МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

КЕМЕРОВСКИЙ ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ ПИЩЕВОЙ ПРОМЫШЛЕННОСТИ

Л.А. ЯКОВЛЕВА

ЭКОНОМЕТРИКА

Комплексное учебное пособие для студентов экономических специальностей заочного дистанционного обучения

КЕМЕРОВО 2002

СОДЕРЖАНИЕ

ЭКОНОМЕТРИКА

Предмет и задачи эконометрики

Рыночная экономика требует улучшения использования эконометрической статистической информации, характеризующей результат экономической деятельности. Эффективная хозяйственная деятельность в условиях рыночной экономики предполагает оценку связей между различными экономическими факторами, выявление их тенденций и разработку экономических нормативов и прогнозов. необходимо выделить факторы, которые положительно или отрицательно влияют на результаты хозяйственной деятельности на микро, мезо, макроуровнях.

Эконометрика и компьютерная грамотность в условиях рыночной экономики являются предпосылками к успешному ее функционированию. Руководитель предприятия, применяющий эконометрические методы, получит дополнительную информацию для обоснования своих решений.

Предметом эконометрики являются факторы, формирующие развитие эконометрических явлений и процессов.

Эконометрические переменные не меняются во времени независимо друг от друга, они имеют множество взаимосвязей. Как установить такие связи, как доказать их взаимосвязь, как оценить их параметры? На эти и другие вопросы дает ответ эконометрика.

Определение: Эконометрика – наука, которая определяет количественное выражение взаимосвязей экономических явлений и процессов.

В любом экономическом процессе присутствует объясняющий фактор (аргумент) и объясняемый фактор (результативный показатель).

Например, формирующийся на рынке спрос на некоторый товар рассматривается как функция его цены. Затраты, связанные с изготовлением какого-либо продукта зависят от объема производства. Потребительские расходы могут быть функциями дохода.

Эконометрика возникла вследствие междисциплинарного подхода к изучению экономики. Эконометрика является синтезом четырех наук: экономической теории, статистики, математики, информатики.

Назначение и цель эконометрики – экономические и социально-экономические приложения.

Все многообразие задач, решаемых с помощью эконометрики можно классифицировать по трем параметрам:

1. по конечным прикладным целям

а) прогноз экономических и социально-экономических показателей, характеризующих состояние и развитие анализируемой системы;

б) имитация различных возможных сценариев социального развития анализируемой системы.

2. по уровню иерархии экономической системы

а) макроуровень (страна в целом);

б) мезоуровень (регионы, отрасли, корпорации);

в) микроуровень (фирмы, предприятия, семьи).

3. по профилю экономического исследования:

проблема рынка, инвестиции, финансы, социальная сфера, ценообразование, распределительные отношения, спрос и потребление.

Этапы эконометрического исследования.

1. Постановка проблемы
2. Получение данных и их качественный анализ, выделение зависимых и независимых переменных
3. Спецификация (выбор вида) модели
4. Оценка параметров модели (проверка ряда гипотез)
5. Интерпретация результатов.

ПАРНАЯ РЕГРЕССИЯ

Парная регрессия - уравнение связи двух переменных y и x:

$$y = \widehat{f}(x)$$

где y - зависимая переменная (результативный признак);

x - независимая, объясняющая переменная (признак-фактор).

Различают линейные и нелинейные регрессии.

Линейная регрессия: $y = a + b * x + \varepsilon$.

где *а, в* – параметры модели;

е – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уровню регрессии.

Случайная величина *е* называется возмущением. Она включает также не учтенные модели факторы, случайные ошибки, влияние особенностей измерения.

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Регрессии, нелинейные по объясняющим переменным:

• полиномы разных степеней $y = a + b_1 * x + b_2 * x^2 + b_3 * x^3 + \varepsilon$;

• равносторонняя гипербола $y = a + \frac{b}{x} + \varepsilon$.

Регрессии, нелинейные по оцениваемым параметрам:

• степенная $y = a * x^b * \varepsilon$

• показательная $y = \dot{a} * b^x * \varepsilon$,

• экспоненциальная $y = e^{a+b*x} * \varepsilon$

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака у от теоретических $\hat{y}_x$ минимальна, т.е.

$$\sum (y - \hat{y}_x)^2 \to \min$$

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно а и b:

$$\begin{cases} na + b\sum x = \sum y \\ a\sum x + b\sum x^2 = \sum yx \end{cases}$$

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

$$a = \bar{y} - b * \bar{x}, \qquad b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\overline{y * x} - \bar{y} * \bar{x}}{\overline{x^2} - \bar{x}^2}.$$

Параметр *в* коэффициент регрессии. Его величина показывает среднее изменение результата *у* с изменением фактора на 1 единицу.

Например, функция издержек при производстве некоторой продукции имеет вид

$$\hat{y}_x = 3000 + 2x,$$

где *у* – издержки (тыс. руб.)

х – количество единиц продукции

Параметр *в*=2, таким образом с увеличением объема продукции на 1 единицу, издержки производства возрастают в среднем на 2 тысячи. Отсюда следует, что дополнительный прирост выпуска продукции на 1 единицу потребует увеличения затрат в среднем на 2 тысячи рублей.

