

## 6 КОРРЕЛЯЦИОННО- РЕГРЕССИОННЫЙ АНАЛИЗ

В математическом анализе зависимость между двумя величинами выражается понятием функции  $y = f(x)$ , где каждому допустимому значению одной переменной соответствует одно и только одно значение другой переменной. Такая зависимость носит название функциональной. Она обнаруживается с помощью строгих логических доказательств и не нуждается в опытной проверке.

В экспериментальных исследованиях, связанных с сельскохозяйственными объектами (растениями, животными, технологическими процессами) чаще всего встречаются такие соотношения между переменными, когда каждому значению признака  $x$  соответствует ни одно, а множество возможных значений признака  $y$ . Такие связи в отличие от функциональных связей, называются стохастическими (вероятностными) или корреляционными. При изучении таких связей возникают два основных вопроса – о тесноте связи и форме связи.

Для измерения тесноты связи используют специальный статистический метод, называемый *корреляцией*, а формы связи – *регрессией*. Корреляция может быть линейной и криволинейной, а по направлению прямой и обратной. Если с увеличением  $x$  увеличивается  $y$  – корреляция прямая, уменьшается  $y$  – обратная. Корреляцию и регрессию называют простой, если исследуется связь между двумя признаками и множественной, когда изучается зависимость между тремя и более признаками. Под регрессией понимается изменение результативного признака  $y$  (функции) при определенном изменении одного или нескольких факториальных признаков (аргументов). Связь между функцией и аргументом выражается уравнением регрессии или корреляционным уравнением.

Уравнение простой регрессии обозначается как

$$y = f(x), \quad (6.1)$$

множественной

$$y = f(x, z, v). \quad (6.2)$$

Если степень связи между признаками велика, то по уравнению регрессии можно предсказать значение результативного признака для определенных значений факториальных признаков. Для оценки тесноты (силы) связи используют коэффициенты корреляции и корреляционное отношение.

Совместное применение методов корреляции, регрессии и дисперсионного анализа для анализа эксперимента получило название ковариационного анализа. Сущность этого метода заключается в следующем: если между результативным признаком  $y$  и сопутствующим эксперименту неизучаемым признаком  $x$  имеет место значимая линейная связь, то методом ковариации можно статистически выравнять условия проведения опыта в отношении признака  $x$ , и тем самым заметно снизить ошибку эксперимента и получить большие информации об изучаемом явлении.

#### Линейная корреляция и регрессия.

Под линейной (прямолинейной) корреляционной зависимостью между двумя признаками  $x$  и  $y$  понимают такую зависимость, которая носит линейный характер и выражается уравнением прямой линии  $y = a + b(x)$ . Это уравнение называется уравнением регрессии  $y$  на  $x$ , а соответствующая ему прямая линия – выборочной линией регрессии  $y$  на  $x$ .

Линейная регрессия – это такая зависимость, когда при любом значении аргумента  $x$  одинаковые приращения его вызывают однокровные изменения функции  $y$ .

Когда при одинаковых приращениях аргумента функция имеет неодинаковые изменения, регрессия называется криволинейной.

Для анализа линейной корреляции между  $x$  и  $y$  проводят  $n$  – независимых парных наблюдений, исходом каждого из которых является пара чисел  $(x_1, y_1)$ ,  $(x_2, y_2)$ , ...,  $(x_n, y_n)$ . По этим значениям определяют выборочные эмпирические коэффициенты корреляции и регрессии, рассчитывают уравнение регрессии, строят теоретическую линию регрессии и оценивают значимость полученных результатов.

В качестве числового показателя простой линейной корреляции, указывающего на тесноту (силы) и направление связи  $x$  с  $y$ , используют коэффициент корреляции  $r$ . Это безразмерная величина, изменяющаяся от -1 до +1, рассчитываемая по формуле

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}, \quad (6.3)$$

или минуя вычисления отклонений и квадратов отклонений по выражению

$$r = \frac{\sum xy - (\sum x \sum y) : n}{\sqrt{\left( \sum x^2 - \frac{\sum x^2}{n} \right) \left( \sum y^2 - \frac{\sum y^2}{n} \right)}}, \quad (6.4)$$

Для вычисления коэффициента корреляции составляют таблицу 6.1.

