Практическое занятие № 11

МНОЖЕСТВЕННАЯ КОРРЕЛЯЦИЯ

Корреляция называется ***множественной,*** если исследуется связь между тремя и более признаками. Наиболее простой формой множественной корреляции является зависимость между тремя признаками. Для определения формы и тесноты связи нескольких факторов применяют частные и множественные коэффициенты корреляции.

Частный коэффициент корреляции для трех признаков (факторов) определяет форму и тесноту связи между двумя при постоянном значении третьего. Обозначается частный коэффициент корреляции как $r_{xy.z}$, $r_{xz.y}$ и $r_{yz.x}$, где до точки указываются факторы, которые изучаются, после точки – фактор, значение которого элиминируется. Вычисляют частные коэффициенты корреляции по формулам:

$$r_{xy.z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2) \cdot (1 - r_{yz}^2)}}$$

$$r_{xz.y} = \frac{r_{xz} - r_{xy} \cdot r_{yz}}{\sqrt{(1 - r_{xy}^2) \cdot (1 - r_{yz}^2)}}$$

$$r_{yz.x} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{(1 - r_{xy}^2) \cdot (1 - r_{xz}^2)}}$$

где r_{xy}, r_{xz} и r_{yz} - парные коэффициенты корреляции.

Частные коэффициенты корреляции изменяются в пределах от –1 до +1.

Множественный коэффициент корреляции показывает тесноту связи одного фактора с совокупностью двух других факторов, то есть зависимость одного фактора от двух других. Обозначается множественный коэффициент корреляции как $R_{x.yz}$, $R_{y.xz}$ и $R_{z.xy}$, где до точки указывается признак, который изучается, после точки – факторы, влияющие на изучаемый признак.

Вычисляют множественные коэффициенты корреляции по формулам:

$$R_{x.yz} = \sqrt{\frac{r_{xy}^2 + r_{xz}^2 - 2r_{xy} \cdot r_{xz} \cdot r_{yz}}{1 - r_{yz}^2}}$$

$$R_{y.xz} = \sqrt{\frac{r_{xy}^2 + r_{yz}^2 - 2r_{xy} \cdot r_{xz} \cdot r_{yz}}{1 - r_{xz}^2}}$$

$$R_{z.xy} = \sqrt{\frac{r_{xz}^2 + r_{xy}^2 - 2r_{xy} \cdot r_{xz} \cdot r_{yz}}{1 - r_{xy}^2}}$$

Коэффициент множественной корреляции всегда положительное число и находится в пределах от 0 до 1.

Пример. Определить зависимость выживаемости диапазирующих особей калифорнийской щитовки от температуры воздуха и длины светового дня (По В.Ф. Пересыпкину, 1989).

Признак	1	2	3	4	5	6	7	8	9	10
Количество живых особей, % (Y)	99.0	95.5	94.0	92.0	79.5	58.5	44.9	26.6	14.4	7.9
Средняя температура воздуха в камере, ^{0}C (X)	-18	-20	-22	-24	-26	-28	-29	-30	-31	-32
Длина светового дня в период реактивации, ч (Z)	17	16	16	15	14	14	13	12	12	10

1. Заполняем расчетную таблицу:

x	y	z	x^2	y^2	z^2	xy	xz	yz
18	99,0	17	324	9801,0	289	1782,0	306	1683,0
20	95,5	16	400	9120,2	256	1910,0	320	1528,0
22	94,0	16	484	8836,0	256	2068,0	352	1504,0
24	92,0	15	576	8464,0	225	2208,0	360	1380,0
26	79,5	14	676	6320,2	196	2067,0	364	1113,0
28	58,5	14	784	3422,2	196	1638,0	392	819,0
29	44,9	13	841	2016,0	169	1302,1	377	583,7
30	26,6	12	900	707,6	144	798,0	360	319,2
31	14,4	12	961	207,4	144	446,4	372	172,8
32	7,9	10	1024	62,4	100	252,8	320	79,0
$\sum x=$ **260**	$\sum y=$ **612,3**	$\sum z=$ **139**	$\sum x^2 =$ **6970**	$\sum y^2 =$ **48957**	$\sum z^2=$ **1975**	$\sum xy =$ **14472,3**	$\sum xz =$ **3523**	$\sum yz =$ **9181,7**

