4.4 Тема. Множественная корреляция

Цель. Знакомство с методами вычисления коэффициента множественной корреляции

Наряду с димерным анализом в биологических исследованиях применяется и многомерный анализ корреляционных связей, когда корреляция измеряется одновременно между несколькими варьирующими признаками. Простейшим случаем множественной корреляционной зависимости является корреляция трех признаков: Y, X и Z. Теснота связи между ними измеряется с помощью коэффициента множественной корреляции, который рассчитывается по следующей формуле:

$$R = \sqrt{\frac{r^2_{xz} + r^2_{yz} - 2r_{xy}r_{xz}r_{yz}}{1 - r^2_{xy}}}, \quad (45)$$

где r_{xy}, r_{xz} и r_{yz} – парные коэффициенты корреляции между признаками X и Y, X и Z, Y и Z.

Теснота связи между признаками X и Y (при постоянстве z) определяется частным коэффициентом корреляции:

$$r_{xy(z)} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r^2_{xz})(1 - r^2_{yz})}}. \quad (46)$$

Соответственно частный коэффициент корреляции между признаками X и Z при исключении влияния на эту связь признака Y равен

$$r_{xz(y)} = \frac{r_{xz} - r_{xy}r_{yz}}{\sqrt{(1 - r^2_{xy})(1 - r^2_{yz})}}. \quad (47)$$

Частный коэффициент корреляции между Y и Z при исключении влияния на эту связь признака X равен

$$r_{yz(x)} = \frac{r_{yz} - r_{xy}r_{xz}}{\sqrt{(1 - r^2_{xy})(1 - r^2_{xz})}}. \quad (48)$$

Частные коэффициенты корреляции имеют тот же смысл и обладают теми же свойствами, что и обыкновенный парный коэффициент корреляции.

t-критерий для проверки гипотезы о независимом варьировании двух признаков при исключении влияния третьего признака выражается в виде следующего отношения:

$$t = \frac{r_{частный}\sqrt{n - m}}{\sqrt{1 - r^2_{частный}}},$$

где n – объем выборки, m – число зависимых признаков, для которых вычисляется частный коэффициент корреляции.

Пример. Из снопа озимой ржи случайным способом было отобрано 10 колосьев. Затем измерялась длина (мм) каждого колоса (X), подсчитывалось число колосков (Y) и количество зерен (Z) в каждом колосе. Собранные данные и их предварительная обработка приведены в таблице 4.4.1

Таблица 4.4.1

X	Y	Z	X^2	Y^2	Z^2	X Y	YZ	XZ
70	18	36	4900	324	1296	1260	648	2520
60	17	29	3600	289	841	1020	493	1740
70	22	40	4900	484	1600	1540	880	2800
46	10	12	2116	100	144	460	120	552
58	16	31	3364	256	961	928	496	1798
69	18	32	4761	324	1024	1242	576	2208
32	9	13	1024	81	169	288	117	416
62	18	35	3844	324	1225	1116	630	2170
46	15	30	2116	225	900	690	450	1380
62	22	36	3844	484	1296	1364	792	2232
575	165	294	34469	2891	9456	9908	5202	17816

Для определения коэффициента множественной корреляции между этими признаками нужно сначала рассчитать парные коэффициенты корреляции. Пользуясь итогами таблицы 4.4.1, находим суммы квадратов отклонений вариант от их средних арифметических:
$\sum(x_i-\overline{X})^2=\sum x^2_i-(\sum x_i)^2/n=34469-575^2/10=34469-33062.5=1406.5$; $\sum(y_i-\overline{y})^2=\sum y^2_i-(\sum y_i)^2/n=2891-165^2/10=2891-2722.5=168.5$;
$\sum(z_i-\overline{z})^2=\sum z^2_i-(\sum z_i)^2/n=9456-294^2/10=9456-8643.6=812.4$; $\sigma_y=\sqrt{168{,}5/10}=4{,}10$; $\sigma_x=\sqrt{812{,}4/10}=9{,}01$.

Далее рассчитываем величины сопряженной величины:
$\sum(y_i-\overline{y})(x_i-\overline{X})=\sum yx-\sum y/n=9908-575\cdot165/10=420{,}5$;
$\sum(y_i-\overline{y})(z_i-\overline{z})=\sum yz--\sum z/n=5202-165\cdot294/10=351{,}0$;
$\sum(x_i-\overline{X})(z_i-\overline{z})=\sum xz--\sum x/n=17816-575\cdot294/10=911{,}0$.

Определим парные коэффициенты корреляции:

$$r_{xy}=\frac{\sum(y_i-\overline{Y}_i)(x_i-\overline{X})}{n\sigma_x\sigma_y}=\frac{420.5}{10\cdot4.1\cdot14.1}=0.727;$$

$$r_{yz}=\frac{351}{10\cdot4.1\cdot9.0}=0.951;\quad r_{xz}=\frac{911.0}{10\cdot14.1\cdot9.0}=0.710.$$

Рассчитываем частные коэффициенты корреляции:
$r_{xy(z)}=0.239$; $r_{yz(x)}=0.900$; $r_{xz(y)}=0.090$.

Определим общий коэффициент множественной корреляции для всех трех признаков:
$R^2=0.531$; $R=0.729$.