Таблица 6.1 Расчет коэффициента корреляции

№ пары наблюдений	Значения признаков		$x^2$	$y^2$	$xy$
	$x$	$y$			
	$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$

Если каждой величине  $x$  соответствует только определенная величина  $y$ , то корреляционная связь переходит в функциональную (частный случай корреляционной). При

полных связях, когда корреляционная связь превращается в функциональную, значение  $r$  для положительных или прямых связей составляет  $r = +1$ , для отрицательных или обратных  $r = -1$ . Когда  $r = 0$  между  $x$  и  $y$  нет линейной связи, но криволинейная зависимость может существовать. Считается, что при  $r < 0,3$  корреляционная зависимость между признаками слабая,  $r = 0,3...0,7$  – средняя, а при  $r > 0,7$  – сильная.

Квадрат коэффициента корреляции ( $r^2$ ) называется коэффициентом детерминации и обозначается  $d_{xy}$ . Он показывает долю (%) тех изменений, которые в данной вариации зависят от случайного фактора.

Для оценки надежности выборочного коэффициента корреляции вычисляют его ошибку и критерий существенности.

Стандартная ошибка коэффициента корреляции определяется по формуле

$$S_r = \sqrt{\frac{1-r^2}{n-2}}, \quad (6.5)$$

где  $S_r$  – ошибка коэффициента корреляции;

$r$  – коэффициент корреляции;

$n$  – число пар значений, по которым вычислен выборочный коэффициент корреляции.

Из выражения (6.5) следует, что коэффициенты корреляции близкие к единице оказываются всегда точнее близких к нулю. С увеличением числа объектов исследования  $S_r$  также будет всегда уменьшаться, а точность в определении  $r$  – возрастать.

Критерий существенности коэффициента корреляции рассчитывается по формуле

$$t_r = \frac{r}{S_r}. \quad (6.6)$$

Если  $t_{r.факт} \geq t_{теор}$  - связь существенна, а при  $t_{r.факт} < t_{теор}$  - несущественна

Число степеней свободы равно  $n - 2$  ( $n$  – число пар).

При малых выборках и значениях  $r$  близких к единице, распределение выборочных коэффициентов корреляции заведомо отличается от нормального распределения. В этих случаях критерий Стьюдента становится ненадежным. Фишер предложил преобразовать  $r$  в величину  $z$ , которая распределена нормально. Для преобразования используются специальные таблицы. Ошибка величины  $z$  равна

$$S_z = \frac{1}{\sqrt{n-3}}. \quad (6.7)$$

где  $n$  – объем выборки.

Критерии значимости  $z$  и разности  $z_1$  и  $z_2$  при сравнении двух коэффициентов корреляции, а также доверительной границы  $z$  определяют по обычным соотношениям

$$t_z = \frac{z}{S_z}, \quad (6.8)$$

$$t_{z_1-z_2} = \frac{z_1 - z_2}{\sqrt{S_{z_1}^2 - S_{z_2}^2}}, \quad (6.9)$$

$$z = t \cdot S_z, \quad (6.10)$$

После определения доверительных границ делают обратный переход от  $z$  к  $r$ .

Проверить нулевую гипотезу  $H_0: r = 0$  можно и без расчетов критерия  $t_z$  непосредственно по таблице 7 приложения. В таблице 7 даны граничные значения коэффициентов корреляции на 5 %-ном и 1 %-ном уровнях значимости. Между  $x$  и  $y$  имеется существенная связь и  $H_0$  (нулевая гипотеза отсутствия связи) отвергается, если  $r_{ф} \geq r_{теор}$ , и не отвечает (нет существенной связи) если  $r_{ф} \leq r_{теор}$ . Рассматривая таблицу 7 легко заметить, что для доказательства зна-

чимости слабых связей необходимо 40-100, средних 12-40 и сильных 6-12 пар наблюдений.

