Международный консорциум «Электронный университет»

Московский государственный университет экономики, статистики и информатики

Евразийский открытый институт

В.С. Мхитарян
М.Ю. Архипова
В.П. Сиротин

Эконометрика

Учебно-методический комплекс

Москва 2008

УДК 519.2
ББК 22.172
М 936

Мхитарян В.С., Архипова М.Ю., Сиротин В.П. **ЭКОНОМЕТРИКА:** Учебно-методический комплекс. – М.: Изд. центр ЕАОИ. 2008. – 144 с.

Рекомендовано Учебно-методическим объединением по образованию в области статистики в качестве учебного пособия для студентов высших учебных заведений, обучающихся по специальности 061700 «Статистика» и другим экономическим специальностям.

ISBN 978-5-374-00053-5

Содержание

Учебное пособие

Введение

В условиях перехода страны к рыночной экономике возрастает интерес и потребность в познании статистических методов анализа и прогнозирования, к количественным оценкам социально-экономических явлений. Как найти связи между переменными, как доказать их значимость и оценить их параметры? На эти вопросы можно ответить с помощью эконометрики, занимающейся применением методов математической статистики в экономическом анализе.

Эконометрика – это дисциплина, объединяющая совокупность теоретических результатов, методов и приемов, экономической теории, экономической статистики и математико-статистического инструментария для количественного решения социально-экономических задач. Курс эконометрики призван научить различным способам выражения связей и закономерностей через эконометрические модели и методы проверки их адекватности, основанные на данных наблюдений. Эконометрический подход характеризует также внимание, которое уделяется в нем вопросу соответствия выбранной модели изучаемому объекту, рассмотрению причин, приводящих к необходимости пересмотра модели на основе более точной системы представлений. Эконометрика занимается, по существу, статистическими выводами, т. е. использованием выборочной информации для получения некоторого представления о свойствах генеральной совокупности.

В данном учебном пособии излагаются основные теоретические положения таких математико-статистических методов, как корреляционный, регрессионный, компонентный и кластерный анализы, а также такие распространенные эконометрические модели, как производственные функции и системы одновременных уравнений.

Значительное внимание в учебном пособии уделяется логическому анализу исходной информации и экономической интерпретации получаемых результатов. Пособие снабжено достаточным количеством экономических примеров и задач для самостоятельного решения.

1. Корреляционный анализ

1.1. Основы корреляционного анализа

Корреляционный анализ, разработанный К.Пирсоном и Дж.Юлом, является одним из методов статистического анализа взаимозависимости нескольких признаков – компонент случайного вектора x. Основная задача корреляционного анализа состоит в оценке степени зависимости между случайными величинами. Степень линейной зависимости между количественными переменными характеризуется с помощью парных, частных и множественных коэффициентов корреляции и детерминации.

Парный коэффициент корреляции характеризует тесноту линейной зависимости между двумя переменными на фоне действия всех остальных показателей, входящих в модель. Частный коэффициент корреляции характеризует тесноту линейной зависимости между двумя переменными при исключении влияния всех остальных показателей, входящих в модель. Данные коэффициенты корреляции изменяются в пределах от -1 до +1, причем чем ближе коэффициент корреляции к ± 1, тем сильнее зависимость между переменными. Если коэффициент корреляции больше 0, то связь положительная, а если меньше нуля – отрицательная.

Множественный коэффициент корреляции характеризует тесноту линейной связи между одной переменной (результативной) и остальными, входящими в модель; изменяется в пределах от 0 до 1. Квадрат множественного коэффициента корреляции называется множественным коэффициентом детерминации. Он характеризует долю дисперсии одной переменной (результативной), обусловленной влиянием всех остальных переменных (аргументов), входящих в модель.

Исходной для анализа является матрица:

$$X=\begin{pmatrix} x_{11}\cdots x_{1j}\cdots x_{1k} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ x_{i1}\cdots x_{ij}\cdots x_{ik} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ x_{n1}\cdots x_{nj}\cdots x_{nk} \end{pmatrix}$$

размерности (n x k), i-я строка которой характеризует i-е наблюдение (объект) по всем k-м показателям (j=1, 2, ..., k).

В корреляционном анализе матрицу X рассматривают как выборку объема n из k-мерной генеральной совокупности, подчиняющейся k-мерному нормальному закону распределения.

По выборке определяют оценки параметров генеральной совокупности, а именно: вектор средних ($\bar{x}$), вектор средне-квадратических отклонений s и корреляционную матрицу (R) порядка ($k\times k$):

$$\bar{x}=\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \bar{x}_k \end{pmatrix},\quad s=\begin{pmatrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ \cdot \\ s_k \end{pmatrix},\quad R=\begin{pmatrix} 1 & r_{12} & . & . & r_{1k} \\ r_{21} & 1 & . & . & r_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k1} & r_{k2} & . & . & 1 \end{pmatrix}.$$

Матрица R является симметричной ($r_{jl} = r_{lj}$) и положительно определенной:

$$\bar{x} = \frac{1}{n}\sum x_{ij}, \quad s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}, \tag{1.1}$$

$$r_{jl} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)\left(x_{il} - \bar{x}_l\right)}{s_j s_l}, \tag{1.2}$$

где x_{ij} – значение *i*-го наблюдения *j*-го фактора; r_{il} – выборочный парный коэффициент корреляции, характеризует тесноту линейной связи между показателями x_j и x_l. При этом r_{jl} является оценкой генерального парного коэффициента корреляции.

Кроме того, находятся точечные оценки частных и множественных коэффициентов корреляции любого порядка. Например, частный коэффициент корреляции (*k-2*)-го порядка между факторами X_1 и X_2 равен:

$$r_{12/3,4\ldots,k} = -\frac{R_{12}}{\sqrt{R_{11}R_{22}}}, \tag{1.3}$$

где R_{jl} – алгебраическое дополнение элемента r_{jl} корреляционной матрицы R. При этом $R_{jl} = (-1)^{j+l} \times M_{jl}$, где M_{jl} – минор, определитель матрицы, получаемой из матрицы R путем вычеркивания *j*-й строки и *l*-го столбца.

Множественный коэффициент корреляции (*k-1*)-го порядка фактора (результативного признака) X_1 определяется по формуле:

$$r_{1/2,3,\ldots,k} = r_1 = \sqrt{1 - \frac{|R|}{R_{11}}}, \tag{1.4}$$

где $|R|$ – определитель матрицы R.

1.2. Двумерный нормальный закон распределения

Рассмотрим генеральную совокупность с двумя признаками x и y, совместное распределение которых задано плотностью двумерного нормального закона

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left\{Q_2(x,y)\right\}, \tag{1.5}$$

где $Q_2(x,y) = \frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x-\mu_x}{\sigma_x}\cdot\frac{y-\mu_y}{\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]$, определяемого пятью параметрами:

$$M(x) = \mu_x, \; D(x) = \sigma_x^2,$$

$$M(y) = \mu_y, \; D(y) = \sigma_y^2,$$

$$M\left[\frac{x-\mu_x}{\sigma_x}\cdot\frac{y-\mu_y}{\sigma_y}\right] = \rho,$$

$$\rho^2 \neq 1.$$

Имея эти параметры, можно получить уравнения линий регрессии, показывающих изменение условных математических ожиданий в зависимости от изменения соответствующих значений случайных аргументов:

$M(y/x) - My = \beta_{yx}(x - Mx)$ – линейное уравнение регрессии y на x;

$M(y/x) - Mx = \beta_{xy}(y - My)$ – линейное уравнение регрессии x на y;

$\beta_{yx} = \rho \dfrac{\sigma_y}{\sigma_x}$ – коэффициент регрессии y на x;

$\beta_{xy} = \rho \dfrac{\sigma_x}{\sigma_y}$ – коэффициент регрессии x на y.

Полезно вспомнить, что квадрат коэффициента корреляции ρ^2, т.е. *коэффициент детерминации, в рассматриваемой модели указывает долю дисперсии одной случайной величины, обусловленную вариацией другой. Коэффициент регрессии* β_{yx} *показывает, на сколько единиц своего измерения увеличится* ($\beta>0$) *или уменьшится* ($\beta<0$) *в среднем y (M(y/x)), если x увеличить на единицу своего измерения.*

Задача двумерного корреляционного анализа состоит, прежде всего, в оценке пяти параметров, определяющих генеральную совокупность.

В качестве точечных оценок неизвестных начальных моментов первого и второго порядка генеральной совокупности берутся соответствующие выборочные моменты.

Точечные же оценки неизвестных других параметров получают с помощью формул, аналогичных формулам вычисления самих параметров через генеральные начальные моменты. Таким образом, будем иметь:

$\bar{x}$ – оценка для μ_x,
$\bar{y}$ – оценка для μ_y,
$\overline{x^2}$ – оценка для $M(x^2)$,
$\overline{y^2}$ – оценка для $M(y^2)$,
$\overline{xy}$ – оценка для $M(xy)$.

Откуда

$s_x^2 = \overline{x^2} - (\bar{x})^2$ – оценка для σ_x^2,

$s_y^2 = \overline{y^2} - (\bar{y})^2$ – оценка для σ_y^2,

$r = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x S_y}$ – оценка для ρ.

Оценки генеральных коэффициентов регрессии β_{yx} и β_{xy} получаются соответственно по формулам:

$$b_{yx} = r\frac{S_y}{S_x}, \qquad b_{xy} = r\frac{s_x}{s_y},$$

откуда оценки уравнений регрессии имеют вид:

$$\overline{y/x} - \bar{y} = b_{yx}(x - \bar{x}), \qquad \overline{x/y} - \bar{x} = b_{xy}(y - \bar{y}).$$

При этом $\overline{y/x}$ и $\overline{x/y}$ – обозначения оценок для условных математических ожиданий *M(y/x)* и *M(x/y)* генеральной совокупности.

Следует отметить, что вышеприведенные точечные оценки являются состоятельными, а $\bar{x}$ и $\bar{y}$ несмещенными и эффективными. Кроме того, распределение выборочных средних ($\bar{x}$, $\bar{y}$) не зависит от распределения (s_x^2, s_y^2, r). Наконец, выборочный коэффициент корреляции r по абсолютной величине не превосходит единицы.

1.3. Проверка значимости параметров связи

В двумерной модели параметрами связи являются коэффициент корреляции ρ (или его квадрат, называемый коэффициентом детерминации) и коэффициенты регрессии β_{yx} и β_{xy}.

Заметим, что в двумерной модели достаточно проверить значимость только коэффициента корреляции. Если коэффициент корреляции незначим, то признаки x и y считаются независимыми в генеральной совокупности.

Значимость частных и парных коэффициентов корреляции, т. е. гипотеза H_0: $\rho = 0$, проверяется по t-критерию Стъюдента. Наблюдаемое значение критерия находится по формуле:

$$t_{\text{набл}} = \frac{r}{\sqrt{1-r^2}}\sqrt{n-l-2} \qquad (1.6)$$

где r – соответственно оценка частного или парного коэффициента корреляции; l – порядок частного коэффициента корреляции, т. е. число фиксируемых факторов. Для парного коэффициента корреляции l=0.

Напомним, что проверяемый коэффициент корреляции считается значимым, т. е. гипотеза H_0: ρ=0 отвергается с вероятностью ошибки α, если $t_{\text{набл}}$ по модулю будет больше, чем $t_{\text{кр}}$, определяемое по таблицам t-распределение (см. приложения) для заданного α и $\nu = n - l - 2$.

Значимость коэффициентов корреляции можно также проверить с помощью таблиц Фишера-Иейтса (табл. 9 приложения).

1.4. Интервальные оценки параметров связи

Для значимых параметров связи имеет смысл найти интервальные оценки.

При определении с надежностью γ доверительного интервала для значимого парного или частного коэффициентов корреляции ρ используют Z-преобразование Фишера и предварительно устанавливают интервальную оценку для Z

$$Z' - t_\gamma \sqrt{\frac{1}{n-l-3}} \le Z \le Z' + t_\gamma\sqrt{\frac{1}{n-l-3}}, \qquad (1.7)$$

где t_γ вычисляют по таблице интегральной функции Лапласа (табл. 1 приложения) из условия

$$\Phi(t_\gamma)=\gamma,$$

$$Z(r) = \frac{1}{2}\ln\frac{1+r}{1-r}.$$

Значение Z' определяют по таблице Z – преобразования (табл. 6 приложения) по найденному значению r. Функция нечетная, т. е.

$$Z'(-r) = -Z'(r).$$

Обратный переход от Z к ρ осуществляют также по таблице Z-преобразования, после использования которой получают интервальную оценку для ρ с надежностью γ :

$$r_{\min} \le \rho \le r_{\max} .$$

Таким образом, с вероятностью γ гарантируется, что генеральный коэффициент корреляции ρ будет находиться в интервале ($r_{\min}$, $r_{\max}$).

1.5. Проверка значимости множественного коэффициента корреляции

Значимость множественного коэффициента корреляции (или его квадрата – коэффициента детерминации) проверяется по F-критерию.

Например, для множественного коэффициента корреляции проверка значимости сводится к проверке гипотезы, что генеральный множественный коэффициент корреляции равен нулю, т. е. H_0: $\rho_{1/2,\ldots,k}=0$. Наблюдаемое значение статистики находится по формуле:

$$F_{\text{набл}} = \frac{\frac{1}{k-1} r^2_{1/2,\ldots,k}}{\frac{1}{n-k}\left(1-r^2_{1/2,\ldots,k}\right)} . \qquad (1.8)$$

Множественный коэффициент корреляции считается значимым, т. е. имеет место линейная статистическая зависимость между X_1 и остальными факторами $X_2,\ldots,X_K$, если: $F_{\text{набл}} > F_{\text{кр}}$ (α, $k-1$, $n-k$), где $F_{\text{кр}}$ определяется по таблице F – распределения для заданных α, $\nu_1 = k-1$, $\nu_2 = n-k$.

1.6. Задачи, решаемые при помощи статистики Фишера

Кроме нахождения интервальной оценки для ρ с помощью преобразования

$$Z_r = \frac{1}{2}\ln\frac{1+r}{1-r}$$

можно решить следующие задачи.

1. Проверить, согласуется ли выборочный коэффициент корреляции r с предполагаемым значением генерального коэффициента корреляции ρ_o. Для этого, взяв уровень значимости α, проверяем, попадает ли абсолютная величина разности $|Z_r - Z_{\rho_o}|$ в интервал $[0, t_{1-\alpha}/\sqrt{n-3}]$. Если попадает, то гипотеза H_o:$\rho=\rho_o$ не отвергается. В противном случае отвергается с вероятностью ошибки α.

2. Проверить гипотезу об однородности коэффициентов корреляции. Пусть $r_1, r_2, \ldots, r_k$ – коэффициенты корреляции, полученные из k нормально распределенных совокупностей по выборкам с объемами $n_1, n_2, \ldots, n_k$. Проверяется гипотеза

$$H_o: \rho_1=\rho_2= \ldots =\rho_k=\rho.$$

Статистика

$$\sum_{i=1}^{k} \frac{(z_{r_i} - z_\rho)^2}{1/(n_i-3)}$$

имеет распределение χ^2 с k степенями свободы. Если заменить z_ρ на среднее арифметическое

$$\bar{z}_r = \frac{\Sigma z_i n_i}{\Sigma n_i},$$

то получим, что

$$\sum_{i=1}^{k} \frac{(z_{r_i} - \bar{z}_r)^2}{1/(n_i - 3)}$$

распределена по закону χ^2 с $\nu=k-1$ степенями свободы.

Если теперь для заданных α и $\nu=k-1$

$$\chi^2_{\text{набл.}} < \sum_{i=1}^{k} \frac{(z_{r_i} - \bar{z}_r)^2}{1/(n_i - 3)},$$

то гипотеза однородности отвергается с вероятностью ошибки α. В противном случае гипотеза H_o не отвергается.

В случае принятия гипотезы однородности предпочтительной точечной оценкой ρ является значение r, полученное обратным преобразованием из z_r.

1.7. Тренировочный пример

Деятельность $n = 8$ карьеров характеризуется себестоимостью 1т. песка (X_1), сменной добычей песка (X_2) и фондоотдачей (X_3). Значения показателей представлены в таблице.

X_1 (тыс.руб)	30	20	40	35	45	25	50	30
X_2 (тыс.руб)	20	30	50	70	80	20	90	25
X_3	20	25	20	15	10	30	10	20

Требуется:

1. Оценить параметры генеральной совокупности, которая предполагается нормально распределенной.

2. При $\alpha = 0.05$ проверить значимость частных коэффициентов корреляции $\rho_{12/3}$, $\rho_{13/2}$ и $\rho_{23/1}$ и при $\gamma = 0.95$ построить интервальную оценку для $\rho_{13/2}$.

3. Найти точечную оценку множественного коэффициента корреляции $\rho_{1/23}$ и при $\alpha = 0.05$ проверить его значимость.

Решение:

1. Найдем значения средних арифметических ($\bar{x}_j$) и средних квадратических отклонений (S_j), где $j = 1, 2, 3$, а также парных коэффициентов корреляции r_{12}, r_{13} и r_{23} по формулам:

$$\bar{x}_1 = \frac{30+20+40+35+45+25+50+30}{8} = 34.375 \text{ тыс. руб.}$$

$\bar{x}_2 = 48.125$

$\bar{x}_3 = 18.75$

$S_1 = 9,49$

$S_2 = 26,68$.

$\mathrm{S}_3 = 6,48$

$$r_{12} = \frac{\overline{x_1 x_2} - \overline{x_1}\,\overline{\mathrm{x}_2}}{\mathrm{S}_1\, S_2} = \frac{1875 - 34.375 \times 48,125}{9.49 \times 26,68} = \frac{220.70}{9.49 \times 26.68} = 0.871,$$

где $\overline{x_1 x_2} = \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{i2} = \frac{1}{8}\left(30\times 20 + 20\times 30 + 40\times 50 + \ldots + 30\times 25\right) = 1875$.

В результате расчетов получим:

$$\overline{X} = \begin{pmatrix} 34.38 \\ 48.12 \\ 18.75 \end{pmatrix};\quad \mathrm{S}=\begin{pmatrix} 9.49 \\ 26.68 \\ 6.48 \end{pmatrix};\quad \mathrm{R}=\begin{pmatrix} 1 & 0.871 & -0.874 \\ 0.871 & 1 & -0.879 \\ -0.874 & -0.879 & 1 \end{pmatrix}.$$

2. Предварительно найдем точечные оценки частных коэффициентов корреляции из выражения

$$\mathrm{r}_{12/3} = -\frac{R_{12}}{\sqrt{R_{11}\times R_{22}}},$$

где R_{12} – алгебраическое дополнение элемента r_{12} корреляционной матрицы R, а R_{11} и R_{22} алгебраические дополнения 1-го и 2-го диагонального элемента этой матрицы

$$R_{12} = (-1)^3 \times \begin{vmatrix} 0.871 & -0.879 \\ -0.874 & 1 \end{vmatrix} = -0.103$$

$$R_{11} = (-1)^2 \times \begin{vmatrix} 1 & -0.879 \\ -0.879 & 1 \end{vmatrix} = 0.227$$

$$R_{22} = (-1)^4 \times \begin{vmatrix} 1 & -0.874 \\ -0.874 & 1 \end{vmatrix} = 0.236$$

$$r_{12/3} = \frac{0.103}{\sqrt{0.227\times 0.236}} = 0.445$$

Аналогично находим: $\mathrm{r}_{13/2}$=-0.462 и $\mathrm{r}_{23/1}$ =-0.494.

Для проверки значимости частных коэффициентов корреляции найдем $\mathrm{r}_{кр.}$ ($\alpha = 0.05$, $\nu = n{-}l{-}2 = 5$) = 0.754, где l – порядок коэффициента корреляции (число фиксированных признаков). В нашем примере $l = 1$.

Так как $|r|<\mathrm{r}_{кр.}$=0.754, то гипотезы H_0: ρ=0 не отвергаются, т. е. предположение о равенстве его нулю не противоречит наблюдениям, но $n = 8$ мало.

Определим интервальную оценку для $\rho_{13/2}$ при γ =0.95. Для этого используем Z-преобразование Фишера и предварительно найдем интервальную оценку для Z из условия:

$$Z \in \left[Z' \pm t_{\gamma}\sqrt{\frac{1}{n-l-3}} \right].$$

По таблице Z-преобразования Фишера для $\mathrm{r}_{13/2}$ = -0.462, учитывая, что Z'(-r) = -Z'(r), будем иметь Z'(-0,462) = -0.497. По таблице нормального закона из условия Φ(t) = 0.95 найдем t = 1.96.

Тогда

$$Z \in \left[-0.497 \pm 1.96\sqrt{\frac{1}{8-4}}\right],$$

откуда $Z \in [-1.477, 0.483]$.

По таблице Z-преобразования для Z_{min}= -1,477 и Z_{max}= 0.483 найдем интервальную оценку для $\rho_{13/2}$:

$\rho_{13/2} \in [-0.9,\ 0.45]$.

Полученная интервальная оценка подтверждает вывод о незначимости частного коэффициента корреляции $\rho_{13/2}$, т. к. ноль находится внутри доверительного интервала.

3. Найдем точечную оценку множественного коэффициента корреляции $\rho_{1/23}$ и при α=0.05 проверим его значимость.

Точечная оценка определяется по формуле:

$r_{1/23} = \sqrt{1 - \frac{|R|}{R_{11}}}$, где $|R|$ – определитель корреляционной матрицы.

$|R|$= 1+0.871(-0.879)(-0.874)+0.871(-0.879)(-0.874) – $(0.874)^2$ - 0.871^2 - $(-0.879)^2$ = =0.043.

$$r_{1/23} = \sqrt{1 - \frac{0.043}{0.227}} = 0.90.$$

Проверим гипотезу H_0: $\rho_{1/23} = 0$

$$F_{набл} = \frac{\frac{1}{k-1} r_{1/23}^2}{\frac{1}{n-k}\left(1 - r_{1/23}^2\right)} = \frac{\frac{1}{2}0.81}{\frac{1}{5}0.19} = 10.66,$$

где l=2.

Критическое значение по таблице F-распределения

$F_{кр}$ (α=0.05, ν_1 =2, ν_2 =5) = 5.79

Т. к. $F_{набл} > F_{кр}$, то гипотеза H_0 отвергается, т. е. множественный коэффициент корреляции не равен нулю ($\rho_{1/23} \neq 0$).

1.8. Задание для самостоятельного решения

По данным n=10 машиностроительных предприятий методами корреляционного анализа исследуется взаимосвязь между следующими показателями: x_1 – рентабельность (%); x_2 – премии и вознаграждения на одного работника (млн.руб.); x_3 – фондоотдача.

N п/п	X_1	X_2	X_3
1	13,26	1,23	1,45
2	10,16	1,04	1,30
3	13,72	1,80	1,37
4	12,82	0,43	1,65
5	10,63	0,88	1,91
6	9,12	0,57	1,68
7	25,83	1,72	1,94
8	23,39	1,70	1,89
9	14,68	0,84	1,94
10	10,05	0,60	2,06

Требуется:

а) рассчитать вектора средних и среднеквадратических отклонений, матрицу парных коэффициентов корреляции $\left(\overline{X},\quad S,\quad R\right)$;

б) проверить при α=0,05 значимость парного коэффициента корреляции ρ_{12} и найти его интервальную оценку с доверительной вероятностью γ=0,95;

в) по корреляционной матрице R рассчитать частный коэффициент корреляции $r_{12/3}$;

г) проверить при α=0,05 значимость частного коэффициента корреляции $\rho_{12/3}$ и определить его интервальную оценку при γ=0,95;

д) по корреляционной матрице R вычислить оценку множественного коэффициента корреляции $r_{1(2,3)}$ и при α=0,05 проверить гипотезу H_0: $r_{1(2,3)}=0$.

Задание выполняется по вариантам. Каждый должен вычеркнуть объект №, соответствующий последней цифре зачетной книжки. Так, например, если последняя цифра номера вашей зачетной книжки равна 2, то вы вычеркиваете второй объект.

2. Регрессионный анализ

2.1. Основы регрессионного анализа

Регрессионный анализ – это статистический метод исследования зависимости случайной величины Y от переменных X_j ($j = 1, 2, ..., k$), рассматриваемых в регрессионном анализе как неслучайные величины независимо от истинного закона распределения X_j.

Обычно предполагается, что случайная величина Y имеет нормальный закон распределения с условным математическим ожиданием $\widetilde{Y} = (x_1,...,x_k)$, являющимся функцией от аргументов x_j, и с постоянной, не зависящей от аргументов дисперсией σ^2.

Для проведения регрессионного анализа из ($k+1$) -мерной генеральной совокупности ($Y,X_1,X_2,...,X_j,...,X_k$) берется выборка объемом n и каждое i-ое наблюдение (объект) характеризуется значениями переменных (y_i , x_{i1}, x_{i2},..., x_{ij} ,..., x_{ik}), где x_{ij} – значение j-ой переменной для i-го наблюдения ($i=1,2,...,n$), y_i – значение результативного признака для i-го наблюдения.

Наиболее часто используемая множественная линейная модель регрессионного анализа имеет вид:

$$y = \beta_0 +\beta_1 x_{i1} +...+\beta_j x_{ij}+...+\beta_k x_{ik}+\varepsilon_I, \qquad (2.1)$$

где ε_i – случайные ошибки наблюдения, независимые между собой, имеют нулевую среднюю и дисперсию σ^2.

Отметим, что модель (2.1) справедлива для всех $i = 1,2,.., n$, линейна относительно неизвестных параметров $\beta_0, \beta_1,..., \beta_j,..., \beta_k$ и аргументов.

Как следует из (2.1) *коэффициент регрессии β_j показывает, на какую величину в среднем изменится результативный признак Y, если переменную X_j увеличить на единицу измерения, т. е. является нормативным коэффициентом.*

В матричной форме регрессионная модель имеет вид:

$$Y = X\beta + \varepsilon \qquad (2.2)$$

где Y – случайный вектор – столбец размерности (n x 1) наблюдаемых значений результативного признака ($y_1, y_2,..., y_n$); X – матрица размерности [n x (k+1)] наблюдаемых значений аргументов. Элемент матрицы x_{ij} рассматривается как неслучайная величина (i =1,2,...,n; j=0,1,2,...k); β – вектор – столбец размерности [(k+1) x 1] неизвестных, подлежащих оценке параметров (коэффициентов регрессии) модели; ε – случайный вектор – столбец размерности (n x 1) ошибок наблюдений (остатков). Компоненты вектора ε_i независимы между собой, имеют нормальный закон распределения с нулевым математическим ожиданием ($M\varepsilon_I = 0$) и неизвестной дисперсией σ^2 ($D\varepsilon_I = \sigma^2$).

На практике рекомендуется, чтобы n превышало k не менее, чем в три раза.

В модели (2.2)

$$X=\begin{pmatrix} 1 & x_{11} & . & . & x_{1k} \\ \hline 1 & x_{i1} & . & . & x_{ik} \\ \hline 1 & x_{n1} & . & . & x_{nk} \end{pmatrix};\ Y=\begin{pmatrix} y_1 \\ \cdots \\ y_i \\ \cdots \\ y_n \end{pmatrix};\ \beta=\begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_j \\ \cdots \\ \beta_k \end{pmatrix}.$$

Единицы в первом столбце матрицы призваны обеспечить наличие свободного члена в модели (2.1). Здесь предполагается, что существует переменная x_0, которая во всех наблюдениях принимает значения, равные 1.

Основная задача регрессионного анализа заключается в нахождении по выборке объемом n оценки неизвестных коэффициентов регрессии β_0, β_1,..., β_k модели (2.1) или вектора β в (2.2).

Так как в регрессионном анализе x_j рассматриваются как неслучайные величины, а $M\varepsilon_i = 0$, то согласно (2.1) уравнение регрессии имеет вид:

$$\tilde{y}_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_j x_{ij} + \ldots + \beta_k x_{ik}, \qquad (2.3)$$

для всех i= 1,2,...,n, или в матричной форме:

$$\tilde{Y} = X\beta, \qquad (2.4)$$

где $\tilde{Y}$ – вектор-столбец с элементами $\tilde{y}_1, \ldots, \tilde{y}_i, \ldots, \tilde{y}_n$.

Для оценки вектора β наиболее часто используют метод наименьших квадратов (МНК), согласно которому в качестве оценки принимают вектор b, который минимизирует сумму квадратов отклонения наблюдаемых значений y_i от модельных значений $\tilde{y}_i$, т. е. квадратичную форму:

$$Q=(Y - X\beta)^T (Y - X\beta) = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \Rightarrow \min_{\beta_0, \beta_1, \ldots, \beta_k}.$$

Наблюдаемые и модельные значения показаны на рис. 2.1.

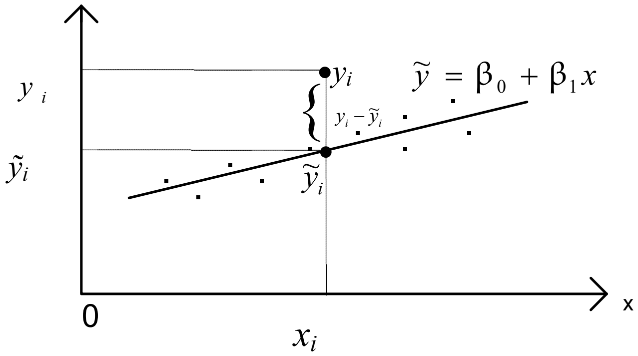


Рис. 2.1. **Наблюдаемые и модельные значения результативной величины y**

Дифференцируя с учетом (2.4) и (2.3) квадратичную форму Q по $\beta_0, \beta_1, ..., \beta_к$ и приравнивая производные нулю, получим систему нормальных уравнений:

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_k} = 0 \end{cases} \qquad \text{для всех } j = 0,1,\ldots, k,$$

решая которую и получаем вектор оценок b, где $b=(b_0\ b_1...b_k)^T$.

Согласно методу наименьших квадратов, вектор оценок коэффициентов регрессии получается по формуле:

$$b = (X^T X)^{-1} X^T Y\,, \tag{2.5}$$

$$b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_j \\ \vdots \\ b_k \end{pmatrix},$$

где X^T – транспортированная матрица X;

$(X^TX)^{-1}$ – матрица, обратная матрице X^TX.

Зная вектор оценок коэффициентов регрессии b, найдем оценку $\hat{y}_i$ уравнения регрессии:

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + ... + b_k\,. \tag{2.6}$$

Или в матричном виде: $y = X\beta$,

где $\hat{y} = (\hat{y}_1, \hat{y}_2, ..., \hat{y}_n)^T$.

Оценка ковариационной матрицы коэффициентов регрессии вектора b определяется из выражения:

$$S(b) = \hat{S}^2 (X^T X)^{-1}\,, \tag{2.7}$$

где $$\hat{S}^2 = \frac{1}{n-k-1}(Y - Xb)^T (Y - Xb)\,. \tag{2.8}$$

Учитывая, что на главной диагонали ковариационной матрицы находятся дисперсии коэффициентов регрессии, имеем:

$$\hat{S}^2{}_{b_{(j-1)}} = \hat{S}^2 \left[(X^T X)^{-1}\right]_{jj} \quad \text{для } j=1,2,...,k,k+1\,. \tag{2.9}$$

2.2. Проверка значимости уравнения регрессии

Значимость уравнения регрессии, т. е. гипотеза H_0: $\beta=0$ $(\beta_0=\beta_1=...=\beta_k=0)$, проверяется по F-критерию, наблюдаемое значение которого определяется по формуле:

$$F_{\text{набл.}} = \frac{Q_R/(k+1)}{Q_{\text{ост.}}/(n-k-1)}, \tag{2.10}$$

где $$Q_R = (Xb)^T (Xb), \;\; Q_{\text{ост}} = (Y - Xb)^T (Y - Xb) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\,. \tag{2.11}$$

По таблице F-распределения для заданных α, $\nu_1=к+1$, $\nu_2=n-к-1$ находят $F_{кр}$.

Гипотеза H_0 отклоняется с вероятностью α, если $F_{набл}>F_{кр}$. Из этого следует, что уравнение является значимым, т. е. хотя бы один из коэффициентов регрессии отличен от нуля.

Для проверки значимости отдельных коэффициентов регрессии, т. е. гипотез H_0: $\beta_j=0$, где $j=1,2,...k$, используют t-критерий и вычисляют: $t_{набл}(b_j)=b_j/\hat{S}_{bj}$. По таблице t-распределения для заданного α и ν= n-k-1, находят $t_{кр}$.

Гипотеза H_0 отвергается с вероятностью α, если $|t_{набл}|>t_{кр}$. Из этого следует, что соответствующий коэффициент регрессии β_j значим, т. е. $\beta_j\neq 0$. В противном случае коэффициент регрессии незначим и соответствующая переменная в модель не включается. Тогда реализуется алгоритм пошагового регрессионного анализа, состоящий в том, что исключается одна из незначимых переменных, которой соответствует минимальное по абсолютной величине значение $t_{набл}$. После этого вновь проводят регрессионный анализ с числом факторов, уменьшенным на единицу. Алгоритм заканчивается получением уравнения регрессии со значимыми коэффициентами.

Существуют и другие алгоритмы пошагового регрессионного анализа, например, с последовательным включением факторов.

2.3. Интервальное оценивание коэффициентов регрессии

Наряду с точечными оценками b_j генеральных коэффициентов регрессии β_j регрессионный анализ позволяет получать и интервальные оценки последних с доверительной вероятностью γ.

Интервальная оценка с доверительной вероятностью γ для параметра β_j имеет вид:

$$b_j - t_\alpha \hat{S}_{b_j} \le \beta_j \le b_j + t_\alpha \hat{S}_{b_j}, \qquad (2.12)$$

где t_α находят по таблице t-распределения при вероятности $\alpha=1-\gamma$ и числе степеней свободы $\nu=n-\kappa-1$.

Интервальная оценка для уравнения регрессий $\tilde{y}$ в точке, определяемой вектором начальных условий $X^0=(1, X_1^0, X_2^0,...,X_k^0)^T$, равна:

$$\tilde{y} \in \left[(X^0)^T b \pm t_\alpha \hat{S}\sqrt{(X^0)^T (X^T X)^{-1} X^0} \right]. \qquad (2.13)$$

А точечная оценка $\tilde{y}(x^o)=(x^o)^T в$.

Интервал оценки предсказания $_1$ с доверительной вероятностью γ определяется как:

$$\tilde{y} \in \left[(X^0)^T b \pm t_\alpha \hat{S}\sqrt{(X^0)^T (X^T X)^{-1} X^0 + 1} \right], \qquad (2.14)$$

где t_α определяется по таблице t-распределения при $\alpha=1-\gamma$ и $\nu=n-\kappa-1$.

По мере удаления вектора начальных условий x^0 от вектора средних $\bar{x}$ ширина доверительного интервала при заданном γ будет увеличиваться (рис. 2.2.), где $\bar{x}=(1,\bar{x}_1,...,\bar{x}_k)$.

$$\delta = t_a \hat{S} \sqrt{(X^0)^T (X^T X)^{-1} X^0}$$

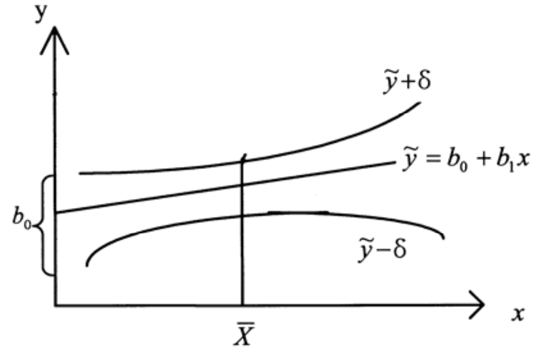


Рис. 2.2. **Точечная $\tilde{y}$ и интервальная оценки $[y-\delta<\tilde{y}<y+\delta]$ уравнения регрессии**

$$\tilde{y} = \beta_0 + \beta_1 x$$

2.4. Мультиколлинеарность

Одним из основных препятствий эффективного применения множественного регрессионного анализа является мультиколлинеарность. Она связана с линейной зависимостью между аргументами $x_1, x_2, ..., x_k$. В результате мультиколлинеарности матрица парных коэффициентов корреляции и матрица ($X^T X$) становятся слабообусловленными, то есть их определители близки к нулю.

Это вызывает неустойчивость оценок коэффициентов регрессии (2.5), большие дисперсии $S^2{}_{b_j}$, оценок этих коэффициентов (2.7), т. к. в их выражения входит обратная матрица $(X^T X)^{-1}$, получение которой связано с делением на определитель матрицы $\left|X^T X\right|$. Отсюда следуют заниженные значения $t(b_j)$. Кроме того, мультиколлинеарность приводит к завышению значения множественного коэффициента корреляции.

На практике о наличии мультиколлинеарности обычно судят по матрице парных коэффициентов корреляции. Если один из элементов матрицы R больше 0.8, т. е. $|r_{jl}| > 0{,}8$, то считают, что имеет место мультиколлинеарность и в уравнение регрессии следует включать только один из показателей x_j или x_l.

Чтобы избавиться от этого негативного явления, обычно используют алгоритм пошагового регрессионного анализа или строят уравнение регрессии на главных компонентах.

2.5. Пример построения регрессионного уравнения

По данным n=20 сельскохозяйственных районов требуется построить регрессионную модель урожайности на основе следующих показателей:

y – урожайность зерновых культур (ц/га);

x_1 – число колесных тракторов (приведенной мощности) на 100 га;

x_2 – число зерноуборочных комбайнов на 100 га;

x_3 – число орудий поверхностной обработки почвы на 100 га;

x_4 – количество удобрений, расходуемых на гектар;

x_5 – количество химических средств оздоровления растений, расходуемых на гектар.

Исходные данные для анализа приведены в таблице.

