Министерство высшего и среднего специального образования Республики Узбекистан

Научно исследовательский университет «Ташкентский институт инженеров ирригации и механизации сельского хозяйства»

Кафедра «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ»

МЕТОДИЧЕСКИЕ ПОСОБИЕ

для проведения практических занятий

по дисциплине

«ПРИКЛАДНАЯ ЭКОНОМЕТРИКА»

Ташкент-2019

Методическое пособие одобрено научно-методическим Советом НИУ «Ташкентского института инженеров ирригации и механизации сельского хозяйства» от «___» «__________» 2022 года, протокол № ___.

В данной работе приведены методические укзания по выполнению заданий по предмету «Прикладная эконометрика» для студентов направлений бакалавриата 60310100 – Экономика (в водном хозяйстве), 60411200 – Менеджмент (в водном хозяйстве), 60410100 – Бухгалтерский учет и аудит (в водном хозяйстве) и 60412300 – Организация и управление водным хозяйством.

Составители: **Г.Шодмонова,** профессор, к.э.н.
С.С. Мирзаев, доцент, к.т.н.

Рецензенты: **Ш.С.Насретдинова**, зав. каф. «Эконометрика и математические методы» НУУ, к.ф-.м.н., доцент
З.С. Абдуллаев, зав.каф. «Информацонные технологии» НИУ «ТИИИМСХ», к.ф.-м. н., доцент

ВВЕДЕНИЕ

Эконометрика и эконометрические модели широко внедряются во всех отраслях знаний, в частности, в такой сложной области знаний как экономика. Создание моделей экономического процесса и исследования, проводимые на построенной эконометрической модели представляет собой сущность эконометрического моделирования экономического процесса. Таким образом, эконометрическое моделирование представляет собой мощный инструмент не только научного познания, но и практического применения полученных результатов в хозяйственной деятельности. С другой стороны широкое использование информационных технологий позволяет упростить процесс эконометрического моделирования в решении задач экономики. Данное методическое пособие тесно связано с курсом лекций “Эконометрика”. Методическое пособие позволят на конкретных задачах практически реализовать те теоретические знания, которые студенты получили, изучая курс лекций “Эконометрика”. Практические задания содержат помимо предлагаемых задач методические указания по их выполнению.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №1

Тема: Основы эконометрического моделирования. Этапы построения эконометрических моделей

1. **Задание.** Для выполнение этого задания целесообразно использование новых педагогических технологий (например, «мозговой штурм»).

Контрольные вопросы:

1. Какие модели называются эконометрическими?
2. Что изучает эконометрика?
3. Почему необходимо использование математики в экономике?
4. Отличие эконометрических моделей от других моделей.
5. Что такое математическая модель?
6. Как строится математическая модель экономических явлений или объекта? Приведите пример построения уточнения модели.
7. Какова связь между математической структурой модели и ее содержательной интерпретацией?
8. Какие переменные модели называются экзогенными, а какие эндогенными?
9. В чем отличие статических моделей от динамических? К какому типу относится приведенная в главе модель?
10. В чем отличие математической экономики от эконометрике?
11. Предположим, перед вами стоит задача обоснования линейной зависимости объема потребления от доходов на основе эмпирических данных. Относится ли решение подобной задачи к математической экономике или к эконометрике?

12. Пусть фирма выпускает один вид продукции, причем объем выпуска зависит от затрат только двух факторов: труда и капитала. Заданы цена единицы труда (ставка заработной платы), цена единицы капитала (ставка процента) и издержки фирмы. Задача состоит в определении объемов затрат труда и капитала, максимизирующих объем выпуска. Постройте соответствующую экономико-математическую модель.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №2

Тема: Элементы математической статистики

Задача №1. Вычислить выборочные характеристики по исходным данным:

№	1	2	3	4	5
x	2 +k	6 +k	10 +k	14 +k	18 +k

Теоретическая часть

Закономерности в экономике выражаются в виде зависимостей экономических показателей и математических моделей их поведения. Такие зависимости и модели могут быть получены только путем обработки реальных статистических данных, с учетом внутренних связей и случайных факторов.