Коэффициент корреляции указывает на направление и степень сопряженности в изменчивости признаков, но не позволяет судить о том, как количественно меняется результативный признак при изменении факториального на единицу измерения, что важно в познавательных и практических целях. В подобных случаях на помощь приходит регрессионный анализ.

### Регрессионный анализ.

Его основная задача - определить формулу корреляционной зависимости, т.е. уравнение прямой линии.

Уравнение линейной регрессии  $y$  по  $x$  имеет вид

$$y = \bar{y} - b_{yx}(x - \bar{x}), \quad (6.11)$$

где  $\bar{x}$  и  $\bar{y}$  - средние арифметические для ряда  $x$  и  $y$ ;

$b_{yx}$  - коэффициент регрессии  $y$  по  $x$ .

Коэффициент регрессии вычисляют по формулам.

$$b_{yx} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}, \quad (6.12)$$

$$b_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{(y - \bar{y})^2}, \quad (6.13)$$

Коэффициент регрессии  $b_{yx}$  показывает, как изменяется  $y$  при изменении  $x$  на единицу измерения и выражается в единицах  $y$ . Коэффициент  $b_{xy}$  указывает регрессию  $x$  на  $y$  и выражается в единицах  $x$ . При исследовании односторонней зависимости, например, корреляции между урожаями  $y$  и количеством выпавших осадков  $x$  вычисляют только один коэффициент регрессии результативного признака  $y$  на факториальный  $x$ , т.е. значение  $b_{yx}$ , так как регрессия  $x$  по  $y$  лишена в подобных случаях логики (нет смысла заказывать дождь под планируемый урожай).

Таким образом, коэффициентом линейной регрессии называется число, показывающее, в каком направлении и на какую величину изменяется в среднем признак  $y$  (функция) при изменении признака  $x$  (аргумента) на единицу измерения. Коэффициенты регрессии имеют знак коэффициента корреляции.

Произведение коэффициентов регрессии равно квадрату коэффициента корреляции

$$\mathbf{b}_{yx} \cdot \mathbf{b}_{xy} = \mathbf{r}^2, \quad (6.14)$$

По этой формуле можно производить проверку вычисления коэффициентов регрессии.

Ошибку коэффициента регрессии определяют по выражению

$$S_{b_{yx}} = S_r \sqrt{\frac{\sum (y - \bar{y})^2}{\sum (x - \bar{x})^2}}, \quad (6.15)$$

$$S_{b_{xy}} = \sqrt{\frac{\sum (x - \bar{x})^2}{\sum (y - \bar{y})^2}}. \quad (6.16)$$

Критерии существенности коэффициента регрессии вычисляют по формуле

$$t_b = \frac{\mathbf{b}}{S_b}. \quad (6.17)$$

Если определен критерий существенности для коэффициента корреляции, то он может быть использован и для оценки значимости коэффициента регрессии, так как  $t_b = t_r$ .

Существенность коэффициента регрессии оценивают по критерию Стьюдента. Число степеней свободы принимают равным

$$\mathbf{v} = \mathbf{n} - 2. \quad (6.18)$$

Корреляция может быть изображена графически в виде линии регрессии. При этом необходимо устранить влияние случайных отклонений и найти положение теоретической

линии регрессии, то есть усредненное течение функции при равномерном увеличении аргумента. Существуют два способа построения: графический (приближенный) и аналитический. Сущность аналитического способа построения состоит в том, что по исходным наблюдениям вычисляют  $\bar{x}$ ,  $\bar{y}$  и  $b_{yx}$ .