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции r_{xy} для линейной регрессии ($-1 \le r_{xy} \le 1$):

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\overline{y * x} - \bar{y} * \bar{x}}{\sigma_x \sigma_y}.$$

и индекс корреляции p_{xy} - для нелинейной регрессии ($0 \le p_{xy} \le 1$):

$$p_{xy} = \sqrt{1 - \frac{\sigma^2_{oct}}{\sigma^2_y}} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}$$

Оценку качества построенной модели даст коэффициент (индекс) детерминации, а также средняя ошибка аппроксимации.

Средняя ошибка аппроксимации - среднее отклонение расчетных значений от фактических:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| * 100\%$$

Допустимый предел значений $\bar{A}$ - не более 8 - 10%. Средний коэффициент эластичности $\bar{Э}$ показывает, на сколько процентов в среднем по совокупности изменится результат у от своей средней величины при изменении фактора *x* на 1% от своего среднего значения:

$$\bar{Э} = f'(x) \frac{\bar{x}}{\bar{y}}$$

Задача дисперсионного анализа состоит в анализе дисперсии зависимой переменной:

$$\sum (y - \bar{y})^2 = \sum (\hat{y}_x - \bar{y})^2 + \sum (y - \hat{y}_x)^2$$

где $\sum (y - \bar{y})^2$ - общая сумма квадратов отклонений;

$\sum (\hat{y}_x - \bar{y})^2$ -сумма квадратов отклонений, обусловленная регрессией («объясненная» или «факторная»);

$\sum (y - \hat{y}_x)^2$ - остаточная сумма квадратов отклонений.

Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака у характеризует коэффициент (индекс) детерминации R^2:

$$R^2 = \frac{\sum (\hat{y}_x - \bar{y})^2}{\sum (y - \bar{y})^2}$$

Коэффициент детерминации - квадрат коэффициента или индекса корреляции.

F-тест - оценивание качества уравнения регрессии - состоит в проверке гипотезы H_o о статистической незначимости уравнения регрессии и показателя тесноты связи. Для этого выполняется сравнение фактического $F_{факт}$ критического (табличного) $F_{табл}$ значений F-критерия Фишера. $F_{факт}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:

$$\mathbf{F_{факт}} = \frac{\sum(\hat{\mathbf{y}} - \bar{\mathbf{y}})^2 / \mathbf{m}}{\sum(\mathbf{y} - \hat{\mathbf{y}})^2 / (\mathbf{n} - \mathbf{m} - 1)} = \frac{\mathbf{r}_{xy}^2}{1 - \mathbf{r}_{xy}^2}(\mathbf{n} - 2)$$

где n - число единиц совокупности;

m - число параметров при переменных *x*.

$F_{табл}$ – это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости *a*. Уровень значимости *a* - вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно, а принимается равной 0,05 или 0,01.

Если $F_{табл} < F_{факт}$, то H_o - гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность. Если $F_{табл} > F_{факт}$ то гипотеза H_o не отклоняется и признается статистическая незначимость, ненадежность уравнения регрессии.

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитываются критерий Стьюдента и доверительные интервалы каждого из показателей. Выдвигается гипотеза H_o о случайной природе показателей, т.е. о незначимом их отличии от нуля. Оценка значимости коэффициентов регрессии и корреляции с помощью t-критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_b = \frac{b}{m_b}; \quad t_a = \frac{a}{m_a}\;; \quad t_r = \frac{r}{m_r}$$

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$\mathbf{m_b} = \sqrt{\frac{\sum(\mathbf{y}-\hat{\mathbf{y}}_\mathbf{x})^2/(\mathbf{n}-2)}{\sum(\mathbf{x}-\bar{\mathbf{x}})^2}} = \sqrt{\frac{\mathbf{S}^2_{\mathbf{ост}}}{\sum(\mathbf{x}-\bar{\mathbf{x}})^2}} = \frac{\mathbf{S}_{\mathbf{ост}}}{\boldsymbol{\sigma}_\mathbf{x}\sqrt{\mathbf{n}}}$$

$$m_a = \sqrt{\frac{\sum(y-\hat{y}_x)^2}{(n-2)}\frac{\sum x^2}{n\sum(x-\bar{x})^2}} = \sqrt{S^2_{ост}\frac{\sum x^2}{n^2\sigma_x^2}} = S_{ост}\frac{\sqrt{\sum x^2}}{n\sigma_x}$$

$$m_{r_{xy}} = \sqrt{\frac{1-r_{xy}^2}{n-2}}$$

Сравнивая фактическое и критическое (табличное) значения t-статистики - $F_{табл}$ и $t_{факт}$ - принимаем или отвергаем гипотезу H_o.

Связь между F-критерием Фишера и t-статистикой Стьюдента выражается равенством

$$t_r^2 = t_b^2 = \sqrt{F}$$

Если $t_{табл} < t_{факт}$ то H_o отклоняется, т.е. a, b и r_{xy} не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора *x*. Если $t_{табл} > t_{факт}$ то гипотеза H_o не отклоняется и признается случайная природа формирования a, b или r_{xy}.

Для расчета доверительного интервала определяем предельную ошибку Δ для каждого показателя:

$$\Delta_a = t_{табл} m_a \qquad \Delta_b = t_{табл} m_b$$

Формулы для расчета доверительных интервалов имеют следующий вид:

$$\gamma_a = a \pm \Delta_a \quad \gamma_{a_{\min}} = a - \Delta_a \quad \gamma_{a_{\max}} = a + \Delta_a$$

$$\gamma_b = b \pm \Delta_b \quad \gamma_{b_{\min}} = a - \Delta_b \quad \gamma_{b_{\max}} = b \pm \Delta_b$$

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.