2. Вычисляем парные коэффициенты корреляции:

$$r_{xy} = \frac{\sum xy - (\sum x \cdot \sum y) \div n}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \cdot \left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$= \frac{14472{,}3 - (260 \cdot 612{,}3) : 10}{\sqrt{(6970 - 260^2 : 10) \cdot (48597 - 612{,}3^2 : 10)}} =$$

$$= \frac{14472{,}3 - 159198 : 10}{\sqrt{(6790 - 67600 : 10) \cdot (48957 - 374911{,}3 : 10)}} =$$

$$= \frac{14472{,}3 - 15919{,}8}{\sqrt{(6970 - 6760) \cdot (48957 - 37491{,}1)}} = \frac{-1447{,}5}{\sqrt{210 \cdot 11465{,}9}} =$$

$$= \frac{-1447{,}5}{\sqrt{2407839}} = \frac{-1447{,}5}{1551{,}7} = -0{,}93$$

$$r_{xz} = \frac{\sum xz - (\sum x \cdot \sum z) \div n}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \cdot \left[\sum z^2 - \frac{(\sum z)^2}{n}\right]}} =$$

$$= \frac{3523 - (260 \cdot 139) \div 10}{\sqrt{210 \cdot (1975 - 139^2 \div 10)}} = \frac{3523 - 36140 \div 10}{\sqrt{210 \cdot (1975 - 19321 \div 10)}} =$$

$$= \frac{3523 - 3614}{\sqrt{210 \cdot (1975 - 1932{,}1)}} = \frac{-91}{\sqrt{210 \cdot 42{,}9}} = \frac{-91}{\sqrt{9009}} = \frac{-91}{94{,}9} = -0{,}95$$

$$r_{yz} = \frac{\sum yz - (\sum y \cdot \sum z) \div n}{\sqrt{\left[\sum y^2 - \frac{(\sum y)^2}{n}\right] \cdot \left[\sum z^2 - \frac{(\sum z)^2}{n}\right]}} =$$

$$= \frac{9181{,}7 - (612{,}3 \cdot 139) \div 10}{\sqrt{11465{,}9 \cdot 42{,}9}} = \frac{9181{,}7 - 85109{,}7 \div 10}{\sqrt{491887{,}1}} =$$

$$= \frac{9181{,}7 - 8511{,}0}{701{,}3} = \frac{670{,}7}{701{,}3} = 0{,}95$$

3. В соответствии с задачей эксперимента вычисляем частные коэффициенты корреляции:

$$r_{xy.z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1-r_{xz}^2)\cdot(1-r_{yz}^2)}} = \frac{-0{,}93-(-0{,}95\cdot 0{,}95)}{\sqrt{\left(1-0{,}95^2\right)\cdot\left(1-0{,}95^2\right)}} =$$

$$= \frac{-0{,}93+0{,}90}{\sqrt{(1-0{,}9)\cdot(1-0{,}9)}} = \frac{-0{,}03}{\sqrt{0{,}1\cdot 0{,}1}} = \frac{-0{,}03}{\sqrt{0{,}01}} = \frac{-0{,}03}{0{,}1} - 0{,}3$$

$$r_{yz.x} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{(1-r_{xy}^2)\cdot(1-r_{xz}^2)}} = \frac{0{,}95-(-0{,}93)\cdot(-0{,}95)}{\sqrt{\left(1-0{,}93^2\right)\cdot\left(1-0{,}95^2\right)}} =$$

$$= \frac{0{,}95-0{,}88}{\sqrt{(1-0{,}86)\cdot(1-0{,}90)}} = \frac{0{,}07}{\sqrt{0{,}14\cdot 0{,}1}} = \frac{0{,}07}{\sqrt{0{,}014}} = \frac{0{,}07}{0{,}12} = 0{,}58$$