Таблица

Исходные данные для анализа

Номер наблюдения	y	x_1	x_2	x_3	x_4	x_5
1	9.70	1.59	0.26	2.05	0.32	0.14
2	8.40	0.34	0.28	0.46	0.59	0.66
3	9.00	2.53	0.31	2.46	0.30	0.31
4	9.90	4.63	0.40	6.44	0.43	0.59
5	9.60	2.16	0.26	2.16	0.39	0.16
6	8.60	2.16	0.30	2.69	0.32	0.17
7	12.50	0.68	0.29	0.73	0.42	0.23
8	7.60	0.35	0.26	0.42	0.21	0.08
9	6.90	0.52	0.24	0.49	0.20	0.08
10	13.50	3.42	0.31	3.02	1.37	0.73
11	9.70	1.78	0.30	3.19	0.73	0.17
12	10.70	2.40	0.32	3.30	0.25	0.14
13	12.10	9.36	0.40	11.51	0.39	0.38
14	9.70	1.72	0.28	2.26	0.82	0.17
15	7.00	0.59	0.29	0.60	0.13	0.35
16	7.20	0.28	0.26	0.30	0.09	0.15
17	8.20	1.64	0.29	1.44	0.20	0.08
18	8.40	0.09	0.22	0.05	0.43	0.20
19	13.10	0.08	0.25	0.03	0.73	0.20
20	8.70	1.36	0.26	1.17	0.99	0.42

Решение. Предварительно, с целью анализа взаимосвязи показателей построена таблица парных коэффициентов корреляции R.

y	x_1	x_2	x_3	x_4	x_5
1.00	0.43	0.37	0.40	0.58	0.33
0.43	1.00	0.85	0.98	0.11	0.34
0.37	0.85	1.00	0.88	0.03	0.46
0.40	0.98	0.88	1.00	0.03	0.28
0.58	0.11	0.03	0.03	1.00	0.57
0.33	0.34	0.46	0.28	0.57	1.00

Анализ матрицы парных коэффициентов корреляции показывает, что результативный показатель наиболее тесно связан с показателем x_4 – количеству удобрений, расходуемых на гектар (r_{y4}=0.58).

В то же время связь между признаками-аргументами достаточно тесная. Так, существует практически функциональная связь между числом колесных тракторов (x_1) и числом орудий поверхностной обработки почвы (x_3) – r_{13}=0.98.

О наличии мультиколлинеарности свидетельствуют также коэффициенты корреляции r_{12}=0.85 и r_{32}=0.88.

Чтобы продемонстрировать отрицательное влияние мультиколлинеарности, рассмотрим регрессионную модель урожайности, включив в нее все исходные показатели:

$$\hat{Y} = 3.515 - 0.006x_1 + 15.542x_2 + 110x_3 + 4.475x_4 - 2.932x_5. \qquad (2.15)$$
$$(-0.01) \quad (0.72) \quad (0.13) \quad (2.90) \quad (-0.95)$$

В скобках указаны $t_{набл}(b_j)$, расчетные значения t – критерия для проверки гипотезы о значимости коэффициента регрессии H_0: $\beta_j=0$, j=1, 2, 3, 4, 5. Критическое значение $t_{кр}=1.76$ найдено по таблице t – распределения при уровне значимости $\alpha=0.1$ и числе степеней свободы $\nu=14$. Из уравнения следует, что статистически значимым является коэффициент регрессии только при x_4, так как $|t_4|=2.90>t_{кр}=1.76$. Не поддаются экономической интерпретации отрицательные знаки коэффициентов регрессии при x_1 и x_5, из чего следует, что повышение насыщенности сельского хозяйства колесными тракторами (x_1) и средствами оздоровления растений (x_5) отрицательно сказывается на урожайности. Таким образом, полученное уравнение регрессии не приемлемо.

После реализации алгоритма пошагового регрессионного анализа с исключением переменных и учетом того, что в уравнение должна войти только одна из трех тесно связанных переменных (x_1, x_2 или x_3), получаем окончательное уравнение регрессии:

$$\hat{Y} = 7.342 + 0.345x_1 + 3.294x_4. \qquad (2.16)$$
$$(11.12) \quad (2.09) \quad (3.02)$$

В уравнение (2.16) включен x_1, как определяющий из трех показателей.

Уравнение значимо при $\alpha=0.05$, т.к. $F_{набл}=266>F_{кр}=3.20$, найденного по таблице F-распределения при $\alpha=0.05$; $\nu_1=3$ и $\nu_2=17$. Значимы и все коэффициенты регрессии β_1 и β_4 в уравнении $|t_j|>t_{кр}$ ($\alpha=0.05$; $\nu=17$) = 2.11. Коэффициент регрессии β_1 следует признать значимым ($\beta_1\neq0$) из экономических соображений, при этом $t_1=2.09$ лишь незначительно меньше $t_{кр}=2.11$. При $\alpha=0.1$ $t_{кр}=1.74$ и β_1 статистически значим.

Из уравнения регрессии следует, что увеличение на 1 числа тракторов на 100 га пашни приводит к росту урожайности зерновых в среднем на 0.345 ц/га ($b_1=0.345$).

Коэффициенты эластичности $Э_1=0.068$ и $Э_4=0.161$ показывают, что при увеличении показателей x_1 и x_4 на 1% урожайность зерновых повышается соответственно на 0.068% и 0.161%, ($Э_j=b_j\frac{\bar{x}_j}{\bar{y}}$).

Множественный коэффициент детерминации $r_y^2=0.469$ свидетельствует о том, что только 46.9% вариации урожайности объясняется вошедшими в модель показателями (X_1 и X_4), то есть насыщенностью растениеводства тракторами и удобрениями. Остальная часть вариации обусловлена действием неучтенных факторов (x_2, x_3, x_5, погодных условий и др.). Средняя относительная ошибка аппроксимации $\bar{\delta}=10.5\%$ характеризует адекватность модели, также как и величина остаточной дисперсии $S^2=1.97$.

2.6. Тренировочный пример

По данным годовых отчетов десяти (n=10) машиностроительных предприятий провести регрессионный анализ зависимости производительности труда y (млн. руб. на чел.) от объема производства x (млрд. руб.). Предполагается линейная модель, т.е. $\tilde{y} = \beta_0 + \beta_1 x$.

Таблица

Исходная информация для анализа и результаты расчетов

y_i	x_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$
2,1	3	2,77	-0,67
2,8	4	3,52	-0,72
3,2	5	4,27	-1,07
4,5	5	4,27	0,23
4,8	5	4,27	0,53
4,9	5	4,27	0,63
5,5	6	5,02	0,48
6,5	7	5,77	0,73
12,1	15	11,75	0,35
15,1	20	15,50	-0,4

Решение: Определим вектор оценок b коэффициентов регрессии. Согласно методу наименьших квадратов, вектор b получается из выражения:

$$b = (x^T x)^{-1} x^T y\,.$$

Воспользовавшись правилами умножения матриц, будем иметь

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 4 & 5 & 5 & 5 & 5 & 6 & 7 & 15 & 20 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 5 \\ 1 & 5 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 15 \\ 1 & 20 \end{pmatrix} = \begin{pmatrix} 10 & 75 \\ 75 & 835 \end{pmatrix}.$$

В матрице $(x^T x)$ число 10, лежащее на пересечении 1-й строки и 1-го столбца, получено как сумма произведений элементов 1-й строки матрицы x^T и 1-го столбца матрицы x, а число 75, лежащее на пересечении 1-й строки и 2-го столбца, как сумма произведений элементов 1-й строки матрицы x^T и 2-го столбца матрицы x и т.д.

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 4 & 5 & 5 & 5 & 5 & 6 & 7 & 15 & 20 \end{pmatrix} \begin{pmatrix} 2{,}1 \\ 2{,}8 \\ 3{,}2 \\ 4{,}5 \\ 4{,}8 \\ 4{,}9 \\ 5{,}5 \\ 6{,}5 \\ 12{,}1 \\ 15{,}1 \end{pmatrix} = \begin{pmatrix} 61{,}4 \\ 664{,}5 \end{pmatrix}.$$

Найдем обратную матрицу

$$\left(x^{T}x\right)^{-1}=\frac{1}{10\cdot 835-(75)^{2}}\begin{pmatrix}835 & -75\\ -75 & 10\end{pmatrix}=\begin{pmatrix}0{,}306422 & -0{,}0275229\\ -0{,}0275227 & 0{,}0036697\end{pmatrix},$$

тогда вектор оценок коэффициентов регрессии равен

$$b=\begin{pmatrix}0{,}306422 & -0{,}0275229\\ -0{,}0275299 & 0{,}0036697\end{pmatrix}\cdot\begin{pmatrix}61{,}4\\ 664{,}5\end{pmatrix}=\begin{pmatrix}0{,}5253430\\ 0{,}7486096\end{pmatrix},$$

а оценка уравнения регрессии будет иметь вид

$$\hat{y}=0{,}52534+0{,}74861x.$$

Перейдем к статистическому анализу полученного уравнения регрессии: проверке значимости уравнения и его коэффициентов, исследованию абсолютных $e_i=y_i-\hat{y}_i$ и относительных $\delta_i=\frac{y_i-\hat{y}_i}{y_i}100\%$ ошибок аппроксимации.

Предварительно определим вектор модельных значений результативного показателя $\hat{y}$:

$$\hat{y}=xb=\begin{pmatrix}1 & 3\\ 1 & 4\\ 1 & 5\\ 1 & 5\\ 1 & 5\\ 1 & 5\\ 1 & 6\\ 1 & 7\\ 1 & 15\\ 1 & 20\end{pmatrix}\cdot\begin{pmatrix}0{,}5253430\\ 0{,}7486096\end{pmatrix}=\begin{pmatrix}2{,}77\\ 3{,}52\\ 4{,}27\\ 4{,}27\\ 4{,}27\\ 4{,}27\\ 5{,}02\\ 5{,}77\\ 11{,}75\\ 15{,}50\end{pmatrix}.$$

Тогда

$$Q=(y-\hat{y})^{T}(y-\hat{y})=\sum_{i=1}^{n}(y_i-\hat{y}_i)^{2}=3{,}9847314.$$

Откуда согласно (2.8.) несмещенная оценка остаточной дисперсии равна:

$$\hat{S}^{2}=\frac{1}{8}\cdot 3{,}9847314=0{,}49809176,$$

а оценка среднего квадратического отклонения

$$\hat{S}=\sqrt{\hat{S}^{2}}=0{,}70575616.$$

Проверим на уровне значимости $\alpha=0{,}05$ значимость уравнения регрессии, т.е. гипотезу H_0:$\beta=0$ ($\beta_0=\beta_1=0$). Для этого вычисляем согласно (2.10.) величину

$$F_{\text{набл}}=\frac{\frac{1}{2}Q_R}{\frac{1}{8}Q_{\text{ост}}}=\frac{264{,}84958}{0{,}49809176}=531{,}72849.$$

По таблице F – распределения для $\alpha=0{,}05$, $\nu_1=2$ и $\nu_2=8$ находим $F_{\text{кр}}=4{,}46$. Так как $F_{\text{набл}}>F_{\text{кр}}$, то уравнение является значимым.

Найдем оценку ковариационной матрицы вектора b:

$$\hat{S}(b) = \hat{s}^2 (X^T X)^{-1} = \begin{pmatrix} 0,306422 & -0,0275299 \\ -0,0275229 & 0,0036697 \end{pmatrix} \cdot 0,49809176 =$$

$$= \begin{pmatrix} 0,15262627 & -0,013712416 \\ -0,013712416 & 0,0018278473 \end{pmatrix}.$$

Отсюда получаем несмещенные оценки дисперсий и среднеквадратических отклонений коэффициентов регрессии:

$$\hat{s}^2_{b_0} = 0,15262627 \qquad \hat{s}_{b_0} = 0,3906741$$

$$\hat{s}^2_{b_1} = 0,0018278473 \qquad \hat{s}_{b_1} = 0,0427527.$$

Для проверки значимости коэффициента регрессии, т.е. гипотезы H_0:β_1=0, находим по таблице t- распределения при α=0,05, ν=8 значение $t_{кр}$=2,31:

$$t(b_1) = \frac{b_1}{\hat{S}_{b_1}} = \frac{0,74861}{0,0427527} = 17,5102.$$

Так как $t_{набл}(b_1)$=17,51 больше $t_{кр}$=2,31, то коэффициент регрессии β_1 значимо отличается от нуля. Таким образом, окончательное уравнение регрессии имеет вид $\hat{y} = 0,52534 + 0,74861$.

Определим интервальные оценки коэффициентов уравнения с доверительной вероятностью γ=0,95.

Из (2.12.) следует:

$\beta_0 \in$[0,525 ± 2,31×0,391], откуда -0,378 ≤ β_0 ≤ 1,428 и

$\beta_1 \in$[0,74861 ± 2,31×0,0428], откуда 0,650 ≤ β_1 ≤0 ,847.

2.7. Задание для самостоятельного решения

На основании данных о темпе прироста (%) внутреннего национального продукта (y) и промышленного производства (x) десяти развитых стран мира за 1992г., приведенных в таблице, и предположения, что генеральное уравнение регрессии имеет вид: $\tilde{y} = \beta_0 + \beta_1 x$.

страны	y	x
Япония	3,5	4,3
США	3,1	4,6
Германия	2,2	2,0
Франция	2,7	3,1
Италия	2,7	3,0
Великобритания	1,6	1,4
Канада	3,1	3,4
Австралия	1,8	2,6
Бельгия	2,3	2,6
Нидерланды	2,3	2,4

Требуется:

а) определить оценки вектора b и остаточной дисперсии $\hat{Y}^2$;

б) при α=0,05 проверить значимость уравнении регрессии;

в) при α=0,05 проверить значимость коэффициентов уравнения;

г) с доверительной вероятностью γ=0,9 построить интервальные оценки β_0 и β_1;

д) с доверительной вероятностью γ=0,9 построить интервальные оценки уравнения регрессии в точках, определяемых вектором начальных условий $x^0=\begin{pmatrix}1\\3\end{pmatrix}$; $x^0=\begin{pmatrix}1\\5\end{pmatrix}$.

Задание выполняется по вариантам. Каждый должен вычеркнуть объект №, соответствующий последней цифре номера зачетной книжки.

3. Методы многомерной классификации. Кластерный анализ

Часто в экономических исследованиях возникает задача анализа неоднородных в некотором смысле данных. Так, например, исследуя зависимость спроса от цены товара, взяв для исследования данные за 1992 и 1997 гг., мы получим следующую зависимость: увеличение цены приводит к росту спроса на товар. Такая зависимость не соответствует реальным экономическим процессам. С чем связана полученная ошибка? Ответ состоит в том, что мы не учли инфляционные процессы в стране, произошедшие за этот период времени и, соответственно, повысившуюся цену на товар.

В таких случаях, прежде чем переходить к построению регрессионных моделей, необходимо выделить однородные группы объектов и уже внутри каждой группы строить регрессионные зависимости. В данном случае необходимо было рассматривать два уравнения, описывающих развитие процесса в 1992 и 1997 гг., раздельно.

3.1. Основные понятия кластерного анализа

В статистических исследованиях группировка первичных данных является основным приемом решения задачи классификации, а поэтому и основой всей дальнейшей работы с собранной информацией.

Традиционно эта задача решается следующим образом. Из множества признаков, описывающих объект, отбирается один, наиболее информативный с точки зрения исследователя, и производится группировка в соответствии со значениями данного признака. Если требуется провести классификацию по нескольким признакам, ранжированным между собой по степени важности, то сначала производится классификация по первому признаку, затем каждый из полученных классов разбивается на подклассы по второму признаку и т.д. Подобным образом строится большинство комбинационных статистических группировок.

В тех случаях, когда не представляется возможным упорядочить классификационные признаки, применяется наиболее простой метод многомерной группировки – создание интегрального показателя (индекса), функционально зависящего от исходных признаков, с последующей классификацией по этому показателю.

Развитием этого подхода является вариант классификации по нескольким обобщающим показателям (главным компонентам), полученным с помощью методов факторного или компонентного анализа.

При наличии нескольких признаков (исходных или обобщенных) задача классификации может быть решена методами *кластерного анализа,* которые отличаются от других методов многомерной классификации отсутствием обучающих выборок, т.е. априорной информации о распределении генеральной совокупности, которая представляет собой вектор X.

Различия между схемами решения задачи по классификации во многом определяются тем, что понимают под понятием «сходство» и «степень сходства».

После того как сформулирована цель работы, естественно попытаться определить критерии качества, целевую функцию, значения которой позволят сопоставить различные схемы классификации.

В экономических исследованиях целевая функция, как правило, должна минимизировать некоторый параметр, определенный на множестве объектов (например, целью классифицировать оборудования может явиться группировка, минимизирующая совокупность затрат времени и средств на ремонтные работы).

В случаях, когда формализовать цель задачи не удается, критерием качества классификации может служить возможность содержательной интерпретации найденных групп.

Рассмотрим следующую задачу. Пусть исследуется совокупность n объектов, каждый из которых характеризуется по k замеренным на нем признакам X. Требуется разбить эту совокупность на однородные в некотором смысле группы (классы).

При этом практически отсутствует априорная информация о характере распределения измерений X внутри классов.

Полученные в результате разбиения группы обычно называются кластерами♣ (таксонами♣♣, образами), методы их нахождения – кластер-анализом (соответственно численной таксономией или распознаванием образов с самообучением).

При этом необходимо с самого начала четко представить, какая из двух задач классификации подлежит решению. Если решается обычная задача типизации, то совокупность наблюдений разбивают на сравнительно небольшое число областей группирования (например, интервальный вариационный ряд в случае одномерных наблюдений) так, чтобы элементы одной такой области находились друг от друга по возможности на небольшом расстоянии.

Решение другой задачи заключается в определении естественного расслоения исходных наблюдений на четко выраженные кластеры, лежащие друг от друга на некотором расстоянии.

Если первая задача типизации всегда имеет решение, то при второй постановке может оказаться, что множество исходных наблюдений не обнаруживает естественного расслоения на кластеры, т.е. образует один кластер.

Хотя многие методы кластерного анализа довольно элементарны, основная часть работ, в которых они были предложены, относится к последнему десятилетию. Это объясняется тем, что эффективное решение задач поиска кластеров требует большого числа арифметических и логических операций и поэтому стало возможным только с возникновением и развитием вычислительной техники.

Обычной формой представления исходных данных в задачах кластерного анализа служит прямоугольная таблица:

$$X=\begin{pmatrix} x_{11}\cdots & x_{1j}\cdots & x_{1k} \\ \vdots & \vdots & \vdots \\ x_{i1}\cdots & x_{ij}\cdots & x_{ik} \\ \vdots & \vdots & \vdots \\ x_{n1}\cdots & x_{nj}\cdots & x_{nk} \end{pmatrix},$$

каждая строка которой представляет результат измерений k рассматриваемых признаков на одном из обследованных объектов. В конкретных ситуациях может представлять интерес как группировка объектов, так и группировка признаков. В тех случаях, когда разница между двумя этими задачами не существенна, например при описании некоторых алгоритмов, мы будем пользоваться только термином «объект», включая в это понятие и «признак».

♣ cluster (англ.) - группа элементов, характеризуемых каким-либо общим свойством.

♣♣ taxon (англ.) - систематизированная группа любой категории.

Матрица X не является единственным способом представления данных в задачах кластерного анализа. Иногда исходная информация задана в виде квадратной матрицы

$$R=(r_{ij}),\ \ i,j=1, 2, ..., n,$$

элемент r_{ij} который определяет степень близости i-го объекта к j-му.

Большинство алгоритмов кластерного анализа полностью исходит из матрицы расстояний (или близостей) либо требует вычисления отдельных ее элементов, поэтому если данные представлены в форме X, то первым этапом решения задачи поиска кластеров будет выбор способа вычисления расстояний, или близости, между объектами или признаками.

Относительно проще решается вопрос об определении близости между признаками. Как правило, кластерный анализ признаков преследует те же цели, что и факторный анализ – выделение групп связанных между собой признаков, отражающих определенную сторону изучаемых объектов. Мерами близости в этом случае служат различные статистические коэффициенты связи, например $/r_{mj}/$, mj =1, 2, ...,к. Элемент r_{mj} определяет степень близости m-го признака к j-му.

3.2. Расстояние между объектами (кластерами) и мера близости

Наиболее трудным и наименее формализованным в задаче классификации является определение понятия однородности объектов.

В общем случае понятие однородности объектов задается либо введением правила вычисления расстояний $\rho(x_i,x_j)$ между любой парой исследуемых объектов $(x_1, x_2, ... , x_n)$, либо заданием некоторой функции $r(x_i,x_j)$, характеризующей степень близости i-го и j-го объектов. Если задана функция $\rho(x_i,x_j)$, то близкие с точки зрения этой метрики объекты считаются однородными, принадлежащими к одному классу. Очевидно, что необходимо при этом сопоставлять $\rho(x_i,x_j)$ с некоторыми пороговыми значениями, определяемыми в каждом конкретном случае по-своему.

Аналогично используется и мера близости $r(x_i,x_j)$, при задании которой мы должны помнить о необходимости выполнения следующих условий: симметрии $r(x_i,x_j)= r(x_j,x_i)$; максимального сходства объекта с самим собой $r(x_i,x_i)= \max\limits_{i\,j}\ r(x_i,x_j)$, при $1\le i,j\le n$, и монотонного убывания $r(x_i,x_j)$ по мере увеличения $\rho(x_i,x_j)$, т.е. из $\rho(x_k,x_l)\ge\rho\ (x_i,x_j)$ должно следовать неравенство $r(x_k,x_l)\le r(x_i,x_j)$.

Выбор метрики или меры близости является узловым моментом исследования, от которого в основном зависит окончательный вариант разбиения объектов на классы при данном алгоритме разбиения. В каждом конкретном случае этот выбор должен производиться по-своему в зависимости от целей исследования, физической и статистической природы вектора наблюдений X, априорных сведений о характере вероятностного распределения X.

Рассмотрим наиболее широко используемые в задачах кластерного анализа расстояния и меры близости.

Обычное Евклидово расстояние

$$\rho_E(x_i, x_j) = \sqrt{\sum_{\ell=1}^{k}(x_{i\ell} - x_{j\ell})^2}, \qquad (3.1)$$

где $x_{i\ell}$, $x_{j\ell}$ – величина ℓ-ой компоненты у i-го (j-го) объекта (ℓ=1,2,...,к; i,j=1,2,..., n)

Использование этого расстояния оправдано в следующих случаях:

а) наблюдения берутся из генеральной совокупности, имеющей многомерное нормальное распределение с ковариационной матрицей вида $\sigma^2 E_к$, т.е. компоненты X взаимно независимы и имеют одну и ту же дисперсию, где $E_к$ – единичная матрица;

б) компоненты вектора наблюдений X однородны по физическому смыслу и одинаково важны для классификации;

в) признаковое пространство совпадает с геометрическим пространством.

Естественное с геометрической точки зрения евклидово пространство может оказаться бессмысленным (с точки зрения содержательной интерпретации), если признаки измерены в разных единицах. Чтобы исправить положение, прибегают к нормированию каждого признака путем деления центрированной величины на среднее квадратическое отклонение и переходят от матрицы X к нормированной матрице с элементами

$$t_{i\ell} = \frac{x_{i\ell} - \bar{x}_\ell}{S_\ell},$$

где $\overline{X_{i\ell}}$ - значение ℓ-го признака у i-го объекта;

$\bar{x}_\ell$ - среднее значение ℓ-го признака;

$S_\ell = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{i\ell} - \bar{x}_\ell)^2}$ – среднее квадратическое отклонение l-го признака.

Однако эта операция может привести к нежелательным последствиям. Если кластеры хорошо разделены по одному признаку и не разделены по другому, то после нормирования дискриминирующие возможности первого признака будут уменьшены в связи с увеличением «шумового» эффекта второго.

«Взвешенное» Евклидово пространство

$$\rho_{BE}(x_{i,} x_\ell) = \sqrt{\sum_{\ell=1}^{k}\omega_\ell (x_{i\ell} - x_{j\ell})^2} \qquad (3.2)$$

применяется в тех случаях, когда каждой компоненте x_ℓ вектора наблюдений X удается приписать некоторый «вес» ω_ℓ, пропорционально степени важности признака в задаче классификации. Обычно принимают $0 \le \omega_\ell \le 1$, где ℓ=1,2,...k.

Определение «весов», как правило, связано с дополнительными исследованиями, например, организацией опроса экспертов и обработкой их мнений. Определение весов ω_ℓ только по данным выборки может привести к ложным выводам.

Хеммингово расстояние

Используется как мера различия объектов, задаваемых дихотомическими признаками. Это расстояние определяется по формуле

$$\rho_H(x_i, x_j) = \sum_{\ell=1}^{k} \left| x_{i\ell} - x_{j\ell} \right| \tag{3.3}$$

и равно числу несовпадений значений соответствующих признаков в рассматриваемых i-м и j-м объектах.

В некоторых задачах классификации объектов в качестве меры близости объектов можно использовать некоторые физические содержательные параметры, так или иначе характеризующие взаимоотношения между объектами. Например, задачу классификации отраслей народного хозяйства с целью агрегирования решают на основе матрицы межотраслевого баланса [1].

В данной задаче объектом классификации является отрасль народного хозяйства, а матрица межотраслевого баланса представлена элементами s_{ij}, характеризующими сумму годовых поставок i-ой отрасли в j-ю в денежном выражении. В качестве меры близости $\{r_{ij}\}$ принимают симметризованную нормированную матрицу межотраслевого баланса. С целью нормирования денежное выражение поставок i-ой отрасли в j-ю заменяют долей этих поставок по отношению ко всем поставкам i-ой отрасли. Симметризацию же нормированной матрицы межотраслевого баланса можно проводить, выразив близость между i-й и j-й отраслями через среднее значение из взаимных поставок, так что в этом случае $r_{ij}=r_{ji}$.

Как правило, решение задач классификации многомерных данных предусматривает в качестве предварительного этапа исследования реализацию методов, позволяющих выбрать из компонент $x_1, x_2, ..., x_к$ наблюдаемых векторов X сравнительно небольшое число наиболее существенно информативных, т.е. уменьшить размерность наблюдаемого пространства.

В ряде процедур классификации (кластер-процедур) используют понятия расстояния между группами объектов и меры близости двух групп объектов.

Пусть s_i- i-я группа (класс, кластер), состоящая из n_i объектов;

$\bar{x}_i$ – среднее арифметическое векторных наблюдений s_i группы, т.е. "центр тяжести" i-й группы;

$\rho(s_\ell, s_m)$ – расстояние между группами s_ℓ и s_m.

Наиболее употребительными расстояниями и мерами близости между классами объектов являются:

– расстояние, измеряемое по принципу «ближайшего соседа»,

$$\rho_{\min}(S_\ell, S_m) = \min_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j); \tag{3.4}$$

– расстояние, измеряемого по принципу «дальнего соседа»,

$$\rho_{\max}(S_\ell, S_m) = \max_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j); \tag{3.5}$$

– расстояние, измеряемое по «центрам тяжести» групп,

$$\rho_{ц.т.}(S_\ell, S_m) = \rho(\bar{x}_\ell, \bar{x}_m); \tag{3.6}$$

– расстояние, измеряемое по принципу «средней связи», определяется как среднее арифметическое всех попарных расстояний между представителями рассматриваемых групп

$$\rho_{ср.}(S_\ell, S_m) = \frac{1}{n_\ell n_m} \sum_{x_i \in S_\ell} \sum_{x_j \in S_m} \rho(x_i, x_j). \tag{3.7}$$

Академиком А.Н. Колмогоровым было предложено «обобщенное расстояние» между классами, которое включает в себя в качестве частных случаев все рассмотренные выше виды расстояний.

Расстояния между группами элементов особенно важно в так называемых агломеративных иерархических кластер-процедурах, так как принцип работы таких алгоритмов состоит в последовательном объединении элементов, а затем и целых групп, сначала самых близких, а затем все более и более отдаленных друг от друга.

При этом расстояние между классами s_{Y_i} и $s_{(m,q)}$, являющиеся объединением двух других классов s_m и s_q, можно определить по формуле

$$\rho_{\ell,(m,q)} = \rho(S_\ell, S_{(m,q)}) = \alpha\rho_{\ell m} + \beta\rho_{\ell q} + \gamma\rho_{mq} + \delta(\rho_{\ell m} - \rho_{\ell q}), \tag{3.8}$$

где $\rho_{\ell m} = \rho(S_\ell, S_m); \quad \rho_{\ell q} = \rho(S_\ell, S_q) \;\; и \;\; \rho_{mq} = \rho(S_m, S_q)$

– расстояния между классами s_ℓ, s_m и s_q;

– α, β, δ и γ – числовые коэффициенты, значения которых определяют специфику процедуры, ее алгоритм.

Например, при $\alpha= \beta=-\delta=1/2$и $\gamma=0$ приходим к расстоянию, построенному по принципу «ближайшего соседа». При $\alpha= \beta=\delta=1/2$ и $\gamma=0$ – расстояние между классами определяется по принципу «дальнего соседа», то есть как расстояние между двумя самыми дальними элементами этих классов.

И, наконец, при

$$\alpha = \frac{n_m}{n_m + n_q}; \quad \beta = \frac{n_q}{n_m + n_q}, \quad \gamma=\delta=0$$

соотношение (3.8) приводит к расстоянию $\rho_{ср}$ между классами, вычисленному как среднее из расстояний между всеми парами элементов, один из которых берется из одного класса, а другой из другого.

3.3. Функционалы качества разбиения

Существует большое количество различных способов разбиения заданной совокупности элементов на классы. Поэтому представляет интерес задача сравнительного анализа качества этих способов разбиения Q(S), определенного на множестве всех возможных разбиений.

Тогда под наилучшим разбиением S^* понимаем такое разбиение, при котором достигается экстремум выбранного функционала качества. Следует отметить, что выбор того или иного функционала качества, как правило, опирается на эмпирические соображения.

Рассмотрим некоторые наиболее распространенные функционалы качества разбиения. Пусть исследованием выбрана метрика ρ в пространстве X и пусть $S=(s_1,s_2,...,s_p)$ – некоторое фиксированное разбиение наблюдений $x_1, ...,x_n$ на заданное число p классов $s_1,s_2,\ldots,s_p$.

За функционал качества берут сумму («взвешенную») внутриклассовых дисперсий

$$Q_1(S)=\sum_{\ell=1}^{\rho}\sum_{x_i\in S_\ell}\rho^2(x_i,x_j). \qquad (3.9)$$

3.4. Иерархические кластер-процедуры

Иерархические (древообразные) процедуры являются наиболее распространенными (в смысле реализации на ЭВМ) алгоритмами кластерного анализа, Они бывают двух типов: агломеративные и дивизимные. В агломеративных процедурах начальным является разбиение, состоящее из n одноэлементных классов, а конечным – из одного класса; в дивизимных – наоборот.

Принцип работы иерархических агломеративных (дивизимных) процедур состоит в последовательном объединении (разделении) групп элементов сначала самых близких (далеких), а затем все более отдаленных (близких) друг от друга. Большинство этих алгоритмов исходит из матрицы расстояний (сходства).

К недостаткам иерархических процедур следует отнести громоздкость их вычислительной реализации. Алгоритмы требуют вычисления на каждом шаге матрицы расстояний, а следовательно, емкой машинной памяти и большого количества времени. В этой связи реализация таких алгоритмов при числе наблюдений, большем нескольких сотен, нецелесообразна, а в ряде случаев и невозможна.

В качестве примера рассмотрим агломеративный иерархический алгоритм. На первом шаге алгоритма каждое наблюдение x_i (i=1,2,...n) рассматривается как отдельный кластер. В дальнейшем на каждом шаге работы алгоритма происходит объединение двух самых близких кластеров и с учетом принятого расстояния по формуле пересчитывается матрица расстояний, размерность которой, очевидно, снижается на единицу. Работа алгоритма заканчивается, когда все наблюдения объединены в один класс.

Большинство программ, реализующих алгоритм иерархической классификации, предусматривает графическое представление результатов классификации в виде дендрограммы.

3.5. Тестовый пример

Провести классификацию n=6 объектов, каждый из которых характеризуется двумя признаками.

Номер объекта (i)	***1***	***2***	***3***	***4***	***5***	***6***
x_{i1}	5	6	5	10	11	10
x_{i2}	10	12	13	9	9	7

Расположение этих точек на плоскости показано на рис. 3.1.

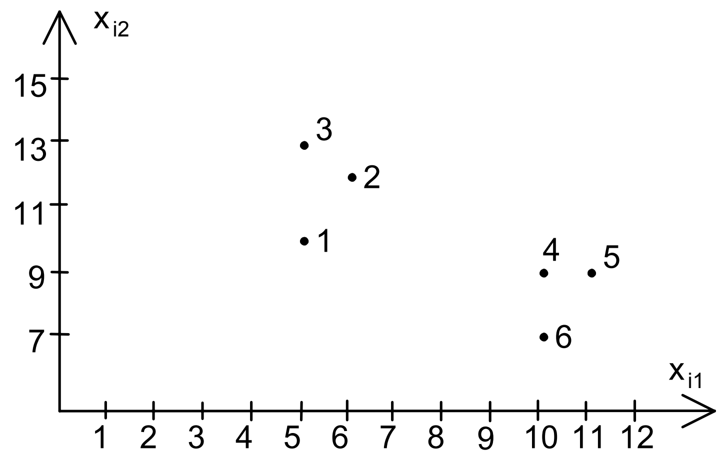


Рис. 3.1

Воспользуемся агломеративным иерархическим алгоритмом классификации. В качестве расстояния между объектами примем обычное евклидово расстояние. Тогда согласно (3.1) расстояние между объектами 1 и 2 равно

$$\rho_{12} = \sqrt{(5-6)^2 + (10-12)^2} = 2{,}24,$$

а между объектами 1 и 3 –

$$\rho_{13} = \sqrt{(5-5)^2 + (10-13)^2} = 3,$$

очевидно, что

$$\rho_{11} = 0.$$

Аналогично находим расстояния между всеми шестью объектами и строим матрицу расстояний

$$R_1 = \{\rho(x_i, x_j)\} = \begin{pmatrix} 0 & 2{,}24 & 3 & 5{,}10 & 6{,}08 & 5{,}83 \\ 2{,}24 & 0 & 1{,}41 & 5 & 5{,}83 & 6{,}40 \\ 3 & 1{,}41 & 0 & 6{,}40 & 7{,}21 & 7{,}81 \\ 5{,}10 & 5 & 6{,}40 & 0 & 1 & 2 \\ 6{,}08 & 5{,}83 & 7{,}21 & 1 & 0 & 2{,}24 \\ 5{,}83 & 6{,}40 & 7{,}81 & 2 & 2{,}24 & 0 \end{pmatrix}.$$

Из матрицы расстояний следует, что объекты 4 и 5 наиболее близки $d_{4,5}$=1,00 и поэтому объединяются в один кластер.

После объединения имеем пять кластеров

Номер кластера	**1**	**2**	**3**	**4**	**5**
Состав кластера	(1)	(2)	(3)	(4,5)	(6)

Расстояние между кластерами будем находить по принципу «ближайшего соседа», воспользовавшись формулой пересчета (3.8). Так, расстояние между объектом s_1 и кластером $s_{(4,5)}$ равно

$$\rho_{1,(4,5)} = \rho\left(S_1, S_{(4,5)}\right) = \frac{1}{2}\rho_{14} + \frac{1}{2}\rho_{15} - \frac{1}{2}\left|\rho_{14} - \rho_{15}\right| = \frac{1}{2}(5{,}10 + 6{,}08) -$$

$$-\frac{1}{2}\left(\left|5{,}10 - 6{,}08\right|\right) = 5{,}10.$$

Мы видим, что расстояние $\rho_{1,(4,5)}$ равно расстоянию от объекта 1 до ближайшего к нему объекта, входящего в кластер $s_{(4,5)}$, т.е. $\rho_{1,(4,5)} = \rho_{1,4} = 5{,}10$. Тогда матрица расстояний равна

$$R_2 = \begin{pmatrix} 0 & 2{,}24 & 3 & 5{,}10 & 5{,}83 \\ 2{,}24 & 0 & 1{,}41 & 5 & 6{,}40 \\ 3 & 1{,}41 & 0 & 6{,}40 & 7{,}81 \\ 5{,}10 & 5 & 6{,}40 & 0 & 2 \\ 5{,}83 & 6{,}40 & 7{,}81 & 2 & 0 \end{pmatrix}.$$

Объединим объекты 2 и 3, имеющие наименьшее расстояние $\rho_{2,3}$=1,41. После объединения имеем четыре кластера:

$$s_{(1)},\quad s_{(2,3)},\quad s_{(4,5)},\quad s_{(6)}.$$

Вновь найдем матрицу расстояний. Для этого необходимо рассчитать расстояние до кластера $s_{(2,3)}$. Для этого воспользуемся матрицей расстояний D_2.

Например, расстояние между кластерами $s_{(4,5)}$ и $s_{(2,3)}$ равно:

$$\rho_{(4,5),(2,3)} = \frac{1}{2}\rho_{(4,5),2} + \frac{1}{2}\rho_{(4,5),3} - \frac{1}{2}\left|\rho_{(4,5),2} - \rho_{(4,5),3}\right| = \frac{5}{2} + \frac{6{,}40}{2} - \frac{1{,}40}{2} = 5.$$

Проведя аналогичные расчеты, получим

$$D_3 = \begin{pmatrix} 0 & 2{,}24 & 5{,}10 & 5{,}83 \\ 2{,}24 & 0 & 5 & 6{,}40 \\ 5{,}10 & 5 & 0 & 2 \\ 5{,}83 & 6{,}40 & 2 & 0 \end{pmatrix}.$$

Объединенные кластеры $s_{(4,5)}$ и $s_{(6)}$, расстояние между которыми согласно матрице R_3 наименьшее: $\rho_{(4,5),6}$=2.

В результате этого получим три кластера

$$s_1,\quad s_{(2,3)}\quad \text{и}\quad s_{(4,5,6)}.$$

Матрица расстояний будет иметь вид

$$R_4 = \begin{pmatrix} 0 & 2{,}24 & 5{,}10 \\ 2{,}24 & 0 & 5 \\ 5{,}10 & 5 & 0 \end{pmatrix}.$$

Объединим теперь кластеры s_1 и $s_{2,3}$, расстояние между которыми равно $\rho_{1,(2,3)} = 2{,}24$. В результате получим два кластера: $s_{(1,2,3)}$ и $s_{(4,5,6)}$, расстояние между которыми, найденное по принципу «ближайшего соседа», равно

$$\rho_{(1,2,3);(4,5,6)} = 5\,.$$

Результаты иерархической классификации объектов представлены на рис. 3.2 в виде дендрограммы.

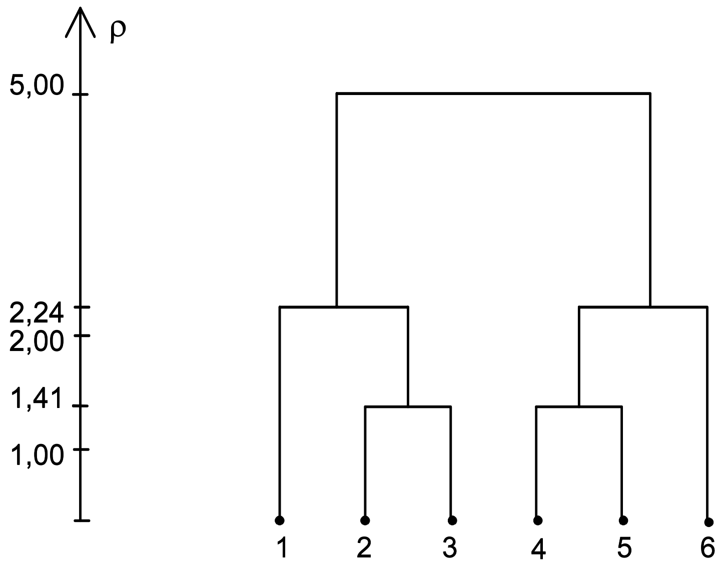


Рис. 3.2. **Дендрограмма**

Слева на рисунке приводится расстояние между объединяемыми на данном этапе кластерами (объектами).