Прогнозное значение y_p определяется путем подстановки в уравнение регрессии y соответствующего (прогнозного) значения x_p. Вычисляется средняя стандартная ошибка прогноза $m_{\hat{y}_p}$:

$$m_{\hat{y}_p} = \sigma_{ост}\sqrt{1 + \frac{1}{m} + \frac{(x_p - \bar{x})^2}{\sum(x - \bar{x})^2}}$$

где $\sigma_{ост} = \sqrt{\frac{\sum(y - \hat{y})^2}{n - m - 1}}$, и строится доверительный интервал прогноза:

$$\gamma_{\hat{y}_p} = \hat{y}_p \pm \Delta_{\hat{y}_p} \qquad \gamma_{\hat{y}_{p\min}} = \hat{y}_p - \Delta_{\hat{y}_p} \qquad \gamma_{\hat{y}_{p\max}} = \hat{y}_p + \Delta_{\hat{y}_p}$$

РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Пример 1

По семи территориям Уральского района за 199Х г. известны значения двух признаков (табл. 1).

Район	Среднедневная заработная плата одного работающего, руб., x	Расходы на покупку продовольственных товаров в общих расходах, %, y
Удмуртская респ.	45,1	68,8
Свердловская обл.	59,0	61,2
Башкортостан	57,2	59,9
Челябинская обл.	61,8	56,7
Пермская обл.	58,8	55,0
Курганская обл.	47,2	54,3
Оренбургская обл.	55.2	49.3

табл. 1

Требуется:

1. Для характеристики зависимости *y* от *x* рассчитать параметры следующих функций:

а) линейной;

б) степенной;

в) показательной;

г) равносторонней гиперболы.

2. Оценить каждую модель через среднюю ошибку аппроксимации $\overline{A}$ и F-критерий Фишера.

Решение

1а. Для расчета параметров a и b линейной регрессии y = a + b * x решаем систему нормальных уравнений относительно a и b:

$$\begin{cases} na + b\sum x = \sum y \\ a\sum x + b\sum x^2 = \sum yx \end{cases}$$

По исходным данным рассчитываем $\sum y, \sum x, \sum yx, \sum x^2, \sum y^2$.

$$b = \frac{\overline{y * x} - \overline{y} * \overline{x}}{\sigma_x^2}.$$

$$a = \overline{y} - b * \overline{x}.$$

Уравнение регрессии: $\hat{y}$ = 76,88 - 0,35 * *x*. С увеличением среднедневной заработной платы на 1 руб. доля расходов на покупку продовольственных товаров снижается в среднем на 0,35 %-ных пункта.

Рассчитаем линейный коэффициент парной корреляции:

$$r_{xy} = b\frac{\sigma_x}{\sigma_y}.$$

Связь умеренная, обратная.

Определим коэффициент детерминации:

$$r_{xy} = 0{,}127$$

Вариация результата на 12,7% объясняется вариацией фактора x. Подставляя в уравнение регрессии фактические значения x, определим теоретические (расчетные) значения $\hat{y}_x$. Найдем величину средней ошибки аппроксимации $\overline{A}$:

$$\overline{A} = \frac{1}{n}\sum A_j = \frac{1}{n}\sum |y - \hat{y}| * 100\% = 8{,}1\%$$

В среднем расчетные значения отклоняются от фактических на 8,1%.

Рассчитаем F-критерий:

$$F_{факт} = 0{,}7$$

поскольку $1 < F < \infty$, следует рассмотреть F^{-1}.

Полученное значение указывает на необходимость принять гипотезу H_o о случайной природе выявленной зависимости и статистической незначимости параметров уравнения и показателя тесноты связи.

1б. Построению степенной модели $y = a * x^б$ предшествует процедура линеаризации переменных. В примере линеаризация производится путем логарифмирования обеих частей уравнения:

$$\lg y = \lg a + b * \lg x$$

$$Y = C + b * X$$

где $Y = \lg y, X = \lg x, C = \lg a$

Для расчетов используем данные табл. 1.

Рассчитаем C и b:

$$b = \frac{\overline{Y * X} - \overline{Y} * \overline{X}}{\sigma_X^2}.$$

$$C = \overline{Y} - b * \overline{X}$$

Получим линейное уравнение: $\hat{Y} = 2{,}278 - 0.298 * X$.

Выполнив его потенцирование, получим: $\hat{Y}$

Подставляя в данное уравнение фактические значения x, получаем теоретические значения результата $\hat{y}_x$. По ним рассчитаем показатели: тесноты связи - индекс корреляции p_{xy} и среднюю ошибку аппроксимации $\overline{A}_j$:

$$p_{xy} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \overline{y})^2}} = 0{,}3758 \qquad \overline{A} = 8{,}0\%$$

Характеристики степенной модели указывают, что она несколько лучше линейной функции описывает взаимосвязь.

1в. Построению уравнения показательной кривой $y = a * b^x$ предшествует процедура линеаризации переменных при логарифмировании обеих частей уравнения:

$$\lg y = \lg a + x \cdot \lg b$$
$$Y = C + B \cdot x$$

где $Y = \lg y,\ C = \lg a,\ B = \lg b$

Для расчетов используем данные табл. 1.