4. Вычисляем ошибки частных коэффициентов корреляции:

$$s_{r_{xy.z}} = \sqrt{\frac{1-r_{xy.z}^2}{n-2}} = \sqrt{\frac{1-0{,}3^2}{10-2}} = \sqrt{\frac{1-0{,}09}{8}} = \sqrt{\frac{0{,}91}{8}} =$$

$$= \sqrt{0{,}11} = 0{,}33$$

$$s_{r_{yz.x}} = \sqrt{\frac{1-r_{yz.x}^2}{n-2}} = \sqrt{\frac{1-0{,}58^2}{10-2}} = \sqrt{\frac{1-0{,}34}{8}} = \sqrt{\frac{0{,}66}{8}} =$$

$$= \sqrt{0{,}08} = 0{,}29$$

5. Определяем критерии значимости частных коэффициентов корреляции и их существенность:

$$t_{xy.z} = \frac{r_{xy.z}}{s_{r_{xy.z}}} = \frac{0{,}30}{0{,}33} = 0{,}91$$

$$t_{yz.x} = \frac{r_{yz.x}}{s_{r_{yz.x}}} = \frac{0{,}58}{0{,}29} = 2{,}0$$

при $\nu = n\text{-}3 = 10\text{-}3 = 7$ $\qquad t_{05} = 2{,}37$

Критерии существенности частных коэффициентов корреляции меньше критерия Стьюдента, следовательно, корреляционная связь между двумя признаками при постоянном значении третьего незначима.

Т.е. при постоянной длине светового дня в период реактивации нет значимой корреляционной зависимости между количеством живых особей и средней температурой воздуха в камере; при постоянной средней температуре воздуха в камере нет значимой корреляционной зависимости между количеством живых особей и длиной светового дня в период реактивации.

6. В соответствии с задачей эксперимента вычисляем коэффициент множественной корреляции:

$$R_{y.xz} = \sqrt{\frac{r_{xy}^2 + r_{yz}^2 - 2r_{xy} \cdot r_{xz} \cdot r_{yz}}{1 - r_{xz}^2}} =$$

$$= \sqrt{\frac{0{,}93^2 + 0{,}95^2 - 2 \cdot (-0{,}93) \cdot (-0{,}95) \cdot 0{,}95}{1 - 0{,}95^2}} =$$

$$= \sqrt{\frac{0{,}86 + 0{,}90 - 2 \cdot 0{,}86 \cdot 0{,}9 \cdot 0{,}9}{1 - 0{,}9}} = \sqrt{\frac{1{,}76 - 1{,}67}{0{,}1}} =$$

$$= \sqrt{\frac{0{,}09}{0{,}1}} = \sqrt{0{,}9} = 0{,}95$$

7. Определяем значимость множественной корреляции по критерию Фишера:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - k}{k - 1} = \frac{0{,}95^2}{1 - 0{,}95^2} \cdot \frac{10 - 3}{3 - 1} =$$

$$= \frac{0{,}9}{0{,}1} \cdot \frac{7}{2} = 9 \cdot 3{,}5 = 31{,}5$$

где: n – объем выборки, k – число изучаемых признаков

при $\nu_2 = n\text{-}k = 7$ и $\nu_1 = k\text{-}1 = 2$ $\qquad F_{05} = 4{,}74$

Фактическое значение критерия Фишера больше теоретического, следовательно между изучаемыми признаками (факторами) есть значимая корреляционная связь.

Вывод: Выживаемость диапаузирующих особей калифорнийской щитовки в значительной степени зависит от температуры воздуха и длины светового дня.

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

Пример 1. Определить зависимость урожайности огурца от нормы выпуска хищной галлицы и количества вносимых фосфорных удобрений (По В.Ф. Пересыпкину, 1989).

.

