пшеница по пласту трав				
	Новоукраинка 84	32,4	33,3	34,8	34,6
	Безостая 4	43,2	44,1	46,8	43,9
	Скороспелка 3	54,3	54,6	58,3	60,5
6	Яровая пшеница по пару				
	Мильтурум 321	29,4	28,6	28,0	27,0
	Альбидиум 3700	27,0	29,0	28,4	30,6
	Цезиум 111	25,4	25,8	24,4	25,4