Значения параметров регрессии A и B составили:

$$B = \frac{\overline{Y * x} - \overline{Y} * \overline{x}}{\sigma_x^2} = -0.0023$$
$$A = \overline{Y} - B * \overline{x}.$$

Получено линейное уравнение: $\hat{Y}$ = 1,887 - 0,0023 * x. Произведем потенцирование полученного уравнения и запишем его в обычной форме:

$$\hat{Y} = 10^{1.887} * 10^{-0.0023\,x} = 77.1 * 0.9947^{x}$$

Тесноту связи оценим через индекс корреляции p_{xy} :

$$P_{xy} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \hat{y})^2}} = \sqrt{1 - \frac{28.27}{32.92}} = 0.3589$$

Связь умеренная.

$\overline{A}$ = 8,0%, что говорит о повышенной ошибке аппроксимации, но в допустимых пределах. Показательная функция чуть хуже, чем степенная, она описывает изучаемую зависимость.

1г. Уравнение равносторонней гиперболы $y = a + b\frac{1}{x}$ линеаризуется при замене: $z = \frac{1}{x}$. Тогда y = a + b * z. Для расчетов используем данные табл. 1.

Значения параметров регрессии a и b составили:

$$a = \overline{y} - b * \overline{x} = 38{,}5$$
$$b = \frac{\overline{y * z} - \overline{y} * \overline{z}}{\sigma_z^2} = 1051.4$$

Получено уравнение: $\hat{y} = 38.5 + 1051.4 \cdot \frac{1}{x}$

Индекс корреляции: $p_{xy} = \sqrt{1 - \frac{27.84}{32.92}} = 0.3944$

$\overline{A}$ = 8,1%. По уравнению равносторонней гиперболы получена наибольшая оценка тесноты связи: $p_{xy} = 0{,}3944$ (по сравнению с линейной, степенной и показательной регрессиями). $\overline{A}$ остается на допустимом уровне:

2. $F_{факт} = \frac{p_{yx}^2}{1 - p_{yx}^2} \cdot \frac{n - m - 1}{m} = \frac{0{,}1555}{0{,}8445} \cdot 5 = 0{,}92$

где $F_{табл}$ = 6,6 > $F_{факт}$, a = 0,05

Следовательно, принимается гипотеза H_o о статистически незначимых параметрах этого уравнения. Этот результат можно объяснить сравнительно невысокой теснотой выявленной зависимости и небольшим числом наблюдений.

<u>Пример 2</u>

По территориям региона приводятся данные за 199X г. (см. табл.).

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, РУб., y
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

табл. 2

<u>Требуется:</u>

1. Построить линейное уравнение парной регрессии y от x.
2. Рассчитать линейный коэффициент парной корреляции и среднюю ошибку аппроксимации.
3. Оценить статистическую значимость параметров регрессии и корреляции.
4. Выполнить прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимумах, составляющем 107% от среднего уровня.
5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

<u>Решение</u>

1. Для расчета параметров уравнения линейной регрессии строим расчетную таблицу (см. табл.).

$$b = \frac{\overline{y * x} - \bar{y} * \bar{x}}{\sum x^2 - (\bar{x})^2} = 0{,}92$$

$$a = \bar{y} - b * \bar{x} = 77$$

Получено уравнение регрессии: y = 77 + 0,92 * x. С увеличением среднедушевого прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,92 руб.

2. Тесноту линейной связи оценит коэффициент корреляции:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} 0{,}721 \qquad r_{xy}^2 = 0{,}52$$

Это означает, что 52% вариации заработной платы (y) объясняется вариацией фактора x - среднедушевого прожиточного минимума. Качество модели определяет средняя ошибка аппроксимации:

$$\overline{A} = \frac{1}{n}\sum A_j = \frac{68{,}9}{12} = 5{,}7\%$$

Качество построенной модели оценивается как хорошее, так как $\overline{A}$ не превышает 8 - 10%.

3. Оценку статистической значимости параметров регрессии проведем с помощью t-статистики Стьюдента и путем расчета доверительного интервала каждого из показателей.

Выдвигаем гипотезу H_O статистически незначимом отличии показателей от нуля: a = b = r_{xy} = 0.

$t_{табл}$ для числа степеней свободы f = n - 2 = 12 – 2 = 10 и a = 0,05 составит 2,23.

Определим случайные ошибки $m_a, m_b, m_{r_{xy}}$

$$m_a = 12{,}6\frac{\sqrt{89907}}{12 \cdot 12{,}95} = 24{,}3 \qquad m_b = \frac{12{,}6}{12{,}95 \cdot \sqrt{12}} = 0{,}281$$

$$m_{r_{xy}} = \sqrt{\frac{1 - 0{,}520}{12 - 2}} = 0{,}219$$

Тогда

$$t_a = \frac{77}{24{,}3} = 3{,}2 \qquad t_b = \frac{0{,}92}{0{,}281} = 3{,}3 \qquad t_{r_{xy}} = \frac{0{,}721}{0{,}219} = 3{,}3$$

Фактические значения t-статистики превосходят табличные значения:

$$t_a = 3{,}2 > t_{табл} = 2{,}3 \qquad t_b = 3{,}3 > t_{табл} = 2{,}3 \qquad t_{r_{xy}} = 3{,}3 > t_{табл} = 2{,}3$$

поэтому гипотеза H_o отклоняется, т.е. a, b и r_{xy} не случайно отличаются от нуля, а статистически значимы.