Признак	1	2	3	4	5	6	7	8	9	10
Урожай, кг/м^2 (Y)	16.2	16.5	17.0	18.9	19.2	23.1	24.0	26.3	28.5	30.1
Норма выпуска галлиц, экз./м^2 (X)	50	50	60	60	70	70	80	80	90	100
Доза P_2O_5, г/м^3 (Z)	30	35	40	45	50	55	50	55	55	60

Пример 2. Определить зависимость площади листовой поверхности сеянцев яблони от численности зеленой яблонной тли и количества вносимых азотных удобрений (По В.Ф. Пересыпкину, 1989).

.

Признак	1	2	3	4	5	6	7	8	9	10
Площадь листовой поверхности, мм^2 (Y)	100.0	85.1	82.0	66.5	63.2	60.0	52.1	28.8	20.5	16.3
Среднее число особей на 10 см побега, шт (X)	0.2	1.8	1.5	1.9	4.9	8.8	17.4	15.6	33.8	46.1
Доза азотных удобрений, кг/га д.в. (Z)	45	45	40	40	35	35	30	30	28	30

Пример 3. Определить зависимость развития пыльной головни пшеницы от относительной влажности воздуха и среднемесячной температуры июня (По В.Ф. Пересыпкину, 1989).

.

Признак	1	2	3	4	5	6	7	8	9	10
Степень развития болезни, % (Y)	0.12	0.15	0.20	0.30	0.33	0.36	0.40	0.40	0.58	0.83
Относительная влажность воздуха, % (X)	45.4	61.6	69.1	63.5	66.1	71.9	67.0	77.5	71.1	87.9
Среднемесячная температура воздуха, ^oC (Z)	17.9	18.4	16.4	16.8	16.5	15.8	15.9	16.1	15.3	15.6

Решение

1. Заполняем расчетную таблицу:

x	y	z	x^2	y^2	z^2	xy	xz	yz
$\sum x$	$\sum y$	$\sum z$	$\sum x^2$	$\sum y^2$	$\sum z^2$	$\sum xy$	$\sum xz$	$\sum yz$

2. Вычисляем парные коэффициенты корреляции:

$$r_{xy} = \frac{\sum xy - (\sum x \cdot \sum y) \div n}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \cdot \left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$$

$$r_{xz} = \frac{\sum xz - (\sum x \cdot \sum z) \div n}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \cdot \left[\sum z^2 - \frac{(\sum z)^2}{n}\right]}}$$

$$r_{yz} = \frac{\sum yz - (\sum y \sum z) \div n}{\sqrt{\left[\sum y^2 - \frac{(\sum y)^2}{n}\right] \cdot \left[\sum z^2 - \frac{(\sum z)^2}{n}\right]}}$$

3. В соответствии с задачей эксперимента вычисляем частные коэффициенты корреляции:

$$r_{xy.z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{yz}^2)}}$$

$$r_{yz.x} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{xz}^2)}}$$

4. Вычисляем ошибки частных коэффициентов корреляции:

$$s_{r_{xy.z}} = \sqrt{\frac{1 - r_{xy.z}^2}{n-2}}; \qquad s_{r_{yz.x}} = \sqrt{\frac{1 - r_{yz.x}^2}{n-2}};$$

5. Определяем критерии значимости частных коэффициентов корреляции и их существенность:

$$t_{xy.z} = \frac{r_{xy.z}}{s_{r_{xy.z}}}; \qquad t_{yz.x} = \frac{r_{yz.x}}{s_{r_{yz.x}}}$$

t_{05} = при $\nu = n-3$

6. В соответствии с задачей эксперимента вычисляем коэффициент множественной корреляции:

$$R_{y.xz} = \sqrt{\frac{r_{xy}^2 + r_{yz}^2 - 2r_{xy} \cdot r_{xz} \cdot r_{yz}}{1 - r_{xz}^2}}$$

7. Определяем значимость множественной корреляции:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1}$$

где: n – объем выборки, k – число изучаемых признаков

F_{05} = при $v_2 = n-k$ и $v_1 = k-1$

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