

Практическое занятие № 10(АБ,ЗР,ПВ,ДС)

КОРРЕЛЯЦИЯ КАЧЕСТВЕННЫХ ПРИЗНАКОВ

При изучении корреляционной зависимости между качественными признаками используют **коэффициент ассоциации** при альтернативной изменчивости и **коэффициент сопряженности**, если изучается зависимость между тремя и более признаками.

Коэффициент ассоциации является одной из модификаций критерия хи-квадрат Пирсона и изменяется в пределах от -1 до +1

Коэффициент ассоциации рассчитывается по формуле Юла:

$$r = \frac{n_1 \cdot n_4 - n_2 \cdot n_3}{\sqrt{N_1 \cdot N_2 \cdot N_3 \cdot N_4}}$$

где n_1 и n_4 - частоты с одинаковыми знаками клеток (++) или (--);
 n_2 и n_3 - частоты с разными знаками клеток (+- или -+);
 N_1 и N_2 – суммы частот по столбцам;
 N_3 и N_4 – суммы частот по строкам корреляционной таблицы.

Пример. Изучали эффективность обработки против парши на яблоне. На контрольном участке из 57 деревьев паршой было поражено 42, на обработанном участке было обследовано 63 дерева, из которых у 3 отмечено развитие болезни. Определить зависимость между обработкой и развитием болезни.

Т.к. в нашем примере определяется альтернативная зависимость между качественными признаками, рассчитываем **коэффициент ассоциации**.

1. Составляем четырехпольную корреляционную таблицу:

Вариант	Деревья		Суммы по рядам
	не пораженные (-)	пораженные (+)	
Без обработки – контроль (-)	15 (--; n_1)	42 (-+; n_2)	57 (N_3)
Обработка фунгицидом(+)	60 (+-; n_3)	3 (++; n_4)	63 (N_4)
Суммы по столбцам	75 (N_1)	45 (N_2)	$n = 120$

2. Вычисляем коэффициент ассоциации:

$$r = \frac{n_1 \cdot n_4 - n_2 \cdot n_3}{\sqrt{N_1 \cdot N_2 \cdot N_3 \cdot N_4}} = \frac{15 \cdot 3 - 42 \cdot 60}{\sqrt{75 \cdot 45 \cdot 57 \cdot 63}} = \frac{45 - 2520}{\sqrt{12119625}} = \frac{-2475}{3481} = -0,711$$

3. Рассчитываем ошибку коэффициента ассоциации:

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(-0,711)^2}{120-2}} = \sqrt{\frac{1-0,502}{118}} =$$

$$= \sqrt{\frac{0,498}{118}} = \sqrt{0,0042} = 0,065$$

4. Вычисляем критерий существенности коэффициента ассоциации:

$$t_r = \frac{r}{s_r} = \frac{-0,711}{0,065} = -10,9$$

Т.к. знак критерия существенности не несет смысловой нагрузки, мы его отбрасываем.

По статистической таблице находим значение критерия Стьюдента при 5% уровне значимости и числе степеней свободы равном $v = n-1=120-119$

$$t_{05} = 1,96$$

Критерий существенности больше критерия Стьюдента, следовательно между признаками существует значимая корреляционная зависимость.

Вывод: Обработка фунгицидом существенно снижает количество деревьев, пораженных паршой

Задание. Получив N примера от преподавателя, определить корреляционную зависимость между изучаемыми признаками

Пример 1. Изучали биологическую эффективность защитных мероприятий против плодовых вредителей в саду. Из 150 плодов, снятых с контрольных деревьев, плодовой вредитель был поврежден 124, из 120 плодов, снятых с обработанных деревьев, 36 было повреждено. Определить корреляционную зависимость между обработкой и поврежденностью плодов плодовой вредительницей.

Пример 2. Изучали влияние азотных удобрений на сохранность клубней картофеля. Из 200 клубней, отобранных с контрольного варианта, спустя 60 дней хранения товарными были 182, из 250 клубней, взятых с опытного варианта, товарными были 179. Определить зависимость между внесением азотных удобрений и сохранностью клубней картофеля.