Рассчитаем доверительный интервал для a и b. Для этого определим предельную ошибку для каждого показателя:

$$\Delta_a = 2{,}23 * 24{,}3 = 54 \qquad \Delta_a = 2{,}23 * 0{,}281 = 0{,}62$$

Доверительные интервалы:

$$\gamma_a = a \pm \Delta_a = 77 \pm 54 \qquad \gamma_{a_{\min}} = 77 - 54 = 23 \qquad \gamma_{a_{\max}} = 77 + 54 = 131$$

$$\gamma_b = b \pm \Delta_b = 0{,}92 \pm 0{,}62 \qquad \gamma_{b_{\min}} = 0{,}92 - 0{,}62 = 0{,}30 \qquad \gamma_{b_{\min}} = 0{,}92 + 0{,}62 = 23$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью p = 1 - a = 0,95 параметры a и b, находясь в указанных границах, не принимают нулевых значений, т.е. не являются статистически незначимыми и существенно отличны от нуля.

4. Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Если прогнозное значение прожиточного минимума составит: $x_p = \bar{x} \cdot 1{,}07 = 85{,}6 \cdot 1{,}07 = 91{,}6$ тыс. руб., тогда прогнозное значение прожиточного минимума составит:

$$\hat{y}_p = 77 + 0{,}92 \cdot 91{,}6 = 161 \text{ тыс. руб.}$$

5. Ошибка прогноза составит:

$$m_{\hat{y}_p} = 12{,}6 \cdot \sqrt{1 + \frac{1}{12} + \frac{(91{,}6 - 85{,}6)^2}{12 \cdot 12{,}95^2}} = 13{,}2 \quad \textit{тыс.руб}$$

Предельная ошибка прогноза, которая в 95% случаев не будет превышена, составит:

$$\Delta_{\hat{y}_p} = t_{табл} \cdot m_{\hat{y}_p} = 2{,}23 \cdot 13{,}2 = 29{,}4$$

Доверительный интервал прогноза:

$$\gamma_{\hat{y}_p} = 91{,}6 \pm 29{,}4$$

$$\gamma_{\hat{y}_{p\min}} = 91{,}6 - 29{,}4 = 62{,}2 \qquad \gamma_{\hat{y}_{p\max}} = 91{,}6 + 29{,}4 = 121$$

Выполненный прогноз среднемесячной заработной платы оказался надежным (p = 1 – a = 1 - 0,05 = 0,95), но неточным, так как диапазон верхней и нижней границ доверительного интервала D_γ составляет 1,95 раза:

$$D_\gamma = \frac{\gamma_{\hat{y}_{\max}}}{\gamma_{\hat{y}_{\min}}} = \frac{121}{62} = 1{,}95$$

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

Множественная регрессия - уравнение связи с несколькими независимыми переменными:

$$y = f(x_1, x_2, ..., x_p)$$

где y - зависимая переменная (результативный признак);

$x_1, x_2, ..., x_p$ независимые переменные (факторы).

Для построения уравнения множественной регрессии чаще используются следующие функции:

- линейная $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p + \varepsilon$
- степенная $y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot x_p^{b_p} \cdot \varepsilon$
- экспонента $y = \exp(a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p + \varepsilon)$
- гипербола $y = \dfrac{1}{a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p + \varepsilon}$

Можно использовать и другие функции, приводимые к линейному виду.

Для оценки параметров уравнения множественной регрессии применяют метод наименьших квадратов (МНК). Для линейных уравнений и нелинейных уравнений, приводимых к линейным, строится следующая система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

$$\begin{cases} \sum y = na + b_1 \sum x_1 + b_2 \sum x_2 + ... + b_p \sum x_p \\ \sum yx_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + ... + b_p \sum x_p x_1 \\ .... \\ \sum yx_p = a \sum x_p + b_1 \sum x_1 x_p + b_2 \sum x_2 x_p + ... + b_p \sum x_p^2 \end{cases}$$

Для ее решения может быть применён метод определителей:

$$a = \frac{\Delta a}{\Delta}, \quad b_1 = \frac{\Delta b_1}{\Delta}, ..., \quad b_p = \frac{\Delta b_p}{\Delta}$$

где $\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 & ... & \sum x_p \\ \sum x_2 & \sum x_1^2 & \sum x_2 x_1 & ... & \sum x_p x_1 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & ... & \sum x_p x_2 \\ .... & & & & \\ \sum x_p & \sum x_1 x_p & \sum x_2 x_p & ... & \sum x_p^2 \end{vmatrix}$ - определитель системы.

$\Delta a, \Delta b_1, ..., \Delta b_p$ - частные определители; которые получаются путем замены соответствующего столбца матрицы определителя системы данными левой части системы.

Другой вид уравнения множественной регрессии - *уравнение регрессии в стандартизованном масштабе*:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + ... + \beta_p t_{x_p}$$

где $t_y = \dfrac{y - \bar{y}}{\sigma_y}$, $t_{x_i} = \dfrac{x_i - \bar{x}_i}{\sigma_{x_i}}$ - стандартизованные переменные

β_i - стандартизованные коэффициенты регрессии.