Далее подставляют найденные значения в уравнение линейной регрессии

$$y = \bar{y} + b_{yx}(x - \bar{x}). \quad (6.19)$$

Определяют коэффициенты уравнения прямой линии, которая принимает вид

$$y = a + bx. \quad (6.20)$$

По уравнению находят теоретически усредненные значения  $y$  для двух крайних значений ряда  $x$ . Найденные точки с координатами  $(x_{\min}, y_{\min})$  и  $(x_{\max}, y_{\max})$  наносят на график и соединяют прямой. Это и будет теоретическая линия регрессии  $y$  по  $x$ .

### Частная и множественная корреляция и регрессия

Корреляция называется множественной, если на величину результативного признака одновременно влияют несколько факториальных. Простой формой множественной связи является линейная зависимость между тремя признаками, когда один из них, например, урожай рассматривается как функция ( $y$ ), а два других – как аргументы ( $x$  и  $z$ ). В качестве меры тесноты линейной связи трех признаков используют частные коэффициенты корреляции, обозначаемые  $r_{xy.z}$ ,  $r_{xz.y}$ ,  $r_{zy.x}$  и множественные коэффициенты корреляции, обозначаемые  $R_{x.yz}$ ;  $R_{y.xz}$ ;  $R_{z.xy}$ .

*Частный коэффициент корреляции* – это показатель, измеряющий степень сопряженности двух признаков. При обозначении частных коэффициентов корреляции в индексах букв перед точкой указывают между какими признака-

ми изучается зависимость, а буква после точки – влияние какого признака исключается.

Математическая статистика позволяет установить корреляцию между двумя признаками при постоянном значении третьего, не ставя специального эксперимента, а используя парные коэффициенты корреляции  $r_{XY}$ ,  $r_{XZ}$ ,  $r_{YZ}$ .

Частные коэффициенты рассчитывают по формулам:

$$r_{XY \cdot Z} = \frac{r_{XY} - r_{XZ} \cdot r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}, \quad (6.21)$$

$$r_{XZ \cdot Y} = \frac{r_{XZ} - r_{XY} \cdot r_{ZY}}{\sqrt{(1 - r_{XY}^2)(1 - r_{YZ}^2)}}, \quad (6.22)$$

$$r_{YZ \cdot X} = \frac{r_{YZ} - r_{XY} \cdot r_{XZ}}{\sqrt{(1 - r_{XY}^2)(1 - r_{XZ}^2)}}, \quad (6.23)$$

Ошибку и критерии значимости частной корреляции определяют по тем же формулам, что и для парной корреляции

$$S_{ZXY \cdot Z} = \sqrt{\frac{1 - r_{XY \cdot Z}^2}{n - 2}}, \quad (6.24)$$

$$t = \frac{r}{S_r}, \quad (6.25)$$

Подобно парным коэффициентам корреляции частные коэффициенты могут принимать значения заключенные между  $-1$  и  $+1$ . Частные коэффициенты детерминации находят путем возведения в квадрат частных коэффициентов корреляции:

$$\left. \begin{aligned} d_{XY \cdot Z} &= r_{XY \cdot Z}^2; \\ d_{XZ \cdot Y} &= r_{XZ \cdot Y}^2; \\ d_{YZ \cdot X} &= r_{YZ \cdot X}^2. \end{aligned} \right\} \quad (6.26)$$

Частные коэффициенты корреляции иногда предоставляют большой интерес. Например, изучается зависимость урожая от осадков. Но результат сильно искажает варьиро-

вание температур в этот период. Частный коэффициент исключая влияние температуру позволяет более четко установить связь между первыми двумя признаками при постоянных значениях третьего.