В задаче предпочтение следует отдать предпоследнему этапу классификации, когда все объекты объединены в два кластера

$$S_{(1,2,3)} \quad \text{и} \quad S_{(4,5,6)},$$

что наглядно видно на рис. 3.1 и 3.2.

3.6. Задание для самостоятельного решения

По иерархическому агломеративному алгоритму провести классификацию n=4 хозяйств, работа которых характеризуется показателями объема реализованной продукции: x_1 – растениеводства и x_2 – животноводства с одного гектара пашни (млн.руб/га). Построить дендрограмму.

номер хозяйства	**1**	**2**	**3**	**4**
X_{i1}	**1**	**7**	**1**	**9**
X_{i2}	**5**	**9**	**3**	**7**

Для этого:

а) в качестве расстояния между объектами принять обычное евклидово расстояние, а расстояние между кластерами измерять по принципу «средней связи»;

б) в качестве расстояния между объектами принять взвешенное евклидово расстояние с «весами» ω_1=0,1, ω_2=0,9, а расстояние между кластерами измерять по принципу «дальнего соседа»;

в) в качестве расстояния между объектами принять обычное евклидово расстояние, а расстояние между кластерами измерять по принципу «центра тяжести».

Задание выполняется по вариантам. Каждому необходимо увеличить значения X_{i1}, X_{i2} на к.

4. Производственные функции

Производственная функция представляет собой математическую модель, характеризующую зависимость объема выпускаемой продукции от объема трудовых и материальных затрат. При этом модель может быть построена как для отдельной фирмы и отрасли, так и всей национальной экономики. Рассмотрим производственную функцию, включающую два фактора производства: затраты капитала (K) и трудовые затраты (L), определяющих объем выпуска Q. Тогда можно записать:

$$Q=f(K,L).$$

Определенного уровня выпуска можно достигнуть с помощью различного сочетания капитальных и трудовых затрат, а кривые, описываемые условиями f(K,L)=const, обычно называют изоквантами. Обычно предполагается, что по мере роста значений одной из независимых переменных предельная норма замещения данного фактора производства уменьшается. Поэтому при сохранении постоянного объема производства экономия одного вида затрат, связанная с увеличением затрат другого фактора, постепенно уменьшается. На примере производственной функции Кобба-Дугласа рассмотрим основные выводы, которые можно получить, исходя из предположений о том или ином виде производственной функции. Производственная функция Кобба-Дугласа, включающая два фактора производства, имеет вид:

$$Q = A \times K^{\alpha} \times L^{\beta}, \qquad (4.1)$$

где A, α, β – параметры модели. Величина A зависит от единиц измерения Q, K и L, а также от эффективности производственного процесса.

При фиксированных значениях K и L функции, характеризующейся большей величиной параметра A, соответствует большее значение Q, следовательно, и производственный процесс, описываемый такой функцией, более эффективен. Описываемая функция однозначна и непрерывна (при положительных K и L). Параметры α и β называют коэффициентами эластичности. Они показывают, на какую величину в среднем изменится Q, если α или β увеличить соответственно на один процент. Рассмотрим поведение функции при изменении масштабов производства. Предположим для этого, что затраты каждого фактора производства увеличились в C раз. Тогда новое значение будет определяться следующим образом:

$$Q_1 = A \times (C \times K)^{\alpha} \times (C \times L)^{\beta} = C^{\alpha+\beta} \times Q. \qquad (4.2)$$

При этом, если $\alpha + \beta = 1$, то уровень эффективности не зависит от масштабов производства. Если $\alpha + \beta < 1$, то средние издержки, рассчитанные на единицу продукции, растут, а при $\alpha + \beta > 1$ – убывают по мере расширения масштабов производства. Следует отметить, что эти свойства не зависят от численных значений K, L и сохраняют силу в любой точке производственной функции. Для определения параметров и вида производственной функции необходимо провести дополнительные наблюдения. Как правило, пользуются двумя видами данных – динамическими рядами и данными одновременных наблюдений (пространственной информацией). Динамические ряды данных характеризуют поведение одной и той же фирмы во времени, тогда как данные второго вида обычно относятся к одному и тому же моменту, но к различным фирмам. В случаях, когда исследователь располагает временным рядом, например, годовыми данными, характеризующими деятельность одной и той же фирмы, возникают трудности, с которыми не пришлось бы столкнуться при работе с пространственными данными. Так, относительные цены со временем становятся иными, а следовательно, меняется и оптимальное сочетание затрат отдельных факторов производства. Кроме того, с течением времени меняется и уровень административного управления. Однако основные проблемы при использовании временных рядов порождают последствия технического процесса, в результате которого меняются нормы затрат производственных факторов, соотношения, в которых они могут замещать друг друга, и параметры эффективности. Отсюда с течением времени могут меняться не только параметры, но и формы производственной функции. Технический прогресс может быть учтен в форме некоторого временного тренда, включаемого в состав производственной функции. Тогда

$$Q_t = \varphi(K_t, L_t, t).$$

Производственная функция Кобба-Дугласа с учетом технического прогресса имеет вид:

$$Q_t = A \times e^{\theta \cdot t} \times K_t^{\alpha} \times L_t^{\beta}. \quad (4.3)$$

В этом выражении параметр θ, с помощью которого характеризуется технический прогресс, показывает, что объем выпускаемой продукции ежегодно увеличивается на θ процентов независимо от изменений в затратах производственных факторов и, в частности, от размера новых инвестиций. Такая форма технического прогресса, не связанная с какими-либо затратами труда или капитала, называется «нематеризованным техническим прогрессом». Однако подобный подход не вполне реалистичен, т. к. новые открытия не могут повлиять на функционирование старых машин, а расширение объема производства возможно только посредством новых инвестиций. При другом подходе к учету технического прогресса для каждой возрастной группы капитала строят свою производственную функцию. В этом случае функция Кобба-Дугласа будет иметь вид:

$$Q_t(\nu) = Ae^{\theta \times \nu} \times K_t^{\alpha}(\nu) \times L_t^{\beta}(\nu), \quad (4.4)$$

где $Q_t(\nu)$ – объем продукции, произведенной в период t на оборудовании, введенном в строй в период ν; $L_t(\nu)$ – труд, занятый в период t обслуживанием оборудования, введенного в строй в период ν, и $K_t(\nu)$ – основной капитал, введенный в строй в период ν и использованный в период t. Параметр ν в такой производственной функции отражает состояние технического прогресса. Затем для периода t строится агрегированная производственная функция, представляющая собой зависимость совокупного объема выпускаемой продукции Q_t от общих затрат труда L_t и капитала K_t на момент t. При использовании для построения производственной функции пространственной информации, т. е. данных нескольких фирм, относящихся к одному и тому же времени, возникают проблемы другого рода. Так как наблюдения относятся к разным фирмам, то при их использовании предполагается, что поведение всех фирм может быть описано с помощью одной и той же функции. Для успешной экономической интерпретации полученной модели желательно, чтобы все эти фирмы принадлежали одной и той же отрасли. Кроме того, предполагается, что они располагают примерно одинаковыми производственными возможностями и уровнями административного управления. Рассмотренные выше производственные функции носили детерминированный характер и не учитывали влияние случайных возмущений, присущих каждому экономическому явлению. Поэтому в каждое уравнение, параметры которого предстоит оценить, необходимо ввести еще случайную переменную ε, которая будет отражать воздействие на процесс производства всех тех факторов, которые не вошли в состав производственной функции в явном виде. Таким образом, в общем виде производственную функцию Кобба-Дугласа можно представить как

$$Q = A \times K^{\alpha} \times L^{\beta} \times e^{\varepsilon}. \qquad (4.5)$$

Мы получили степенную регрессионную модель, оценки параметров которой А, α и β можно найти с помощью метода наименьших квадратов, лишь прибегнув предварительно к логарифмическому преобразованию. Тогда для i-го наблюдения имеем:

$$\ln Q_i = \ln A + \alpha \times \ln K_i + \beta \times \ln L_i + \varepsilon_i, \qquad (4.6)$$

где Q_i, K_i и L_i – соответственно объемы выпуска, капитальных и трудовых затрат для i-го наблюдения (i=1,2,...,n), а n-объем выборки, число наблюдений, используемых для получения оценок $\ln\hat{A}$, $\hat{\alpha}$ и $\hat{\beta}$ параметров производственной функции. Относительно ε_i обычно предполагается, что они взаимно независимы между собой и $\varepsilon_i \in N(0,\sigma)$. Исходя из априорных соображений значения α и β должны удовлетворять условиям: 0<α<1 и 0<β<1. Если предположить, что с изменением масштабов производства уровень эффективности остается постоянным, то, приняв $\beta = 1-\alpha$, имеем

$$Q = A \times K^{\alpha} \times L^{1-\alpha} \times e^{\varepsilon} = A \times (\tfrac{K}{L})^{\alpha} \times L \times e^{\varepsilon}, \qquad (4.7)$$

или

$$\tfrac{Q}{L} = A \times (\tfrac{K}{L})^{\alpha} \times e^{\varepsilon}$$

и

$$\ln(Q/L) = \ln A + \alpha \times \ln(K/L) + \varepsilon. \qquad (4.8)$$

Прибегнув к такой форме выражения производственной функции можно устранить влияние мультиколлинеарности между lnK и lnL. В качестве примера приведем полученную на основе данных о 180 предприятиях, выпускающих верхнюю одежду, модель Кобба-Дугласа:

$$\ln(Q/L) = 1{,}43 + \underset{(4{,}67)}{0{,}14} \ln L + \underset{(3{,}80)}{0{,}19} \ln(K/L).$$

В скобках указаны значения t-критерия для коэффициентов регрессии уравнения. При этом множественный коэффициент детерминации и расчетное значение статистики F-критерия соответственно равны: $r^2 = 0{,}16$ и F=12,7. Расчетное значение F указывает на то, что полученное значение не носит случайный характер. Оценки параметров α и β функции Кобба-Дугласа соответственно равны $\hat{\alpha} = 0{,}19$ и $\hat{\beta} = 0{,}95$ (1-0,19+0,14). Так как $\hat{\alpha} + \hat{\beta} = 1{,}14 > 1$, то можно предположить некоторое повышение эффективности по мере расширения масштабов производства. Параметры модели показывают также, что при увеличении капитала K на 1%, объем выпуска увеличивается в среднем на 0,19%, а при увеличении трудовых затрат L на 1% объем выпуска в среднем увеличится на 0,95%.

5. Система одновременных эконометрических уравнений

Систему взаимосвязанных тождеств и регрессионных уравнений, в которой переменные могут одновременно выступать как результирующие в одних уравнениях и как объясняющие в других, принято называть системой одновременных (эконометрических) уравнений. При этом в соотношения могут входить переменные, относящиеся не только к моменту времени t , но и к предшествующим моментам. Такие переменные называются лаговыми (запаздывающими). Тождества относятся к функциональной связи переменных и вытекают из содержательного смысла этих переменных. Техника оценивания параметров системы эконометрических уравнений имеет свои особенности. Это связано с тем, что в регрессионных уравнениях системы независимые переменные и случайные погрешности оказываются коррелированы между собой. Достаточно хорошо изучены статистические свойства и вопросы оценивания систем линейных уравнений. Будем рассматривать линейную модель вида

$$\beta_{i1}y_{1t}+\beta_{i2}y_{2t}+\ldots+\beta_{ig}y_{gt}+\gamma_{i1}x_{1t}+\ldots+\gamma_{ik}x_{kt}=u_{it}, \qquad (5.1)$$

где t=1,2,...,n; i=1,2,...,g;

Y_i – значение эндогенной (результирующей) переменной в момент времени t;

x_{jt} – значение предопределенной переменной, т. е. экзогенной (объясняющей) переменной в момент t или лаговой эндогенной переменной;

u_{it} – случайные возмущения, имеющие нулевые средние.

Совокупность равенств (4.9) называется системой одновременных уравнений в структурной форме. Наличие априорных ограничений, связанных, например, с тем, что часть коэффициентов считаются равными нулю, обеспечивает возможность статистического оценивания оставшихся. В матричном виде систему уравнений можно представить как

$$By_t+\Gamma x_t=\varepsilon_t\ , \qquad (5.2)$$

где B-матрица порядка $g\times g$, состоящая из коэффициентов при текущих значениях эндогенных переменных;

Г-матрица порядка $g\times k$, состоящая из коэффициентов экзогенных переменных.

$y_t=(y_{1t},\ldots,y_{gt})^T$; $x_t=(x_{1t},\ldots,x_{Kt})^T$; $\varepsilon_t=(\varepsilon_{1t},\ldots,\varepsilon_{gt})^T$ – векторы-столбцы значений соответственно эндогенных и экзогенных переменных, случайных ошибок. При этом $M\varepsilon_t=0$; $\Sigma_{(\varepsilon)}=M\varepsilon_t\varepsilon_t^T=\sigma_t^2E_n$, где E_n – единичная матрица. Таким образом, если $M\varepsilon_{t_1}\varepsilon_{t_2}=0$ при $t_1\neq t_2$ и $t_1,t_2=1,2,\ldots,n$, то случайные ошибки независимы между собой. Если дисперсия ошибки постоянна $M\varepsilon_t^2=\sigma_t^2=\sigma^2$ и не зависит от t и x_t, то это свидетельствует о гомоскедастичности остатков. Условием гетероскедастичности является зависимость значений $M\varepsilon_t^2=\sigma_t^2$ от t и x_t. Умножив все элементы уравнения слева на обратную матрицу s_j, получим приведенную форму системы одновременных уравнений:

$$y_t=B^{-1}\Gamma x_t+B^{-1}\varepsilon_t\,. \qquad (5.3)$$

Среди систем одновременных уравнений наиболее простыми являются рекурсивные системы, для оценивания коэффициентов которых можно использовать метод наименьших квадратов. Систему (5.3) одновременных уравнений называют рекурсивной, если выполняются следующие условия:

1) Матрица значений эндогенных переменных

$$B=\begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ \beta_{21} & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{i1} & \beta_{i2} & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{g1} & \beta_{g2} & \cdots & \beta_{gj} & \cdots & 1 \end{pmatrix}$$

является нижней треугольной матрицей, т. е. $\beta_{ij}=0$ при j>i и $\beta_{ii}=1$;

2) случайные ошибки независимы между собой, т. е. $\sigma_{ii}>0, \sigma_{ij}=0$ при $i\neq j$, где i,j=1,2,...,g. Отсюда следует, что ковариационная матрица ошибок $M(\varepsilon_t\varepsilon_t^T)=\Sigma_{(\varepsilon)}$ диагональна;

3) каждое ограничение на структурные коэффициенты относится к отдельному уравнению. Процедура оценивания коэффициентов рекурсивной системы с помощью метода наименьших квадратов, примененного к отдельному уравнению, приводит к состоятельным оценкам. В качестве примера рассмотрим ситуацию, которая приводит к рекурсивной системе уравнений. Предположим, что цены на рынке P_t в день t зависят от объема продаж в предыдущий день q_{t-1}, а объем покупок q_t в день t зависит от цены товара в день t. Математически систему уравнений можно представить в виде

$$P_t=\alpha_0+\alpha_1 q_{t-1}+\varepsilon_t,$$
$$q_t=\beta_0+\beta_1 P_t+\xi_t.$$

Случайные возмущения ε_t и ξ_t можно считать независимыми. Мы получили рекурсивную систему двух уравнений, причем в правую часть первого уравнения входит предопределенная переменная q_{t-1}, а второго – эндогенная переменная P_t.

Применение метода наименьших квадратов для получения оценок системы одновременных уравнений приводит к смещенным и несостоятельным оценкам, поэтому область его применения ограничена рекурсивными системами. Для оценивания систем одновременных уравнений в настоящее время наиболее часто используют двухшаговый метод наименьших квадратов, применяемый к каждому уравнению системы в отдельности, и трехшаговый метод наименьших квадратов, предназначенный для оценивания всей системы в целом. Двухшаговый метод наименьших квадратов (2 МНК) применяют для оценки отдельного уравнения системы одновременных уравнений. Сущность этого метода состоит в том, что для оценивания параметров структурного уравнения метод наименьших квадратов применяют в два этапа. Он дает состоятельные, но в общем случае смещенные оценки коэффициентов уравнения, является достаточно простым с теоретической точки зрения и удобным для вычисления. Запишем исходное i-е структурное уравнение системы в виде

$$y_i=Y_i\beta_i+X_i\gamma_i+\varepsilon_i$$

где y_i – вектор n наблюдений над i-й эндогенной переменной;

Y_i – матрица порядка $(n\times q_i)$ значений эндогенных переменных, входящих в i-е уравнение (кроме y_i-й);

β_i – вектор размерности ($q_i \times 1$) значений структурных коэффициентов эндогенных переменных из матрицы Y_i;

X_i – матрица порядка ($\mathrm{n} \times k_i$) значений экзогенных переменных, входящих в уравнение;

γ_i – вектор размерности ($k_i \times 1$) коэффициентов, относящихся к переменным X_i; ε_i – вектор случайных возмущений, имеющий размерность ($\mathrm{n} \times 1$), причем $M\varepsilon_i = 0$; $\Sigma_{(\varepsilon)} = \sigma_i^2 \mathrm{E}_n$.

Непосредственно применить в данном случае метод наименьших квадратов нельзя, так как эндогенные переменные, содержащиеся в матрице Y_i коррелированы со случайными составляющими ε_i.

В связи с этим представим эндогенные переменные Y_i, входящие в уравнение, как функцию всех содержащихся в модели экзогенных переменных (X). Найдем оценку матрицы Y_i, которая согласно методу наименьших квадратов определяется из выражения

$$\hat{Y}_i = X_i (X_i^T X_i)^{-1} X_i^T Y_i \ .$$

Тогда

$Y_i = \hat{Y}_i + \hat{U}$, где $\hat{U}$ – матрица оценок остаточных величин преобразованной системы. Исходное структурное уравнение может быть преобразовано к виду:

$$y_i = \hat{Y}_i \beta_i + X_i \gamma + \mathrm{v}_i \ ,$$

где

$$\mathrm{v}_i = \varepsilon_i + \hat{U}\beta_i .$$

Применяя метод наименьших квадратов для нахождения оценок параметров вновь полученного уравнения регрессии, будем иметь:

$$d = \begin{pmatrix} \hat{\beta}_i \\ \hat{\gamma}_i \end{pmatrix} = \begin{pmatrix} \hat{Y}_i^T \hat{Y}_i; & \hat{Y}_i^T X_i \\ X_i^T \hat{Y}_i; & X_i^T X_i \end{pmatrix}^{-1} \begin{pmatrix} \hat{Y}_i^T y_i \\ X_i^T y_i \end{pmatrix},$$

где d – вектор оценок коэффициентов размерности $((q_i + k_i) \times 1)$. Перейдя к исходным переменным, получим:

$$d = \begin{pmatrix} \hat{\beta}_i \\ \hat{\gamma}_i \end{pmatrix} = \begin{pmatrix} Y_i^T X_i (X_i^T X_i)^{-1} X_i^T Y_i; & Y_i^T X_i \\ X_i^T Y_i; & X_i^T X_i \end{pmatrix} \begin{pmatrix} Y_i^T X_i (X_i^T X)^{-1} X_i^T y_i \\ X_i^T y_i \end{pmatrix}.$$

Полученная оценка и носит название оценки двухшагового метода наименьших квадратов параметров β и γ.

Таким образом, двухшаговый метод наименьших квадратов состоит в замене матрицы Y_i расчетной матрицей $\hat{Y}_i$, после чего оцениваются коэффициенты обыкновенного уравнения регрессии y_i на $\hat{Y}_i$ и X_i. Согласно алгоритму трехшагового метода наименьших квадратов первоначально с целью оценки коэффициентов каждого структурного уравнения применяют двухшаговый метод наименьших квадратов, а затем определяют оценку для ковариационной матрицы случайных возмущений. После этого с целью оценивания коэффициентов всей системы применяется обобщенный метод наименьших квадратов. Рассмотрим систему одновременных уравнений, содержащую g эндогенных и к экзогенных переменных, принимаемых как неслучайные. Преобразуем i-е уравнение (5.3) к виду:

$$y_i = Z_i\delta_i + \varepsilon_i ,$$

где

$$Z_i = (Y_i X_i) \ , \quad \delta_i = \begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix} .$$

Умножив левую и правую части уравнения слева на транспонированную матрицу X^T значений всех экзогенных переменных модели, получим

$$X^T y_i = X^T Z_i \delta_i + X^T \varepsilon_i .$$

Записав таким образом все уравнения системы, получим

$$\begin{pmatrix} X^T y_1 \\ \vdots \\ X^T y_i \\ \vdots \\ X^T y_g \end{pmatrix} = \begin{pmatrix} X^T Z_1 & 0 & \cdots & \cdots & 0 \\ 0 & X^T Z_2 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & X^T Z_g \end{pmatrix} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_i \\ \vdots \\ \delta_G \end{pmatrix} + \begin{pmatrix} X^T \varepsilon_1 \\ \vdots \\ X^T \varepsilon_i \\ \vdots \\ X^T \varepsilon_g \end{pmatrix} .$$

Для применения обобщенного метода наименьших квадратов построим ковариационную матрицу вектора возмущений

$$\Sigma_{(U)} = \begin{pmatrix} \sigma_{11} X^T X & \cdots & \sigma_{1i} X^T X & \cdots & \sigma_{1g} X^T X \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{i1} X^T X & \cdots & \sigma_{ii} X^T X & \cdots & \sigma_{2g} X^T X \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{g1} X^T X & \cdots & \sigma_{gi} X^T X & \cdots & \sigma_{gg} X^T X \end{pmatrix} = \Sigma \otimes X^T X \quad ^*.$$

Заменив матрицу $\Sigma = (\sigma_{ij})$ ее оценкой $\mathrm{S} = (\mathrm{s_{ij}})$, получим оценку ковариационной матрицы вектора возмущений

$$S_{(U)} = S(X^T X)$$

и соответствующую обратную матрицу

$$S_{(U)}^{-1} = S^{-1} \otimes (X^T X)^{-1} .$$

Тогда искомая оценка трехшагового метода наименьших квадратов имеет вид:

$$\hat{\delta} = (A^T S_{(U)}^{-1} A)^{-1} A^T S_{(U)}^{-1} Z \ ,$$

---

* Прямое (кронекерово) произведение матрицы A={a_{ij}}, i=1, 2,…, n и j=1, 2,…,m и матрицы B размерности (k×l) обозначается A ⊗ B и представляет собой матрицу размерности (nk×ml) вида:

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1j}B & \cdots & a_{1m}B \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{i1}B & \cdots & a_{ij}B & \cdots & a_{im}B \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1}B & \cdots & a_{nj}B & \cdots & a_{nm}B \end{pmatrix}$$

Каждый элемент этой матрицы представляет собой матрицу $a_{ij}B$ размерности k×l.

где

$$A = \begin{pmatrix} X^T Z_1 & 0 & \cdots & 0 \\ 0 & X^T Z_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & X^T Z_g \end{pmatrix} \quad ; \quad Z = \begin{pmatrix} X^T y_1 \\ \vdots \\ X^T y_i \\ \vdots \\ X^T y_g \end{pmatrix} .$$

В случае, когда матрица Σ не является диагональной, т. е. когда возмущения, входящие в различные структурные уравнения, зависимы, трехшаговая процедура имеет лучшую асимптотическую эффективность по сравнению с двухшаговой.

5.1. Тренировочный пример

Построение эконометрической модели мирового рынка нефти

Очевидно, что модель должна отражать взаимосвязь между тремя основными элементами рыночного механизма – спросом, ценой и предложением (эндогенными переменными). В свою очередь состояние указанных элементов в каждый момент времени можно охарактеризовать с помощью системы объясняющих, экзогенных переменных.

Система включает общехозяйственные и товарно-рыночные показатели. Общехозяйственные показатели отражают экономические процессы, происходящие в мире и отдельных странах, и дают представление о фоне, на котором происходит развитие рынка. Вторая группа показателей отражает явления, которые характерны для рынка нефти. Особый интерес представляют показатели, обладающие опережающим эффектом (временным лагом) по отношению к динамике эндогенных переменных конъюнктуры рынка нефти.

При выборе экзогенных переменных учитывалось, что состояние рынка нефти в любой момент времени определяется не только его внутренними факторами, но и состоянием внешней среды, т.е. общехозяйственной конъюнктуры всего мирового хозяйства, и, в первую очередь, динамикой воспроизводственного цикла, состоянием деловой активности в отраслях-потребителях, положением в кредитно-денежной и валютно-финансовой сферах экономики.

Завершающим этапом разработки модели исследуемого рынка является ее реализация. На данном этапе математическая модель формируется в общем виде, оцениваются ее параметры, проводится содержательная экономическая интерпретация, выясняются статистические и прогностические свойства модели.

При построении модели использовалась система показателей, основанная на ежеквартальных динамических рядах за последние 15 лет, которая характеризует основные стороны рынка нефти в экономическом, временном и географическом аспектах.

Использование корреляционного анализа на этапе предварительной обработки данных позволило ограничить круг используемых показателей (первоначально их было более 100), выбрать для дальнейшего анализа такие, которые отражают воздействие основных факторов на рынок нефти и наиболее тесно связаны с динамикой показателей конъюнктуры. При этом решалась также задача исключения влияния мультиколлинеарности.

Модель строилась исходя из предпосылки, что величина спроса играет более активную роль, чем факторы предложения и цены. Рекурсивная модель включает линейные регрессионные уравнения для следующих эндогенных переменных в момент времени t:

$y_{1,t}$ – экспорт нефти из стран ОПЕК;

$y_{2,t}$ – добыча нефти в странах ОПЕК;

$y_{3,t}$ – цена на нефть легкую аравийскую.

В модель вошли предопределенные переменные:

$y_{3,t-1}$ – цена на нефть легкую аравийскую с лагом в 1 квартал;

$x_{6,t}$ – поставки нефти на переработку в Японию;

$x_{7,t-1}$ – поставки нефти на переработку в США в момент t-1;

$x_{9,t}$ – коммерческие запасы нефти в странах Западной Европы;

$x_{10,t-1}$ – коммерческие запасы нефти в США с лагом в 1 квартал;

$x_{12,t}$ – экспорт нефти из бывшего СССР в развитые страны;

$x_{20,t-2}$ – индекс экспортных цен ООН на топливо с лагом в 2 квартала, а $x_{20,t-3}$ – в 3 квартала;

$x_{23,t-1}$ – загрузка производственных мощностей обрабатывающей промышленности США;

$\dfrac{y_{1,t}}{y_{2,t}}$ – показатель, учитывающий дисбаланс на рынке нефти в момент времени t.

Эконометрическая модель конъюнктуры рынка нефти имеет вид:

$$\begin{cases} \hat{y}_{1,t} = \underset{(8,5)}{4{,}2}x_{6,t} + \underset{(9,7)}{0{,}8}x_{7,t-1} + \underset{(9,7)}{1{,}5}x_{9,t} - \underset{(-9,0)}{0{,}6}x_{10,t} + \underset{(9,0)}{2{,}1}x_{12,t} - \underset{(-9,4)}{0{,}4}x_{20,t-2} - \underset{(-2,5)}{169{,}2} \\ \hat{y}_{2,t} = \underset{(12,0)}{0{,}9}y_{1,t} + \underset{(2,4)}{0{,}8}x_{7,t-1} + \underset{(1,8)}{0{,}3}x_{20,t-3} - \underset{(-1,1)}{64{,}0} \\ \hat{y}_{3,t} = \underset{(5,1)}{0{,}5}y_{3,t-1} + \underset{(1,4)}{16{,}2}\dfrac{y_{1,t}}{y_{2,t}} + \underset{(4,1)}{0{,}2}x_{20,t-3} + \underset{(4,1)}{0{,}3}x_{23,t-1} - \underset{(-2,0)}{32{,}6} \end{cases}.$$

Анализ статистических характеристик модели показал, что в целом она адекватно описывает рынок нефти – все уравнения значимы, объясняют от 67% до 92% дисперсии эндогенных переменных и характеризуются незначительными отклонениями расчетных значений эндогенных переменных от фактических. Значимость коэффициентов модели проверялась по t-критерию, расчетные значения которых указаны в скобках под соответствующими коэффициентами.

Построенная модель позволяет анализировать различные ситуации развития рынка нефти.

Выводы

В учебном пособии рассматриваются основные теоретические положения наиболее часто встречаемых в практике экономического анализа математико-статистических методов (корреляционный, регрессионный анализы), типологическая регрессия (кластерный анализ), а также производственные функции и системы одновременных уравнений.

Значительное внимание уделяется логическому анализу исходной информации и экономической интерпретации получаемых результатов. Пособие снабжено достаточным числом подробно разработанных типовых примеров и сквозных задач, взятых из экономической практики и решенных с использованием ЭВМ.

Сквозные примеры иллюстрируют необходимость комплексного применения математико-статистических методов при решении социально-экономических задач. При этом корреляционный анализ, с одной стороны, используется на этапе предварительного анализа для выявления мультиколлинеарности, а с другой – при оценке адекватности регрессионной модели. При окончательном выборе модели в работе рекомендуется использовать как экономические, так и статистические критерии.

В пособии также рассматриваются производственные функции и системы одновременных эконометрических уравнений.

Учебное пособие предназначено для студентов, изучающих многомерные статистические методы, и специалистов, желающих повысить свою квалификацию по современным эконометрическим методам.

6. Список рекомендуемой литературы

1. Айвазян С.А., Бухштабер В. М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерностей. – М.: Финансы и статистика, 1989. – 607 с.

2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. – М.: Финансы и статистика, 1985. – 487с.

3. Айвазян С.А. Енюков, И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных. – М.: Финансы и статистика, 1983. – 471с.

4. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. В 2 т. – 2-е изд. – М.: ЮНИТИ, 2001.

5. Джонстон Дж. Эконометрические методы. – М.: Статистика, 1980. – 446 с.

6. Доугерти Кр. Введение в эконометрику/ Пер. с англ. – М.: МГУ; ИНФРА-М, 2003.

7. Дубров А.М., Мхитарян В.С., Трошин Л.И. Математическая статистика для бизнесменов и менеджеров. – М.: МЭСИ, 2004. – 140 с.

8. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистика, 2003.

9. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс, 3-е изд. – М.: Дело, 2005. – 503 с.

10. Мхитарян В.С., Дубров А.М., Трошин Л.И. Многомерный статистический анализ в экономике. – М.: МЭСИ, 1995. – 149 с.

11. Мхитарян В.С., Трошин Л.И. Исследование зависимостей методами корреляции и регрессии. – М.: МЭСИ, 2004. – 51 с.

12. Эконометрика: Учебник./ Под ред. И.И. Елисеевой. – 2-е изд.– М.: Финансы и статистика, 2005. – 276 с.

7. Приложения

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

Методические указания к использованию некоторых таблиц

В таблице 1 протабулирована функция

$$\Phi(t) = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx,$$

f(t) – плотность нормированной нормально распределенной случайной величины $T \in N(0,1)$.

Вероятность попадания случайной величины T в интервал от t_1 до t_2 вычисляется по формуле

$$P(t_1 < T < t_2) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)].$$

Φ(t) обладает следующими свойствами:

$$\Phi(-t) = -\Phi(t); \quad \Phi(\infty) = 1; \quad \Phi(3) = 0{,}9973.$$

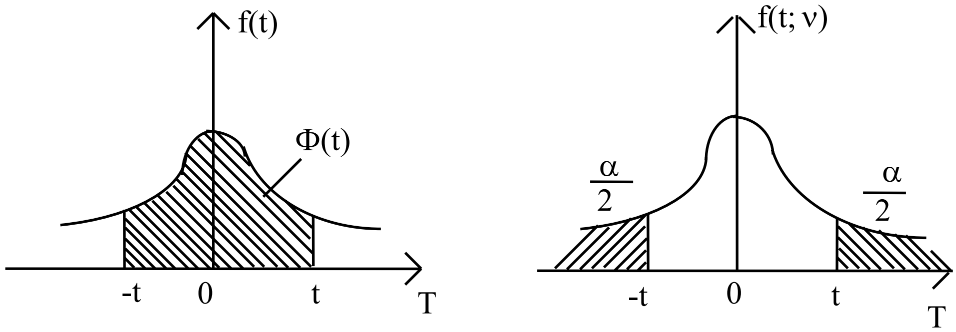
Пример:

$$P(-1{,}36 < T < 2{,}15) = \frac{1}{2}[\Phi(2{,}15)\text{-}\Phi(\text{-}1{,}36)] = \frac{1}{2}[0{,}9684+0{,}8262] = 0{,}8973.$$

В таблице 2 протабулирована вероятность выхода за пределы интервала от -t до +t случайной величины, имеющей распределение Стьюдента (t–распределение) с числом степеней свободы ν.

$$\alpha = St(t;\nu) = P(|T| > t)$$

$f(t;\ \nu)$ – плотность распределения Стьюдента с числом степеней свободы ν.



Вероятность попадания случайной величины T в интервал от t_1 до t_2 вычисляется по формуле

$$P(t_1 < T < t_2) = \frac{1}{2}[St(t_1) - St(t_2)].$$

Функция St(t) обладает следующими свойствами: St(-t) = 2 – St(t);
St (∞) = 0; St (-∞) = 2; St (0) = 1.

Пример: при $\nu = 10$ определить

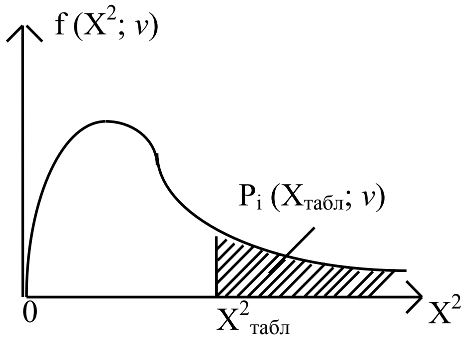
$$P(-1{,}36 < T < 2{,}15) = \frac{1}{2}[St(-1{,}36) - St(2{,}15)] = \frac{1}{2}[2 - St(1{,}36) -$$

$$- St(2{,}15)] \cong \frac{1}{2}[2 - St(1{,}372) - St(2{,}228)] = \frac{1}{2}[2 - 0{,}2 - 0{,}05] = 0{,}875.$$

Чтобы не прибегать к интерполяции, в строке, соответствующей ν=10, мы взяли ближайшие к заданным значениям 1,36 и 2,15.

Каждая строка таблицы отвечает t-распределению, с соответствующим числом степеней свободы $\bar{y}$.

В таблице 3 протабулирована вероятность того, что наблюдаемое значение случайной величины χ^2, имеющей распределение Пирсона (хи-квадрат распределение) с числом степеней свободы ν, превысит табличное значение $\chi^2_{табл}$.



На рис. 3 представлен график функции $f(X^2_{табл})$ – плотности χ^2 – распределения с числом степеней свободы ν.

$f(\chi^2_{табл}; y)$ – плотность χ^2 - распределения с числом степеней свободы ν.

Вероятность попадания случайной величины χ^2 в интервал от χ_1^2 до χ_2^2 вычисляется по формуле

$$P(\chi_1^2 < \chi^2 < \chi_2^2) = P(\chi^2 > \chi_1^2) - P(\chi^2 > \chi_2^2) = P_i(\chi_1^2) - P_i(\chi_2^2).$$

Функция $P_i(\chi^2_{табл})$ обладает следующими свойствами:

$P_i(0) = 1$; $P_i(\infty) = 0$.

Пример: при ν=10 определить

$P(2{,}5 < \chi^2 < 19{,}0) = P_i(2{,}5) - P_i(19{,}0) \cong P_i(2{,}558) - P_i(18{,}307) =$

$= 0{,}99 - 0{,}05 = 0{,}94.$

Чтобы не прибегать к интерполяции в строке таблицы, соответствующей ν=10, мы взяли ближайшие к заданным значениям 2,5 и 19,0.

Каждая строка таблицы отвечает χ^2- распределению с соответствующим числом степеней свободы ν.

В таблице 4 для случайной величины F, имеющей закон распределения Фишера-Снедекора (F-распределение), с числами степеней свободы числителя ν_1 и знаменателя ν_2, протабулированы три табличных значения, соответствующие трем вероятностям (уровням значимости):

$$\alpha = P(F > F_{табл}) = 0{,}05; \quad 0{,}01 \quad и \quad 0{,}001.$$

Пример. Уровню значимости α = 0,01 и числам степеней свободы числителя ν_1=5 и знаменателя ν_2=7 соответствует $F_{табл}$=7,46.

Статистика F строится таким образом, чтобы наблюдаемое значение было не меньше единицы.