К уравнению множественной регрессии в стандартизованном масштабе применим МНК. Стандартизованные коэффициенты регрессии (β-коэффициенты) определяются из следующей системы уравнений:

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_2x_1} + \beta_3 r_{x_3x_1} + \ ... \ + \beta_p r_{x_px_1}, \\ r_{yx_2} = \beta_1 r_{x_2x_1} + \beta_2 + \beta_3 r_{x_3x_2} + \ ... \ + \beta_p r_{x_px_2}, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ r_{yx_p} = \beta_1 r_{x_px_1} + \beta_2 r_{x_px_2} + \beta_3 r_{x_3x_p} + \ ... \ + \beta_p \end{cases}.$$

Связь коэффициентов множественной регрессии b_1, со стандартизованными коэффициентами β_i, описывается соотношением

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}$$

Параметр а определяется как $a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - ... - b_p\bar{x}_p$

Средние коэффициенты эластичности для линейной регрессии рассчитываются по формуле

$$\bar{Э}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}}$$

Для расчета частных коэффициентов эластичности применяется следующая формула:

$$Э_{y_{xi}} = b_i \frac{x_i}{\hat{y}_{x_i \cdot x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p}}$$

Тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции:

$$R_{yx_1x_2,...,x_p} = \sqrt{1-\frac{\sigma^2_{y_{ост}}}{\sigma^2_y}}$$

Значение индекса множественной корреляции лежит в пределах от 0 до 1 и должно быть больше или равно максимальному парному индексу корреляции:

$$R_{yx_1x_2,...,x_p} \geq r_{yx_i} \qquad (i=\overline{1,p})$$

Индекс множественной корреляции для уравнения в стандартизованном масштабе можно записать в виде

$$R_{yx_1x_2,...,x_p} = \sqrt{\sum \beta_i r_{yx_i}}$$

При линейной зависимости коэффициент множественной корреляции можно определить через матрицу парных коэффициентов корреляции:

$$R_{yx_1x_2,...,x_p} = \sqrt{1-\frac{\Delta r}{\Delta r_{11}}}$$

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & .... & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & .... & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & .... & r_{x_2x_p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & .... & 1 \end{vmatrix} - \textit{определитель матрицы}$$

-парных коэффициентов корреляции;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & .... & r_{x_1x_p} \\ r_{x_2x_1} & 1 & .... & r_{x_2x_p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{x_px_1} & r_{x_px_2} & .... & 1 \end{vmatrix} - \textit{определитель матрицы}$$

-межфакторной корреляции.

Скорректированный индекс множественной детерминации содержит поправку на число степеней свободы и рассчитывается по формуле

$$\widehat{R}^2 = 1-(1-R^2)\frac{(n-1)}{(n-m-1)}$$

где n - число наблюдений;

m - число факторов.

Значимость уравнения множественной регрессии в целом оценивается с помощью F-критерия Фишера:

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

Частный F-критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении. В общем виде для фактора x_i частный F-критерий определится как

$$F_{част_{x_i}} = \frac{R^2_{yx_1\ldots x_i\ldots x_p} - R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_p}}{1-R^2_{yx_1\ldots x_i\ldots x_p}} \cdot \frac{n-m-1}{1}$$

Оценка значимости коэффициентов чистой регрессии с помощью t-критерия Стьюдента сводится к вычислению значения

$$t_{b_1} = \frac{b_i}{m_{b_i}} = \sqrt{F_{x_i}}$$

где m_{b_i} - средняя квадратическая ошибка коэффициента регрессии b_i она может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1-R^2_{yx_1\ldots x_p}}}{\sigma_{x_i} \cdot \sqrt{1-R^2_{x_ix_1\ldots x_p}}} \cdot \frac{1}{\sqrt{n-m-1}}$$

При построении уравнения множественной регрессии может возникнуть проблема мультиколлинеарности факторов, их тесной линейной связанности.

Считается, что две переменные явно коллинеарны, т.е. находятся между собой в линейной зависимости, если $r_{x_ix_j} > 0,7$.

РЕШЕНИЕ С ПОМОЩЬЮ ППП EXCEL

1.Статистическая функция **ЛИНЕЙН** определяет параметры линейной регрессии $y = a + b \cdot x$

. Порядок вычисления следующий:

1) введите исходные данные или откройте существующий файл, содержащий анализируемые данные;

2) выделите область пустых ячеек 5×2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики или область 1×2 – для получения только оценок коэффициентов регрессии;

3) активизируйте Мастер функций любым способом:

а) в главном меню выберите **Вставка/Функция;**

б) на панели инструментов **Стандартная** щелкните по кнопке **Вставка функции;**

4) в окне Категория выберите **Статистические,** в окне Функция – **ЛИНЕЙН**. Щелкните по кнопке **ОК;**

5) заполните аргументы функции:

Известные_значения_у – диапазон, содержащий данные результативного признака;

Известные_значения_х – диапазон, содержащий данные факторов независимого признака;

Константа – логическое значение, которое указывает на наличие или отсутствие свободного члена в уравнении; если *Константа* = 1, то свободный член рассчитывается обычным образом, если *Константа* = 0, то свободный член равен 0;

Статистика - логическое значение, которое указывает, выводить дополнительную информацию по регрессивному анализу или нет. Если *Статистика* = 1, то дополнительная информация выводится, если *Статистика* = 0, то выводятся только щценки параметров уравнения.

Щелкните по кнопке **OK;**

6) в левой верхней ячейке выделеннной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу <F2>, а затем – на комбинацию клавиш <CTRL>+<SHIFT>+<ENTER>.

Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

Значение коэффициента b	Значение коэффициента a
Среднеквадратическое отклонение b	Среднеквадратическое отклонение a
Коэффициент детерминации R^2	Среднеквадратическое отклонение y
F-статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

2. Для вычисления параметров экспоненциальной кривой $y = \alpha \cdot \beta^2$ в MS Excel применяется встроенная статистическая функция **ЛГРФПРИБЛ.**

Порядок вычисления аналогичен функции **ЛИНЕЙН**.

3.С помощью инструмента анализа данных Регрессия, помимо результатов регрессионной статистики, дисперсионного анализа и доверительных интервалов, можно получить остатки и **графики**

подбора линии регрессии и нормальной вероятности. Порядок действий следующий:

1) проверьте доступ к пакету анализа. В главном меню последовательно выберите **Сервис/Надстройки**. Установите флажок **Пакет анализа**;

2) в главном меню выберите **Сервис/Анализ данных/Регрессия**. Щелкните по кнопке **ОК**;

3) Заполните диалоговое окно ввода данных и параметров вывода:

Входной интервал Y – диапазон, содержащий данные результативного признака;

Входной интервал X - диапазон, содержащий данные факторов независимого признака;

Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Константа – ноль – флажок, указывающий, на наличие или отсутствие свободного члена в уравнении;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно задать произвольное имя нового листа.

Если необходимо получить информацию и графики остатков, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке **ОК.**

АППРОКСИМАЦИЯ ДАННЫХ И ПОЛУЧЕНИЕ ФУНКЦИОНАЛЬНЫХ ЗАВИСИМОСТЕЙ С ПОМОЩЬЮ ЛИНИЙ ТРЕНДА

Средства деловой графики позволяют найти уравнения регрессии (до 6 включительно) и не прибегая к вычислениям. Построить кривую функции Y(x) (при этом выбрать тип диаграммы - **Точечная**), щелкнуть на ней правой кнопкой мыши, в появившемся контекстном меню можно выбрать пункт **Добавить линию тренда**, который предъявляет окно **Линия тренда**. Здесь можно выбрать вид уравнения аппроксимации и его степень, а если во вкладыше **Параметры** установить флаг **Показывать уравнение на диаграмме**, то на графике мы увидим не только линию тренда, но и его уравнение. Если во вкладыше **Параметры** установить флаг **поместить на диаграмму величину достоверности аппроксимации,** то можно увидеть значение коэффициента детерминации. Здесь можно визуально оценить поведение анализируемого процесса в будущем/прошлом, если установить **Прогноз вперед/назад** на заданное число единиц независимого аргумента X.

ЛИТЕРАТУРА

1. *Айвазян С.А., Мхитарян* В.С. Прикладная статистика и основы эконометрики: Учебник. – М.: ЮНИТИ, 2001.
2. *Джонстон Дж*. Эконометрические методы. – М.: Статистика, 1980.
3. *Данилов Н.Н.* Курс математической экономики. – Новосибирск: Издательство СОРАН, 2002.
4. *Доугерти. К.* Введение в эконометрику. – М.: Финансы и статистика, 1999.

ЗАДАНИЯ ДЛЯ КОНТРОЛЬНОЙ РАБОТЫ

1. Рассчитайте параметры уравнений линейной, степенной парной регрессии. Поясните смысл коэффициентов.
2. Оцените тесноту связи с помощью показателей корреляции и детерминации.
3. Оцените качество уравнений с помощью средней ошибки аппроксимации.
4. Оцените с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования, выберете лучшее уравнение регрессии и дайте обоснование.
5. Рассчитайте прогнозное значение от среднего значения параметра x.
6. Оцените полученные результаты и сделайте выводы.

Требования к оформлению:

В контрольной работе должен быть представлен № варианта, текст задания, кратко описан порядок выполнения работы (со ссылками на № таблиц, № страниц).

Необходимо представить таблицу с итоговыми расчетами в режиме чисел, в режиме формул с указанием заголовков строк и столбцов.

Отчет представить на бумаге и дискете.

Вариант задания выбирается по двум последним цифрам зачетной книжки: первые 15 номеров совпадают с номером варианта, для последующих номеров две последние цифры номера зачетки надо поделить на 15, остаток от деления = № варианта , (остаток 0 соответствует 15 варианту).

Вариант 1.

Район	Среднемесячная начисленная заработная плата, тыс. руб., х	Доля денежных доходов, направленных на покупку валюты, в общей сумме среднедушевого денежного дохода, %, у
Брянская обл.	289	6,9
Владимирская обл.	334	8,7
Ивановская обл.	300	6,4
Калужская обл.	343	8,4
Костромская обл.	356	6,1
Орловская обл.	289	9,4
Рязанская обл.	341	11,0
Смоленская обл.	327	6,4
Тверская обл.	357	9,3
Тульская обл.	352	8,2

Вариант 2

Район	Прожиточный минимум в среднем на одного пенсионера в месяц, тыс. руб., x	Средний размер назначенных ежемесячных пенсий, тыс. руб., y
Брянская обл.	178	240
Владимирская обл.	202	226
Ивановская обл.	197	221
Калужская обл.	201	226
Костромская обл.	189	220
г. Москва	302	250
Московская обл.	215	237
Орловская обл.	166	232
Рязанская обл.	199	215
Смоленская обл.	180	220