Любые экономические данные представляют собой характеристики какого-либо экономического объекта. Они формируются под воздействием множества факторов, не все из которых доступны внешнему контролю. Неконтролируемые (неучтенные) факторы обусловливают случайность данных, которые они определяют. Поскольку экономические данные имеют статистическую природу, для их анализа и обработки необходимо применять специальные методы математической статистики.

Методическое указание

Процесс решения задачи №1 покажем в случае k=0: Исходные данные (x) и расчетные показатели (x^2) представим в виде расчетной таблицы:

№	x	x^2
1	2	4
2	6	36
3	10	100
4	14	196
5	18	324
Всего	50	660

Среднее	10	132
		$\overline{x^2}$

Выборочные характеристики:

Для вычисления выборочной средней и выборочной дисперсии в Excel можно использовать функции

$\bar{x} = СРЗНАЧ(массив\ x), \quad var(x) = ДИСПР(массив\ x)\bar{x} = СРЗНАЧ(массив\ x), \quad var(x) = ДИСПР(массив\ x)$.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №3
Тема: Ковариация и правила ее вычисления

Задача №1. По приведенным ниже исходным данным вычислить ковариацию и коэффициент корреляции между переменными *x, y,* установить его значимость:

Таблица №1

№	1	2	3	4	5
x	2+k	6+k	10+k	14+k	18+k
y	1	2	4	11	12

Задача №2. В следующей задаче анализировать связи между двумя переменными : в некоторой бюрократической стране годовой доход каждого индивида y определяется по формуле:

$$y = 10000 + 500s + 200t\ ,$$

где s- число лет обучения индивида; t- трудовой стаж (в годах); x- возраст индивида. Рассчитайте Cov (x,y), Cov (x,s), Cov (x,t) для выборки из пяти индивидов, описанной ниже, и проверьте, что

$$\text{Cov}(x,y) = 500\ \text{Cov}(x,s) + 200\ \text{Cov}(x,t).$$

Объясните аналитически, почему так происходит.

Выборочная ковариация между X и Y определяется следующей формулой:

$$\text{Cov (x,y)} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$$

Таблица №2

Индивид	Возраст (годы)	Годы обучения	Трудовой стаж	Доход
1	18	11	1	15700+k
2	29	14	6	18200+k
3	33	12	8	17600+k
4	35	16	10	20000+k
5	45	12	5	17000+k

Здесь к- номер студента по журналу.

Теоретическая часть

Выборочная ковариация между X и Y определяется как:

$$\operatorname{cov}(x,y) = \frac{1}{n}\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = \frac{1}{n}[(x_1 - \bar{x})(y_1 - \bar{y}) + \ldots + (x_n - \bar{x})(y_n - \bar{y})] \quad (14)$$

Вводится следующие обозначения: $Cov\,(x,y)$ - выборочная ковариация, $pop \cdot Cov\,(x,y)$ ковариация генеральной совокупности X и Y, $Var\,(x)$ -выборочная дисперсия, $pop \cdot Var\,(x)$ дисперсия для генеральной совокупности.

Основные правила расчета ковариации

Есть несколько важных правил, которые вытекают непосредственно из определения ковариации:

Правило 1. Если $y = v + w$, *то* $Cov\,(x,y) = Cov\,(x,v) + Cov\,(x,w)$.

Правило 2. Если $y = a \cdot z$, *то* $Cov\,(x,y) = aCov\,(x,z)$.

Правило 3. Если $y = a$, *то* $Cov\,(x,y) = 0$ где *a*- константа.

Методическое указание

Последовательность решения задачи №1 покажем, в случае к=0: Представим исходные данные и расчетные показатели в виде следующей расчетной таблицы:

№	x	y	x^2	xy	y^2
1	2	1	4	2	1
2	6	2	36	12	4
3	10	4	100	40	16
4	14	11	196	154	121
5	18	12	324	216	144
Итого	50	30	660	424	286
Среднее	10	6	132	84,8	57,2
			$\overline{x^2}$	$\overline{xy}$	$\overline{y^2}$

Окончательно имеем:

$$var(x) = \overline{x^2} - (\bar{x})^2 = 132 - 100 = 32,$$

$$var(y) = \overline{y^2} - (\bar{y})^2 = 57{,}2 - 36 = 21{,}2,$$

$$cov(x, y) = \overline{xy} - \bar{x}\ \bar{y} = 84{,}8 - 60 = 24{,}8,$$

$$R = \frac{cov(x, y)}{\sqrt{var(x)var(y)}} = \frac{24{,}8}{\sqrt{32\ \ 21{,}2}} = 0{,}952\,.$$