Нормальный закон распределения

Таблица 1

Значение функции $\Phi(t) = P(|T| \leq t_{табл})$

Целые и десятичные доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2960	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7994	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	9488	9500	9512	9523	9534

Окончание табл. 1

Целые и десятичные доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9910	9912	9915	9917	9920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9998	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,0	0,999936	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,5	0,999994	-	-	-	-	-	-	-	-	-
5,0	0,99999994	-	-	-	-	-	-	-	-	-

Таблица 2

Распределение Стьюдента (t–распределение)

ν	Вероятность $\alpha = St(t) = P(\lvert T\rvert > t_{табл})$												
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,043	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,327	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,583
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,833
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850

Окончание табл. 2

ν	Вероятность $\alpha = St(t) = P(\lvert T \rvert > t_{табл})$												
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,868	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,402	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Р а с п р е д е л е н и е   П и р с о н а   (χ^2-распределение) *Таблица 3*

Значения $\chi^2_{табл}$ для вероятностей P ($\chi^2 > \chi^2_{табл}$)

ν	Вероятность										
	0,999	0,995	0,99	0,98	0,975	0,95	0,90	0,80	0,75	0,70	0,50
1	0,05157	0,04393	0,03157	0,03628	0,03982	0,00393	0,0158	0,0642	0,102	0,148	0,455
2	0,00200	0,0100	0,0201	0,0404	0,0506	0,103	0,211	0,446	0,575	0,713	1,386
3	0,0243	0,0717	0,115	0,185	0,216	0,352	0,584	1,005	1,213	1,424	2,366
4	0,0908	0,207	0,297	0,429	0,484	0,711	1,064	1,649	1,923	2,195	3,357
5	0,210	0,412	0,554	0,752	0,831	1,145	1,610	2,343	2,675	3,000	4,351
6	0,381	0,676	0,872	1,134	1,237	1,635	2,204	3,070	3,455	3,828	5,348
7	0,598	0,989	1,239	1,564	1,690	2,167	2,833	3,822	4,255	4,671	6,346
8	0,857	1,344	1,646	2,032	2,180	2,733	3,490	4,594	5,071	5,527	7,344
9	1,152	1,735	2,088	2,532	2,700	3,325	4,168	5,380	5,899	6,393	8,343
10	1,479	2,156	2,558	3,059	3,247	3,240	4,865	6,179	6,737	7,267	9,342
11	1,834	2,603	3,053	3,609	3,816	4,575	5,578	6,989	7,584	8,148	10,341
12	2,214	3,074	3,571	4,178	4,404	5,226	6,304	7,807	8,438	9,034	11,340
13	2,617	3,565	4,107	4,765	5,009	5,892	7,042	8,634	9,299	9,926	12,340
14	3,041	4,075	4,660	5,368	5,629	6,571	7,790	9,467	10,165	10,821	13,339
15	3,483	4,601	5,229	5,985	6,262	7,261	8,547	10,307	11,036	11,721	14,339
16	3,942	5,142	5,812	6,614	6,908	7,962	9,312	11,152	11,912	12,624	15,338
17	4,416	5,697	6,408	7,255	7,564	8,672	10,085	12,002	12,892	13,531	16,338
18	4,905	6,265	7,015	7,906	8,231	9,390	10,865	12,857	13,675	14,440	17,338
19	5,407	6,844	7,633	8,567	8,907	10,117	11,651	13,716	14,562	15,352	18,338
20	5,921	7,434	8,260	9,237	9,591	10,871	12,443	14,578	15,452	16,266	19,337
21	6,447	8,034	8,897	9,915	10,283	11,591	13,240	15,445	16,344	17,182	20,337
22	6,983	8,643	9,542	10,600	10,982	12,338	14,041	16,314	17,240	18,101	21,337
23	7,529	9,260	10,196	11,293	11,688	13,091	14,848	17,187	18,137	19,021	22,337
24	8,035	9,886	10,856	11,992	12,401	13,848	15,659	18,062	19,037	19,943	23,337
25	8,649	10,520	11,524	12,697	13,120	14,611	16,173	18,940	19,939	20,887	24,337
26	9,222	11,160	12,198	13,409	13,844	15,379	17,292	19,820	20,843	21,792	25,336
27	9,803	11,808	12,879	14,125	14,573	16,151	18,114	20,703	21,749	22,719	26,136
28	10,391	12,461	13,565	14,847	15,308	16,928	18,937	21,588	22,657	23,617	27,336
29	10,986	13,121	14,256	15,574	16,047	17,708	19,768	22,475	23,567	24,577	28,336
30	11,588	13,787	14,953	16,306	16,791	18,493	20,599	23,364	24,478	25,508	29,336

Окончание табл. 3

ν	Вероятность									
	0,30	0,25	0,20	0,10	0,05	0,025	0,02	0,01	0,005	0,001
1	1,074	1,323	1,642	2,706	3,841	5,024	5,412	6,635	7,879	10,827
2	2,408	2,773	3,219	4,605	5,991	7,378	7,824	9,210	10,597	13,815
3	3,665	4,108	4,642	6,251	7,815	9,348	9,837	11,345	12,838	16,268
4	4,878	5,385	5,989	7,779	9,488	11,143	11,668	13,277	14,860	18,465
5	6,064	6,626	7,289	9,236	11,070	12,839	13,388	15,086	16,750	20,517
6	7,231	7,841	8,558	10,645	12,592	14,449	15,033	16,812	18,548	22,457
7	8,383	9,037	9,803	12,017	14,067	16,013	16,622	18,475	20,278	24,322
8	9,524	10,219	11,030	13,362	15,507	17,535	18,168	20,090	21,955	26,125
9	10,656	11,389	12,242	14,684	16,919	19,023	19,679	21,666	23,589	27,877
10	11,781	12,549	13,412	15,987	18,307	20,483	21,161	23,209	25,188	29,588
11	12,899	13,701	14,631	17,275	19,675	21,920	22,618	24,725	26,757	31,264
12	14,011	14,845	15,812	18,549	21,026	23,337	24,054	26,217	28,300	32,909
13	15,119	15,984	16,985	19,812	22,362	24,736	25,472	27,688	29,819	34,528
14	16,222	17,117	18,151	21,064	23,685	26,119	26,873	29,141	31,319	36,123
15	17,322	18,245	19,311	22,307	24,996	27,488	28,259	30,578	32,801	37,697
16	18,418	19,369	20,465	23,542	26,296	28,845	29,633	32,000	34,267	39,252
17	19,511	20,489	21,615	24,769	27,587	30,191	30,995	33,409	35,718	40,790
18	20,601	21,605	22,760	25,989	28,869	31,526	32,346	34,805	37,156	42,312
19	21,689	22,718	23,900	27,204	30,144	32,852	33,687	36,191	38,582	43,820
20	22,775	23,828	25,038	28,412	31,410	34,170	35,020	37,566	39,997	45,315
21	23,858	24,935	26,171	29,615	32,671	35,479	36,343	38,932	41,401	46,797
22	24,939	26,039	27,301	30,813	33,924	36,781	37,659	40,289	42,796	48,268
23	26,018	27,141	28,429	32,007	35,172	38,076	38,968	41,638	44,181	49,728
24	27,096	28,241	29,553	33,196	36,415	39,364	40,270	42,980	45,558	51,170
25	28,172	29,339	30,675	34,382	37,652	40,046	41,566	44,314	46,928	52,620
26	29,246	30,434	31,795	35,563	38,885	41,923	42,856	45,642	48,290	54,052
27	30,319	31,528	32,912	36,741	40,113	43,194	44,140	46,963	49,645	55,476
28	31,391	32,620	34,027	37,916	41,337	44,461	45,419	48,278	50,993	56,893
29	32,461	33,711	35,139	39,087	42,557	45,722	46,693	49,588	52,336	58,302
30	33,530	34,800	36,250	40,256	43,773	46,979	47,962	50,892	53,672	59,703

Таблица 4

Р а с п р е д е л е н и е   Ф и ш е р а – С н е д е к о р а (F-распределение)

Значения $F_{табл}$, удовлетворяющие условию P (F > $F_{табл}$). Первое значение соответствует вероятности 0,05; второе – вероятности 0,01 и третье – вероятности 0,001; ν_1 – число степеней свободы числителя; ν_2 – знаменателя.

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞	t
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	253,3	12,71
	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366	63,66
	406523	500016	536700	562527	576449	585953	598149	610598	623432	636535	636,2
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50	4,30
	98,49	99,01	00,17	99,25	99,30	99,33	99,36	99,42	99,46	99,50	9,92
	998,46	999,00	999,20	999,20	999,20	999,20	999,40	999,60	999,40	999,40	31,00
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53	3,18
	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12	5,84
	67,47	148,51	141,10	137,10	134,60	132,90	130,60	128,30	125,90	123,50	12,94
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63	2,78
	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46	4,60
	74,13	61,24	56,18	53,43	51,71	50,52	49,00	47,41	45,77	44,05	8,61
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36	2,57
	16,26	13,27	12,06	11,39	10,97	10,67	10,27	9,89	9,47	9,02	4,03
	47,04	36,61	33,20	31,09	20,75	28,83	27,64	26,42	25,14	23,78	6,86
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67	2,45
	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88	3,71
	35,51	26,99	23,70	21,90	20,81	20,03	19,03	17,99	16,89	15,75	5,96
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23	2,36
	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65	3,50
	29,22	21,69	18,77	17,19	16,21	15,52	14,63	13,71	12,73	11,70	5,40

Продолжение табл. 4

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞	t
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,99	2,31
	11,26	8,65	7,59	7,10	6,63	6,37	6,03	5,67	5,28	4,86	3,36
	25,42	18,49	15,83	14,39	13,49	12,86	12,04	11,19	10,30	9,35	5,04
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71	2,26
	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31	3,25
	22,86	16,39	13,90	12,56	11,71	11,13	10,37	9,57	8,72	7,81	4,78
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54	2,23
	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91	3,17
	21,04	14,91	12,55	11,28	10,48	9,92	9,20	8,45	7,64	6,77	4,59
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40	2,20
	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60	3,11
	19,69	13,81	11,56	10,35	9,58	9,05	8,35	7,62	6,85	6,00	4,49
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30	2,18
	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36	3,06
	18,64	12,98	10,81	9,63	8,89	8,38	7,71	7,00	6,25	5,42	4,32
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21	2,16
	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16	3,01
	17,81	12,31	10,21	9,07	8,35	7,86	7,21	6,52	5,78	4,97	4,12
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13	2,14
	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00	2,98
	17,14	11,78	9,73	8,62	7,92	7,44	6,80	6,13	5,41	4,60	4,14
15	4,45	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07	2,13
	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87	2,95
	16,59	11,34	9,34	8,25	7,57	7,09	6,47	5,81	5,10	4,31	4,07
16	4,41	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01	2,12
	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75	2,92
	16,12	10,97	9,01	7,94	7,27	6,80	6,20	5,55	4,85	4,06	4,02

Продолжение табл. 4

ν_1 / ν_2	1	2	3	4	5	6	8	12	24	∞	t
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96	2,11
	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65	2,90
	15,72	10,66	8,73	7,68	7,02	6,56	5,96	5,32	4,63	3,85	3,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92	2,10
	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,01	2,57	2,88
	15,38	10,39	8,49	7,46	6,81	6,35	5,76	5,13	4,45	3,67	3,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88	2,09
	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49	2,86
	15,08	10,16	8,28	7,26	6,61	6,18	5,59	4,97	4,29	3,52	3,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84	2,09
	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42	2,84
	14,82	9,95	8,10	7,10	6,46	6,02	5,44	4,82	4,15	3,38	3,85
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,82	2,08
	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36	2,83
	14,62	9,77	7,94	6,95	6,32	5,88	5,31	4,70	4,03	3,26	3,82
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78	2,07
	7,94	5,72	4,82	4,31	3,99	3,75	3,45	3,12	2,75	2,30	2,82
	14,38	9,61	7,80	6,81	6,19	5,76	5,19	4,58	3,92	3,15	3,79
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76	2,07
	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26	2,81
	14,19	9,46	7,67	6,70	6,08	5,56	5,09	4,48	3,82	3,05	3,77
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73	2,06
	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21	2,80
	14,03	9,34	7,55	6,59	5,98	5,55	4,99	4,39	3,74	2,97	3,75

Окончание табл. 4

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞	t
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71	2,06
	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17	2,79
	13,88	9,22	7,45	6,49	5,89	5,46	4,91	4,31	3,66	2,89	3,72
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69	2,06
	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13	2,78
	13,74	9,12	7,36	6,41	5,80	5,38	4,83	4,24	3,59	2,82	3,71
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67	2,05
	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10	2,77
	13,61	9,02	7,27	6,33	5,73	5,31	4,76	4,17	3,52	2,76	3,69
28	4,19	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65	2,05
	7,64	5,54	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06	2,76
	13,50	8,93	7,18	6,25	5,66	5,24	4,69	4,11	3,46	2,70	3,67
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64	2,05
	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03	2,76
	13,39	8,85	7,12	6,19	5,59	5,18	4,65	4,05	3,41	2,64	3,66
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62	2,04
	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01	2,75
	13,29	8,77	7,05	6,12	5,53	5,12	4,58	4,00	3,36	2,59	3,64
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39	2,00
	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60	2,66
	11,97	7,76	6,17	5,31	4,76	4,37	3,87	3,31	2,76	1,90	3,36
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,03	1,96
	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,04	2,58
	10,83	6,91	5,42	4,62	4,10	3,74	3,27	2,74	2,13	1,05	3,29

Таблица 5

G – распределение

Пяти- и однопроцентное пределы для отношения G наибольшей выборочной дисперсии к сумме L выборочных дисперсий, полученных из L независимых выборок объемом n. Первое значение соответствует уровню значимости $\alpha = 0{,}05$, а второе – $\alpha = 0{,}01$

L \ n-1	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,998 0,999	0,975 0,995	0,939 0,979	0,906 0,959	0,877 0,937	0,853 0,917	0,838 0,809	0,816 0,882	0,801 0,867	0,788 0,854	0,734 0,795	0,660 0,700	0,518 0,606	0,500 0,500
3	0,967 0,993	0,871 0,942	0,798 0,883	0,746 0,834	0,707 0,903	0,677 0,761	0,653 0,734	0,633 0,711	0,617 0,691	0,603 0,674	0,547 0,606	0,475 0,515	0,403 0,423	0,333 0,333
4	0,906 0,968	0,768 0,864	0,684 0,781	0,629 0,721	0,590 0,676	0,560 0,641	0,537 0,613	0,518 0,590	0,502 0,570	0,488 0,554	0,437 0,488	0,372 0,406	0,309 0,325	0,250 0,250
5	0,841 0,928	0,684 0,789	0,598 0,696	0,544 0,633	0,507 0,588	0,478 0,553	0,456 0,526	0,439 0,504	0,424 0,485	0,412 0,470	0,365 0,409	0,307 0,335	0,251 0,254	0,200 0,200
6	0,781 0,883	0,616 0,722	0,532 0,626	0,480 0,564	0,445 0,520	0,418 0,487	0,398 0,461	0,382 0,440	0,368 0,423	0,357 0,408	0,314 0,353	0,261 0,286	0,212 0,223	0,167 0,167
7	0,727 0,838	0,561 0,664	0,480 0,569	0,431 0,508	0,397 0,466	0,373 0,435	0,354 0,411	0,338 0,391	0,326 0,375	0,315 0,362	0,276 0,311	0,228 0,249	0,183 0,193	0,143 0,143
8	0,680 0,795	0,516	0,438 0,521	0,391 0,463	0,360 0,423	0,336 0,393	0,319 0,370	0,304 0,352	0,293 0,337	0,283 0,325	0,246 0,278	0,202 0,221	0,162 0,170	0,15 0,125
9	0,639 0,754	0,478 0,573	0,403 0,481	0,358 0,425	0,329 0,387	0,307 0,359	0,290 0,338	0,277 0,321	0,266 0,307	0,257 0,295	0,223 0,251	0,182 0,199	0,145 0,152	0,111 0,111

Таблица 6

Таблица Z–преобразования Фишера

$$Z = \frac{1}{2}\{\ln(1 + r) - \ln(1 - r)\}$$

r	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0101	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
1	0,1003	0,1104	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2661	0,2769	0,2877	0,2986
3	0,3095	0,3205	0,3316	0,3428	0,3541	0,3654	0,3767	0,3884	0,4001	0,4118
4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
5	0,5493	0,5627	0,5764	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
6	0,6932	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
7	0,8673	0,8872	0,9077	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
9	1,4722	1,5275	1,5890	1,6584	1,7381	1,8318	1,9459	2,0923	2,2976	2,6467
0,99	2,6466	2,6996	2,7587	2,8257	2,9031	2,9945	3,1063	3,2504	3,4534	3,8002

Таблица 7

Значение плотности $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ **значение для нормированного нормального закона распределения f(–t) = f(t)**

Целые и десятые доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3525	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3-56	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2631	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	2179	2155	2131	2107	2083	2059	2036	3012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0762	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551

Таблица 8

Т а б л и ц а   Ф и ш е р а – И е й т с а

Зачения $r_{кр}$, найденные для уровня значимости α и чисел степеней свободы $\nu = n - 2$ в случае парной корреляции и $\nu = n - l - 2$, где l – число исключенных величин в случае частной корреляции

ν	Двусторонние границы				ν	Двусторонние границы			
	0,05	0,02	0,01	0,001		0,05	0,02	0,01	0,001
1	0,997	1,000	1,000	1,000	16	0,468	0,543	0,590	0,708
2	0,950	0,980	0,990	0,999	17	0,456	0,529	0,575	0,693
3	0,878	0,934	0,959	0,991	18	0,444	0,516	0,561	0,679
4	0,811	0,882	0,917	0,974	19	0,433	0,503	0,549	0,665
5	0,754	0,833	0,875	0,951	20	0,423	0,492	0,537	0,652
6	0,707	0,789	0,834	0,925	25	0,381	0,445	0,487	0,597
7	0,666	0,750	0,798	0,898	30	0,349	0,409	0,449	0,554
8	0,632	0,715	0,765	0,872	35	0,325	0,381	0,418	0,519
9	0,602	0,685	0,735	0,847	40	0,304	0,358	0,393	0,490
10	0,576	0,658	0,708	0,823	45	0,288	0,338	0,372	0,465
11	0,553	0,634	0,684	0,801	50	0,273	0,322	0,354	0,443
12	0,532	0,612	0,661	0,780	60	0,250	0,295	0,325	0,408
13	0,514	0,592	0,641	0,760	70	0,232	0,274	0,302	0,380
14	0,497	0,574	0,623	0,742	80	0,217	0,257	0,283	0,338
15	0,482	0,558	0,606	0,725	90	0,205	0,242	0,267	0,338
					100	0,195	0,230	0,254	0,321
ν	0,025	0,01	0,005	0,0005	ν	0,025	0,01	0,005	0,0005
	Односторонние границы					Односторонние границы			

Руководство по изучению дисциплины

1. Цели, задачи изучения, сфера профессионального применения

Цель преподавания курса – дать студентам научное представление о методах, моделях и приемах, позволяющих получать количественные выражения закономерностей экономической теории на базе экономической статистики с использованием математико-статистического инструментария. Современные социально-экономические процессы и явления зависят от большого количества факторов, их определяющих. В связи с этим квалифицированному специалисту необходимо не только иметь четкие представления об основных направлениях развития экономики, но и уметь учитывать сложное взаимосвязанное многообразие факторов, оказывающих существенное влияние на изучаемый процесс. Такие исследования не возможно проводить без знания основ теории вероятностей, математической статистики, многомерных статистических методов и эконометрики, т.е. дисциплин, позволяющих исследователю разобраться в огромном количестве стохастической информации и среди множества различных вероятностных моделей выбрать единственную, наилучшим образом отражающую изучаемый процесс или явление. Курс рассчитан на 64 часа.

2. Необходимый объем знаний для изучения курса

Для изучения курса эконометрики студентам необходимо знание основ:

- теории статистики, в которой сформулированы общие методы и принципы определения количественных характеристик массовых процессов и явлений;
- экономической статистики, дающей представление о направлениях развития экономики, о темпах роста цен и занятости, о тенденциях развития и эффективности использования ресурсов в отдельных отраслях и секторах экономики;
- линейной алгебры для проведения расчетов с матрицами;
- высшей математики, обучающей приемам интегрирования и дифференцирования;
- математической статистики, определяющей генеральную и выборочную совокупность, вариационные ряды и их характеристики; методы статистического оценивания параметров и статистической проверки гипотез (статистические критерии); методы корреляционно-регрессионного анализа для исследования взаимосвязи между зависимой переменной и группой влияющих на нее показателей;
- многомерных статистических методов, позволяющих выделять латентные факторы, сжимать признаковое пространство и сопоставлять изучаемые процессы в пространстве латентных факторов, проводить многомерную классификацию;
- владеть приемами статистического анализа нечисловой информации.

В свою очередь данный курс является основой для дисциплины «Эконометрическое моделирование».

Задачи курса.

Научиться строить экономические модели и оценивать их параметры;

научиться проверять гипотезы о свойствах экономических показателей и формах их связи.

В соответствии с целью курса студенты должны освоить методы количественной оценки социально-экономических процессов, научиться содержательно интерпретировать формальные результаты, моделировать с помощью пакетов прикладных программ STATISTICA, SPSS и др.

В курсе описываются формы и типы классических эконометрических моделей, соотношения между ними, их достоинства и недостатки. Рассматриваются предпосылки построения моделей, а также задачи их спецификации и идентификации.

3. Основная информация о курсе и его структура

Курс включает в себя изучение пяти тем:

I. Основы эконометрики. Корреляционный анализ.
II. Регрессионный анализ.
III. Методы многомерной классификации. Кластерный анализ.
IV. Производственные функции.
V. Системы одновременных экономических уравнений.

Курс изучается в форме лекций (2 час/нед) и практических занятий (2 час/нед).

Практические занятия проводятся как в аудитории, так и в компьютерных классах. Цель компьютерных занятий – овладение студентами методов анализа и обработки данных с использованием пакетов прикладных программ. Рекомендуемые программные средства и пакеты прикладных программ: EXCEL, OLIMP, STATISTICA, SPSS.

После прохождения каждой темы предусматривается соответствующая лабораторная работа, по которой сдается и защищается отчет, а также тестирование и самостоятельная подготовка студентов, которые выполняют индивидуальные компьютерные исследования, в конце семестра сдают зачет.

4. Перечень основных тем и подтем

Тема 1. Основы эконометрики. Корреляционный анализ

Изучение данной темы раскрывает студентам задачи эконометрики в области социально-экономических исследований, показывает необходимость использования информационных технологий на базе персональных ЭВМ в эконометрических исследованиях, также подготавливает их к пониманию основных тем курса, таких как:

- классификация переменных в эконометрических моделях;
- понятия спецификации и идентифицикации модели и др.

Изучение этой темы должно подготовить студентов к пониманию следующих тем данного курса.

Основные задачи эконометрики

В рамках данной темы рассматриваются различные классы моделей для анализа и (или) прогноза:

Модели временных рядов

Регрессионные модели

- линейные;
- нелинейные.

Системы одновременных уравнений.

Приводятся основные типы данных (пространственные и временные).

В рамках данной темы изучаются основные вопросы корреляционного анализа.

Знания, умения, навыки по теме 1

Изучив тему 1, студент должен знать:

- Основные задачи эконометрики.
- Понимать логику эконометрического анализа.
- Различать типы эконометрических уравнений; разбираться в спецификациях моделей.
- Алгоритм отбора факторов для построения модели.
- Теорию эконометрического анализа при решении профессиональных задач.
- Основные виды коэффициентов корреляции, их особенности.
- Проверка значимости коэффициентов корреляции. Интервальное оценивание.

***уметь*:**

- ✓ Применять методы эконометрики в экономических исследованиях.
- ✓ Использовать различные типы эконометрических уравнений.
- ✓ Определять типы переменных, используемых при построении эконометрических моделей.
- ✓ Производить идентификацию и спецификацию моделей.
- ✓ Видеть возможности использования эконометрики в профессиональной деятельности.
- ✓ Рассчитывать коэффициенты корреляции. Знать их особенности.
- ✓ Проверять значимость коэффициентов корреляции.
- ✓ Строить интервальные оценки параметров связи.
- ✓ Применять информационные технологии при построении эконометрических моделей.

Ссылки на учебный материал

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998.
2. Джонстон Дж. Эконометрические методы. – М.: Статистика, 1980.
3. Доугерти Кристофер Введение в эконометрику. – М.: ИНФРА-М, 1997.
4. Кейн Э. Экономическая статистика и эконометрия. – М.: Статистика, 1997.
5. Stewart J. and L. Gill Econometrics. 2nd ed., Prentice Hall, 1998.

Интернет-ресурсы:

1. http://upereslavl.botik.ru/UP/ECON/econometrics/top1/tsld006.htm
2. http://www.nsu.ru/ef/tsy/ecmr/study.htm
3. http://www.nsu.ru/ef/tsy/ecmr/index.htm
4. http://www.statsoft.ru/home/textbook/glossary/default.htm
5. http://www.dataforce.net/~antl/article/econometric
6. http://www.nsu.ru/ef/tsy/ecmr/study.htm
7. http://www.tvp.ru/vnizd/mathem4.htm

Задания для самооценки

Выполните задания и ответьте на вопросы:

- Назовите основные задачи эконометрики.
- Приведите простейшие примеры эконометрических моделей.

- Назовите области экономических наук, в которых используется эконометрика.
- В чем заключается процесс эконометрического исследования?
- Какие Вы знаете типы эконометрических моделей?
- Какими свойствами должна обладать построенная модель?
- Назовите основные виды коэффициентов корреляции.
- Как проверить значимость коэффициентов корреляции?

План практических занятий по Теме 1

Занятие 1.

Тема: Задачи эконометрики. Модель предложения и спроса на конкурентном рынке (2 часа).

Решение задач.

Занятие 2.

Тема: Расчет коэффициентов корреляции (2 часа).

Тема 2. Регрессионный анализ

Данная тема знакомит студентов с понятиями линейной модели множественной регрессии.

Для оценки неизвестных параметров уравнения регрессии чаще всего используется метод наименьших квадратов, позволяющий получить несмещенные оценки, состоятельные и эффективные оценки параметров уравнения.

В случаях, когда с точностью до неизвестных значений параметров известен общий вид распределения вероятностей имеющихся выборочных данных, может быть применен метод максимального правдоподобия.

Особое внимание при изучении данной темы необходимо также уделить статистическим свойствам оценок параметров регрессионной модели.

Мультиколлинеарность и отбор наиболее существенных объясняющих переменных в КЛММР

Мультиколлинеарность возникает в случаях существования достаточно тесных линейных статистических связей между объясняющими переменными. Точных количественных критериев для определения наличия (отсутствия) мультиколлинеарности не существует. Однако существуют некоторые рекомендации по выявлению мультиколлинеарности, на которые следует обратить внимание.

Знания, умения, навыки по теме 2

Изучив тему 2, студенты должны знать:

- Особенности линейной модели множественной регрессии.
- Свойства оценок обобщенного метода наименьших квадратов.
- Методы определения мультиколлинеарности.
- Метод наименьших квадратов (МНК).

уметь:

- ✓ Получать оценки регрессионных моделей.
- ✓ Применять статистические пакеты для решения практических задач.

Ссылки на учебный материал

1. Айвазян С.А., Мхитарян В.С., Зехин В.А. Практикум по эконометрике. – М.: МЭСИ, 2005.

2. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерный статистический анализ в экономических исследованиях. – М.: МЭСИ, 1988.

3. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистка, 2003.

4. Мхитарян В.С, Трошин Л.И. Статистический анализ многомерных совокупностей. Учебное пособие. – М.: МЭСИ, 1992.

Интернет-ресурсы:

1. http://www.kgtu.runnet.ru/WD/TUTOR/textbook/modules/stmulreg.html
2. http://www.shpargalka.ru/statis.ru/doc/shpr_e31.htm
3. http://dsmu.donetsk.ua/~statbook/modules/stmulreg.html#cunique
4. http://www3.unicor.ac.ru/d024/p011993.htm
5. http://www.gauss.ru/educat/systemat/butenkov/.asp
6. http://crow.academy.ru/econometrics/seminars_/sem_08_/sem_08.htm
7. http://crow.academy.ru/econometrics/lectures_/lect_03_/index.htm
8. http://u-pereslavl.botik.ru/UP/ECON/econometrics/
9. http://crow.academy.ru/econometrics

Задания для самооценки

Выполните задания и ответьте на вопросы:

- Какие критерии адекватности Вы знаете. Их преимущества и недостатки.
- Назовите свойства оценок метода наименьших квадратов.
- Какими свойствами обладают оценки уравнений регрессии, полученные с помощью МНК?
- Доказать, что при гетероскедастичности остатков ОМНК-оценки вектора β более эффективны, чем МНК-оценки.
- Какие проблемы возникают в практике регрессионного анализа?
- Какие существуют критерии для выбора регрессионной модели?
- Назовите основные признаки мультиколлинеарности.
- Какие методы устранения мультиколлинеарности Вы знаете?
- Как проверить значимость уравнения регрессии и его коэффициентов?

План практических занятий по теме 2

Занятие 1.

Тема: Расчет основных характеристик модели (2 часа).

Студентам предлагается на конкретном статистическом материале с использованием EXCEL, пакета программ STATISTICA построить регрессионную модель среднедушевых сбережений.

Тема 3. Методы многомерной классификации. Кластерный анализ

Данная тема знакомит студентов с типологическими регрессионными моделями. Часто в экономических исследованиях возникает задача анализа неоднородных в некотором смысле данных. В этом случае, прежде чем переходить к построению регрессионных моделей, необходимо выделить однородные группы объектов. Затем, в выделенных группах, уже можно переходить к исследованию зависимостей.

Знания, умения, навыки по теме 3

Изучив тему 3, студенты должны знать:

- Как определить расстояние между объектами.
- Как определить расстояние между кластерами.
- Функционалы качества разбиения.
- Иерархические кластер-процедуры.

уметь:

- ✓ Выделять однородные группы объектов.
- ✓ Использовать различные методы определения расстояний между объектами (кластерами).
- ✓ Проводить сравнительный анализ различных способов разбиения.
- ✓ Применять статистические пакеты при решении практических задач.

Интернет-ресурсы:

1. http://www.econ.msu.ru/kaf/DEI/books/prognoz/lec10.html
2. http://ns.econ.msu.ru/dei/books/prognoz/lec10.html
3. 3http://soc-gw.univ.kiev.ua/EDUCAT/BASIC/MMPS/LABS/LOGREG.HTM
4. http://www.nsu.ru/ef/tsy/ecmr/some_mle/contents.htm
5. http://www.doktor.ru/doctor/biometr/sp/contents4.htm
6. http://www.kgtu.runnet.ru/WD/TUTOR/textbook/modules/stnonlin.html
7. http://dsmu.donetsk.ua/~statbook/modules/stnonlin.html
8. http://softline.perm.ru/statistica/www-page_STATISTICA_for_Windows.html
9. http://eco.rea.ru/LSpace/EconTh.nsf/136acc8cc4a429f5c325654d004b4fc2
10. http://lanserv2.kemsu.ru/departs/matekon/Chapter4/par4_4.html
11. http://www.dvgu.ru/pin/math/for_students/eco/node4.html
12. http://www.hse.ru/rectorat/grebnev/economics/glava13.htm
13. http://ecfor.rssi.ru/0497_r_k.htm
14. http://www.socionet.ru:8100/RuPEc/data/Articles/rusrssicf4_1997article4.html -
15. http://www.mstu.edu.ru/publish/conf/section5/section5_7.html

Задания для самооценки

Выполните задания и ответьте на вопросы:

- Когда используется типологическая регрессия?
- В каких случаях используется Хеммингово расстояние?
- В каких случаях следует применять «взвешенное» Евклидово расстояние?
- Для чего используются иерархические кластер-процедуры?

План практических занятий по теме 3

Занятие 1.
Разбиение неоднородной совокупности на кластеры (2 часа).
Решение задач.
Занятие 2.
Построение регрессионных моделей в выделенных кластерах.

Решение задач.

Практические занятия проходят в компьютерных классах и посвящены построению регрессионных моделей с использованием фиктивных переменных в пакетах программ STATISTICA и SPSS. Например, можно предложить построить регрессионную модель веса новорожденного, включающую фиктивные переменные, рассмотреть потребление в мирное и военное время или зависимость стоимости жилья от различных качественных факторов.

Тема 4. Производственные функции

В данном разделе вводится понятие и дается определение производственной функции; рассматривается производственная функция Кобба-Дугласа.

Также при изучении данной темы необходимо обратить внимание на особенности оценивания параметров производственной функции Кобба-Дугласа.

Производственная функция Кобба-Дугласа

Производственная функция отражает зависимость между количеством применяемых ресурсов и максимально возможным объемом выпускаемой продукции в единицу времени; описывает всю совокупность технически эффективных способов производства (технологий).

Оценивание параметров производственной функции Кобба-Дугласа

При изучении данной темы необходимо обратить внимание на особенности оценивания параметров производственной функции Кобба-Дугласа по пространственной и временной информации.

Знания, умения, навыки по теме 4

Изучив тему 4, студенты должны знать:

- Что такое производственная функция.
- Как оценивать параметры производственной функции Кобба-Дугласа по пространственной информации.
- Как оценивать параметры производственной функции Кобба-Дугласа по временной информации.

уметь:

✓ Оценивать параметры производственной функции Кобба-Дугласа по пространственной и временной информации.

✓ Решать задачи с использованием теоретического материала.

✓ Использовать статистические пакеты для решения поставленных задач.

Ссылки на учебный материал

1. Айвазян С.А., Мхитарян В.С. Практикум по прикладной статистике и эконометрике. – М.: МЭСИ, 2003.

2.Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998.

3. Боровиков В.П., Ивченко Г.И. Statistica. Прогнозирование в системе WINDOWS. – М.: Финансы и статистика, 1999.

4.Боровиков Г.И. Statistica. Анализ и обработка данных в системе WINDOWS. – М.: Финансы и статистика, 1998.

5.Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистка, 2003.

6.Лукашин Ю.П. Регрессионные и адаптивные методы прогнозирования. Учебное пособие. – М.: МЭСИ, 1997.

Интернет-ресурсы:

1. http://www.kgtu.runnet.ru/WD/TUTOR/textbook/modules/stnonlin.html
2. http://softline.perm.ru/statistica/www-page_STATISTICA_for_Windows.html
3. http://ecfor.rssi.ru/0497_r_k.htm
4. http://www.socionet.ru:8100/RuPEc/data/Articles/rusrssicf4_1997article4.html -
5. http://www.mstu.edu.ru/publish/conf/section5/section5_7.html

Задания для самооценки

Выполните задания и ответьте на вопросы:

- Что представляет собой производственная функция Кобба-Дугласа?
- Как оценивать параметры производственной функции Кобба-Дугласа по пространственной и временной информации?

План практических занятий по теме 4

Занятие 1. Производственная функция Кобба-Дугласса

Решение задач.

Занятие 2.

Производственные функции, оценка параметров (2 часа).

Практические занятия проходят в компьютерных классах и посвящены построению и оцениванию параметров производственной функции Кобба-Дугласа с использованием пакетов программ STATISTICA и SPSS.

Тема 5. Системы одновременных эконометрических уравнений

Системы одновременных эконометрических уравнений являются третьим основным классом моделей, которые применяются для анализа и (или) прогноза. Эти модели описываются системами уравнений, которые могут состоять из тождеств и регрессионных уравнений, каждое из которых может, кроме объясняющих переменных, включать в себя также объясняемые переменные из других уравнений системы. Т.е. набор объясняемых переменных связан через уравнения системы.

Системы одновременных эконометрических уравнений

Данная тема знакомит студентов с основными понятиями систем одновременных эконометрических уравнений. Рассматриваются различные формы модели систем одновременных уравнений. Основными понятиями при изучении данного раздела являются структурная и приведенная формы модели систем одновременных уравнений, а также рекурсивные системы одновременных уравнений.

Применение МНК для оценки параметров систем одновременных уравнений

В данном разделе рассматривается двухшаговый и трехшаговый метод наименьших квадратов и их использование для оценивания параметров систем одновременных уравнений. Цель и особенности применения данных методов. Изучение данной темы базируется на знании студентами метода наименьших квадратов (МНК), умении использовать МНК для оценки параметров систем одновременных эконометрических уравнений.

Знания, умения, навыки по теме 5

Изучив тему 5, студенты должны знать:

- Что собой представляют систем одновременных эконометрических уравнений
- Что такое структурная и приведенная форма системы уравнений
- Как оценивать параметры рекурсивной системы одновременных эконометрических уравнений.
- Двухшаговый метод наименьших квадратов.
- Особенности применения трехшагового МНК.

уметь:

✓ Рассчитывать параметры систем одновременных эконометрических уравнений.

✓ Использовать двухшаговый и трехшаговый МНК для оценки параметров систем одновременных эконометрических уравнений.

✓ Применять рекурсивные системы одновременных эконометрических уравнений.

Получить навыки применения систем одновременных эконометрических уравнений в профессиональной деятельности.

Ссылки на учебный материал

1. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Методы исследования зависимостей. – М.: Финансы и статистика, 1983. – Т.1.

2.Айвазян С.А., Мхитарян В.С. Практикум по прикладной статистике и эконометрике. – М.: МЭСИ, 2000.

3.Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998.

4.Боровиков В.П., Ивченко Г.И. Statistica. Прогнозирование в системе WINDOWS. – М.: Финансы и статистика, 1999.

5. Боровиков Г.И. Statistica. Анализ и обработка данных в системе WINDOWS. – М.: Финансы и статистика, 1998.

Интернет-ресурсы:

1. http://subscribe.ru/archive/science.humanity.econometrika/200007/17050500.html
2. http://www.cemi.rssi.ru/rus/publicat/e-pubs/ep97001/1.htm
3. http://www.softlist.ru/cgi-bin/program.cgi?id=1988 - 7K
4. http://web.ido.ru/WWW/Courses.nsf/CoursesList?Open&About=055 - 2K
5. http://www.freeware32.ru/download.php3?id=1335 - 17K
6. http://www.nes.ru/Acad_year_2001/Prob_Stat.htm

План практических занятий по теме 5

Занятие 1.
Тема: Системы одновременных эконометрических уравнений.
Решение задач.

Условия идентифицируемости уравнений системы. Идентификация рекурсивных систем.

Занятие 2.

Тема: Модель спроса – предложения как пример системы одновременных уравнений. Основные структурные характеристики моделей.

На практических занятиях строится рекурсивная модель мирового рынка нефти, отражающая взаимосвязь между тремя основными элементами рыночного механизма – спросом, ценой и предложением (эндогенными переменными), которые, в свою очередь, определяются в каждый момент времени с помощью системы объясняющих, экзогенных переменных.

Особый интерес представляют показатели, обладающие опережающим эффектом (временным лагом) по отношению к динамике эндогенных переменных конъюнктуры рынка нефти.

При построении модели используется система показателей, основанная на ежеквартальных динамических рядах за 15 лет, которая характеризует основные стороны рынка нефти в экономическом, временном и географическом аспектах.

Задания для самооценки

Выполните задания и ответьте на вопросы:

- Почему возникла необходимость изучения данной темы?
- Какие системы одновременных эконометрических уравнений Вы знаете?
- Целесообразно ли применять методы корреляционного анализа на этапе предварительной обработки данных?
- В чем заключаются особенности использования трехшагового МНК для оценки параметров одновременных эконометрических уравнений?

5. Итоговый контроль знаний по курсу

Итоговая оценка знаний складывается по результатам тестирований, проводимых как по отдельным темам, так и по результатам итогового тестирования. Принимается во внимание также своевременность и качество выполнения студентами текущих заданий, активность участия на практических занятиях.

6. Список рекомендуемой литературы

Основная

1. Айвазян С.А., Мхитарян В.С. Практикум по прикладной статистике и эконометрике. – М.: МЭСИ, 2003.

2. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 2001.

3. Боровиков В.П., Ивченко Г.И. Statistica. Прогнозирование в системе WINDOWS. – М.: Финансы и статистика, 1999.

4. Боровиков Г.И. Statistica. Анализ и обработка данных в системе WINDOWS. – М.: Финансы и статистика, 1998.

5. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистка, 2003.

6. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. – М.: Статистка, 2003.

7. Лукашин Ю.П. Регрессионные и адаптивные методы прогнозирования: Учебное пособие. – М.: МЭСИ, 1997.

8. Мхитарян В.С., Дуброва Т.А., Ткачев А.В. Многомерная классификация с использованием пакета программ Statistica. Методическое пособие. – М.: МЭСИ, 1997.

9. Мхитарян В.С, Трошин Л.И. Статистический анализ многомерных совокупностей. Учебное пособие. – М.: МЭСИ, 1992.

Дополнительная

1. Айвазян С.А., Бежаева Э.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика, 1974.

2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности. – М.: Финансы и статистика, 1985, т. 2.

3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Методы исследования зависимостей. – М.: Финансы и статистика, 1983, т. 1.

4. Джонстон Дж. Эконометрические методы. – М.: Статистика, 1980.

5. Доугерти Кристофер Введение в эконометрику. – М.: ИНФРА-М, 1997.

6. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерный статистический анализ в экономических исследованиях. – М.: МЭСИ, 1988.

7. Иберла К. Факторный анализ. – М.: Статистика, 1980.

8. Иванова В.М. Эконометрика. – М.: Соминтек, 1991.

9. Клейнер Г. Производственные функции. – М.: ФиС, 1986.

10. Корнилов И.А. Исследование зависимостей с помощью пакетов программ статистического анализа для ЕС ЭВМ. – М.: МЭСИ, 1988.

11. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. – М.: Дело, 1997.

12. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988.

13. Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания. – М.: Статистика, 1980.

INTERNET-ресурсы

1. http://upereslavl.botik.ru/UP/ECON/econometrics/top1/tsld006.htm
2. http://www.nsu.ru/ef/tsy/ecmr/study.htm
3. http://www.nsu.ru/ef/tsy/ecmr/index.htm
4. http://www.statsoft.ru/home/textbook/def ault.htm
5. http://www.nsu.ru/ef/tsy/ecmr/study.htm
6. http://www.dataforce.net/~antl/article/econometric
7. http://www.nsu.ru/ef/tsy/ecmr/study.htm
8. http://www.tvp.ru/vnizd/mathem4.htm
9. http://www.kgtu.runnet.ru/WD/TUTOR/textbook/modules/stmulreg.html
10. http://www.shpargalka.ru/statis.ru/doc/shpr_e31.htm
11. http://www3.unicor.ac.ru/d024/p011993.htm
12. http://www.gauss.ru/educat/systemat/butenkov/.asp
13. http://crow.academy.ru/econometrics/seminars_/sem_08_/sem_08.htm
14. http://crow.academy.ru/econometrics/lectures_/lect_03_/index.htm
15. http://u-pereslavl.botik.ru/UP/ECON/econometrics/
16. http://crow.academy.ru/econometrics
17. http://www.econ.msu.ru/kaf/DEI/books/prognoz.html
18. http://socgw.univ.kiev.ua/EDUCAT/BASIC/MMPS/LABS/LOGREG.HTM
19. http://www.nsu.ru/ef/tsy/ecmr/some_mle/contents.htm
20. http://www.doktor.ru/doctor/biometr/sp/contents4.htm
21. http://www.kgtu.runnet.ru/WD/TUTOR/textbook/modules/stnonlin.html
22. http://dsmu.donetsk.ua/~statbook/modules/stnonlin.html
23. http://softline.perm.ru/statistica/wwwpage_STATISTICA_for_Windows.html
24. http://eco.rea.ru/LSpace/EconTh.nsf/136acc8cc4a429f5c325654d004b4fc2
25. http://lanserv2.kemsu.ru/departs/matekon/Chapter4/par4_4.html
26. http://www.dvgu.ru/pin/math/for_students/eco/node4.html
27. http://www.hse.ru/rectorat/grebnev/economics/glava13.htm
28. http://ecfor.rssi.ru/0497_r_k.htm
29. http://www.mstu.edu.ru/publish/conf/section5/section5_7.html
30. http://www.econ.msu.ru/kaf/DEI/books/prognoz/lec13.html
31. http://www.bsu.unibel.by/fpmi/bsa/ppp.htm
32. http://vega.math.spbu.ru/caterpillar/en/intro.html
33. http://iai.dn.ua/general/ai_annot.php
34. http://www.biophys.msu.ru/scripts/trans.pl/rus/cyrillic/awse/CONFER
35. http://ecfor.rssi.ru/0497_r_k.htm
36. http://www.codenet.al.ru/progr/packing/arithm/arithm10.htm
37. http://vm.fesma.ru/Gloss/Ag.htm
38. http://bytic.ttk.ru/cue99M/cz586tufhu.html
39. http://www.econ.msu.ru/kaf/DEI/books/prognoz/lec 13.html
40. http://www.bsu.unibel.by/fpmi/bsa/ppp.htm
41. http://www.tvp.ru/ourizd/oppm1996.htm
42. http://vega.math.spbu.ru/caterpillar/en/intro.html
43. http://www.tstu.ru/koi/tgtu/publ/1996/w96_35.htm
44. http://comsci.dsu.dp.ua/russian/curriculums/content/T_chos27.htm
45. http://www.nes.ru/~sanatoly/CV2000rus.htm
46. http://subscribe.ru/archive/science.humanity.econometrika/200007/17050500.html

47. http://www.cemi.rssi.ru/rus/publicat/e-pubs/ep97001/1.htm
48. http://www.freeware32.ru/download.php3?id=1335 – 17К
49. http://www.nes.ru/Acad_year_2001/Prob_Stat.htm
50. http://comsci.dsu.dp.ua/bgv/articles/statistd.htm
51. http://www.eu.spb.ru/econ/courses/s8-00.htm
52. http://www.math.dcn-asu.ru/md/k5/material/f3-3_ru.html
53. http://www.book.ru/cgi/expo.cgi?book=13141 – 2К
54. http://infovisor.ivanovo.ru:8000/rus/press/paper04.htm

7. Глоссарий

АДАПТИВНЫЕ МОДЕЛИ – математические модели, используемые в сочетании с человеко-машинными процедурами в принятии решений, в которых основываются лишь на предположении о существовании некоего обобщенного критерия задачи многокритериальной оптимизации, а необходимая дополнительная информация получается лицом, принимающим решение (ЛПР) последовательно, одновременно с анализом множества альтернатив. Применение А.м. целесообразно, когда ЛПР затрудняется в оценке вклада частных критериев в интегральный критерий. А.м. выгодны тем, что в процессе выработки решения используется информация, поступающая как от ЭВМ, так и от самого ЛПР. Важным преимуществом является и то, что перед специалистом последовательно проходит развитие модели многокритериальной ситуации от начального состояния к некоторому промежуточному (или окончательному) решению, что способствует более объективной оценке возможности улучшения значений обобщенных критериев. Существуют различные подходы к построению адаптивных человеко-машинных процедур.

АДДИТИВНЫЙ КРИТЕРИЙ (критерий справедливой абсолютной уступки) – критерий оптимальности, относящийся к группе прямых априорных методов многокритериальной оптимизации, сформулированный в виде суммирования выходных параметров (критериев оценки) исследуемого объекта. В А.к. используется следующий прием свертки нескольких критериев k_{ij} в обобщенный критерий (тенденция которого может быть к min или к max):

$$u(x_i) = \sum_{j=1}^{m} \omega_j \bar{k}_{ij}; \qquad i = \overline{1,n}$$
,

или в матричной форме – U = K ґ W, где w_j – коэффициент важности (вес) j-го критерия;

U = [u (x_i)] – матрица-столбец критерия u; $K = \left[\overline{k_{ij}}\right]$– нормализованная критериальная матрица; W = [w_j] – матрица-столбец коэффициентов веса. Принцип справедливой абсолютной уступки приводит к утверждению, что оптимальное решение означает максимизацию суммы нормированных частных критериев. Метод имеет строгое математическое обоснование. Однако введение весовых коэффициентов создает существенные трудности, один из путей преодоления которых состоит в применении экспертных оценок.

АЛГОРИТМ – 1) совокупность предписаний, необходимая и достаточная для решения какой-либо конкретной задачи; 2) совокупность правил, определяющих эффективную процедуру решения любой задачи из некоторого заданного класса задач. Понятие А. использовалось в математике давно, но как математический объект исследуется в связи с решением ряда проблем оснований математики с 30-х гг. XX в. Тогда же были разработаны основные понятия теории алгоритмов. В связи с развитием ЭВМ и их широким применением понятие А. стало одним из центральных в прикладной математике. АЛГОРИТМИЧЕСКАЯ МОДЕЛЬ – математическая модель, представленная в форме алгоритма, перерабатывающего заданный набор входных данных в заданный набор выходных данных. А.м. применяют, когда использование аналитических (расчетных) моделей затруднено либо нецелесообразно. Частным видом А.м. являются имитационные модели.

АНАЛИЗ – 1) изучение, научное исследование чего-либо, основанное на расчленении целого на составные части; 2) исследование объектов и явлений окружающего мира, основанное на изучении их внутренней структуры, закономерностей поведения или внешнего проявления их свойств. Анализ в САПР – проектная процедура или группа проектных процедур, имеющая целью получение информации о свойствах заданного проектируемого объекта; 3) функция управления, предназначенная для изучения, систематизации, обобщения и оценки достигнутых результатов. На основании данных анализа выявляются узкие места в деятельности организации, оцениваются конечные результаты производственной деятельности, обосновываются управленческие решения.

АНАЛИТИЧЕСКАЯ МОДЕЛЬ – математическая модель, представляющая собой совокупность аналитических выражений и зависимостей, позволяющих оценивать определенные свойства моделируемого объекта. Аналитические модели могут относиться к функциональным моделям (совокупность явных зависимостей выходных величин от входных), геометрическим (совокупность уравнений поверхности и (или) линий, задающих геометрическую форму моделируемого объекта), к обеспечению программному.

АПРИОРНЫЕ МОДЕЛИ – математические модели, используемые в принятии многокритериальных решений, в которых структура и вид обобщенного критерия постулируются вначале, т.е. вся информация, позволяющая определить наилучшее решение, скрыта в формальной модели задачи.

АРХИВАТОР – программа или программный пакет, предназначенный для «сжатия» (архивации) файла или группы файлов с целью уменьшения занимаемого файлами дискового пространства.

ВРЕМЕННОЙ РЯД – это последовательность наблюдений, упорядоченных во времени (или пространстве).

Если какое-нибудь явление наблюдают на протяжении некоторого времени, имеет смысл представить данные в том порядке, в котором они возникали, из-за того, в частности, что последовательные наблюдения могут быть зависимыми.

В. р. хорошо представлять на диаграмме рассеяния. Значение ряда X откладывают по вертикальной оси, а время t – по горизонтальной. Время называют независимой переменной.

Существует два типа временных рядов:

1. Непрерывные, в которых мы имеем наблюдения в каждый момент времени, например, показатели детектора лжи, электрокардиограммы. Их обозначают как наблюдение X в момент t, X(t).

2. Дискретные, в которых наблюдения делаются через некоторые (обычно одинаковые) интервалы времени. Их обозначают Xi.

Примеры

1. Экономические: недельные цены на акции; месячные прибыли.

2. Метеорологические: дневные осадки; скорость ветра; температура.

3. Социологические, показатели преступности (например, число арестов), показатели безработицы.

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ – это (как правило, лишь воображаемое) полное собрание объектов (людей, животных, растений или вещей), являющееся источником данных. Она представляет все множество статистических единиц (группу интересующих нас предметов).

Информацию о генеральной совокупности мы получаем, изучая выборки из нее; из каждой совокупности можно сделать много разных выборок. По выборке мы получаем информацию об интересующих нас параметрах совокупности.

Например, выборочное среднее дает информацию о среднем всей совокупности. Важно, чтобы перед формированием выборки исследователь тщательно и полно определил генеральную совокупность, а также способ извлечения выборки. Выборка должна быть репрезентативной.

ГЕТЕРОСКЕДАСТИЧНОСТЬ – условие, когда дисперсии регрессионных остатков не отвечают условию гомоскедастичности.

См. гомоскедастичность дисперсии.

ГИСТОГРАММА – это способ представления данных, измеренных в интервальной шкале (как дискретных, так и непрерывных). Часто используется в разведочном анализе данных для иллюстрации основных характеристик распределения. Гистограмма делит диапазон возможных значений множества данных на классы, или группы. Каждой группе соответствует прямоугольник, длина которого равна диапазону значений в заданной группе, а площадь пропорциональна числу наблюдений в этой группе. Это означает, что прямоугольники скорее всего будут различаться по высоте.

Гистограмма годится только для числовых переменных, измеренных в номинальной шкале. Как правило, она используется для больших множеств данных (>100 наблюдений), когда не хотят строить диаграммы ствол-лист.

Гистограммы помогают выявить необычные наблюдения (выбросы) и пропуски в множестве данных.

ГОМОСКЕДАСТИЧНОСТЬ – условие постоянства дисперсий регрессионных остатков.

КОРРЕЛЯЦИЯ – когда говорят, что две случайные переменные коррелированны, имеют в виду, как правило, что они друг с другом как-то связаны. Стандартной мерой связи переменных является коэффициент корреляции. Следует, однако, помнить, что он измеряет лишь силу линейной связи.

КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ – меняется в пределах от -1 до 1, измеряет степень линейной связи двух случайных переменных. Положительное значение коэффициента корреляции означает, что с ростом одной из переменных другая также растет, с убыванием одной из них убывает и другая. Отрицательное значение означает, что с ростом одной из переменных другая убывает, с убыванием одной из них другая растет. Коэффициент корреляции, равный нулю, означает, что между нашими переменными отсутствует линейная связь.

Обратите внимание: даже если коэффициент корреляции равен 1 по абсолютной величине и, следовательно, переменные функционально связаны (линейно), ничего нельзя сказать о причинно-следственной связи между ними.

В статистической практике в ходу два коэффициента корреляции: для числовых переменных используется коэффициент корреляции Пирсона, для ранговых— коэффициент корреляции Спирмена.

КРИТЕРИЙ СОГЛАСИЯ – проверяют гипотезу о совпадении наблюденной эмпирической функции распределения с теоретической функцией постулируемого распределения. Критерий согласия хи-квадрат делает это путем сравнения наблюденных и ожидаемых частот. Критерий Колмогорова – Смирнова основывается на максимальной разности между эмпирической и постулируемой функциями распределения.

ЛИНЕЙНАЯ РЕГРЕССИЯ – в линейной регрессии модельное (теоретическое, предсказанное) значение Y является линейной комбинацией значений одного или более предикторов:

$$У = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k.$$

МЕДИАНА выборки – это точка, по обе стороны которой располагается одинаковое количество элементов выборки. Если объем выборки нечетен и равен 2n + 1, то медиана равна элементу вариационного ряда с номером n + 1. Если объем выборки четен и равен 2n, то медиана равна полусумме элементов вариационного ряда с номерами n и n + 1.

Пример:

Для нечетного количества данных, скажем, 21, имеем:

Данные 96 48 27 72 39 70 7 68 99 36 95 4 6 13 34 74 65 42 28 54 69

Вариационный ряд: 4 6 7 13 27 28 34 36 39 42 48 54 65 68 69 70 72 74 95 96 99

Медиана равна 48, 10 значений ляжет выше нее, и 10 – ниже.

Для четного количества данных, скажем, 20, мы имеем:

Данные: 57 55 85 24 33 49 94 2 8 51 71 30 91 6 47 50 65 43 41 7

Вариационный ряд 2 6 7 8 24 30 33 41 43 47 49 50 51 55 57 65 71 85 91 94

Медианой в данном случае является среднее двух «серединных» точек, в данном случае среднее между 47 и 49 = 48.

Медиана распределения – это точка m, определяемая аналогичным условием: вероятность того, что случайная величина примет значение, не превосходящее m, равна 1/2.

Медиана выборки является оценкой медианы распределения.

Медиана является робастной оценкой центральной тенденции.

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ – это распространенный метод оценивания параметров. Ищутся оценки, минимизирующие сумму квадратов отклонений между смоделированными (предсказанными) и наблюденными значениями.

МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ – это общий метод вычисления оценок параметров. Ищутся такие оценки, чтобы функция правдоподобия выборки, равная произведению значений функции распределения для каждого наблюденного значения данных, была как можно большей.

М. м. п. лучше работает на больших выборках, где он, как правило, дает оценки с минимальной дисперсией. На маленьких выборках оценки максимального правдоподобия часто оказываются смещенными.

МУЛЬТИКОЛЛИНЕАРНОСТЬ – два или более предиктора коллинеарны, если сильна линейная связь между ними; их можно представить в виде линейной комбинации друг друга. Мультиколлинеарность может сделать проводимые для линейной регрессии вычисления неустойчивыми, а то и невозможными, поскольку в этом случае матрицы плохо обусловлены. Кроме того, она может вызвать завышенные оценки стандартных ошибок для коэффициентов при предсказывающих переменных.

НЕЗАВИСИМОСТЬ – две случайные переменные независимы, если их совместная плотность распределения равна произведению отдельных (маргинальных) плотностей. Менее формально: две случайные переменные A и B независимы, если информация о значении B не влияет на распределение вероятностей значений A, и наоборот. Выборка взаимно независимых случайных переменных называется независимой выборкой.

НЕЗАВИСИМАЯ ПЕРЕМЕННАЯ – переменная, используемая для объяснения зависимой переменной.

Синонимы: предиктор, объясняющая переменная.

Смотрите также зависимую переменную.

НЕЛИНЕЙНАЯ РЕГРЕССИЯ – предполагается, что зависимость отклика от предикторов является нелинейной функцией предикторов.

ОДНОРОДНОСТЬ – равенство дисперсий переменной, подсчитанных в пределах разных групп. Является стандартным требованием в таких, например, методах, как регрессионный и дисперсионный анализы.

Синоним: гомоскедастичность.

Антоним: гетероскедастичность.

ПРЕОБРАЗОВАНИЕ значений данных – производится путем применения одной и той же функции ко всем значениям переменной; важно то, что аргументами такой функции могут являться только значения переменных текущего наблюдения.

Распространенными примерами таких операций являются: добавление константы, умножение на константу, взятие логарифма.

ПРОИЗВОДСТВЕННАЯ ФУНКЦИЯ (production function) – отражает зависимость между количеством применяемых ресурсов и максимально возможным объемом выпускаемой продукции в единицу времени; описывает всю совокупность технически эффективных способов производства (технологий).

СТАНДАРТИЗОВАННАЯ ПЕРЕМЕННАЯ –

Переход от переменной x к переменной $z = \frac{x - \bar{x}}{S}$,

где $\bar{x}$ – среднее значение, а s – среднее квадратическое отклонение, называется стандартизацией переменной x, а результат – стандартизованной переменной z . Часто говорят также о z-преобразовании и даже z-значениях переменной x.

Название восходит к стандартному нормальному распределению N(0,1): ведь если x подчиняется нормальному распределению, то для больших выборок ее стандартизация приводит к z – подчиняющейся стандартному нормальному распределению.

СЕЗОННАЯ КОМПОНЕНТА – один из способов описания временного ряда – разложение его на компоненты: тренд, периодическую и случайную. Когда временная ось связана с датами, а период – с месяцами или кварталами, периодическую компоненту называют сезонной.

СГЛАЖИВАНИЕ, ФИЛЬТРАЦИЯ – сглаживание применяется для уменьшения иррегулярности (случайных изменений) временных рядов. Распространенным методом сглаживания является сглаживание простым скользящим средним (хотя существуют и другие способы). Способ сглаживания определяется свойствами ряда и целями его обработки.

СТАТИСТИКА – это функция элементов выборки. Дает информацию о неизвестных значениях параметров генеральной совокупности. Например, среднее выборки является, как правило, оценкой среднего совокупности, из которой была взята выборка.

СТАТИСТИЧЕСКАЯ НЕЗАВИСИМОСТЬ – отсутствие связи между переменными. Независимость двух непрерывных переменных часто ошибочно отождествляют с равенством нулю их корреляции (ковариации), однако это верно, только если они подчиняются двумерному нормальному распределению.

СТАТИСТИЧЕСКИЙ КРИТЕРИЙ – статистический критерий состоит из следующих компонент: пара гипотез – нулевая и альтернативная, статистика критерия и уровень значимости; по ним находится критическая область.

Проверка гипотезы начинается с вычисления статистики. Если значение попадает в критическую область, мы отвергаем нулевую гипотезу и считаем истинной ее альтернативу. В противном случае у нас нет оснований отвергнуть нулевую гипотезу.

Из генеральной совокупности можно сделать много разных выборок, причем значение статистики в общем случае будет меняться от выборки к выборке; другими словами, выборка является случайной, а значит, случайной величиной является и статистика. Например, выборочные средние для разных выборок из одной и той же совокупности могут различаться между собой.

Статистики обычно обозначают латинскими буквами, а оцениваемые ими параметры – греческими.

СТАЦИОНАРНЫЕ показатели – показатели, среднее которых можно считать неизменным; нестационарными – называются показатели, среднее которых изменяется со временем.

Системы одновременных эконометрических уравнений являются третьим основным классом моделей, которые применяются для анализа и (или) прогноза. Эти модели описываются системами уравнений, которые могут состоять из тождеств и регрессионных уравнений, каждое из которых может, кроме объясняющих переменных, включать в себя также объясняемые переменные из других уравнений системы. Т.е. набор объясняемых переменных связан через уравнения системы.

ТАБЛИЦА СОПРЯЖЕННОСТИ – таблица (ТС), каждый элемент (клетка) которой соответствует клетке кросс-табуляции. В случае двух факторов клетки ТС располагают так, чтобы элементы одной строки соответствовали одному и тому же значению одного фактора, а элементы одного столбца – одному и тому же значению другого фактора; говорят, что уровни одного фактора расположены по строкам, а другого – по столбцам. Такие таблицы часто обозначают rxc, где r – количество уровней фактора, соответствующего строкам, с – столбцам.

В случае трех факторов считают, что ТС состоит из совокупности ТС, каждая из которых соответствует значению третьего фактора, являясь при этом (условной) ТС первых двух факторов. Можно, конечно, построить ТС и для большего числа факторов.

В каждой клетке ТС стоит количество элементов соответствующей клетки кросс-табуляции.

ТС – не слишком удобный способ представления данных для их визуального анализа, если велико количество уровней факторов, тем более, если велико количество факторов.

Для проверки гипотезы о независимости факторов, по которым построена кросс-табуляция, используется критерий независимости хи-квадрат Пирсона. Для таблиц 2х2 (два фактора, по два уровня у каждого) используется также точный критерий Фишера.

Общий метод анализа таблиц сопряженности – лог-линейный анализ.

ТРЕНД – для лучшего понимания временного ряда мы выделяем его основные характеристики. Одной из таких характеристик является тренд.

Тренд – это долговременное изменение временного ряда. Это направление (тенденция к повышению или снижению) и скорость изменения временного ряда, при сделанных допущениях о других компонентах.

ЦИКЛИЧЕСКАЯ КОМПОНЕНТА – чтобы лучше понять поведение временного ряда, выделяют его основные характеристики. Одной из таких характеристик является циклическая компонента. В недельных или месячных данных циклическая компонента описывает любые регулярные колебания. Это не сезонная компонента, изменение которой подчиняются некоторому распознаваемому циклу.

ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ – метод сглаживания временного ряда, используемый для уменьшения иррегулярности (случайных колебаний) временного ряда, что позволяет получить более ясное представление о лежащих в основе этого ряда закономерностях. Используется также для прогнозирования значения ряда (для 1-2 шагов) прогноза.

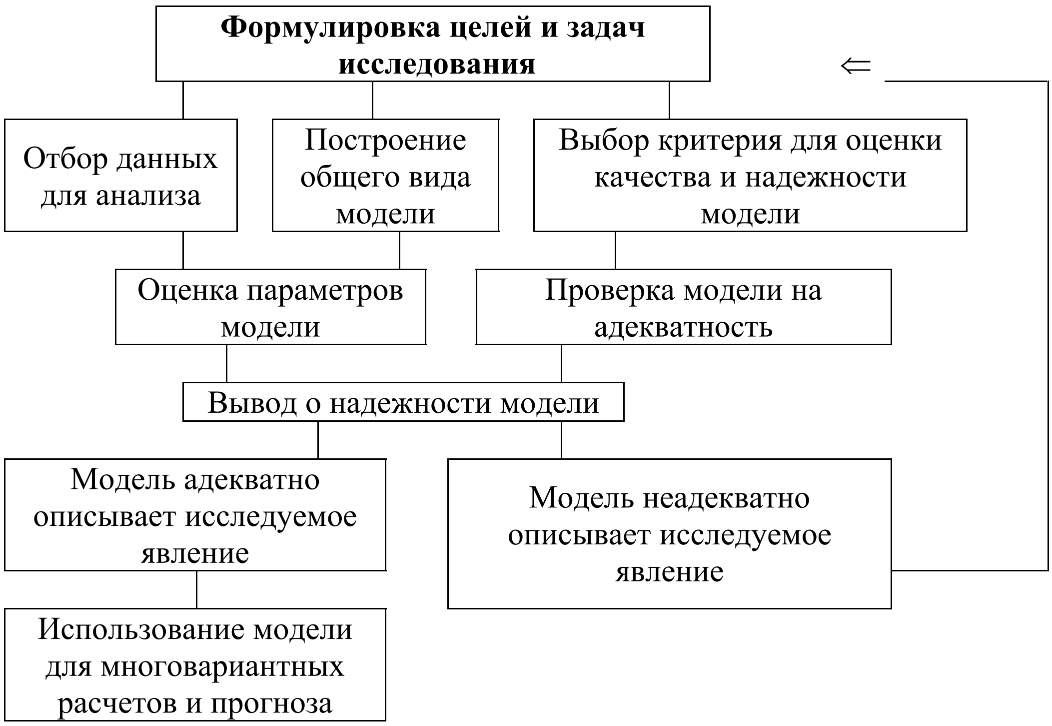
ЭКСТРАПОЛЯЦИЯ – предсказание значения переменной за пределами интервала анализа. Термин применяется, как правило, при анализе временных рядов. Для коротких промежутков времени применяются количественные предсказания, интерполяции.

Количественное предсказание далекого будущего, как правило, менее полезно и применяется для указания на необходимость изменения построенной модели.

Методические указания и задания для студентов-заочников

Введение

Эконометрика – это дисциплина, объединяющая совокупность теоретических результатов, методов и приемов, позволяющих на базе экономической теории, экономической статистики и математико-статистического инструментария получать количественное выражение качественных закономерностей. Эконометрический подход характеризуется большим вниманием, в вопросе соответствия выбранной модели изучаемому объекту, рассмотрению причин, приводящих к необходимости пересмотра модели на основе более точной системы представлений. Эконометрика занимается, по существу, статистическими выводами, т. е. использованием выборочной информации для получения некоторого представления о свойствах генеральной совокупности. Процесс эконометрического исследования в общем виде можно представить в виде следующей схемы.



Данный курс посвящен изучению основных разделов эконометрики, наиболее часто используемых на практике.

Пособие содержит программу курса, список основной и дополнительной литературы, подробно разобранные варианты решения контрольных задач, глоссарий, а также задания для самостоятельного решения.

Перед выполнением задания необходимо изучить учебное пособие «Эконометрика» и проверить уровень приобретенных знаний с помощью тестовых заданий.

1. Цели, задачи изучения, сфера профессионального применения

Цель преподавания курса – дать студентам научное представление о методах, моделях и приемах, позволяющих получать количественные выражения закономерностей экономической теории на базе экономической статистики с использованием математико-статистического инструментария. Современные социально-экономические процессы и явления зависят от большого количества факторов, их определяющих. В связи с этим квалифицированному специалисту необходимо не только иметь четкие представления об основных направлениях развития экономики, но и уметь учитывать сложное взаимосвязанное многообразие факторов, оказывающих существенное влияние на изучаемый процесс. Такие исследования невозможно проводить без знания основ теории вероятностей, математической статистики, многомерных статистических методов и эконометрики, т.е. дисциплин, позволяющих исследователю разобраться в огромном количестве стохастической информации и среди множества различных вероятностных моделей выбрать единственную, наилучшим образом отражающую изучаемый процесс или явление.

Задачи курса. Научиться строить экономические модели и оценивать их параметры;

научиться проверять гипотезы о свойствах экономических показателей и формах их связи.

В соответствии с целью студенты должны освоить методы количественной оценки социально-экономических процессов, научиться содержательно интерпретировать формальные результаты, моделировать с помощью пакетов прикладных программ STATISTICA, SPSS и др.

В курсе описываются формы и типы классических эконометрических моделей, соотношения между ними, их достоинства и недостатки. Рассматриваются предпосылки построения моделей, а также задачи их спецификации и идентификации.

Связь с другими дисциплинами. Для изучения курса эконометрики студентам необходимо знание основ:

теории статистики, в которой сформулированы общие методы и принципы определения количественных характеристик массовых процессов и явлений;

экономической статистики, дающей представление о направлениях развития экономики, о темпах роста цен и занятости, о тенденциях развития и эффективности использования ресурсов в отдельных отраслях и секторах экономики;

линейной алгебры для проведения расчетов над матрицами;

высшей математики, обучающей приемам интегрирования и дифференцирования;

математической статистики, определяющей генеральную и выборочную совокупность, вариационные ряды и их характеристики; методы статистического оценивания параметров и статистической проверки гипотез (статистические критерии); методы корреляционно-регрессионного анализа для исследования взаимосвязи между зависимой переменной и группой влияющих на нее показателей;

многомерных статистических методов, позволяющих выделять латентные факторы, сжимать признаковое пространство и сопоставлять изучаемые процессы в пространстве латентных факторов, проводить многомерную классификацию;

владеть приемами статистического анализа нечисловой информации.

В свою очередь данный курс является основой для дисциплины «Эконометрическое моделирование».

2. Содержание курса

Тема 1. Корреляционный анализ

Основная задача корреляционного анализа состоит в оценке корреляционной матрицы генеральной совокупности по выборке и определении на ее основе оценок частных и множественных коэффициентов корреляции и детерминации. Корреляционный анализ является этапом, предшествующим регрессионному анализу.

Тема 2. Регрессионный анализ

Изучая данную тему, необходимо акцентировать внимание на понятиях линейной модели множественной регрессии (ОЛММР); обобщенного метода наименьших квадратов (ОМНК); обобщенной линейной модели множественной регрессии (ОЛММР) с гомоскедастичными и гетероскедастичными случайными остатками, а также обобщенной линейной модели множественной регрессии с автокоррелированными остатками.

Также в данной теме необходимо обратить внимание на вопросы использования в регрессионных моделях переменных, не носящих количественный характер. Выяснить причины применения таких переменных в экономических моделях, специфику их нахождения.

Тема 3. Нелинейные модели регрессии и их линеаризация

Данная тема посвящена рассмотрению нелинейных моделей, используемых для описания взаимосвязи экономических показателей. В регрессионном анализе вид уравнения регрессии выбирают исходя из анализа физической сущности изучаемого явления и результатов наблюдения. При изучении данной темы следует акцентировать внимание на различных вариантах уравнений регрессии и способах их преобразования к линейному виду.

Тема 4. Методы многомерной классификации. Кластерный анализ

Изучая данную тему, необходимо акцентировать внимание на задачах анализа неоднородных в некотором смысле данных. В таких случаях, прежде чем переходить к построению регрессионных моделей, необходимо выделить однородные группы объектов и уже внутри каждой группы строить регрессионные зависимости.

Тема 5. Производственные функции. Система одновременных эконометрических уравнений

Изучая данную тему, необходимо акцентировать внимание на анализе систем одновременных уравнений, примерах использования таких систем для моделирования различных экономических взаимосвязей. Необходимо выяснить причины невозможности использования стандартных методов оценки характерных для индивидуальных уравнений.

3. Список рекомендуемой литературы

Основная литература

1. Айвазян С.А., Мхитарян В.С. Практикум по прикладной статистике и эконометрике. – М.: МЭСИ, 2003.
2. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 2001.
3. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистка, 2003.
4. Мхитарян В.С., Архипова М.Ю. Эконометрика. – М.: МЭСИ, 2005.

Дополнительная литература

1. Айвазян С.А., Бежаева Э.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика, 1974.
2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности. – М.: Финансы и статистика, 1985, т. 2.

3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Методы исследования зависимостей. – М.: Финансы и статистика, 1983, т. 1.
4. Джонстон Дж. Эконометрические методы. – М.: Статистика, 1980
5. Доугерти Кристофер Введение в эконометрику. – М.: ИНФРА-М, 1997.
6. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерный статистический анализ в экономических исследованиях. – М.: МЭСИ, 1988.
7. Иберла К. Факторный анализ. – М.: Статистика, 1980.
8. Иванова В.М. Эконометрика. – М.: Соминтек, 1991.
9. Катышев П.К., Магнус Я.Р., Пересецкий А.А. Сборник задач к начальному курсу Эконометрики. – М.: Дело, 2002.
10. Клейнер Г. Производственные функции. – М.: ФиС, 1986.
11. Корнилов И.А. Исследование зависимостей с помощью пакетов программ статистического анализа для ЕС ЭВМ. – М.: МЭСИ, 1988.
12. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. – М.: Дело, 1997.
13. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988.
14. Практикум по эконометрике/ Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002.
15. Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания. – М.: Статистика, 1980.

4. Примеры решения типовых задач

4.1. Корреляционный анализ

Анализ взаимосвязи социально-экономических показателей группы стран

В ходе корреляционного анализа выявляется статистическая взаимосвязь между признаками и отбираются переменные для включения в регрессионную модель. Предпосылками корреляционного анализа являются случайность признаков и нормальный многомерный закон их совместного распределения. Поэтому необходимым условием для его проведения является однородность выборки, простейший способ обеспечения которой – группировка объектов по общности их основных свойств.

По данным 1995 года о 20 бывших и нынешних социалистических странах, взятых из таблицы ПРИЛОЖЕНИЯ 1, рассчитана матрица выборочных парных коэффициентов корреляции

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,879	-0,758	-0,556	0,767	-0,600	0,826	-0,580	0,698
x_2		1	0,817	0,710	-0,591	0,631	-0,676	0,406	-0,514
x_3			1	0,717	-0,515	0,664	-0,615	0,433	-0,466
x_4				1	-0,249	0,624	-0,329	0,313	-0,057
x_5					1	-0,604	0,963	-0,865	0,851
x_6						1	-0,658	0,612	-0,419
x_7							1	-0,833	0,906
x_8								1	-0,637
x_9									1

Исследуемый признак – x_1 – детская смертность (число умерших младенцев на 1000 новорожденных).

Требуется:

1) Проверить значимость каждого из коэффициентов на уровне значимости $\alpha = 0{,}05$.

2) Определить признаки, наиболее важные для объяснения вариации исследуемой переменной, рассчитать выборочные частные коэффициенты корреляции исследуемого признака с каждым из признаков при фиксированном значении остальных. Найти интервальные оценки частных коэффициентов корреляции, определить значимость коэффициентов. Сравнить частные коэффициенты корреляции с соответствующими парными и сделать выводы относительно роли исключенной переменной в изменении степени тесноты статистической связи, характеризуемой этими коэффициентами корреляции.

3) Рассчитать значение множественного коэффициента корреляции исследуемого признака с выбранными в п.2 признаками. Найти коэффициент детерминации, проверить его значимость.

Решение:

1) Определим по таблице Фишера-Йейтса критическое значение $r_{кр}$ для одного из наиболее часто использующихся уровней значимости $\alpha=0{,}05$. С учетом объема выборки n=20 находим число степеней свободы $\nu=n-2=18$. По данным таблицы получаем $r_{кр} = 0{,}444$.

Для выборочных парных коэффициентов корреляции r_{ij}, абсолютная величина которых превосходит критическое значение, отвергается гипотеза о равенстве нулю соответствующих им истинных коэффициентов корреляции (H_0: $\rho_{ij}=0$), и они считаются значимыми. Остальные истинные значения коэффициентов корреляции от нуля существенно не отличаются. Подчеркнем значимые коэффициенты корреляции

	X_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	$\underline{-0{,}879}$	$\underline{-0{,}758}$	$\underline{-0{,}556}$	$\underline{0{,}767}$	$\underline{-0{,}600}$	$\underline{0{,}826}$	$\underline{-0{,}580}$	$\underline{0{,}698}$
x_2		1	$\underline{0{,}817}$	$\underline{0{,}710}$	$\underline{-0{,}591}$	$\underline{0{,}631}$	$\underline{-0{,}676}$	0,406	$\underline{-0{,}514}$
x_3			1	$\underline{0{,}717}$	$\underline{-0{,}515}$	$\underline{0{,}664}$	$\underline{-0{,}615}$	0,433	$\underline{-0{,}466}$
x_4				1	-0,249	$\underline{0{,}624}$	-0,329	0,313	-0,057
x_5					1	$\underline{-0{,}604}$	$\underline{0{,}963}$	$\underline{-0{,}865}$	$\underline{0{,}851}$
x_6						1	$\underline{-0{,}658}$	$\underline{0{,}612}$	-0,419
x_7							1	$\underline{-0{,}833}$	$\underline{0{,}906}$
x_8								1	$\underline{-0{,}637}$
x_9									1

С вероятностью $1-\alpha=0{,}95$ можно утверждать наличие статистически значимой связи между i-м и j-м признаками, выборочный парный коэффициент корреляции которых r_{ij} значим. Связь между другими признаками с такой мерой уверенности не установлена (что, впрочем, не дает оснований говорить о ее отсутствии).

2) Среди признаков, которые могут обусловливать вариацию детской смертности, выделим уровень грамотности населения (x_4) и среднее число детей в семье (x_9). Соответствующие парные коэффициенты корреляции значимы и свидетельствуют о наличии существенной связи между этими переменными и исследуемой переменной. Ограничив корреляционную модель исследуемой переменной и двумя выбранными признаками, запишем для нее матрицу парных коэффициентов корреляции, взяв значения коэффициентов из общей корреляционной матрицы

	x_1	x_4	x_9
x_1	1	-0,556	0,698
x_4	-0,556	1	-0,057
x_9	0,698	-0,057	1

Вычислим выборочные частные коэффициенты корреляции $r_{14(9)}$ и $r_{19(4)}$ по формуле

$$r_{14(9)} = \frac{-R_{14}}{\sqrt{R_{11}R_{44}}},$$

где A_{ij} – алгебраическое дополнение элемента r_{ij} матрицы выборочных парных коэффициентов корреляции.