Пример 3. Изучали эффективность защитных мероприятий против пероноспороза на луке. На участке, обработанном, фунгицидами, насчитывалось 12% пораженных растений, на контрольном – 89. Определить зависимость между обработками фунгицидами и поражаемостью растений пероноспорозом.

Решение:

1. Составляем четырехпольную корреляционную таблицу:

Вариант			Суммы по рядам
Суммы по столбцам			n =

2. Вычисляем коэффициент ассоциации:

$$r = \frac{n_1 \cdot n_4 - n_2 \cdot n_3}{\sqrt{N_1 \cdot N_2 \cdot N_3 \cdot N_4}}$$

3. Рассчитываем ошибку коэффициента ассоциации:

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

4. Вычисляем критерий существенности коэффициента ассоциации:

$$t_r = \frac{r}{s_r}$$

$$t_{05} = \quad \text{при } v = n - 1$$

Вывод:

При измерении тесноты связи между качественными альтернативными признаками и непрерывно варьирующим количественным признаком используется **биссерийальный коэффициент корреляции** (r_{bs}). Вычисляется коэффициент по формуле:

$$r_{bs} = \frac{\bar{x}_1 - \bar{x}_2}{s} \cdot \sqrt{\frac{n_1 \cdot n_2}{N(N-1)}}$$

где \bar{x}_1 и \bar{x}_2 - средние арифметические альтернативных групп, s - стандартное отклонение, n_1 и n_2 - объемы альтернативных групп, $N = (n_1+n_2)$ – общее число наблюдений.

Изменяется биссерийальный коэффициент корреляции в пределах от -1 до $+1$; при $\bar{x}_1 = \bar{x}_2$ $r_{bs} = 0$. Как и для коэффициента ассоциации, знак для биссерийального коэффициента не несет смысловой нагрузки.

Пример. Изучали влияние пораженности ботвы фитофторозом на урожайность картофеля сорта Приекульский ранний. Определить существует ли корреляционная зависимость между урожайностью картофеля и пораженностью ботвы фитофторозом.

Урожай, кг с 1 куста (X)		0,7	0,6	0,5	0,4	0,3	0,2
Число кустов, шт	всего (f)	12	15	18	13	9	6
	в т.ч. пораженных (f_1)	0	4	9	10	7	6

1. Составляем расчетную таблицу.

X	f_1	f_2	$f=f_1+f_2$	f_1X	f_2X	fX	X^2	fX^2
0,7	0	12	12	0	8,4	8,4	0,49	5,88
0,6	4	11	15	2,4	6,6	9,0	0,36	5,40
0,5	9	9	18	4,5	4,5	9,0	0,25	4,50
0,4	10	3	13	4,0	1,2	5,2	0,16	2,08
0,3	7	2	9	2,1	0,6	2,7	0,09	0,81
0,2	6	0	6	1,2	0	1,2	0,02	0,12
Суммы	$n_1=36$	$n_2=37$	$N=73$	$\sum f_1X=14,2$	$\sum f_2X=21,3$	$\sum fX=35,5$	-	$\sum fX^2=18,79$

2. Вычисляем средние для альтернативных групп:

$$\bar{x}_1 = \frac{\sum f_1X}{n_1} = \frac{14,2}{36} = 0,39$$

$$\bar{x}_2 = \frac{\sum f_2 X}{n_2} = \frac{21,3}{37} = 0,58$$

3. Вычисляем стандартное отклонение:

$$\begin{aligned} s &= \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{N}}{N-1}} = \sqrt{\frac{18,79 - 35,5^2 \div 73}{73-1}} = \\ &= \sqrt{\frac{18,79 - 1260,25 \div 73}{72}} = \sqrt{\frac{18,79 - 17,26}{72}} = \sqrt{\frac{1,53}{72}} = \\ &= \sqrt{0,021} = 0,14 \end{aligned}$$