Метод частной корреляции дает возможность вычислить частный коэффициент корреляции второго порядка. Этот коэффициент указывает на взаимосвязь между первым и вторым признаком при постоянном значении третьего и четвертого. Частный коэффициент корреляции второго порядка определяют на основании частных коэффициентов первого порядка по формуле

$$r_{XY \cdot ZV} = \frac{r_{XY \cdot V} - r_{XZ \cdot V} \cdot r_{Y \cdot V}}{\sqrt{(1 - r_{XY \cdot V}^2)(1 - r_{YZ \cdot V}^2)}}, \quad (6.27)$$

где  $r_{XY \cdot V}$ ,  $r_{XZ \cdot V}$ ,  $r_{YZ \cdot V}$  - частные коэффициенты первого порядка определяют по формуле частного коэффициента, используя коэффициенты парной корреляции  $r_{XY}$ ,  $r_{XZ}$ ,  $r_{XV}$ ,  $r_{YZ}$ ,  $r_{YV}$ ,  $r_{XV}$ .

Множественный коэффициент корреляции трех переменных – это показатель тесноты линейной связи между одним из признаков (буква индекса перед точкой) и совокупностью двух других признаков (буквы индекса после точки)

$$R_{X \cdot YZ} = \sqrt{\frac{r_{XY}^2 + r_{XZ}^2 - 2r_{XY} \cdot r_{XZ} \cdot r_{YZ}}{1 - r_{YZ}^2}}, \quad (6.28)$$

$$R_{Y \cdot XZ} = \sqrt{\frac{r_{XY}^2 + r_{YZ}^2 - 2r_{XY} \cdot r_{XZ} \cdot r_{YZ}}{1 - r_{XZ}^2}}, \quad (6.29)$$

$$R_{Z \cdot XY} = \sqrt{\frac{r_{XZ}^2 + r_{YZ}^2 - 2r_{XY} \cdot r_{XZ} \cdot r_{YZ}}{1 - r_{XY}^2}}, \quad (6.30)$$

Эти формулы позволяют легко вычислить множественные коэффициенты корреляции при известных значениях парной корреляции  $r_{XY}$ ,  $r_{XZ}$ ,  $r_{YZ}$ .

$1 \leq R > 0$  между коэффициентом множественной корреляции, например,  $R_{Y.XZ}$  и двумя коэффициентами парной корреляции  $r_{YX}$  и  $r_{YZ}$  существует следующее соотношение: каждый из парных коэффициентов не может превышать по абсолютной величине  $R_{Y.XZ}$ .

Квадрат коэффициента множественной корреляции  $R^2$  называется коэффициентом множественной детерминации. Он показывает долю вариации зависимой переменной под воздействием изучаемых факторов.

Значимость множественной корреляции оценивается по  $F$  критерию.

$$F = \frac{R^2}{1 - R^2} \left( \frac{n - k}{k - 1} \right), \quad (6.31)$$

где  $n$  – объем выборки;

$k$  – число признаков.

Теоретическое значение  $F$  – критерия берут из таблиц для  $\nu = (k - 1)$  и  $\gamma_2 = n - k$  степеней свободы. Нулевая гипотеза ( $H_0 : R = 0$ ) о равенстве множественного коэффициента в совокупности нулю принимается, если  $F_{\text{факт}} < F_{\text{теор}}$  и отвергается, если  $F_{\text{факт}} > F_{\text{теор}}$ .

**Пример:** При изучении методики селекционного отбора у репчатого лука второго года жизни получены коэффициенты парной корреляции ( $n=15$ ) между урожаем ( $y$ ) и средней массой луковицы ( $x$ ),  $r_{yx} = 0,6$  между урожаем ( $y$ ) и гнездой (з)  $r_{yz} = 0,3$ . Между средней массой луковицы ( $x$ ) и гнездой (з)  $r_{xz} = -0,2$ .