В данном случае, благодаря небольшой размерности матрицы, несложно получить расчетное соотношение в аналитическом виде

$$r_{14(9)} = \frac{r_{14} - r_{19}r_{49}}{\sqrt{(1-r_{19}^2)(1-r_{49}^2)}}.$$

После подстановки значений получаем

$$r_{14(9)} = \frac{-0{,}556 - 0{,}698\cdot(-0{,}057)}{\sqrt{(1-0{,}698^2)(1-(-0{,}057)^2)}} = -0{,}722.$$

Аналогично определяем другой выборочный частный коэффициент корреляции

$$r_{19(4)} = \frac{r_{19} - r_{14}r_{49}}{\sqrt{(1-r_{14}^2)(1-r_{49}^2)}};\ r_{19(4)} = 0{,}803.$$

Выборочные частные коэффициенты корреляции $r_{14(9)}$ и $r_{19(4)}$ не отличаются по знаку от соответствующих парных коэффициентов r_{14} и r_{19}, но превосходят их по абсолютной величине. Следовательно, исключаемый признак x_9 ослабляет взаимосвязь между признаками x_1 и x_4, а признак x_4 ослабляет связь признаков x_1 и x_9.

Рассчитаем интервальные оценки парных коэффициентов корреляции. Определяемая значением выборочного коэффициента корреляции величина $z = \frac{1}{2}\ln\frac{1+r}{1-r}$, называемая z-преобразованием Фишера, распределена приближенно нормально с математическим ожиданием $z = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}$ и дисперсией $\sigma^2 = \frac{1}{n-m-2}$, где m – число исключенных величин, ρ – истинное значение коэффициента корреляции. Интервальная оценка для нормально распределенной величины определяется выражением

$$P\left(Z' - t_\gamma\sqrt{\frac{1}{n-l-3}} \le Z \le Z' + t_\gamma\sqrt{\frac{1}{n-l-3}}\right) = \Phi(t_\gamma),$$

где $\Phi(t_\gamma)$ – интеграл Лапласа,

$z' = \frac{1}{2}ln\frac{1+r}{1-r} + \frac{r}{2(n-1)}$ – несмещенная оценка математического ожидания.

Для выборочного частного коэффициента корреляции $r_{14(9)} = -0{,}722$ получаем $z' = -0{,}931$. Можно использовать приближенное значение без поправки на несмещенность, определяемое по таблице z-преобразования Фишера, $z' \approx -0{,}91$. Используя последнее значение и определив по таблице нормального закона распределения для $\Phi(t_\gamma)=1-\alpha=0{,}95$ величину $t_\gamma=1{,}96$, получаем

$$P(-1{,}4 < z < -0{,}42) = 0{,}95.$$

По таблице z-преобразования Фишера находим значения коэффициента корреляции ρ, соответствующие границам интервала величины z, и определяем его интервальную оценку

$$P(-0{,}89 < \rho_{14(9)} < -0{,}40) = 0{,}95.$$

В интервале возможных значений частного коэффициента корреляции нуль не содержится, поэтому с вероятностью 0,95 можно утверждать, что частный коэффициент корреляции нулю не равен. Диапазон возможных значения частного коэффициента корреляции показывает, что между детской смертностью и уровнем грамотности взрослого населения существует обратная линейная статистическая зависимость, степень тесноты которой либо умеренная, либо сильная.

Аналогично получим интервальную оценку для другого частного коэффициента корреляции

$$P(0{,}71 < \rho_{19(4)} < 0{,}92) = 0{,}95.$$

Этот коэффициент также является значимым, а диапазон его значений указывает на прямую зависимость детской смертности от среднего числа детей в семье.

Рассчитаем значение выборочного множественного коэффициента корреляции исследуемого признака x_1 по формуле

$R_1 = \sqrt{1 - \dfrac{|R|}{R_{11}}}$, где $|R|$ – определитель матрицы выборочных парных коэффициентов корреляции.

Расчетное аналитическое соотношение будет иметь вид

$$R_1 = \sqrt{\frac{r_{14}^2 + r_{19}^2 - 2r_{14}r_{49}r_{19}}{1 - r_{49}^2}}.$$

Подставим значения выборочных парных коэффициентов корреляции и получим

$$R_1 = \sqrt{\frac{(-0{,}556)^2 + 0{,}698^2 - 2\cdot(-0{,}556)\cdot 0{,}698\cdot(-0{,}057)}{1-(-0{,}057)^2}} = 0{,}869.$$

Рассчитанный коэффициент является выборочным значением множественного коэффициента корреляции – максимального среди взятых по модулю парных коэффициентов корреляции переменной x_1 с линейными комбинациями признаков x_4 и x_9. Квадрат множественного коэффициента корреляции – коэффициент детерминации $\rho_{1(49)}^2$ – показывает долю дисперсии исследуемой случайной переменной, обусловленную вариацией включенных в модель признаков. Выборочное значение коэффициента детерминации $R_1^2 = r_{1(49)}^2 = 0{,}755$. Остальные 24,5% дисперсии исследуемой переменной обусловлены действием признаков, не включенных в модель. С помощью F-критерия определим значимость коэффициента детерминации, проверив гипотезу H_0: $\rho^2_{1(49)} = 0$. Вычислим значение F-статистики

$$F_н = \frac{r^2_{1(49)}/2}{(1-r^2_{1(49)})/(20-3)}$$

Рассчитанное значение $F_н = 26{,}16$ сравним с критическим $F_{кр} = 3{,}59$, найденным по таблице Фишера – Снедекора для уровня значимости α = 0,05 и числа степеней свободы числителя $\nu_1 = 2$ и знаменателя $\nu_2 = n-3 = 17$.

Так как рассчитанное значение превышает критическое, проверяемая гипотеза отвергается, и с вероятностью 1-α=0,95 можно утверждать, что множественный коэффициент корреляции $\rho_{1(49)}$ не равен нулю. Следовательно, существует статистически значимая связь детской смертности с уровнем грамотности взрослого населения и средним числом детей в семье.

4.2. Регрессионный анализ

Регрессионная модель уровня детской смертности

В ходе регрессионного анализа выявляется форма и параметры зависимости одного из признаков, называемого зависимой переменной, от других – объясняющих переменных, считающихся неслучайными величинами. Зависимая переменная представляет собой наиболее важный из практических соображений признак. Отбор признаков для использования в качестве объясняющих переменных производится на основе анализа их содержательной сущности и результатов корреляционного анализа. При этом из признаков, связанных зависимостью, близкой к неслучайной функциональной, выбирают какой-либо один во избежание эффекта мультиколлинеарности объясняющих переменных. Выбор вида уравнения регрессии определяется сущностью изучаемого явления. Простейшей из регрессионных моделей является линейная. Оценка параметров уравнения входит в число важнейших задач регрессионного анализа. Наряду с нахождением значений параметров оценивается их точность, проверяется значимость уравнения и его коэффициентов.

По данным 1995 года о 20 бывших и нынешних социалистических странах, взятых из таблицы ПРИЛОЖЕНИЯ 1, наряду с приведенной выше матрицей выборочных парных коэффициентов корреляции, построены уравнения регрессии. В этих уравнениях зависимой переменной является социально значимый признак x_1 – детская смертность (число умерших младенцев на 1000 новорожденных). В качестве объясняющих переменных использованы признаки в различных комбинациях

$$\underset{}{\hat{y}} = \underset{(42,430)}{99{,}891} - \underset{(0,200)}{0{,}225x_3} - \underset{(0,564)}{0{,}957x_4} + \underset{(1,005)}{0{,}215x_6} + \underset{(3,738)}{12{,}994x_9}; \qquad R^2=0{,}774;\ F=12{,}883;$$

$$\hat{y} = \underset{(12,652)}{31{,}134} - \underset{(0,128)}{0{,}497x_3} + \underset{(3,241)}{9{,}939x_9}; \qquad R^2=0{,}726;\ F=22{,}556;$$

$$\hat{y} = \underset{(12,945)}{30{,}980} - \underset{(0,161)}{0{,}445x_3} - \underset{(0,989)}{0{,}493x_6} + \underset{(3,362)}{9{,}661x_9}; \qquad R^2=0{,}730;\ F=14{,}455;$$

$$\hat{y} = \underset{(31,207)}{121,093} - \underset{(0,314)}{1,354}x_4 + \underset{(2,718)}{15,099}x_9; \qquad R^2=0,775; \quad F=26,159.$$

Для каждого уравнения рассчитаны значения коэффициентов детерминации и F-статистик. Под коэффициентами приведены значения их выборочных средних квадратических отклонений.

Требуется:

1) Используя критерий Фишера, проверить на уровне α=0,05 значимость каждого из уравнений регрессии. В значимых уравнениях рассчитать значения t-статистик всех коэффициентов. Переписать уравнения регрессии, указывая под коэффициентами значения t-статистик.

2) По таблице распределения Стьюдента определить $t_{кр}$ – критическое значение t-статистики для каждого из уравнений на уровне значимости α=0,05. Проверить значимость коэффициентов уравнения регрессии.

3) Выбрать из предложенных уравнений наилучшее. Рассчитать интервальные оценки его коэффициентов. Произвести анализ уравнения.

Решение:

1) Для каждого из уравнения определим $F_{кр}$ – критическое значение F-статистики по таблице Фишера – Снедекора при уровне значимости α = 0,05 и числе степеней свободы числителя p, а знаменателя ν=n-p-1, где p – число регрессоров в уравнении. Получаем

$$F_{кр1}(0,05;4;15)=3,06; \qquad F_{кр2}(0,05;2;17)= F_{кр4}=3,59; \qquad F_{кр3}(0,05;3;16)=3,24.$$

Значения F-статистик всех уравнений превышают соответствующие критические значения. Следовательно, все уравнения являются статистически значимыми.

Для проверки значимости коэффициентов проверим гипотезу о равенстве нулю каждого истинного значения β каждого из них H_0: β=0. Для этого вычислим по выборочному значению b каждого коэффициента и его выборочному среднему квадратическому отклонению S статистику

$$t_н = \frac{b-0}{S} = \frac{b}{S}.$$

Для первого коэффициента первого уравнения $t_н = 99,891/42,430 = 2,354$.

Вычислим значения остальных t-статистик и запишем уравнения с указанием их значений

$$\hat{y} = \underset{(2,354)}{99,891} - \underset{(-1,125)}{0,225}x_3 - \underset{(-1,696)}{0,957}x_4 + \underset{(0,210)}{0,215}x_6 + \underset{(3,476)}{12,994}x_9; \qquad R^2=0,774; F=12,883;$$

$$\hat{y} = \underset{(2,461)}{31,134} - \underset{(-3,856)}{0,497}x_3 + \underset{(3,067)}{9,939}x_9; \qquad R^2=0,726; F=22,556;$$

$$\hat{y} = \underset{(2,393)}{30,980} - \underset{(-2,770)}{0,445}x_3 - \underset{(-0,499)}{0,493}x_6 + \underset{(2,871)}{9,661}x_9; \qquad R^2=0,730; \quad F=14,455;$$

$$\hat{y} = \underset{(3,880)}{121,093} - \underset{(-4,309)}{1,354}x_4 + \underset{(5,554)}{15,099}x_9; \qquad R^2=0,775; \quad F=26,159.$$

2) Критические значения t-статистик обычно лежат в интервале от 2 до 3. Рассчитаем их для каждого уравнения по таблице распределения Стьюдента для уровня значимости α=0,05 и числа степеней свободы ν=n–p-1, где p – число регрессоров в уравнении.

$t_{кр1}(0{,}05;15) = 2{,}131$; $t_{кр2}(0{,}05;17) = t_{кр4}(0{,}05;17) = 2{,}110$; $t_{кр3}(0{,}05;16) = 2{,}120$.

Сравним абсолютные величины t-статистик с критическими значениями.

Если $|t_н| > t_{кр}$, то с вероятностью $1-\alpha=0{,}95$ истинный коэффициент уравнения регрессии нулю не равен, и соответствующий признак влияет на вариацию зависимой переменной. В противном случае предположение о нулевом значении коэффициента и, следовательно, об отсутствии влияния регрессора на поведение зависимой переменной не противоречит имеющимся данным, и такой коэффициент считается незначимым.

Выделим значимые коэффициенты в каждом уравнении

$$\hat{y} = \underset{(\underline{2,354})}{\mathbf{99,891}} - \underset{(-1,125)}{0,225}x_3 - \underset{(-1,696)}{0,957}x_4 + \underset{(0,210)}{0,215}x_6 + \underset{(\underline{3,476})}{\mathbf{12,994}}x_9; \quad t_{кр}=2,131;\ R^2=0,774;\ F=12,883;$$

$$\hat{y} = \underset{(\underline{2,461})}{\mathbf{31,134}} - \underset{(\underline{-3,856})}{\mathbf{0,497}}x_3 + \underset{(\underline{3,067})}{\mathbf{9,939}}x_9; \quad t_{кр}=2,110;\ R^2=0,726;\ F=22,556;$$

$$\hat{y} = \underset{(\underline{2,393})}{\mathbf{30,980}} - \underset{(\underline{-2,770})}{\mathbf{0,445}}x_3 - \underset{(-0,499)}{0,493}x_6 + \underset{(\underline{2,871})}{\mathbf{9,661}}x_9; \quad t_{кр}=2,120;\ R^2=0,730;\ F=14,455;$$

$$\hat{y} = \underset{(\underline{3,880})}{\mathbf{121,093}} - \underset{(\underline{-4,309})}{\mathbf{1,354}}x_4 + \underset{(\underline{5,554})}{\mathbf{15,099}}x_9; \quad t_{кр}=2,110;\ R^2=0,775;\ F=26,159.$$

Во втором и четвертом уравнениях все коэффициенты значимы.

3) Для практического использования пригодны лишь уравнения со значимыми коэффициентами при регрессорах. Выберем из соответствующих данному условию уравнений то, которое характеризуется наибольшей величиной коэффициента детерминации R^2,

$$\hat{y} = \underset{(\underline{3,880})}{\mathbf{121,093}} - \underset{(\underline{-4,309})}{\mathbf{1,354}}x_4 + \underset{(\underline{5,554})}{\mathbf{15,099}}x_9; \quad t_{кр}=2,110;\ R^2=0,775;\ F=26,159.$$

Рассчитаем интервальные оценки его коэффициентов

$$P(b - t_\alpha S_b < \beta < b + t_\alpha S_b) = \gamma.$$

По таблице распределения Стьюдента для доверительной вероятности $\gamma=1-\alpha=0{,}95$ найдем с учетом числа степеней свободы $\nu=n-k-1$ значение $t_\alpha = t_{0,05} = 2{,}110$. С учетом приведенных в исходных данных значений выборочных средних квадратических отклонений S_b коэффициентов определим интервальную оценку коэффициента b_0

$$P(b_0 - t_\alpha S_{b0} < \beta < b_0 + t_\alpha S_{b0}) = \gamma,$$

$$P(121{,}093 - 2{,}110 \cdot 31{,}207 < \beta_0 < 121{,}093 + 2{,}110 \cdot 31{,}207) = 0{,}95,$$

$$P(55{,}246 < \beta_0 < 186{,}940) = 0{,}95$$

и остальных коэффициентов

$$P(-2{,}017 < \beta_1 < -0{,}691) = 0{,}95,$$

$$P(9{,}364 < \beta_2 < 20{,}834) = 0{,}95\,.$$

Нуль не содержится ни в одном из рассчитанных интервалов возможных значений коэффициентов уравнения регрессии, что еще раз свидетельствует о значимости каждого из коэффициентов.

С увеличением уровня грамотности населения на один процент детская смертность снижается в среднем на 1,354 событий на 1000 новорожденных, при этом с вероятностью 0,95 снижение составляет в худшем случае 0,691, а лучшем случае 2,017 событий на 1000 новорожденных.

Рост среднего числа детей в семье на одного ребенка сопровождается увеличением детской смертности в среднем на 15,099 событий на 1000 новорожденных. Соответственно, с вероятностью 0,95 в худшем случае это увеличение произойдет на 20,834, а в лучшем – на 9,364 событий на 1000 новорожденных.

4.3. Нелинейные регрессионные модели

При анализе расходов на продовольственные товары в общих расходах (%) семи семей Москвы в зависимости от среднемесячной заработной платы (тыс. руб.) получены следующие данные, представленные в табл.4.3.1

Таблица 4.3.1

Номер семьи	Расходы на покупку продовольственных товаров в общих расходах (%)	Среднемесячная заработная плата работающих (тыс. руб.)
1	61,2	59,0
2	59,9	57,2
3	56,7	61,8
4	55,0	58,8
5	54,3	47,2
6	49,3	55,2
7	68,8	45,1

Требуется найти оценки параметров следующих функций:

a. *равносторонней гиперболы*
b. *степенной*
c. *показательной*

Применение метода наименьших квадратов возможно в случае использования функций с линейными параметрами. Поэтому прежде, чем переходить к нахождению неизвестных коэффициентов уравнения регрессии, необходимо привести функцию с нелинейными параметрами к линейному виду.

1. Уравнение равносторонней гиперболы $\boldsymbol{y=a+b} \cdot \frac{1}{x}$ приводится к линейному виду путем замены переменных: $z = \frac{1}{x}$.

Тогда исходное уравнение принимает следующий вид: $y = a + b \cdot z$.

Для расчетов используем вспомогательную табл. 4.3.2.

Таблица 4.3.2

	y	z	yz	z^2	y^2	$\hat{y}_x$	$y-\hat{y}_x$	$(y-\hat{y}_x)^2$
1	61,2	0,0169	1,0373	0,000287	3745,44	56,3	4,9	24,01
2	59,9	0,0175	1,0472	0,000306	3588,01	56,9	3,0	9,00
3	56,7	0,0162	0,9175	0,000262	3214,89	55,5	1,2	1,44
4	55	0,0170	0,9354	0,000289	3025,00	56,4	-1,4	1,96
5	54,3	0,0212	1,1504	0,000449	2948,49	60,8	-6,5	42,25
6	49,3	0,0181	0,8931	0,000328	2430,49	57,5	-8,2	67,24
7	68,8	0,0222	1,5255	0,000492	4733,44	61,8	7,0	49,00
Итого	405,2	0,1291	7,5064	0,002413	23685,76	405,2	0,0	194,90
Среднее значение	57,9	0,0184	1,0723	0,000345	3383,68	-	-	27,84
S	5,74	0,002145	-	-	-	-	-	-
S^2	32,9476	0,000005	-	-	-	-	-	-

Значения параметров регрессии a и b получим, используя следующие формулы:

$$b = \frac{\overline{yz} - \overline{y} \times \overline{z}}{s_Z^2} = \frac{1.0723 - 57.9 \cdot 0.0184}{0.002145^2} \approx 1051.4,$$

$$a = \overline{y} - b \cdot \overline{z} \ = 57,89 - 1051,4.0,0184 = 38,5.$$

Искомое уравнение регрессии будет иметь следующий вид: $\hat{y} = 38,5 + 1051,4 \cdot \frac{1}{x}$.

2. Для построения степенной модели $\boldsymbol{y = a \cdot x^b}$ сначала необходимо провести линеаризацию переменных, путем логарифмирования обеих частей уравнения:

$$lgy = lga + b \cdot lgx;$$

Обозначив через $Y=lgy$, $X=lgx$, $C=lga$,

получим: $Y=C+b \cdot X$.

Для расчетов используем данные вспомогательной табл. 4.3.3.

Таблица 4.3.3

	Y	X	YX	Y^2	X^2	$\hat{y}_x$	$y-\hat{y}_x$	$(y-\hat{y}_x)^2$
1	1,7868	1,7709	3,1642	3,1927	3,1361	56,3	4,9	24,0
2	1,7774	1,7574	3,1236	3,1592	3,0885	56,8	3,1	9,6
3	1,7536	1,7910	3,1407	3,0751	3,2077	55,5	1,2	1,4
4	1,7404	1,7694	3,0795	3,0290	3,1308	56,3	-1,3	1,7
5	1,7348	1,6739	2,9039	3,0095	2,8019	60,2	-5,9	34,8
6	1,6928	1,7419	2,9487	2,8656	3,0342	57,4	-8,1	65,6
7	1,8376	1,6542	3,0398	3,3768	2,7364	61,0	7,8	60,8
Итого	12,3234	12,1587	21,4003	21,7078	21,1355	403,5	1,7	197,9
Среднее значение	1,7605	1,7370	3,0572	3,1011	3,0194	-	-	28,27
s	0,0425	0,0484	-	-	-	-	-	-
s^2	0,0018	0,0023	-	-	-	-	-	-

Рассчитаем C и b по формулам:

$$b = \frac{\overline{YX} - \overline{Y} \times \overline{X}}{s_x^{\,2}} = \frac{3{,}0572 - 1.7605 \cdot 1.7370}{0.0484^2} \approx -0.298\,,$$

$$C = \overline{Y} - b \cdot \overline{X} = 1{,}7605 + 0{,}298 \cdot 1{,}7370 = 2{,}278.$$

Получим линейное уравнение: $\hat{Y} = 2{,}278 - 0{,}298 \cdot X$.

Далее, выполнив его потенцирование, получим:

$$\hat{y} = 10^{2.278} \cdot x^{-0.298} = 189.7 \cdot x^{-0.298}.$$

3. Построению уравнения показательной кривой $y = a \cdot b^x$ предшествует процедура линеаризации переменных при логарифмировании обеих частей уравнения

$$lgy = lga + x \cdot lgb$$

Обозначив Y= lg y, C= lg a, B = lg b,

получим: Y=C+B $\cdot$ x.

Для расчетов используем данные табл. 4.3.4.

Таблица 4.3.4

	y	x	Yx	Y^2	X^2	$\hat{y}_x$	$y-\hat{y}_x$	$(y-\hat{y}_x)^2$
1	1,7868	59,0	105,4212	3,1927	3481,00	56,4	4,8	23,04
2	1,7774	57,2	101,6673	3,1592	3271,84	56,9	3,0	9,00
3	1,7536	61,8	108,3725	3,0751	3819,24	55,5	1,2	1,44
4	1,7404	58,8	102,3355	3,0290	3457,44	56,4	-1,4	1,96
5	1,7348	47,2	81,8826	3,0095	2227,84	60,0	-5,7	32,49
6	1,6928	55,2	93,4426	2,8656	3047,04	57,5	-8,2	67,24
7	1,8376	45,1	82,8758	3,3768	2034,01	60,7	8,1	65,61
Итого	12,3234	384,3	675,9974	21,7078	21338,41	403,4	-1,8	200,78
Среднее значение	1,7605	54,9	96,5711	3,1011	3048,34	-	-	28,68
s	0,0425	5,86	-	-	-	-	-	-
s^2	0,0018	34,3396	-	-	-	-	-	-

Значения параметров регрессии A и B составили:

$$B = \frac{\overline{Y \cdot x} - \overline{Y} \cdot \overline{x}}{s^2{}_X} = \frac{96.5711 - 1.7605 \cdot 54.9}{5.86^2} \approx -0.0023,$$

$$A = \overline{Y} - B \cdot \overline{x} = 1{,}7605 + 0{,}0023 \cdot 54{,}9 = 1{,}887.$$

Получено линейное уравнение: $\hat{Y} = 1{,}887 - 0{,}0023 \cdot x$.

Произведем потенцирование полученного уравнения и запишем его в обычной форме:

$$\hat{y} = 10^{1.887} \,.\, 10^{-0.0023x} = 77{,}1 \cdot 0{,}9947^{x}.$$

4.4. Методы многомерной классификации. Кластерный анализ

Классификация семей по анализируемой структуре расходов

По данным, представленным в табл. 4.4.1, провести классификацию n=5 семей по двум показателям: уровень расходов (млн. руб.) за летние месяцы на культурные нужды, спорт и отдых - $x^{(1)}$ и питание $x^{(2)}$:

Таблица 4.4.1.

№ семьи (i)	1	2	3	4	5
$x_i^{(1)}$	2	4	8	12	13
$x_i^{(2)}$	10	7	6	11	9

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного и взвешенного (w_1=0,05; w_2=0,95) евклидова расстояния, а также принципов: “ближайшего” и “дальнего” соседа, центра тяжести и средней связи.

Сравнить полученные результаты и обосновать выбор окончательного варианта классификации.

Примечание: На основании предварительного качественного анализа было выдвинуто предположение, что по потребительскому поведению три первые семьи принадлежат одной типологической группе, а две последние (4 и 5) – другой, что согласуется с расположением пяти наблюдений на плоскости, представленных на рис.1.

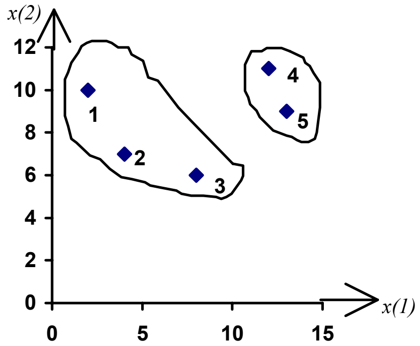


Рис.1. **Исходные данные для классификации**

Решение. *1. Проведем классификацию, выбрав при обычном евклидовом расстоянии принцип “ближайшего соседа”.*

Согласно обычной евклидовой метрике расстояние между наблюдениями 1 и 2 равно

$$\rho_{1,2} = \sqrt{\sum_{j=1}^{p}(x_1^{(j)} - x_2^{(j)})^2} = \sqrt{(2-4)^2 + (10-7)^2} = 3.61, \qquad (1)$$

очевидно, что $\rho_{1,1}$=0.

Аналогично находим расстояния между всеми пятью наблюдениями и строим матрицу расстояний

$$\mathrm{R}_1 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}05 & 11{,}05 \\ 3{,}61 & 0 & 4{,}12 & 8{,}94 & 9{,}22 \\ 7{,}21 & 4{,}12 & 0 & 6{,}40 & 5{,}83 \\ 10{,}05 & 8{,}94 & 6{,}40 & 0 & 2{,}24 \\ 11{,}05 & 9{,}22 & 5{,}83 & 2{,}24 & 0 \end{pmatrix}.$$

Из матрицы расстояний следует, что объекты 4 и 5 наиболее близки $\rho_{4,5}$=2,24 и поэтому объединим их в один кластер. После объединения объектов имеем четыре кластера: S_1, S_2, S_3, $S_{(4,5)}$.

Расстояние между кластерами будем находить по принципу “ближайшего соседа”, воспользовавшись формулой пересчета. Так, расстояние между кластером S_1 и кластером $S_{(4,5)}$ равно:

$$\rho_{1,(4,5)} = \rho(S_1, S_{(4,5)}) = \frac{1}{2}\rho_{1,4} + \frac{1}{2}\rho_{1,5} - \frac{1}{2}\left|\rho_{1,4} - \rho_{1,5}\right| = \\ = \frac{1}{2}\left(10{,}05 + 11{,}05\right) - \frac{1}{2}\left|10{,}05 - 11{,}05\right| = 10{,}05. \quad (2)$$

Мы видим, что расстояние $\rho_{1,(4,5)}$ равно расстоянию от объекта 1 до ближайшего к нему объекта, входящего в кластер $S_{(4,5)}$, т.е. $\rho_{1,(4,5)} = \rho_{1,4}$=10,05. Проводя аналогичные расчеты, получим матрицу расстояний

$$R_2 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}05 \\ 3{,}61 & 0 & 4{,}12 & 8{,}94 \\ 7{,}21 & 4{,}12 & 0 & 5{,}83 \\ 10{,}05 & 8{,}94 & 5{,}83 & 0 \end{pmatrix}.$$

Объединим наблюдения 1 и 2, имеющие наименьшее расстояние $\rho_{1,2}$=3,61. После объединения имеем три кластера $S_{(1,2)}$, S_3 и $S_{(4,5)}$.

Вновь строим матрицу расстояний. Для этого необходимо рассчитать расстояние до кластера $S_{(1,2)}$. Воспользуемся матрицей расстояний R_2.

Например, расстояние между кластерами $S_{(4,5)}$ и $S_{(1,2)}$ равно:

$$\rho_{(4,5),(2,3)} = \frac{1}{2}\rho_{(4,5),1} + \frac{1}{2}\rho_{(4,5),2} - \frac{1}{2}\left|\rho_{(4,5),1} - \rho_{(4,5),2}\right| = \\ = \frac{1}{2}10{,}05 + \frac{1}{2}8{,}94 - \frac{1}{2}\left(10{,}05 - 8{,}94\right) = 8{,}94.$$

Как видим, оно равно расстоянию от кластера $S_{(4,5)}$ до ближайшего объекта, входящего в кластер $S_{(1,2)}$, то есть $\rho_{(4,5),(1,2)} = \rho_{(4,5),2}$=8,94.

Проведя аналогичные расчеты, получим матрицу расстояний

$$R_3 = \begin{pmatrix} 0 & 4{,}12 & 8{,}94 \\ 4{,}12 & 0 & 5{,}83 \\ 8{,}94 & 5{,}83 & 0 \end{pmatrix}.$$

Объединим кластеры $S_{(1,2)}$ и $S_{(3)}$, расстояние между которыми, согласно матрице $\mathbf{R}_3$, минимально $\rho_{(1,2),3}$=4,12. В результате этого получим два кластера: $S_{(1,2,3)}$ и $S_{(4,5)}$. Матрица расстояний будет иметь вид:

$$R_4 = \begin{pmatrix} 0 & 5{,}83 \\ 5{,}83 & 0 \end{pmatrix}.$$

Из матрицы $\mathbf{R}_4$ следует, что на расстоянии $\rho_{(1,2,3),(4,5)}$=5,83 все пять наблюдений объединяются в один кластер. Результаты кластерного анализа представим графически в виде дендрограммы (рис.2).

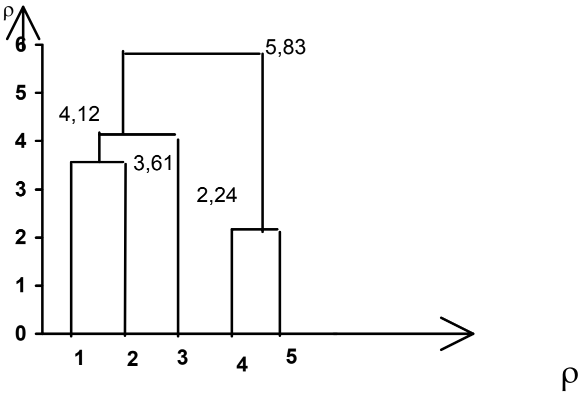


Рис. 2. **Дендрограмма**
(обычное евклидово расстояние, “ближайший сосед”)

На основании графического представления результатов кластерного анализа (рис.4.2) можно сделать вывод, что наилучшим является разбиение пяти семей на два кластера: $S_{(1,2,3)}$ и $S_{(4,5)}$, когда пороговое расстояние находится в интервале 4,12< $\rho_{пор}$<5,83.

Напомним, что выводы нами сделаны для случая, когда для измерения расстояния выбрано “обычное евклидово расстояние” и принцип “ближайшего соседа”.

2. Проведем классификацию, выбрав при обычном евклидовом расстоянии принцип “дальнего соседа”.

Как и в случае (1), мы используем обычное евклидово расстояние, поэтому матрица $\mathbf{R}_1$ остается без изменения. Согласно агломеративному алгоритму объединяются в один кластер объекты 4 и 5, как наиболее близкие $\rho_{4,5}$=2,24. После объединения имеем четыре кластера: S_1, S_2, S_3 и $S_{(4,5)}$.

В виду того, что расстояние между кластерами измеряем по принципу “дальнего соседа”, воспользуемся формулой пересчета, приняв δ=1/2, а не -1/2, как в случае (1). Тогда, например, расстояние между кластером S_1 и кластером $S_{(4,5)}$ определяется по формуле

$$\rho_{1,(4,5)} = \rho(S_1, S_{(4,5)}) = \frac{1}{2}\rho_{1,4} + \frac{1}{2}\rho_{1,5} + \frac{1}{2}\left|\rho_{1,4} - \rho_{1,5}\right| =$$
$$= \frac{1}{2}(10{,}05 + 11{,}05) + \frac{1}{2}\left|10{,}05 - 11{,}05\right| = 11{,}05.$$

Таким образом, расстояние $\rho_{1,(4,5)}$ равно расстоянию от объекта 1 до наиболее отдаленного от него объекта, входящего в кластер $S_{(4,5)}$, то есть $\rho_{1,(4,5)}$= $\rho_{1,5}$=11,05.

Аналогично рассматриваются все остальные элементы матрицы расстояния

$$R_2 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 11{,}05 \\ 3{,}61 & 0 & 4{,}12 & 9{,}22 \\ 7{,}21 & 4{,}12 & 0 & 6{,}40 \\ 11{,}05 & 9{,}22 & 6{,}40 & 0 \end{pmatrix}.$$

Объединим объекты 1 и 2 в один кластер, как наиболее близкие (согласно матрице R_2), $\rho_{1,2}$=3,61.

После объединения имеем три кластера: $S_{(1,2)}$, S_3 и $S_{(4,5)}$.

Строим матрицу расстояний $\mathbf{R}_3$, воспользовавшись принципом “дальнего соседа”.

$$R_3 = \begin{pmatrix} 0 & 7{,}21 & 11{,}05 \\ 7{,}21 & 0 & 6{,}40 \\ 11{,}05 & 6{,}40 & 0 \end{pmatrix}.$$

Объединим кластеры S_3 и $S_{(4,5)}$, расстояние между которыми $\rho_{3,(4,5)}$=6,40 минимально, и получим два кластера: $S_{(1,2)}$ и $S_{(3,4,5)}$, расстояние между которыми определяется по матрице $\mathbf{R}_4 = \begin{pmatrix} 0 & 11{,}05 \\ 11{,}05 & 0 \end{pmatrix}$ и равно $\rho_{(1,2),(3,4,5)}$=11,05.

Графические результаты классификации представлены на рис.3.

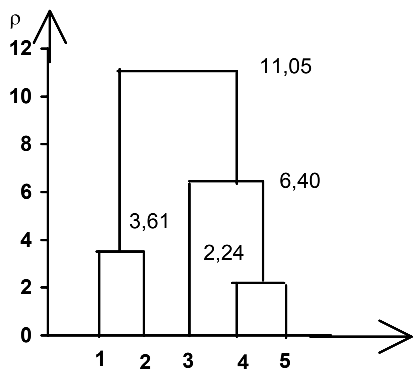


Рис. 3. **Дендрограмма**
(обычное евклидово расстояние, “дальний сосед”)

Как и в предыдущем случае наилучшим является разбиение семей на два кластера (рис. 4.3): $S_{(1,2)}$ и $S_{(3,4,5)}$, после предпоследнего шага классификации, когда интервал измерения расстояния объединения наибольший 6,40≤ $\rho_{пор}$≤11,05.

Таким образом, используя принцип “дальнего соседа”, мы получили разбиение регионов на два кластера $S_{(1,2)}$ и $S_{(3,4,5)}$, которое отличается от разбиения по принципу “ближайшего соседа”: $S_{(1,2,3)}$ и $S_{(4,5)}$.

3. Классификация на основе обычного евклидова расстояния и принципа “центра тяжести”.

Так как мы используем обычное евклидово расстояние, то матрица $\mathbf{R}_1$ остается без изменения. Согласно агломеративному алгоритму объединяются в кластер $S_{(4,5)}$ объекты 4 и 5, как наиболее близкие $\rho_{4,5}$=2,24.

Кластер $S_{(4,5)}$ характеризуется в дальнейшем его центром тяжести, определяемым вектором средних $\overline{X}_{(4,5)} = \begin{pmatrix} 12{,}5 \\ 10 \end{pmatrix}$.

Расстояние от этого кластера до первого наблюдения равно

$$\rho_{(4,5),1} = \sqrt{(12{,}5-2)^2 + (10-10)^2} = 10{,}50\,.$$

Тогда матрица расстояний примет вид

$$R_2 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}50 \\ 3{,}61 & 0 & 4{,}12 & 9{,}01 \\ 7{,}21 & 4{,}12 & 0 & 6{,}02 \\ 10{,}50 & 9{,}01 & 6{,}02 & 0 \end{pmatrix}.$$

Объединим объекты 1 и 2, расстояние между которыми $\rho_{1,2}$=3,61 минимальное. Кластер характеризуется центром тяжести $\overline{X}_{1,2} = \begin{pmatrix} 3 \\ 8{,}5 \end{pmatrix}$, расстояния от которого до кластера $S_{(4,5)}$ равно

$$\rho_{(1,2),(4,5)} = \sqrt{(3-12{,}5)^2 + (8{,}5-10)^2} = 9{,}62\,.$$

Тогда матрица расстояний примет вид

$$R_3 = \begin{pmatrix} 0 & 5{,}59 & 9{,}62 \\ 5{,}59 & 0 & 6{,}02 \\ 9{,}62 & 6{,}02 & 0 \end{pmatrix}.$$

В матрице $\mathbf{R}_3$ минимальное расстояние $\rho_{(1,2),3}$=5,59, поэтому образуем кластер $S_{(1,2,3)}$ и определим его вектор средних $\overline{X}_{(1,2,3)} = \begin{pmatrix} \frac{2+4+8}{3} \\ \frac{10+7+6}{3} \end{pmatrix} = \begin{pmatrix} 4{,}67 \\ 7{,}67 \end{pmatrix}$.

Найдем расстояние между $S_{(1,2,3)}$ и $S_{(4,5)}$

$$\rho_{(1,2,3),(4,5)} = \sqrt{(4{,}67-12{,}5)^2 + (7{,}67-10)^2} = 8{,}17\,,$$

на котором все пять объектов объединяются в один кластер. Графически результаты классификации представлены на рис. 4.

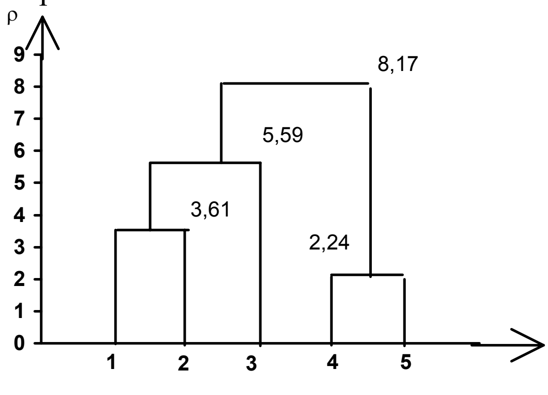


Рис. 4. **Дендрограмма**
(обычное евклидово расстояние, принцип “центра тяжести”)

Из рис.4. видно, что наибольший скачок в расстояниях объединения ρ имеет место на последнем шаге, поэтому целесообразно выбрать разбиение на два кластера $S_{(1,2,3)}$ и $S_{(4,5)}$, что совпадает со случаем (1).

4. Классификация на основе обычного евклидова расстояния и принципа "средней связи".

Используя матрицу $\mathbf{R}_1$, согласно агломеративному алгоритму объединим кластеры S_4 и S_5 в один $S_{(4,5)}$, так как расстояние между ними $\rho_{4,5}=2{,}24$ минимально.