Вариант 3

Район	Прожиточный минимум в среднем на душу населения, тыс. руб., x	Средняя заработная плата и выплаты социального характера, тыс. руб., y
Брянская обл.	289	615
Владимирская обл.	338	727
Ивановская обл.	287	584
Калужская обл.	324	753
Костромская обл.	307	707
Орловская обл.	304	657
Рязанская обл.	307	654
Смоленская обл.	290	693
Тверская обл.	314	704
Тульская обл.	304	780

Вариант 4

Район	Средняя заработная плата и выплаты социального характера, тыс. руб., x	Потребительские расходы в расчете на душу населения, тыс. руб., y
Респ. Марий Эл	554	302
Респ. Мордовия	560	360
Чувашская Респ.	545	310
Кировская обл.	672	415
Нижегородская обл.	796	452
Белгородская обл.	777	502
Воронежская обл.	632	355
Курская обл.	416	688
Липецкая обл.	501	833
Тамбовская обл.	403	577

Вариант 5

Район	Денежные доходы на душу населения, тыс. руб., *x*	Потребительские расходы на душу , населения, тыс. руб., y
Респ. Карелия	913	596
Респ. Коми	1095	417
Архангельская обл.	606	354
Вологодская обл.	876	526
Мурманская обл.	1314	934
Ленинградская обл.	593	412
Новгородская обл.	754	525
Псковская обл.	528	367
Брянская обл.	520	364
Владимирская обл.	539	336

Вариант 6

Район	Денежные доходы на душу населения, тыс. руб., х	Потребительские расходы на душу населения, тыс. руб., у
Респ. Бурятия	524	408
Респ. Тува	371	249
Респ. Хакасия	453	253
Красноярский край	1006	580
Иркутская обл.	997	651
Бурятский авт. округ	217	139
Читинская обл.	486	322
Респ. Саха (Якутия)	1989	899
Еврейская авт. обл.	595	330
Чукотский авт. округ	1550	446

Вариант 7

Район	Денежные доходы на душу населения, тыс. руб., x	Потребительские расходы на душу населения, тыс. руб., y
Респ. Башкортостан	632	461
Удмуртская Респ.	738	524
Курганская обл.	515	298
Оренбургская обл.	640	351
Пермская обл.	942	624
Свердловская обл.	888	584
Челябинская обл.	704	425
Респ. Алтай	277	603
Кемеровская обл.	573	985
Новосибирская обл.	576	735

Вариант 8

Район	Средняя заработная плата и выплаты социального характера, тыс. руб., x	Потребительские расходы на душу населения, тыс. руб., y
Респ. Башкортостан	912	461
Удмуртская Респ.	809	524
Курганская обл.	748	298
Оренбургская обл.	847	351
Пермская обл.	1087	624
Свердловская обл.	1074	584
Респ. Алтай	682	277
Алтайский край	697	321
Кемеровская обл.	1251	573
Новосибирская обл.	967	576

Вариант 9

Страна	Индекс человеческой бедности (ИЧБ), х	Душевой доход, долл., *y*
Объединенные Арабские Эмираты	14,9	1600
Таиланд	11,7	7100
Уругвай	11,7	6750
Ливия	18,8	6130
Колумбия	10,7	6110
Иордания	10,9	4190
Египет	34,8	3850
Марокко	41,7	3680
Перу	22,8	3650
Шри-Ланка	20,7	3280

Вариант 10

Зависимость уровня рентабельности торговой деятельности от скорости товарооборота

№ торгового предприятия	Число оборотов, х	Уровень рентабельности, %, у
1	5,49	0,78
2	4,68	0,38
3	4.67	0,21
4	4,54	0,51
5	5,56	0,95
6	6,02	1,05
7	5,72	0,83
8	5,43	0,98
9	6,18	0,66
10	5,11	0,81

Вариант 11

№ п/п	Удельный вес продовольственных товаров в товарообороте, %. х	Уровень рентабельности торговой деятельности, %, У
1	74,2	3,62
2	73,5	3,80
3	77,0	2,77
4	84,3	2,12
5	67,3	4,33
6	70,1	4,01
7	83,1	2,01
8	70,5	3,06
9	81,1	2,17
10	76,4	3,99

Вариант 12

№ п.п.	Возраст рабочих, лет, х	Средняя выработка, шт, у
1	20	50
2	25	60
3	30	70
4	35	80
5	40	100
6	45	80
7	50	60
8	55	50
9	60	70
10	65	50

Вариант 13

№ п.п.	Удельный вес рабочих со специальной подготовкой, x	Удельный вес бракованной продукции, %, y
1	15	18
2	25	12
3	35	10
4	45	8
5	55	6
6	65	5
7	70	4
8	80	3
9	90	2
10	95	1

Вариант 14

№ п.п.	Стаж работы, лет, x	Средняя выработка, шт., y
1	4	117
2	4	141
3	10	148
4	13	172
5	14	171
6	18	182
7	18	150
8	23	163
9	24	152
10	33	217

Вариант 15

№ п.п.	Урожайность зерновых культур с 1 га, ц, x	Себестоимость 1 ц зерна, т.руб, y
1	9,1	5,42
2	10,2	4,5
3	12,3	3,6
4	14,4	3,1
5	17,4	2,74
6	19,1	2,64
7	19,4	2,43
8	20,8	2,86
9	21,1	2,57
10	22,4	2,42