Необходимо выяснить зависимость урожая ( $y$ ) от средней массы луковицы ( $x$ ) и гнезды (з), то есть рассчитать коэффициент множественной корреляции.

$$R_{Y.XZ} = \sqrt{\frac{0,6^2 + 0,3^2 - 2 \cdot 0,6 \cdot 0,3(-0,2)}{1 - 0,2^2}} = 0,74$$

$$F_{\text{ФАКТ}} = \frac{0,74^2}{1 - 0,74^2} \left( \frac{15 - 3}{3 - 1} \right) = 7,33$$

Табличное значение  $F$  при 2 и 12 степенях свободы  $F_{0,5} = 3,89$  и  $F_{0,1} = 6,93$ .

Так как  $F_{\text{ФАКТ}} > F_{\text{ТЕОР}}$  ( $7,33 > 6,93$ ) взаимосвязь между урожаем средней массой луковицы и гнездойостью значима на 1 %-ном уровне значимости. Судя по коэффициенту детерминации ( $R^2 = 0,74^2 = 0,55$ ) вариация урожайности лука на 55 % связана с действием изучаемых факторов и 45 % вариации не может быть объяснено влиянием этих переменных.

Математическое уравнение для прямолинейной зависимости между тремя переменными называется множественным линейным уравнением плоскости регрессии. В общем случае оно имеет вид

$$y = a + b_1x + b_2z, \quad (6.32)$$

где  $y$  – зависимая переменная;

$x$  и  $z$  – независимые переменные;

$a$  – общее начало отсчета;

$b_1$  и  $b_2$  – коэффициенты частной регрессии, указывающие на какую величину указывает  $y$  при увеличении на одну единицу значения  $x$  или  $z$  при постоянном значении другого независимого переменного.

Параметры  $a$ ,  $b_1$ ,  $b_2$  вычисляют методом наименьших квадратов.

Для определения параметров уравнения используют соотношения:

$$\left. \begin{aligned} b_1 &= \frac{\sum (z - \bar{z})^2 \cdot \sum (x - \bar{x})(y - \bar{y}) - \sum (x - \bar{x})(z - \bar{z}) \sum (y - \bar{y})(z - \bar{z})}{\sum (x - \bar{x})^2 \sum (z - \bar{z})^2 - [\sum (x - \bar{x})(z - \bar{z})]^2} \\ b_2 &= \frac{\sum (x - \bar{x})^2 \sum (y - \bar{y})(z - \bar{z}) - \sum (x - \bar{x})(z - \bar{z}) \sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2 \sum (z - \bar{z})^2 - [\sum (x - \bar{x})(z - \bar{z})]^2} \\ a &= \bar{y} - b_1\bar{x} - b_2\bar{z} \end{aligned} \right\} \quad (6.33)$$

Изображается графически в виде линии и плоскости.

### Криволинейная корреляция и регрессия

Если связь между изучаемыми явлениями существенно отличается от линейной, что легко установить по точечному графику, то коэффициент корреляции непригоден в качестве меры связи. Он может указать на отсутствие сопряженности там, где налицо сильная криволинейная зависимость.

Показателем, который правильно измеряет степень криволинейной зависимости, является корреляционное отношение, которое обозначается буквой  $\eta$  (эта). Оно измеряет степень корреляции при любой ее форме. Корреляционное отношение при малом числе наблюдений вычисляется по формуле

$$\eta_{yx} = \sqrt{\frac{\sum (y - \bar{y})^2 - \sum (y - \bar{y}_x)^2}{\sum (y - \bar{y})^2}}, \quad (6.34)$$

где  $\sum (y - \bar{y})^2$  - сумма квадратов отклонений индивидуальных значений  $y$  от общей среднеарифметической  $\bar{y}$ .

$\sum (y - \bar{y}_x)^2$  - сумма квадратов отклонений вариантов от частных средних  $\bar{y}_x$ , соответствующих определенным фиксированным значениям независимой переменной  $x$ .