4. Вычисляем бисериальный коэффициент корреляции:

$$\begin{aligned} r_{bs} &= \frac{\bar{x}_1 - \bar{x}_2}{s} \cdot \sqrt{\frac{n_1 \cdot n_2}{N(N-1)}} = \frac{0,39 - 0,58}{0,14} \cdot \sqrt{\frac{36 \cdot 37}{73(73-1)}} = \\ &= \frac{-0,19}{0,14} \cdot \sqrt{\frac{1332}{73 \cdot 72}} = -1,36 \cdot \sqrt{\frac{1332}{5256}} = -1,36 \cdot \sqrt{0,25} = \\ &= -1,36 \cdot 0,5 = -0,68 \end{aligned}$$

5. Рассчитываем ошибку бисериального коэффициента корреляции:

$$\begin{aligned} s_r &= \sqrt{\frac{1-r^2}{N-2}} = \sqrt{\frac{1-(-0,68)^2}{73-2}} = \sqrt{\frac{1-0,462}{71}} = \sqrt{\frac{0,538}{71}} = \\ &= \sqrt{0,0076} = 0,09 \end{aligned}$$

6. Рассчитываем критерий существенности бисериального коэффициента корреляции:

$$t_r = \frac{r}{s_r} = \frac{-0,68}{0,009} = -7,56$$

Т.к. знак критерия существенности не несет смысловой нагрузки, мы его отбрасываем.

По статистической таблице находим значение критерия Стьюдента при 5% уровне значимости и числе степеней свободы равном $v = N-2=73-2=71$

$$t_{05} = 2,01$$

Критерий существенности больше критерия Стьюдента, следовательно между признаками существует значимая корреляционная зависимость.

Вывод: С увеличением пораженности ботвы фитофторозом урожайность картофеля сорта Приекульский ранний существенно снижается.

Пример 4. Изучали влияние пораженности зерна шведской мухой на урожайность озимой пшеницы сорта Обрий. Определить существует ли корреляционная зависимость между урожайностью пшеницы и пораженностью зерен шведской мухой.

Урожай, ц/га		48	42	35	31	25	19
Число зерен в колосе, шт	всего	39	37	36	35	35	32
	в т.ч. пораженных	0	3	8	11	16	18

Решение:

1. Составляем расчетную таблицу.

X	f ₁	f ₂	f=f ₁ +f ₂	f ₁ X	f ₂ X	fX	X ²	fX ²
Суммы	n ₁ =	n ₂ =	N=	∑ f ₁ X	∑ f ₂ X	∑ fX	-	∑ fX ²

2. Вычисляем средние для альтернативных групп:

$$\bar{x}_1 = \frac{\sum f_1 X}{n_1} \qquad \bar{x}_2 = \frac{\sum f_2 X}{n_2}$$

3. Вычисляем стандартное отклонение:

$$s = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{N}}{N-1}}$$

4. Вычисляем бисериальный коэффициент корреляции:

$$r_{bs} = \frac{\bar{x}_1 - \bar{x}_2}{s} \cdot \sqrt{\frac{n_1 \cdot n_2}{N(N-1)}}$$

5. Рассчитываем ошибку бисериального коэффициента корреляции:

$$s_r = \sqrt{\frac{1 - r^2}{N - 2}}$$

6. Рассчитываем критерий существенности бисериального коэффициента корреляции:

$$t_r = \frac{r}{s_r}$$

$t_{05} =$ при $v = N - 2$

Вывод:

Значение критерия t на 5 и 1% уровне значимости

Число степеней свободы	Уровень значимости	
	05	01
1	12.71	63.33
2	4.30	9.93
3	3.18	5.84
4	2.78	4.60
5	2.57	4.03
6	2.45	3.71
7	2.37	3.50
8	2.31	3.36
9	2.26	3.25
10	2.23	3.17
11	2.20	3.11
12	2.18	3.06
13	2.16	3.01
14	2.15	2.98
15	2.13	2.95
16	2.12	2.92
17	2.11	2.90
18	2.10	2.88
19	2.09	2.86
21	2.08	2.83
22	2.07	2.82
23	2.07	2.81
24	2.06	2.80
25	2.06	2.79
26	2.06	2.78
27	2.05	2.77
28	2.05	2.76
29	2.05	2.76
30	2.04	2.75
50	2.01	2.68
100	1.98	2.63
более 100	1.96	2.58