Расстояние от кластера $S_{(4,5)}$ до остальных кластеров определим по принципу "средней связи" на основе матрицы $\mathbf{R}_1$. Например,

$$\rho_{(4,5),1}=\frac{1}{2}(\rho_{4,1}+\rho_{5,1})=\frac{1}{2}(10{,}05+11{,}05)=10{,}55\,.$$

Тогда матрица расстояний имеет вид

$$R_2=\begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}55 \\ 3{,}61 & 0 & 4{,}12 & 9{,}08 \\ 7{,}21 & 4{,}12 & 0 & 6{,}12 \\ 10{,}55 & 9{,}08 & 6{,}12 & 0 \end{pmatrix}.$$

Объединим, как наиболее близкие $\rho_{1,2}=3{,}61$, кластеры S_1 и S_2. Тогда расстояния от $S_{(1,2)}$ до остальных кластеров $S_{(4,5)}$ равны

$$\rho_{(1,2),(4,5)}=\frac{1}{4}(\rho_{1,4}+\rho_{1,5}+\rho_{2,4}+\rho_{2,5})=9{,}82\,,$$

а матрица расстояний имеет вид

$$R_3=\begin{pmatrix} 0 & 5{,}67 & 9{,}82 \\ 5{,}67 & 0 & 6{,}12 \\ 9{,}82 & 6{,}12 & 0 \end{pmatrix}.$$

Объединим, как наиболее близкие $\rho_{(1,2),3}=5{,}41$, кластеры $S_{(1,2)}$ и S_3 и определим расстояние от $S_{(1,2,3)}$ до $S_{(4,5)}$

$$\rho_{(1,2,3),(4,5)}=\frac{1}{6}(\rho_{1,4}+\rho_{1,5}+\rho_{2,4}+\rho_{2,5}+\rho_{3,4}+\rho_{3,5})=8{,}58\,,$$

на котором все пять объектов объединились в один кластер. Графически результаты классификации представлены на рис. 5

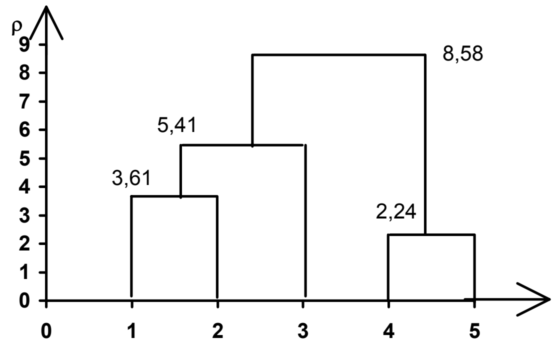


Рис. 5. **Дендрограмма**
(обычное евклидово расстояние, принцип "средней связи")

Анализ рис. 5 показывает, что целесообразным является разбиение на два кластера $S_{(1,2,3)}$ и $S_{(4,5)}$.

Таким образом, сравнивая результаты 4-х разбиений пяти семей на однородные группы, можно отметить, что наиболее устойчивым, а отсюда и предпочтительным, является разбиение на два кластера $S_{(1,2,3)}$ и $S_{(4,5)}$, что согласуется с рис. 4.1. Только в одном случае из четырех, при использовании принципа “дальнего соседа” получено разбиение $S_{(1,2)}$ и $S_{(3,4,5)}$.

Во всех предыдущих алгоритмах классификации мы предполагали, что оба показателя $x^{(1)}$ и $x^{(2)}$ одинаково значимы (использовалось обычное евклидово расстояние). Теперь откажемся от этого предположения.

5. Классификация на основе “взвешенного евклидова расстояния” и принципа “ближайшего соседа”.

Предположим, что показатель $x^{(1)}$ менее важен для классификации, чем $x^{(2)}$. В этой связи припишем им “веса” ω_1=0,05 и ω_2=0,9. Напомним, что взвешенное евклидово расстояние между i-м и l-м наблюдениями определяется по формуле

$$\rho_{BE}(x_i, x_l) = \sqrt{\sum_{j=1}^{p} (x_i^{(j)} - x_l^{(j)})^2 \cdot \omega_j} \,. \qquad (3)$$

Тогда расстояние между объектами 1 и 2 равно

$$\rho_{1,2} = \sqrt{(2-4)^2 \cdot 0{,}05 + (10-7)^2 \cdot 0{,}95} = 2{,}96\,.$$

Аналогично находим все остальные расстояния и строим матрицу расстояний

$$R_1 = \begin{pmatrix} 0 & 2{,}96 & 4{,}12 & 2{,}44 & 2{,}65 \\ 2{,}96 & 0 & 1{,}32 & 4{,}29 & 2{,}80 \\ 4{,}12 & 1{,}32 & 0 & 4{,}95 & 3{,}13 \\ 2{,}44 & 4{,}29 & 4{,}95 & 0 & 1{,}96 \\ 2{,}65 & 2{,}80 & 3{,}13 & 1{,}96 & 0 \end{pmatrix}.$$

Объединив S_2 и S_3, имеющих минимальное расстояние $\rho_{2,3}$=1,32, в кластер $S_{2,3}$ и применив принцип “ближайшего соседа”, получим матрицу расстояний R

$$R_2 = \begin{pmatrix} 0 & 2{,}96 & 2{,}44 & 2{,}65 \\ 2{,}96 & 0 & 4{,}29 & 2{,}80 \\ 2{,}44 & 4{,}29 & 0 & 1{,}96 \\ 2{,}65 & 2{,}80 & 1{,}96 & 0 \end{pmatrix}.$$

Образовав на расстоянии $\rho_{4,5}$=1,96 кластер $S_{4,5}$, вновь построим матрицу расстояний

$$R_3 = \begin{pmatrix} 0 & 2{,}96 & 2{,}44 \\ 2{,}96 & 0 & 2{,}80 \\ 2{,}44 & 2{,}80 & 0 \end{pmatrix}.$$

Объединим S_1 и $S_{4,5}$, имеющих минимальное расстояние $\rho_{(4,5)1}$=2,43, в кластер $S_{(1,4,5)}$ и получим матрицу расстояний

$$R_4 = \begin{pmatrix} 0 & 2{,}80 \\ 2{,}80 & 0 \end{pmatrix}.$$

Следовательно, на расстояние $\rho_{(1,4,5),(2,3)}$=2,80 объединяются кластеры $S_{(1,4,5)}$ и $S_{(2,3)}$ и все пять объектов образуют один кластер.

Результаты классификации представлены графически на рис.6.

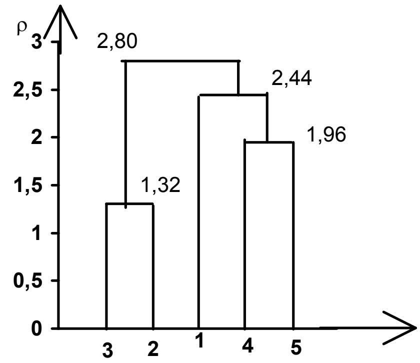


Рис. 6. **Дендрограмма**
(взвешенное евклидово расстояние, принцип "ближайшего соседа")

Как и прежде, отдадим предпочтение разбиению на два кластера; мы получаем третий вариант разбиения, а именно $S_{(2,3)}$ и $S_{(1,4,5)}$.

Таким образом, использовав пять алгоритмов кластерного анализа, мы получили три варианта разбиения пяти семей на две статистически однородные группы. С одной стороны, это свидетельствует о гибкости (возможностях) методов кластерного анализа, а с другой – о необходимости использования экономических (содержательных) и статистических критериев для выбора наилучшего варианта классификации. При этом часто бывает полезной априорная информация об исследуемом явлении. В нашем примере, окончательно следует остановиться на разбиении $S_{(1,2,3)}$ и $S_{(4,5)}$, как наиболее устойчивом. Это разбиение получено по трем алгоритмам из пяти. Кроме того, оно согласуется с данными априорного, качественного анализа.

5. Контрольные задания

Задание №1.

Тема « Корреляционный анализ»

1. По данным своего варианта определите критическое значение $r_{кр}$ для выборочных парных коэффициентов корреляции, представленных в корреляционной матрице, по таблице Фишера-Йейтса и проверьте значимость каждого из коэффициентов на уровне значимости $\alpha = 0,05$.

2. Определите два признака, с Вашей точки зрения наиболее важные для объяснения вариации исследуемого признака. Рассчитайте выборочные частные коэффициенты корреляции исследуемого признака с каждым из них при фиксированном значении другого. Найдите интервальные оценки частных коэффициентов корреляции, определите значимость коэффициентов. Сравните частные коэффициенты корреляции с соответствующими парными. Сделайте выводы относительно роли исключенной переменной в изменении степени тесноты статистической связи, характеризуемой этими коэффициентами корреляции.

3. Рассчитайте значение множественного коэффициента корреляции исследуемого признака с выбранными в п.2 признаками Определите коэффициент детерминации, проверьте его значимость.

Задание №2.

Тема « Регрессионный анализ»

1. Используя критерий Фишера, проверьте на уровне значимости α=0,05 значимость каждого уравнения регрессии из исходных данных для Вашего варианта. В значимых уравнениях рассчитайте значения t-статистик всех коэффициентов, используя значения выборочных средних квадратических отклонений, приведенных под каждым из коэффициентов. Перепишите уравнения регрессии, указывая под коэффициентами значения t-статистик.

2. По таблице распределения Стьюдента определите $t_{кр}$ – критическое значение t-статистики для каждого из уравнений на уровне значимости α=0,05. Проверьте значимость коэффициентов уравнения регрессии.

3. Выберите из предложенных уравнений наилучшее. Рассчитайте интервальные оценки его коэффициентов. Произведите анализ уравнения.

Задание 3.

Тема «Нелинейные регрессионные модели»

По представленным в таблице данным требуется найти оценки параметров нелинейных функций.

Задание 4.

Тема « Методы многомерной классификации объектов»

По представленным в таблице данным провести классификацию n = 4 предприятий по двум показателям. Классификацию провести по иерархическому агломеративному алгоритму, используя обычное евклидово расстояние. Расстояние между кластерами определять по принципу, указанному для каждого варианта.

6. Варианты заданий

Номер варианта определяется последней цифрой номера зачетной книжки

ВАРИАНТ 1

Задание 1,2.

Матрица парных коэффициентов корреляции по данным о 19 странах Африки из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак (зависимая переменная) x_2 – средняя продолжительность жизни женщин.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,783	-0,399	-0,744	0,121	-0,420	0,700	0,758	0,621
x_2		1	0,479	0,341	-0,043	0,565	-0,808	-0,943	-0,667
x_3			1	0,290	-0,152	0,437	-0,572	-0,542	-0,680
x_4				1	-0,295	0,421	-0,497	-0,326	-0,427
x_5					1	-0,486	0,561	-0,158	0,593
x_6						1	-0,763	-0,463	-0,641
x_7							1	0,708	0,907
x_8								1	0,592
x_9									1

Уравнения регрессии

$\hat{y} = \underset{(8,443)}{96,786} - \underset{(0,053)}{0,330}x_1 - \underset{(0,085)}{0,242}x_4$; $R^2=0,744$; $F=23,193$;

$\hat{y} = \underset{(10,300)}{91,008} - \underset{(0,051)}{0,301}x_1 + \underset{(0,075)}{0,049}x_3 - \underset{(0,072)}{0,280}x_4 + \underset{(0,953)}{2,374}x_6 + \underset{(1,336)}{0,170}x_9$; $R^2=0,858$; $F=15,764$;

$\hat{y} = \underset{(4,621)}{48,158} + \underset{(0,101)}{0,056}x_4 + \underset{(1,526)}{3,466}x_6$; $R^2=0,332$; $F=3,972$;

$\hat{y} = \underset{(3,398)}{46,848} + \underset{(0,115)}{0,152}x_3 + \underset{(2,975)}{2,975}x_6$; $R^2=0,385$; $F=5,017$.

Под значениями коэффициентов приведены их средние квадратические отклонения.
По приведенным данным выполнить Задания№1 и №2.

Задание 3.

Зависимость объема производства у (тыс. ед) от численности занятых х (чел.) некоторой фирмы приводятся в таблице.

x	10	12	15	17	21	23
y	29	31	37	39	43	58

Для характеристики зависимости у от х рассчитайте параметры степенной функции $y=a \cdot x^b$.

Задание 4.

По данным, представленным в таблице, провести классификацию n = 4 предприятий по двум показателям

Номер предприятия	1	2	3	4
$x_i^{(1)}$	9	6	2	8
$x_i^{(2)}$	6	10	4	9

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «ближайшего соседа» и центра тяжести.

ВАРИАНТ 2

Задание1, 2.

Матрица парных коэффициентов корреляции по данным о 17 странах Ближнего Востока и Средней Азии из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак (зависимая переменная) x_2 – средняя продолжительность жизни женщин.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,909	-0,408	-0,563	-0,125	-0,656	0,549	0,717	0,437
x_2		1	0,425	0,743	-0,055	0,518	-0,671	-0,600	-0,584
x_3			1	0,130	0,229	0,389	-0,256	-0,294	-0,227
x_4				1	-0,533	0,108	-0,667	-0,137	-0,691
x_5					1	-0,325	0,518	-0,559	0,591
x_6						1	-0,307	-0,519	-0,136
x_7							1	0,167	0,964
x_8								1	0,043
x_9									1

Уравнения регрессии

$\hat{y}$ = 71,980 + 0,026x_3 + 0,101x_4 - 0,145x_5 – 0,09x_6 - 0,301x_7 + 1,404x_8 ; R^2=0,926; F=20,729;
(5,932) (0,021) (0,041) (0,037) (0,159) (0,249) (1,324)

$\hat{y}$ = 71,278 - 0,168x_1 + 0,098x_6 ; R^2=0,904; F=66,046;
(2,837) (0,023) (0,029)

$\hat{y}$ = 54,224 + 0,201x_4 + 0,492x_6 ; R^2=0,746;
F=20,611;
(2,980) (0,039) (0,150)

$\hat{y}$ = 76,371 + 0,496x_6 – 1,533x_9 ; R^2=0,537; F=8,122.
(2,968) (0,204) (0,538)

Под значениями приведены их средние квадратические отклонения.
По приведенным данным выполнить Задания №1 и №2.

Задание 3.

Зависимость объема производства y (тыс. ед) от численности занятых x (чел.) некоторой фирмы приводятся в таблице.

x	10	12	15	17	21	23
y	29	31	37	39	43	58

Для характеристики зависимости y от x рассчитайте параметры показательной функции $y= a \cdot b^x$.

Задание 4.

По данным, представленным в таблице, провести классификацию n = 4 предприятий по двум показателям

Номер предприятия	1	2	3	4
$x_i^{(1)}$	9	6	2	8
$x_i^{(2)}$	6	10	4	9

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «дальнего соседа» и средней связи.

ВАРИАНТ 3

Задание 1, 2.

Матрица парных коэффициентов корреляции по данным о 21 стране Латинской Америки из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак (зависимая переменная) x_2 – средняя продолжительность жизни женщин.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,965	-0,452	-0,779	0,273	-0,497	0,662	0,696	0,714
x_2		1	0,500	0,808	-0,238	0,550	-0,684	-0,784	-0,758
x_3			1	0,560	-0,342	0,285	-0,651	-0,259	-0,642
x_4				1	-0,508	0,608	-0,830	-0,470	-0,839
x_5					1	-0,682	0,791	-0,316	0,663
x_6						1	-0,658	-0,113	-0,600
x_7							1	0,237	0,953
x_8								1	0,424
x_9									1

Уравнения регрессии

$$\hat{y} = \underset{(4,662)}{76,316} - \underset{(0,023)}{0,251}x_1 + \underset{(0,030)}{0,033}x_3 + \underset{(0,373)}{0,631}x_6 + \underset{(0,215)}{0,445}x_7 - \underset{(1,356)}{2,995}x_9; \qquad R^2=0,957; \quad F=67,018;$$

$$\hat{y} = \underset{(5,068)}{75,396} - \underset{(0,028)}{0,258}x_1 + \underset{(0,050)}{0,078}x_4; \qquad R^2=0,940; \; F=141,328;$$

$$\hat{y} = \underset{(1,335)}{81,667} - \underset{(0,020)}{0,277}x_1 + \underset{(0,335)}{0,464}x_6; \qquad R^2=0,939; \; F=137,786;$$

$$\hat{y} = \underset{(4,697)}{57,912} + \underset{(0,076)}{0,154}x_3 + \underset{(0,925)}{2,212}x_6; \qquad R^2=0,431; \quad F=6,815.$$

Под значениями коэффициентов приведены их средние квадратические отклонения.
По приведенным данным выполнить Задания №1 и №2.

Задание 3.

Зависимость объема производства y (тыс. ед) от численности занятых x (чел.) некоторой фирмы приводятся в таблице.

x	11	14	16	19	22	25
y	54	43	39	34	29	27

Для характеристики зависимости y от x рассчитайте параметры равносторонней гиперболы $y = a + b \cdot \frac{1}{x}$.

Задание 4.

По данным, представленным в таблице, провести классификацию n = 4 предприятий по двум показателям

Номер предприятия	1	2	3	4
$x_i^{(1)}$	7	4	5	3
$x_i^{(2)}$	5	9	4	7

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «ближайшего соседа» и центра тяжести.

ВАРИАНТ 4

Задание 1,2.
Матрица парных коэффициентов корреляции по данным о 17 странах Азиатско-Тихоокеанского региона из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак (зависимая переменная) x_2 – средняя продолжительность жизни женщин.

	x_1	x_2	x_3	x_4	x_5	X_6	x_7	x_8	x_9
x_1	1	-0,987	-0,766	-0,889	0,804	-0,648	0,937	0,910	0,917
x_2		1	0,800	0,894	-0,847	0,708	-0,942	-0,874	-0,922
x_3			1	0,585	-0,738	0,859	-0,716	-0,557	-0,572
x_4				1	-0,713	0,448	-0,865	-0,812	-0,398
x_5					1	-0,760	0,919	0,631	0,894
x_6						1	-0,661	-0,372	-0,584
x_7							1	0,851	0,974
x_8								1	0,823
x_9									1

Уравнения регрессии

$$\hat{y} = \underset{(8,716)}{55,472} + \underset{(0,057)}{0,078}x_3 + \underset{(0,071)}{0,184}x_4 + \underset{(0,243)}{0,279}x_6 - \underset{(1,085)}{1,878}x_9; \qquad R^2=0,942;\ F=49,091;$$

$$\hat{y} = \underset{(8,729)}{56,912} + \underset{(0,036)}{0,130}x_3 + \underset{(0,069)}{0,162}x_4 - \underset{(1,067)}{2,172}x_9; \qquad R^2=0,936;\ F=63,479;$$

$$\hat{y} = \underset{(3,789)}{75,857} + \underset{(0,041)}{0,125}x_3 - \underset{(0,663)}{4,276}x_9; \qquad R^2=0,909;\ F=70,000;$$

$$\hat{y} = \underset{(3,938)}{54,352} + \underset{(0,119)}{0,279}x_3 + \underset{(0,541)}{0,139}x_6; \qquad R^2=0,641;\ F=12,494.$$

Под значениями коэффициентов уравнений регрессии, оцененных методом наименьших квадратов, в скобках приведены их средние квадратические отклонения.

По приведенным данным выполнить Задания №1 и №2.

Задание 3.

Зависимость объема производства y (тыс. ед) от численности занятых x (чел.) некоторой фирмы приводятся в таблице.

x	11	14	16	19	22	25
y	27	29	34	39	43	51

Для характеристики зависимости y от x рассчитайте параметры показательной функции $y = a \cdot b^x$.

Задание 4.

По данным, представленным в таблице, провести классификацию n = 4 предприятий по двум показателям

Номер предприятия	1	2	3	4
$x_i^{(1)}$	2	4	12	13
$x_i^{(2)}$	10	7	11	9

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «дальнего соседа» и средней связи.

ВАРИАНТ 5

Задание 1,2.

Матрица парных коэффициентов корреляции по данным о 14 странах Восточной Европы из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак (зависимая переменная) x_2 – средняя продолжительность жизни женщин.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,687	0,345	0,316	0,394	0,044	0,419	0,053	0,532
x_2		1	-0,571	-0,322	0,279	0,020	0,108	-0,567	-0,136
x_3			1	0,726	-0,224	0,560	-0,048	0,752	0,392
x_4				1	-0,113	0,580	0,019	0,588	0,531
x_5					1	0,130	0,759	-0,626	0,593
x_6						1	0,036	0,234	0,282
x_7							1	-0,539	0,795
x_8								1	-0,008
x_9									1

Уравнения регрессии

$$\hat{y} = \underset{(9,814)}{80,165} - \underset{(0,047)}{0,100}x_3 - \underset{(0,151)}{0,020}x_4 - \underset{(1,766)}{0,345}x_5 + \underset{(0,240)}{0,340}x_6 + \underset{(4,528)}{1,346}x_9; \quad R^2=0,500;\ F=1,601;$$

$\hat{y} = \underset{(9,346)}{89,813} - \underset{(0,102)}{0,153}x_4 + \underset{(0,219)}{0,203}x_6$; $R^2=0,169$; $F=1,118$;

$\hat{y} = \underset{(8,898)}{75,050} - \underset{(0,038)}{0,08}x_3 + \underset{(0,108)}{0,059}x_4$; $R^2=0,344$; $F=2,878$;

$\hat{y} = \underset{(1,467)}{80,053} - \underset{(0,028)}{0,09}x_3 + \underset{(0,168)}{0,321}x_6$; $R^2=0,494$; $F=5,366$.

Под значениями коэффициентов уравнений приведены их средние квадратические отклонения.

По приведенным данным выполнить Задания №1 и №2.

Задание 3.

Изучалась зависимость условно-постоянных затрат, приходящихся на единицу продукции y (ед.) от объема производства x (тыс. шт.) по 6 однородным организациям

Номер предприятия	1	2	3	4	5	6
y	100	200	300	400	500	600
x	9	6	5	3	3,7	3,6

Для характеристики зависимости y от x рассчитайте параметры параболической функции $y=a_0+a_1x+a_2x^2$.

Задание 4.

По данным, представленным в таблице, провести классификацию n = 4 предприятий по двум показателям

Номер предприятия	1	2	3	4
$x_i^{(1)}$	2	4	12	13
$x_i^{(2)}$	10	7	11	9

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «ближайшего соседа» и центра тяжести.

ВАРИАНТ 6

Задание 1, 2.

Матрица парных коэффициентов корреляции по данным о 19 странах Африки из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак (зависимая переменная) x_1 – детская смертность (число умерших младенцев на 1000 новорожденных).

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,783	-0,399	-0,744	0,121	-0,420	0,700	0,758	0,621
x_2		1	0,479	0,341	-0,043	0,565	-0,808	-0,943	-0,667
x_3			1	0,290	-0,152	0,437	-0,572	-0,542	-0,680
x_4				1	-0,295	0,421	-0,497	-0,326	-0,427
x_5					1	-0,486	0,561	-0,158	0,593
x_6						1	-0,763	-0,463	-0,641
x_7							1	0,708	0,907
x_8								1	0,592
x_9									1

Уравнения регрессии

$$\hat{y} = \underset{(66,621)}{288,978} - \underset{(0,534)}{2,659}x_2 - \underset{(0,154)}{0,970}x_4 + \underset{(3,211)}{5,237}x_6 - \underset{(1,408)}{1,025}x_7 + \underset{(5,270)}{5,452}x_9; \quad R^2=0,912;\ F=27,043;$$

$$\hat{y} = \underset{(13,354)}{149,612} - \underset{(0,291)}{1,107}x_4 - \underset{(4,409)}{3,131}x_6; \quad R^2=0,567;\ F=10,497;$$

$$\hat{y} = \underset{(17,732)}{251,786} - \underset{(0,344)}{2,150}x_2 - \underset{(0,153)}{0,866}x_4; \quad R^2=0,871;\ F=53,788;$$

$$\hat{y} = \underset{(94,537)}{180,01} - \underset{(0,932)}{2,251}x_2 - \underset{(1,160)}{0,863}x_7; \quad R^2=0,626;\ F=13,371.$$

Под значениями приведены их средние квадратические отклонения.
По приведенным данным выполнить Задания №1 и №2.

Задание 3.

Изучалась зависимость выпуска продукции y (ед.) от потребления материалов на единицу продукции x (кг) по 6 однородным организациям

Номер предприятия	1	2	3	4	5	6
y	100	200	300	400	500	700
x	9	6	5	3	2,7	2,5

Для характеристики зависимости y от x рассчитайте параметры показательной функции $y = a \cdot b^x$.

Задание 4.

По данным, представленным в таблице, провести классификацию n = 4 предприятий по двум показателям

Номер предприятия	1	2	3	4
$x_i^{(1)}$	9	6	2	8
$x_i^{(2)}$	6	10	4	9

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «дальнего соседа» и средней связи.

ВАРИАНТ 7

Задание 1, 2.

Матрица парных коэффициентов корреляции по данным о 17 странах Ближнего Востока и Средней Азии из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак (зависимая переменная) x_1 – детская смертность (число умерших младенцев на 1000 новорожденных).

	x_1	x_2	x_3	X_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,909	-0,408	-0,563	-0,125	-0,656	0,549	0,717	0,437
x_2		1	0,425	0,743	-0,055	0,518	-0,671	-0,600	-0,584
x_3			1	0,130	0,229	0,389	-0,256	-0,294	-0,227
x_4				1	-0,533	0,108	-0,667	-0,137	-0,691
x_5					1	-0,325	0,518	-0,559	0,591
x_6						1	-0,307	-0,519	-0,136
x_7							1	0,167	0,964
x_8								1	0,043
x_9									1

Уравнения регрессии

$$\hat{y} = \underset{(48,580)}{312,506} - \underset{(0,906)}{3,937}x_2 + \underset{(0,104)}{0,0429}x_3 + \underset{(0,217)}{0,173}x_4 - \underset{(0,636)}{0,999}x_6; \qquad R^2=0,879;\ F=21,877;$$

$$\hat{y} = \underset{(32,812)}{284,461} - \underset{(0,476)}{3,312}x_2 - \underset{(0,528)}{1,197}x_6; \qquad R^2=0,873;\ F=47,904;$$

$$\hat{y} = \underset{(14,363)}{101,057} - \underset{(0,189)}{0,614}x_4 - \underset{(0,724)}{2,847}x_6; \qquad R^2=0,676;\ F=14,577;$$

$$\hat{y} = \underset{(12,079)}{65,679} - \underset{(0,189)}{0,160}x_3 - \underset{(1,009)}{2,771}x_6; \qquad R^2=0,458;\ F=5,921.$$

Под значениями коэффициентов приведены их средние квадратические отклонения.

По приведенным данным выполнить Задания №1 и №2.

Задание 3.

Изучалась зависимость выпуска продукции y (ед.) от потребления материалов на единицу продукции x (кг) по 6 однородным организациям

Номер предприятия	1	2	3	4	5	6
y	200	300	400	500	600	700
x	6	5	3	3,7	3,6	3,3

Для характеристики зависимости y от x рассчитайте параметры равносторонней гиперболы $y=a+b\cdot\frac{1}{x}$.

Задание 4.

По данным, представленным в таблице, провести классификацию n = 4 семей по двум показателям

Номер семьи	1	2	3	4
$x_i^{(1)}$	8	4	10	2
$x_i^{(2)}$	6	2	5	4

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «ближайшего соседа» и центра тяжести.

ВАРИАНТ 8

Задание 1,2.

Матрица парных коэффициентов корреляции по данным о 21 стране Латинской Америки из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак (зависимая переменная) x_1 – детская смертность (число умерших младенцев на 1000 новорожденных).

	x_1	x_2	x_3	X_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,965	-0,452	-0,779	0,273	-0,497	0,662	0,696	0,714
x_2		1	0,500	0,808	-0,238	0,550	-0,684	-0,784	-0,758
x_3			1	0,560	-0,342	0,285	-0,651	-0,259	-0,642
x_4				1	-0,508	0,608	-0,830	-0,470	-0,839
x_5					1	-0,682	0,791	-0,316	0,663
x_6						1	-0,658	-0,113	-0,600
x_7							1	0,237	0,953
x_8								1	0,424
x_9									1

Уравнения регрессии

$$\hat{y} = \underset{(56,020)}{67,517} - \underset{(0,227)}{0,015}x_3 - \underset{(0,453)}{0,688}x_4 - \underset{(2,815)}{2,451}x_6 + \underset{(1,150)}{3,532}x_8 + \underset{(6,044)}{2,519}x_9; \quad R^2=0,776; \quad F=9,829;$$

$$\hat{y} = \underset{(5,068)}{114,449} - \underset{(0,487)}{1,103}x_4 + \underset{(6,221)}{0,078}x_9; \quad R^2=0,618; \; F=14,589;$$

$$\hat{y} = \underset{(1,335)}{80,688} - \underset{(0,020)}{0,461}x_3 - \underset{(0,335)}{6,636}x_6; \quad R^2=0,352; \; F=4,894;$$

$$\hat{y} = \underset{(24,944)}{101,946} - \underset{(0,245)}{1,053}x_4 + \underset{(1,049)}{3,307}x_8; \quad R^2=0,746; \; F=26,473.$$

Под значениями коэффициентов приведены их средние квадратические отклонения.

По приведенным данным выполнить Задания №1 и №2.

Задание 3.

Изучалась зависимость выпуска продукции y (ед.) от потребления материалов на единицу продукции x (кг) по 6 однородным организациям

Номер предприятия	1	2	3	4	5	6
y	300	400	500	600	700	150
x	2,5	3	3,7	3,6	5,3	6

Для характеристики зависимости y от x рассчитайте параметры степенной функции $y=a\cdot x^b$.

Задание 4.

По данным, представленным в таблице, провести классификацию n = 4 семей по двум показателям

Номер семьи	1	2	3	4
$x_i^{(1)}$	8	4	10	2
$x_i^{(2)}$	6	2	5	4

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «дальнего соседа» и средней связи.

ВАРИАНТ 9

Задание 1, 2.

Матрица парных коэффициентов корреляции по данным о 17 странах Азиатско-Тихоокеанского региона из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак x_1 – детская смертность (число умерших младенцев на 1000 новорожденных).

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,987	-0,766	-0,889	0,804	-0,648	0,937	0,910	0,917
x_2		1	0,800	0,894	-0,847	0,708	-0,942	-0,874	-0,922
x_3			1	0,585	-0,738	0,859	-0,716	-0,557	-0,572
x_4				1	-0,713	0,448	-0,865	-0,812	-0,398
x_5					1	-0,760	0,919	0,631	0,894
x_6						1	-0,661	-0,372	-0,584
x_7							1	0,851	0,974
x_8								1	0,823
x_9									1

Уравнения регрессии

$$\hat{y} = \underset{(46,969)}{90,951} - \underset{(0,310)}{0,426}x_3 - \underset{(0,382)}{0,690}x_4 - \underset{(1,309)}{0,210}x_6 + \underset{(5,847)}{10,109}x_9; \qquad R^2=0,908;\ F=29,646;$$

$$\hat{y} = \underset{(13,016)}{171,316} - \underset{(0,184)}{0,609}x_3 - \underset{(0,208)}{1,237}x_4; \qquad R^2=0,883;\ F=52,614;$$

$\hat{y} = \underset{(18,393)}{11,163} - \underset{(0,201)}{0,445}x_3 + \underset{(3,220)}{19,069}x_9;$ $R^2=0,882;\ F=52,473;$

$\hat{y} = \underset{(17,999)}{110,909} - \underset{(0,545)}{1,302}x_3 + \underset{(2,472)}{0,293}x_6;$ $R^2=0,588;\ F=9,981.$

Под значениями коэффициентов приведены их средние квадратические отклонения.
По приведенным данным выполнить Задания №1 и №2.

Задание 3.

Изучалась зависимость выпуска продукции y (ед.) от потребления материалов на единицу продукции x (кг) по 6 однородным организациям

Номер предприятия	1	2	3	4	5	6
y	100	200	300	400	500	700
x	9	12	14	13	8	4

Для характеристики зависимости y от x рассчитайте параметры параболической функции $y=a_0+a_1x+a_2x^2$.

Задание 4

По данным, представленным в табл., провести классификацию n = 4 предприятий по двум показателям

Номер предприятия	1	2	3	4
$x_i^{(1)}$	3	5	5	8
$x_i^{(2)}$	6	2	7	3

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «ближайшего соседа» и центра тяжести.

ВАРИАНТ 10

Задание 1,2.

Матрица парных коэффициентов корреляции по данным о 14 странах Восточной Европы из таблицы 1 ПРИЛОЖЕНИЯ.

Исследуемый признак (зависимая переменная) x_1 – детская смертность (число умерших младенцев на 1000 новорожденных).

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	-0,687	0,345	0,316	0,394	0,044	0,419	0,053	0,532
x_2		1	-0,571	-0,322	0,279	0,020	0,108	-0,567	-0,136
x_3			1	0,726	-0,224	0,560	-0,048	0,752	0,392
x_4				1	-0,113	0,580	0,019	0,588	0,531
x_5					1	0,130	0,759	-0,626	0,593

x_6						1	0,036	0,234	0,282
x_7							1	-0,539	0,795
x_8								1	-0,008
x_9									1

Уравнения регрессии

$$\underset{}{\hat{y}} = \underset{(105,207)}{325,293} - \underset{(1,249)}{4,364}x_2 - \underset{(0,183)}{0,236}x_3 - \underset{(0,444)}{0,069}x_4 + \underset{(0,832)}{0,576}x_6 + \underset{(8,460)}{22,308}x_9; \quad R^2=0,744; \quad F=4,640;$$

$$\hat{y} = \underset{(68,982)}{218,607} - \underset{(0,868)}{3,093}x_2 + \underset{(6,999)}{17,798}x_9; \quad R^2=0,667; F=11,025;$$

$$\hat{y} = \underset{(50,684)}{-11,517} + \underset{(0,217)}{0,129}x_3 + \underset{(0,615)}{0,209}x_4; \quad R^2=0,128; \quad F=0,808;$$

$$\hat{y} = \underset{(79,439)}{286,226} - \underset{(0,999)}{3,943}x_2 - \underset{(0,116)}{0,173}x_3 + \underset{(7,191)}{21,957}x_9; \quad R^2=0,728; \quad F=8,931.$$

Под значениями коэффициентов приведены их средние квадратические отклонения.
По приведенным данным выполнить Задания №1 и №2.

<u>*Задание 3.*</u>

Изучалась зависимость выпуска продукции y (ед.) от потребления материалов на единицу продукции x (кг) по 6 однородным организациям

Номер предприятия	1	2	3	4	5	6
y	200	300	400	500	600	700
x	6	5	3	3,7	3,6	5,3

Для характеристики зависимости y от x найти параметры уравнения $y = b_0 + b_1x^2$.

<u>*Задание 4.*</u>

По данным, представленным в таблице, провести классификацию n = 4 предприятий по двум показателям

Номер предприятия	1	2	3	4
$x_i^{(1)}$	2	5	5	8
$x_i^{(2)}$	6	2	7	3

Классификацию провести по иерархическому агломеративному алгоритму с использованием обычного евклидова расстояния, а расстояние между кластерами определять по принципу «дальнего соседа» и средней связи.