Для вычисления корреляционного отношения, значения независимого признака  $x$  располагают по ранжиру в возрастающем порядке, и разбивают весь ряд наблюдений на 4 – 7 групп с таким расчетом, чтобы в каждой группе по ряду  $x$  было не менее двух наблюдений. Затем определяют общую среднюю  $\bar{y}$ , групповые средние  $\bar{y}_x$ , соответствующие каждой фиксированной группе  $x$  и суммы квадратов отклонений для общего  $\sum (y - \bar{y})^2$  и группового  $\sum (y - \bar{y}_x)^2$  варьирования признака  $y$ .

При большом объеме наблюдений ( $n > 30$ ) обработка проводится в форме корреляционной таблицы. После груп-

пировки и разности данных определяют сумму квадратов отклонений общего варьирования  $\sum f \cdot (\bar{y}_x - \bar{y})^2$  и сумму квадратов отклонений общего варьирования  $\sum f \cdot (y - \bar{y})^2$  и вычисляют корреляционное отношение по формуле

$$\eta_{yx} = \sqrt{\frac{\sum f (y_x - \bar{y})^2}{\sum f (y - \bar{y})^2}}, \quad (6.35)$$

Сумма квадратов отклонений групповых средних  $\bar{y}_x$  от общей средней  $\bar{y}$  (групповое варьирование) характеризует ту часть варьирования признака  $y$ , которая связана с изменчивостью признака  $x$ , а сумма квадратов разностей между каждой данной и общей средней  $\bar{y}$ , то есть  $\sum f \cdot (y - \bar{y})^2$  - характеризует общее варьирование признака  $y$ .

При функциональной зависимости  $y$  от  $x$  корреляционное отношение равно единице. Если оно равно нулю, то корреляция отсутствует. При промежуточных значениях  $x$

$$0 < \eta_x < 1. \quad (6.36)$$

Величина  $\eta_{yx}^2$  - отношение сумм квадратов группового варьирования к общему имеет самостоятельное значение. Оно показывает ту долю варьирования признака  $y$ , которая обусловлена степенью колебания  $x$  и называется *индексом детерминации*, который определяет процент вариации  $y$  под влиянием  $x$ .

Ошибку и критерии существенности рассчитывают по формулам

$$S_\eta = \sqrt{\frac{1 - \eta^2}{n - 2}}, \quad (6.37)$$

$$t_\eta = \frac{\eta}{S_\eta}. \quad (6.38)$$

### Критерий линейности корреляции

Для определения степени приближения криволинейной зависимости к прямолинейной используется критерий Фишера, определяемый по выражению

$$F = \frac{(\eta^2 - r^2)(n - k)}{(r - \eta^2)(k_x - 2)}, \quad (6.39)$$

где  $\eta^2$  - квадрат корреляционного отношения  $y$  по  $x$ ;

$r^2$  - квадрат коэффициента линейной корреляции;

$n$  - объем выборки;

$k$  - число групп по ряду  $x$ .

Связь можно практически принять за линейную, если  $F_\Phi < F_T$  и определять показатели для прямолинейной корреляции и регрессии. Корреляция нелинейна, если  $F_\Phi > F_T$ .

Криволинейная регрессия – это такая зависимость, когда при одинаковых приращениях независимой переменной  $x$  зависимая переменная  $y$  имеет неодинаковые приращения.

Эмпирические точки корреляционного поля при криволинейной корреляции располагаются около кривых различного типа – парабол, гипербол, логарифмических кривых.

Статистическая обработка экспериментального материала часто приводит к построению уравнений близких к квадратической параболе

$$y = a + b_1x + b_2x^2, \quad (6.40)$$

Уравнение для квадратической параболы можно рассчитать по соотношению:

$$y = \bar{y} + \frac{\sum (x - \bar{x})y}{\sum (x - \bar{x})^2} (x - \bar{x}) + \left[ \frac{\sum (x - \bar{x})^2 y - n \cdot C_y}{\sum (x - \bar{x})^4 - n \cdot C^2} \right] \times [(x - \bar{x})^2 - C], \quad (6.41)$$

где

$$C = \frac{\sum (x - \bar{x})^2}{n}. \quad (6.42)$$