8. Приложение

Таблица 1

СТРАНЫ МИРА В 1995 ГОДУ

	z	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Афганистан	3	168	44	18	29	2,8	0,205	53	22	6,9
Аргентина	6	25,6	75	86	95	1,3	3,408	20	9	2,8
Армения	5	27	75	68	98	1,4	5,0	23	6	3,19
Австралия	1	7,3	80	85	100	1,38	16,848	15	8	1,9
Австрия	1	6,7	79	58	99	0,2	18,396	12	11	1,5
Азербайджан	5	35	75	54	98	1,4	3,0	23	7	2,8
Бахрейн	5	25	74	83	77	2,4	7,875	29	4	3,96
Бангладеш	3	106	53	16	35	2,4	0,202	35	11	4,7
Барбадос	6	20,3	78	45	99	0,21	6,95	16	8	1,78
Беларусь	2	19	76	65	99	0,32	6,5	13	11	1,88
Бельгия	1	7,2	79	96	99	0,2	17,912	12	11	1,7
Боливия	6	75	64	51	78	2,7	0,73	34	9	4,21
Босния	2	12,7	78	36	86	0,7	3,098	14	6	1,7
Ботсвана	4	39,3	66	25	72	2,7	2,677	32	8	5,1
Бразилия	6	66	67	75	81	1,28	2,354	21	9	2,7
Болгария	2	12	75	68	93	-0,2	3,831	13	12	1,8
Буркина-Фасо	4	118	50	15	18	2,81	0,357	47	18	6,94
Бурунди	4	105	50	5	50	2,26	0,208	44	21	6,8
Камбоджа	3	112	52	12	35	2,9	0,26	45	16	5,81
Камерун	4	77	58	40	54	2,9	0,993	41	12	5,7
Канада	1	6,8	81	77	97	0,7	19,904	14	8	1,8
Цент. Афр. Респ.	4	137	44	47	27	2,4	0,457	44	21	5,42
Чили	6	14,6	78	85	93	1,7	2,591	23	6	2,5
Китай	3	52	69	26	78	1,1	0,377	21	7	1,84
Колумбия	6	28	75	70	87	2	1,538	24	6	2,47
Коста-Рика	6	11	79	47	93	2,3	2,031	26	4	3,1
Хорватия	2	8,7	77	51	97	-0,1	5,487	11	11	1,65
Куба	6	10,2	78	74	94	0,95	1,382	17	7	1,9
Чехия	2	9,3	77	69	99	0,21	7,311	13	11	1,84
Дания	1	6,6	79	85	99	0,1	18,277	12	12	1,7
Доминик. Респ.	6	51,5	70	60	83	1,8	1,034	25	6	2,8
Эквадор	6	39	73	56	88	2,01	1,085	26	6	3,08
Египет	5	76,4	63	44	48	1,95	0,748	29	9	3,77
Сальвадор	6	41	69	44	73	2,04	1,078	33	7	3,78
Эстония	2	19	76	72	99	0,52	6,0	14	12	2
Эфиопия	4	110	54	12	24	3,1	0,122	45	14	6,81
Финляндия	1	5,3	80	60	100	0,3	15,877	13	10	1,8
Франция	1	6,7	82	73	99	0,47	18,944	13	9	1,8
Габон	4	94	58	46	61	1,46	4,283	28	14	3,97

	z	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Гамбия	4	124	52	23	27	3,1	0,351	46	16	6,29
Грузия	2	23	76	56	99	0,8	4,5	16	9	2,18
Германия	1	6,5	79	85	99	0,36	17,539	11	11	1,47
Греция	1	8,2	80	63	93	0,84	8,06	10	10	1,5
Гватемала	6	57	67	39	55	2,58	1,342	35	8	4,76
Гаити	6	109	47	29	53	1,63	0,383	40	19	5,94
Гондурас	6	45	70	44	73	2,73	1,03	35	6	4,9
Гонконг	3	5,8	80	94	77	-0,09	14,641	13	6	1,4
Венгрия	2	12,5	76	64	99	-0,3	5,249	12	13	1,8
Исландия	1	4	81	91	100	1,1	17,241	16	7	2,11
Индия	3	79	59	26	52	1,9	0,275	29	10	4,48
Индонезия	3	68	65	29	77	1,6	0,681	24	9	2,8
Иран	5	60	67	57	54	3,46	1,5	42	8	6,33
Ирак	5	67	68	72	60	3,7	1,955	44	7	6,71
Ирландия	1	7,4	78	57	98	0,3	12,17	14	9	1,99
Израиль	5	8,6	80	92	92	2,22	13,066	21	7	2,83
Италия	1	7,6	81	69	97	0,21	17,5	11	10	1,3
Япония	3	4,4	82	77	99	0,3	19,86	11	7	1,55
Иордания	5	34	74	68	80	3,3	1,157	39	5	5,64
Кения	4	74	55	24	69	3,07	0,323	42	11	5,91
Кувейт	5	12,5	78	96	73	5,24	6,818	28	2	4
Латвия	2	21,5	75	71	99	0,5	7,4	14	12	2
Ливан	5	39,5	71	84	80	2	1,429	27	7	3,39
Либерия	4	113	57	45	40	3,3	0,409	43	12	6,8
Ливия	5	63	65	82	64	3,7	5,91	45	8	6,4
Литва	2	17	77	69	99	0,3	6,71	15	10	2
Малайзия	3	25,6	72	43	78	2,3	2,995	29	5	3,51
Мексика	6	35	77	73	87	1,9	3,604	28	5	3,2
Марокко	4	50	70	46	50	2,12	1,062	29	6	3,83
Сев. Корея	3	27,7	73	60	99	1,83	1,0	24	6	2,4
Нидерланды	1	6,3	81	89	99	0,58	17,245	13	9	1,58
Новая Зел.	1	8,9	80	84	99	0,57	14,381	16	8	2,03
Никарагуа	6	52,5	67	60	57	2,68	0,447	35	7	4,33
Нигерия	4	75	57	35	51	3,1	0,282	44	12	6,4
Норвегия	1	6,3	81	75	99	0,4	17,755	13	10	2
Оман	5	36,7	70	11	65	3,46	7,467	40	5	6,53
Пакистан	3	101	58	32	35	2,8	0,406	42	10	6,43
Панама	6	16,5	78	53	88	1,94	2,397	25	5	2,9
Парагвай	6	25,2	75	48	90	2,7	1,5	33	5	4,3
Перу	6	54	67	70	85	2	1,107	26	7	3,11
Филиппины	3	51	68	43	90	1,92	0,867	27	7	3,35
Польша	2	13,8	77	62	99	0,3	4,429	14	10	1,94
Португалия	1	9,2	78	34	85	0,36	9,0	12	10	1,5

	z	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Румыния	2	20,3	75	54	96	0,06	2,702	14	10	1,82
Россия	2	27	74	74	99	0,2	6,68	13	11	1,83
Руанда	4	117	46	6	50	2,8	0,292	49	21	8,19
Южная Корея	3	21,7	74	72	96	1	6,627	16	6	1,65
Сауд. Арав.	5	52	70	77	62	3,2	6,651	38	6	6,67
Сенегал	4	76	58	40	38	3,1	0,744	43	12	6,1
Сингапур	3	5,7	79	100	88	1,2	14,99	16	6	1,88
Сомали	4	126	55	24	24	3,2	2,126	46	13	7,25
ЮАР	4	47,1	68	49	76	2,6	3,128	34	8	4,37
Испания	1	6,9	81	78	95	0,25	13,047	11	9	1,4
Швеция	1	5,7	81	84	99	0,52	16,9	14	11	2,1
Швейцария	1	6,2	82	62	99	0,7	22,384	12	9	1,6
Сирия	5	43	68	50	64	3,7	2,436	44	6	6,65
Тайвань	3	5,1	78	71	91	0,92	7,055	15,6	6	1,7
Танзания	4	110	45	21	46	2,5	0,263	46	19	6,2
Таиланд	3	37	72	22	93	1,4	1,8	19	6	2,1
Турция	5	49	73	61	81	2,02	3,721	26	6	3,21
ОАЭ	5	22	74	81	68	4,8	14,193	28	3	4,5
Великобритания.	1	7,2	80	89	99	0,2	15,974	13	11	1,83
США	1	8,11	79	75	97	0,99	23,474	15	9	2,06
Уганда	4	112	43	11	48	2,42	0,325	49	24	6,77
Украина	2	20,7	75	67	97	0,05	2,34	12	13	1,82
Уругвай	6	17	77	89	96	0,8	3,131	17	10	2,44
Узбекистан	5	53	72	41	97	2,13	1,35	30	7	3,73
Венесуэла	6	28	76	91	88	2,16	2,829	26	5	3,05
Вьетнам	3	46	68	20	88	1,78	0,23	27	8	3,33
Замбия	4	85	45	42	73	2,8	0,573	46	18	6,68

z – регион или экономическая группа:

1 – группа развитых стран
2 – Восточная Европа
3 – Азиатско-Тихоокеанский регион
4 – Африка
5 – Ближний Восток
6 – Латинская Америка

x_1 – детская смертность (количество умерших младенцев на 1000 новорожденных);
x_2 – средняя продолжительность жизни женщин;
x_3 – доля городского населения, %;
x_4 – уровень грамотности населения, %;
x_5 – прирост населения, %;
x_6 – ВВП на душу населения, тыс.долл.США;
x_7 – уровень рождаемости (число родившихся на 1000 жителей);
x_8 – уровень смертности (число умерших на 1000 жителей);
x_9 – среднее число детей в семье.

Тесты по дисциплине

ТЕСТ №1

1. Коэффициент корреляции, равный нулю, означает, что между переменными:

а) линейная связь отсутствует;
б) существует линейная связь;
в) ситуация не определена.

2. Коэффициент корреляции, равный - 1, означает, что между переменными:

а) линейная связь отсутствует;
б) существует линейная связь;
в) функциональная зависимость;
г) ситуация не определенна.

3. В регрессионном анализе обычно предполагается, что случайная величина Y имеет нормальный закон распределения с условным математическим ожиданием $\widetilde{Y} = \varphi(x_1,...,x_k)$, *являющимся функцией от аргументов xj, и с постоянной, ……… от аргументов дисперсией* σ^2 *:*

а) не зависящей;
б) зависящей.

4. Статистика Дарбина – Уотсона (DW) вычисляется по формуле:

а)

$$\mathbf{DW} = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2};$$

б)

$$\mathbf{DW} = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} y_t^2};$$

с)

$$\mathbf{DW} = \frac{\sum_{t=2}^{n} (y_t - y_{t-1})^2}{\sum_{t=1}^{n} y_t^2}.$$

5. В модели $lnY = \beta_0 + \beta X + \varepsilon$ коэффициент β имеет смысл:

а) абсолютного прироста;
б) темпа роста;
в) темпа прироста.

6. При анализе эластичности спроса по цене целесообразно использовать следующую модель:

а) линейную;
б) полиномиальную;
в) логарифмическую;
г) степенную;
д) экспоненциальную.

7. Использование обычного Евклидова расстояния оправдано в следующих случаях (выберите необходимые варианты):

а) наблюдения берутся из генеральной совокупности, имеющей многомерное нормальное распределение с ковариационной матрицей вида $\sigma^2 E_к$, т.е. компоненты Х взаимно независимы и имеют одну и ту же дисперсию, где $E_к$ - единичная матрица;

б) наблюдения берутся из генеральной совокупности, имеющей биномиальное распределение;

в) компоненты вектора наблюдений Х неоднородны по физическому смыслу и при классификации используются с определенным весом;

г) компоненты вектора наблюдений Х однородны по физическому смыслу и одинаково важны для классификации;

д) признаковое пространство совпадает с геометрическим пространством;

е) совпадение признакового пространства с геометрическим пространством необязательно.

8. Академиком А.Н.Колмогоровым было предложено:

а) “обобщенное расстояние” между классами;
б) расстояние, измеряемое по принципу “ближайшего соседа”;
в) расстояние, измеряемое по принципу “дальнего соседа”;
г) расстояние, измеряемое по “центрам тяжести” групп;
д) расстояние, измеряемое по принципу “средней связи”.

9. Производственная функция Кобба – Дугласа с учетом технического прогресса имеет вид:

а) $Q_t = A \times e^{\theta \cdot t} \times K_t^{\alpha} \times L_t^{\beta}$;
б) $Q = A \times K^{\alpha} \times L^{\beta} \times e^{\varepsilon}$;
в) $Q = A \times K^{\alpha} \times L^{1-\alpha} \times e^{\varepsilon} = A \times (\frac{K}{L})^{\alpha} \times L \times e^{\varepsilon}$.

10. Оценки неизвестных параметров А, α и β в производственной функции Кобба – Дугласа можно найти с помощью:

а) метода наименьших квадратов;
б) принципа “ближнего соседа”;
а) дисконтированием множителей.

ТЕСТ №2

1. Двумерная корреляционная модель определяется ……… параметрами (вставьте необходимое слово):

а) тремя;
б) пятью;
в) семью.

2. Коэффициент регрессии определяется по формуле:

а) $\beta_{yx} = \rho \dfrac{\sigma_y}{\sigma_x}$ - коэффициент регрессии y на x;

б) $M[\dfrac{x-\mu_x}{\sigma_x} \cdot \dfrac{y-\mu_y}{\sigma_y}] = \rho,$;

в) $r_{12/3,4...,k} = - \dfrac{R_{12}}{\sqrt{R_{11} R_{22}}}$;

г) $r_1 = \sqrt{1 - \dfrac{|R|}{R_{11}}}$.

3. Если вектор ошибок имеет постоянную дисперсию, то это явление называется:

а) гомоскедастичностью;
б) гетероскедастичностью;
в) ситуация не определена.

4. С увеличением объема выборки:

а) увеличивается точность оценок;
б) уменьшается ошибка регрессии;
в) расширяются интервальные оценки;
г) уменьшается коэффициент детерминации.

5. При анализе издержек Y от объемы выпуска X целесообразно использовать следующую модель:

а) линейную;
б) полиномиальную;
в) логарифмическую;
г) степенную;
д) экспоненциальную.

6. Модель $Y = \beta_0 + \beta \ln X + \varepsilon$ используется, когда необходимо исследовать влияние:

а) процентного изменения независимой переменной на абсолютное изменение зависимой переменной;

б) процентного изменения независимой переменной на процентное изменение зависимой переменной;

в) абсолютное изменения независимой переменной на абсолютное изменение зависимой переменной.

7. Наиболее употребительными расстояниями и мерами близости между классами объектов являются (выберите необходимый вариант):

а) расстояние, измеряемое по принципу “ближайшего соседа”;
б) расстояние, измеряемого по принципу “дальнего соседа”;
в) расстояние, измеряемое по принципу “родственной связи”;
г) расстояние, измеряемое по “центрам тяжести” групп;
д) расстояние, измеряемое по принципу “незначимой связи”;
е) расстояние, измеряемое по принципу “средней связи”;
ж) расстояние, измеряемое по принципу “значимой связи”.

8. Расстояние, измеряемое по принципу “ближайшего соседа” находится по формуле:

а) $\rho_{\varepsilon}(x_i, x_j) = \sqrt{\sum_{\ell=1}^{k}(x_{i\ell} - x_{j\ell})^2}$;

б) $\rho_{\min}(S_\ell, S_m) = \min\limits_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j)$;

в) $\rho_{\max}(S_\ell, S_m) = \max\limits_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j)$;

г) $\rho_{Ц.Т}(S_\ell, S_m) = \rho(\bar{x}_\ell, \bar{x}_m)$;

д) $\rho_{\ell,(m,q)} = \rho(S_\ell, S_{(m,q)}) = \alpha\rho_{\ell m} + \beta\rho_{\ell q} + \gamma\rho_{mq} + \delta(\rho_{\ell m} - \rho_{\ell q})$,

е) $\rho_{\text{ср}}(S_\ell, S_m) = \frac{1}{n_\ell n_m} \sum_{x_i \in S_\ell} \sum_{x_j \in S_m} \rho(x_i, x_j)$.

9. Параметры α и β в производственной функции Кобба – Дугласа называют:

а) коэффициентами эластичности;
б) коэффициентами корреляции;
в) коэффициентами автокорреляции.

10. Коэффициенты эластичности показывают, на какую величину в среднем изменится Q, если α или β увеличить соответственно:

а) на один процент;
б) на единицу своего измерения.

ТЕСТ №3

1. Коэффициент регрессии показывает:

а) на сколько единиц своего измерения увеличится (β>0) или уменьшится (β<0) в среднем y(My/x), если x увеличить на единицу своего измерения;

б) долю дисперсии одной случайной величины, обусловленную вариацией другой;

в) на сколько % увеличится (β>0) или уменьшится (β<0) в среднем y(My/x), если x увеличится на 1 %.

2. Коэффициент регрессии изменяется в пределах от:

а) –1 до 1;
б) 0 до 1;
в) принимает любое значение.

3. Квадратичная форма

$$Q=(Y - X\beta)^T (Y - X\beta) = \sum_{i=1}^{n}\left(y_i - \tilde{y}_i\right)^2 \Rightarrow \min_{\beta_0,\beta_1,\ldots,\beta_k}$$

соответствует:

а) методу максимального правдоподобия;
б) методу наименьших квадратов;
в) методу “дальнего соседа”;
г) методу “средней связи”;
д) двухшаговому методу наименьших квадратов.

4. На главной диагонали ковариационной матрицы в выражении $S(b) = \hat{S}^2 (X^T X)^{-1}$ *находятся:*

а) дисперсии коэффициентов регрессии;
б) средние значения коэффициентов регрессии;
в) коэффициенты корреляции;
г) квадраты коэффициентов корреляции.

5. При анализе производственной функции целесообразно использовать следующую модель:

а) линейную;
б) полиномиальную;
в) логарифмическую;
г) степенную;
д) экспоненциальную.

6. Модели $\ln Y = \beta_0 + \beta X + \varepsilon$

$Y = \beta_0 + \beta \ln X + \varepsilon$

называются:

а) линейными;
б) полулогарифмическими;
в) логарифмическими.

7. Расстояние, измеряемое по принципу “дальнего соседа”, находится по формуле:

а) $\rho_\varepsilon(x_i, x_j) = \sqrt{\sum_{\ell=1}^{k}(x_{i\ell} - x_{j\ell})^2}$;

б) $\rho_{\min}(S_\ell, S_m) = \min_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j)$;

в) $\rho_{\max}(S_\ell, S_m) = \max_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j)$;

г) $\rho_{Ц.Т.}(S_\ell, S_m) = \rho(\overline{x}_\ell, \overline{x}_m)$;

д) $\rho_{\ell,(m,q)} = \rho(S_\ell, S_{(m,q)}) = \alpha\rho_{\ell m} + \beta\rho_{\ell q} + \gamma\rho_{mq} + \delta(\rho_{\ell m} - \rho_{\ell q})$;

е) $\rho_{\text{ср}}(S_\ell, S_m) = \dfrac{1}{n_\ell n_m} \sum\limits_{x_i \in S_\ell} \sum\limits_{x_j \in S_m} \rho(x_i, x_j)$.

8. Расстояние, измеряемое по “центрам тяжести” групп, находится по формуле:

а) $\rho_\varepsilon(x_i, x_j) = \sqrt{\sum\limits_{\ell=1}^{k} (x_{i\ell} - x_{j\ell})^2}$;

б) $\rho_{\min}(S_\ell, S_m) = \min\limits_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j)$;

в) $\rho_{\max}(S_\ell, S_m) = \max\limits_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j)$;

г) $\rho_{Ц.Т.}(S_\ell, S_m) = \rho(\bar{x}_\ell, \bar{x}_m)$;

д) $\rho_{\ell,(m,q)} = \rho(S_\ell, S_{(m,q)}) = \alpha\rho_{\ell m} + \beta\rho_{\ell q} + \gamma\rho_{mq} + \delta(\rho_{\ell m} - \rho_{\ell q})$;

е) $\rho_{\text{ср}}(S_\ell, S_m) = \dfrac{1}{n_\ell n_m} \sum\limits_{x_i \in S_\ell} \sum\limits_{x_j \in S_m} \rho(x_i, x_j)$.

9. Если $\alpha + \beta = 1$, то уровень эффективности:

а) не зависит от масштабов производства;
б) зависит от масштабов производства.

10. Если $\alpha + \beta < 1$, то средние издержки, рассчитанные на единицу продукции:

а) растут по мере расширения масштабов производства;
б) убывают по мере расширения масштабов производства.

ТЕСТ №4

1. В двумерной модели для вывода о независимости признаков x и y в генеральной совокупности достаточно проверить значимость:

а) только коэффициента корреляции;
б) коэффициента корреляции и регрессии;
в) коэффициента корреляции, детерминации и регрессии.

2. Значимость частных и парных коэффициентов корреляции проверяется с помощью:

а) нормального закона распределения;
б) t-критерия Стъюдента;
в) F–критерия;
г) таблицы Фишера – Иейтса.

3. В регрессионном анализе x_j рассматриваются как:

а) неслучайные величины;
б) случайные величины;
в) любые величины.

4. Для оценки вектора β наиболее часто используют метод наименьших квадратов (МНК), согласно которому в качестве оценки принимают вектор b, который минимизирует:

а) сумму отклонений наблюдаемых значений y_i от модельных значений $\tilde{y}_i$;

б) сумму квадратов отклонения наблюдаемых значений y_i от модельных значений $\tilde{y}_i$.

5. Если в модели $Y = \beta_0 + \beta \ln X + \varepsilon$ положить Y = GNP (валовой национальный продукт), а X=M (денежная масса), то из формулы:

$GNP = \beta_0 + \beta \ln M + \varepsilon$, следует, что если увеличить предложение денег M на …….., то ВНП вырастет на 0,01 β:

а) 1%;

б) 1 измерения.

6. Для получения качественных оценок уравнений регрессии необходимо выполнение следующих предпосылок МНК (выберите необходимые пункты):

а) отклонения ε_i должны быть нормально распределенными случайными величинами с нулевым математическим ожиданием и постоянной дисперсией;

б) отклонения ε_i не должны коррелировать друг с другом;

в) отклонения ε_i должны иметь показательный закон распределения.

7. Расстояние, измеряемое по принципу "средней связи", находится по формуле:

а) $\rho_{\varepsilon}(x_i, x_j) = \sqrt{\sum_{\ell=1}^{k}(x_{i\ell} - x_{j\ell})^2}$;

б) $\rho_{\min}(S_\ell, S_m) = \min_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j)$;

в) $\rho_{\max}(S_\ell, S_m) = \max_{x_i \in S_\ell, x_j \in S_m} \rho(x_i, x_j)$;

г) $\rho_{Ц.Т.}(S_\ell, S_m) = \rho(\bar{x}_\ell, \bar{x}_m)$;

д) $\rho_{\ell,(m,q)} = \rho(S_\ell, S_{(m,q)}) = \alpha\rho_{\ell m} + \beta\rho_{\ell q} + \gamma\rho_{mq} + \delta(\rho_{\ell m} - \rho_{\ell q})$,

е) $\rho_{ср.}(S_\ell, S_m) = \frac{1}{n_\ell n_m} \sum_{x_i \in S_\ell} \sum_{x_j \in S_m} \rho(x_i, x_j)$.

8. Кластерный анализ позволяет проводить:

а) группировку объектов;

б) группировку признаков;

в) группировку объектов и группировку признаков.

9. Если $\alpha + \beta > 1$, то средние издержки, рассчитанные на единицу продукции:

а) растут по мере расширения масштабов производства;

б) убывают по мере расширения масштабов производства.

10. Исходя из априорных соображений значения α и β должны удовлетворять условиям:

а) $0<\alpha<1$ и $0<\beta<1$;

б) $-1<\alpha<1$ и $-1<\beta<1$;

в) $-1<\alpha<1$ и $0<\beta<1$.

ТЕСТ №5

1. Коэффициент корреляции считается значимым с вероятностью ошибки α, если:

а) $t_{набл}$ по модулю будет больше, чем $t_{кр}$,

б) не имеет значения;

в) $t_{набл}$ по модулю будет меньше, чем $t_{кр.}$

2. Матрица R парных коэффициентов корреляции является (выберите необходимые пункты):

а) обратной;

б) транспонированной;

в) симметричной;

г) положительно определенной.

3. В каких пределах изменяется множественный коэффициент корреляции:

а) от 0 до 1;

б) от –1 до 0;

в) от –1 до 1;

г) от 0 до 10.

4. В каких пределах изменяется коэффициент детерминации:

а) от 0 до 1;

б) от –1 до 0;

в) от –1 до 1;

г) от 0 до 10.

5. В хорошо подобранной модели остатки должны (выберите необходимые пункты):

а) иметь нормальный закон распределения с нулевым математическим ожиданием и постоянной дисперсией;

б) не коррелировать друг с другом;

в) иметь экспоненциальный закон распределения;

г) хаотично разбросаны;

д) форма и вид распределения не важен.

6. Неправильный выбор функциональной формы или объясняющих переменных называется:

а) ошибками спецификации;

б) ошибками прогноза;

в) гетероскедастичностью.

7. С какой целью производят нормирование признаков:

а) с целью устранения влияния различных единиц измерения;
б) с целью уменьшить признаковое пространство;
в) с целью упрощения расчетов.

8. Хеммингово расстояние вычисляется по формуле:

а) $\rho_\varepsilon(x_i, x_j) = \sqrt{\sum_{\ell=1}^{k}(x_{i\ell} - x_{j\ell})^2}$;

б) $\rho_{B\varepsilon}(x_i, x_e) = \sqrt{\sum_{\ell=1}^{k}\omega_\ell(x_{i\ell} - x_{j\ell})^2}$;

в) $\upsilon_H(x_i, x_j) = \sum_{\ell=1}^{k}\left|x_{i\ell} - x_{j\ell}\right|$.

9. Коэффициент α интерпретируется как:

а) эластичность по труду;
б) эластичность по капиталу;
в) эластичность замещения.

10. Для определения параметров и вида производственной функции пользуются следующими видами данных:

а) динамическими рядами;
б) данными одновременных наблюдений (пространственной информацией);
в) динамическими рядами и пространственной информацией.

ТЕСТ №6

1. С помощью данной формулы $r_{12/3,4...,k} = -\frac{R_{12}}{\sqrt{R_{11}R_{22}}}$ *можно определить:*

а) множественный коэффициент корреляции (*k-2*)-го порядка между факторами X1 и X2;
б) частный коэффициент корреляции (*k-2*)-го порядка между факторами X1 и X2;
в) парный коэффициент корреляции (*k-2*)-го порядка между факторами X1 и X2.

2. С помощью данной формулы $r_{1/2,3,...,k} = r_1 = \sqrt{1 - \frac{|R|}{R_{11}}}$ *можно определить:*

а) множественный коэффициент корреляции (*k-2*)-го порядка между факторами X1 и X2;
б) частный коэффициент корреляции (*k-2*)-го порядка между факторами X1 и X2;
в) парный коэффициент корреляции (*k-2*)-го порядка между факторами X1 и X2.

3. Коэффициент детерминации – это:

а) квадрат парного коэффициента корреляции;
б) квадрат частного коэффициента корреляции;
в) квадрат множественного коэффициента корреляции.

4. Метод максимального правдоподобия лучше работает на…, где он, как правило, дает оценки с минимальной дисперсией:

а) больших выборках;
б) малых выборках;
в) любых выборках.

5. Модель вида $Y = AK^{\alpha}L^{\beta}$ *носит название:*

а) функции Энгеля;
б) функции Кобба – Дугласа;
в) лог-линейной модели;
г) степенной модели.

6. Модель вида $Y_t = Y_0(1+r)^t$ *носит название:*

а) функции Энгеля;
б) функции Кобба – Дугласа;
в) лог-линейной модели;
г) степенной модели.

7. В задаче классификации данное расстояние применяется в тех случаях, когда каждой компоненте x_l *вектора наблюдений* X *удается приписать некоторый “вес”, пропорционально степени важности признака.*

а) Хеммингово расстояние;
б) “взвешенное” Евклидово пространство;
в) обычное Евклидово расстояние.

8. Иерархические (древообразные) процедуры являются наиболее распространенными (в смысле реализации на ЭВМ) алгоритмами кластерного анализа, они бывают … типов:

а) 2;
б) 3;
в) 5;
г) любых.

9. Если производство, эффективность которого не зависит от масштабов и описывается производственной функцией Кобба – Дугласа, то с ростом параметра α *параметр* β*:*

а) растет;
б) уменьшается;
в) остается неизменным;
г) растет или уменьшается.

10. Если производство, эффективность которого растет по мере его укрупнения, описывается производственной функцией Кобба – Дугласа, то параметры модели удовлетворяют соотношению:

а) $\alpha+\beta<1$;
б) $\alpha+\beta=1$;
в) $\alpha+\beta=0$;
г) $\alpha+\beta>1$.

ТЕСТ №7

1. Уравнение $My/x - My = \beta_{yx}(x - Mx)$:

а) прямая регрессии *y* на *x;*
б) прямая регрессии *x на y.*

2. Квадрат какого коэффициента указывает долю дисперсии одной случайной величины, обусловленную вариацией другой:

а) коэффициент детерминации;
б) парный коэффициент корреляции;
в) частный коэффициент корреляции;
г) множественный коэффициент корреляции.

3. Оценки максимального правдоподобия и метода наименьших квадратов:

а) могут не совпадать;
б) совпадают;
в) никогда не совпадают.

4. В матричной форме регрессионная модель имеет вид: $Y = X\beta + \varepsilon$, *где Y:*

а) матрица, размерности [n x (k+1)];
б) случайный вектор-столбец размерности (n x 1).

5. Какой смысл у коэффициентов регрессии в логарифмических регрессионных моделях:

а) показывают процентное изменение Y для данного процентного изменения X;
б) показывают абсолютное изменение Y для данного процентного изменения X;
в) показывают процентное изменение Y для данного абсолютного изменения X.

6. Изменяются ли свойства случайного отклонения при преобразовании уравнения регрессии:

а) да;
б) нет;
в) случайное отклонение не зависит от вида уравнения регрессии

7. В … процедурах начальным является разбиение, состоящее из n одноэлементных классов, а конечным - из одного класса; в …… – наоборот (вставьте необходимые буквы):

а) агломеративных, дивизимных;
б) дивизимных, агломеративных;
в) дисконтированных, агломеративных.

8. Большинство программ, реализующих алгоритм иерархической классификации, предусматривает графическое представление результатов классификации в виде:

а) дендрограммы;
б) длок-схемы;
в) графиков показателей.

9. В задачах многомерной классификации объектов α, β, δ и γ являются:

а) числовыми коэффициентами;
б) коэффициентами эластичности.

10. В производственной функции Кобба-Дугласа параметр β соответствует коэффициенту:

а) корреляции;
б) вариации;
в) эластичности;
г) детерминации.

ТЕСТ №8

1. Величина, рассчитанная по формуле $r = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x s_y}$*, является оценкой:*

а) коэффициента детерминации;
б) парного коэффициента корреляции;
в) частного коэффициента корреляции;
г) множественного коэффициента корреляции.

2. Выборочный коэффициент корреляции r по абсолютной величине:

а) не превосходит единицы;
б) не превосходит нуля;
в) принимает любые значения.

3. В матричной форме регрессионная модель имеет вид: $Y = X\beta + \varepsilon$, *где X:*

а) матрица, размерности [n x (k+1)];
б) случайный вектор-столбец размерности (n x 1).

4. В матричной форме регрессионная модель имеет вид: $Y = X\beta + \varepsilon$, *где* ε*:*

а) матрица, размерности [n x (k+1)] ошибок наблюдений (остатков);
б) случайный вектор-столбец размерности (n x 1) ошибок наблюдений (остатков).

5. Отметьте основные виды ошибок спецификации:

а) отбрасывание значимой переменной;
б) добавление незначимой переменной;
в) низкое значение коэффициента детерминации;
г) выбор неправильной формы модели.

6. Можно ли обнаружить ошибки спецификации с помощью исследования остаточного члена:

а) да;

б) нет;

в) ситуация не определена.

7. В задачах многомерной классификации объектов при $\alpha=\beta=-\delta=1/2$ и $\gamma=0$ расстояние между классами определяется по принципу:

а) “дальнего соседа”;

б) “средней связи”;

в) “ближайшего соседа”.

8. В задачах многомерной классификации объектов при $\alpha=\beta=\delta=1/2$ и $\gamma=0$ – расстояние между классами определяется по принципу:

а) “дальнего соседа”;

б) “средней связи”;

в) “ближайшего соседа”.

9. Получены две производственные функции Кобба – Дугласа, имеющие равные значения параметров α и β, но различающиеся по параметру А. В каком случае первое производство более эффективно, чем второе:

а) $A_1<A_2$;

б) $A_1>A_2$;

в) $A_1=A_2$;

г) $A_1\neq A_2$.

10. В матричном виде структурная формы системы одновременных эконометрических уравнений имеет следующий вид: $By_t + \Gamma x_t = \varepsilon_t$ *:*

а) да, это так;

б) нет;

в) данное уравнение не является структурной формой системы одновременных эконометрических уравнений.

ТЕСТ №9

1. Есть ли необходимость при определении с надежностью γ доверительного интервала для значимого парного или частного коэффициентов корреляции использовать Z-преобразование Фишера и предварительно устанавливать интервальную оценку для Z:

а) нет;

б) да;

в) ситуация не определена.

2. Для проверки значимости какого коэффициента $F_{набл} = \dfrac{\frac{1}{k-1} r^2_{1/2,\dots,k}}{\frac{1}{n-k}\left(1 - r^2_{1/2,\dots,k}\right)}$ *рассчитывают:*

а) коэффициента детерминации;
б) парного коэффициента корреляции;
в) частного коэффициента корреляции;
г) множественного коэффициента корреляции.

3. Компоненты вектора ε_i *:*

а) независимы между собой;
б) зависимы между собой;
в) имеют нормальный закон распределения с нулевым математическим ожиданием ($M\varepsilon_i = 0$) и неизвестной дисперсией σ^2 ($D\varepsilon_i = \sigma^2$).

4. На практике при построении регрессионных моделей рекомендуется, чтобы n превышало k не менее, чем:

а) в два раза;
б) в три раза;
в) не имеет значения.

5. Если в уравнении регрессии имеется несущественная переменная, то она обнаруживает себя по низкому значению:

а) t-статистики;
б) F-статистики;
в) коэффициента детерминации.

6. Какие требования в модели регрессионного анализа предъявляются к распределению ошибок наблюдения ε_i*, а именно, к их математическому ожиданию* $M\varepsilon_i$ *и дисперсии* $D\varepsilon_i$*:*

а) $M\varepsilon_i=1$; $D\varepsilon_i=\sigma^2$;
б) $M\varepsilon_i=0$; $D\varepsilon_i=1$;
в) $M\varepsilon_i=0$; $D\varepsilon_i=\sigma^2$;
г) $M\varepsilon_i=1$; $D\varepsilon_i=0$

7. В задачах многомерной классификации объектов при $\alpha=\beta=\delta=1/2$ *и* $\gamma=0$ *– расстояние между классами определяется по принципу:*

а) “дальнего соседа”;
б) “средней связи”;
в) “ближайшего соседа”.

8. В кластер S_1 входят 4 объекта, расстояние от которых до объекта №5 составляет соответственно: 2, 5, 6, 7. Чему равно расстояние от объекта №5 до кластера S_1, если исходить из принципа “ближайшего соседа”:

а) 2; в) 6;
б) 5; г) 7.

9. Если $M\varepsilon_{t_1}\varepsilon_{t_2} = 0$ *при* $t_1 \neq t_2$ *и* $t_1, t_2 = 1,2,\ldots,n$*, то случайные ошибки регрессии:*

а) зависимы между собой;
б) независимы между собой;
с) ситуация не определена.

10. Если дисперсия ошибки постоянна $M\varepsilon_t^2 = \sigma_t^2 = \sigma^2$ *и не зависит от* t *и* x_t*, то это свидетельствует о:*

а) гомоскедастичности остатков;
б) гетероскедастичности остатков.

ТЕСТ №10

1. Известно, что при фиксированном значении X_3 между величинами X_1 и X_2 существует положительная связь. Какое значение может принять частный коэффициент корреляции $\rho_{12/3}$?

а) -0,8;
б) 0;
в) 0,4;
г) 1,3.

2. По результатам n=20 наблюдений получен частный коэффициент корреляции $r_{12/3}$=0,8. Определите, чему при уровне значимости α=0,05 равна разность между наблюдаемым ($r_{12/3}$) и критическим ($r_{кр}$) значениями коэффициентов корреляции:

а) -0,513;
б) 0, 357;
в) 0, 700;
г) 0,133.

3. На практике о наличии мультиколлинеарности обычно судят по матрице парных коэффициентов корреляции. Если один из элементов матрицы R больше ……., то считают, что имеет место мультиколлинеарность и в уравнение регрессии следует включать только один из показателей x_j или x_e. Вставьте недостающее значение.

а) 0,3;
б) 0,5;
в) 0,6;5;
г) 0,8;
д) 0,9;
е) другое значение.

4. Для проверки значимости отдельных коэффициентов регрессии, т.е. гипотез H_0: $\beta_j=0$, где $j=1,2,...k$, используют:

а) нормальный закон распределения;
б) t-критерий;
в) распределение Фишера.

5. Двойная логарифмическая модель является линейной относительно ее переменных:

а) утверждение истинно;
б) утверждение ложно;
в) утверждение не определено.

6. Коэффициенты двойной логарифмической модели определяют эластичность зависимой переменной по соответствующим определяющим переменным:

а) утверждение истинно;
б) утверждение ложно;
в) утверждение не определено.

7. В кластер S_1 входят 4 объекта, расстояние от которых до объекта №5 составляет соответственно: 2, 5, 6, 7. Чему равно расстояние от объекта №5 до кластера S_1, если исходить из принципа “дальнего соседа”:

а) 2; в) 6;
б) 5; г) 7.

8. В условиях гетероскедастичности случаных остатков оценки коэффициентов, полученные по методу наименьших квадратов, будут:

а) несмещенными; в) эффективными; д) надежными;
б) смещенными; г) неэффективными; е) ненадежными.

9. Условием гетероскедастичности является:

а) независимость значений $M\varepsilon_t^2 = \sigma_t^2$ от t и x_t;
б) зависимость значений $M\varepsilon_t^2 = \sigma_t^2$ от t и x_t;
в) ситуация не определена.

10. Систему $y_t = B^{-1}\Gamma x_t + B^{-1}\varepsilon_t$ одновременных уравнений называют рекурсивной, если выполняются следующие условия (выберите необходимые условия):

а) Матрица значений эндогенных переменных является нижней треугольной матрицей, т. е. $\beta_{ij} = 0$ при j>i и $\beta_{ii} = 1$;

б) случайные ошибки независимы между собой, т. е. $\sigma_{ii} > 0, \sigma_{ij} = 0$ при $i \neq j$, где i,j=1,2,...,G;

в) каждое ограничение на структурные коэффициенты относится к отдельному уравнению.

Учебная программа

Сведения об авторах

Мхитарян Владимир Сергеевич – доктор экономических наук, профессор, академик Международной академии наук Высшей школы, заведующий кафедрой математической статистики и эконометрики.

Архипова Марина Юрьевна – кандидат экономических наук, доцент кафедры экономической статистики и эконометрики, зав. сектором Института проблем информатизации РАН.

Сиротин Вячеслав Павлович – кандидат технических наук, доцент, профессор кафедры экономической статистики и эконометрики.

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цель преподавания курса – дать студентам научное представление о методах, моделях и приемах, позволяющих получать количественные выражения закономерностям экономической теории на базе экономической статистики с использованием математико-статистического инструментария. Курс рассчитан на 64 часа.

Задачи курса. В соответствии с целью курса студенты должны усвоить методы количественной оценки социально-экономических процессов, научиться содержательно интерпретировать формальные результаты.

*Связь с другими дисциплинами***.** Курс базируется на дисциплинах «Теория вероятностей» и «Математическая статистика». В свою очередь является основой для дисциплины «Статистические методы прогнозирования».

Курс изучается в форме лекций (2 час/нед) и практических занятий (2 час/нед). Предусмотрена самостоятельная подготовка студентов. Они выполняют индивидуальные компьютерные исследования, сдают экзамен.

2. СОДЕРЖАНИЕ КУРСА

Тема 1. *Корреляционный анализ.*

Задачи эконометрики в области социально-экономических исследований. Информационные технологии на базе ПЭВМ в эконометрическом исследованиях. Классификация переменных в эконометрических исследованиях. Понятия спецификации и идентифицируемости модели. Примеры эконометрических моделей (модель предложения и спроса на конкурентном рынке).

Оценка ковариационной (корреляционной) матрицы. Оценки частных и множественных коэффициентов корреляции. Проверка их значимости. Построение доверительных интервалов для частных коэффициентов.

Тема 2. *Регрессионный анализ.*

Основные задачи регрессионного анализа. Особенности классической линейной модели множественной регрессии. Выбор адекватного уравнения регрессии.

Оценка вектора коэффициентов уравнения регрессии и остаточной дисперсии с помощью метода наименьших квадратов (МНК).

Проверка значимости и интервальное оценивание коэффициентов и уравнения регрессии, проблема мультиколлинеарности, пошаговые алгоритмы регрессионного анализа.

Тема 3. *Методы многомерной классификации. Кластерный анализ.*

Построение типологических регрессионных моделей по неоднородным данным. Основные понятия кластерного анализа, расстояния между объектами и мера близости. Функционалы качества разбиения. Иерархические кластер-процедуры.

Метод К-средних. Классификация больших совокупностей объектов методами параллельных процедур.

Построение множественных регрессионных моделей по типологическим группам.

Тема 4. *Производственные функции.* Производственная функция Кобба –- Дугласа.

Оценивание параметров производственной функции Кобба – Дугласа по пространственной и временной информации.

Тема 5. *Системы одновременных эконометрических уравнений.*

Структурная и приведенная формы модели систем одновременных уравнений. Рекурсивные системы одновременных уравнений. Модель спроса – предложения как пример системы одновременных уравнений. Основные структурные характеристики моделей. Условия идентифицируемости уравнений системы. Идентификация рекурсивных систем.