Замечание. В Excel можно по исходным данным получить коэффициент корреляции и его квадрат (т.е., дисперсию) с помощью функции:

$r = ПИРСОН\left(массив\, x; массив\, y\right),$

$r^2 = КВПИРСОН\left(массив\, x; массив\, y\right)\,.$

Проверим значимость выборочного коэффициента корреляции. Наблюдаемое значение критерия

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}952\sqrt{3}}{\sqrt{0{,}0937}} = 5{,}5\,.$$

Выполним проверку значимости ***r*** двумя способами.

1. При $\alpha = 0{,}05$ $и$ $\nu = 3$ по таблице или с помощью функции СТЬЮДРАСПОВР (α; ν) находим

П о с к о л ь к у $\quad |t| = 5{,}5 > t_{кр} = 3{,}18,$ то $r = 0{,}952$

значим при 5% – ном уровне.

2. Наблюдаемому (расчетному) значению критерия *t=5,5* соответствует значимость *t=0,0124*, которая может быть определена в Excel с помощью функции

Значимость t=СТЬЮДРАСП(t;v;2),

где, *v=n-2* - число степеней свободы.

Поскольку значимость t=0,0124<0,05, то коэффициент r=0,952 значим при 5%-ном уровне, следовательно, имеется линейная зависимость между переменными.

Используя данные задачи №1, докажите правильность этого способа.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №4

Тема: Правила расчета дисперсии

Задача №1. Найти вероятность, среднее значение и дисперсию данной дискретной случайной величины:

x_i	p_i	$x_i - \mu_i$	$(x_i - \mu)^2$	$(x_i - \mu)^2 \cdot p_i$
0	0.1+k			
1	0.3+k			
2	0.25+k			
3	0.2+k			
4	0.15+k			

Здесь μ -среднее значение.

Теоретическая часть

Для выборки из *n* наблюдений $x_1,\dots,x_n$ выборочная дисперсия определяется как среднеквадратичное отклонение в выборке:

$$\operatorname{var}(x) = \frac{1}{n}\sum\left(x_i - \bar{x}\right)^2 .$$

Существует несколько простых и очен полезных правил для расчета дисперсии, являющихся аналогами правил для ковариации. Эти правила в равной степени можно использовать как для выборочной, так и для теоретической дисперсии.

Правило 1. Если $y = v + w$, *то* $Var\ (y) = Var\ (v) + Var\ (w) + 2Cov\ (v,w)$

Правило 2 . Если $y = az$, *то* $\operatorname{var}(y) = a^2 \operatorname{var}(z)$, здесь *а является* постоянной.

Правило 3. Если $\mathrm{y} = a$, *то* $\operatorname{var}(\mathrm{y}) = 0$, здесь *а* является постоянной.

Правило 4. Если $y = v + a$, *то* $\operatorname{var}(y) = \operatorname{var}(v)$, здесь *а* является постоянной.

Заметим, что дисперсия переменной x может рассматриваться как ковариация между двумя величинами *x*:

$$\operatorname{var}(x) = \frac{1}{n}\sum_{i-1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{1}{n}\sum_{i-1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right) = Cov(x, x).$$

Методическое указание

Теоретическая (генеральная) дисперсия случайной величины определяется как математическое ожидание квадрата отклонения случайной величины X относительно ее средней, т.е.

$$\sigma_x^2 = D(X) = M(X - \mu_x)^2.$$

Замечание. Если ясно, о какой переменной идет речь, нижний индекс в μ_x *или* σ_x^2 *можно не указывать.*

Для вычисления дисперсии часто используется другое выражение, получаемое из определения дисперсии:

$(X) = M\left(X^2\right) - \mu_x^2$Дисперсия является мерой рассеяния случайной величины X.

Среднеквадратичное (стандартное) отклонение случайной величины X определяется как корень квадратный из ее дисперсии, т.е. $\sigma_x = \sqrt{\mathrm{D(X)}}$.

Среднеквадратичное отклонение показывает, насколько в среднем отклоняется случайная величина в совокупности относительно средней (центра).

Свойства дисперсии:

1. $\boldsymbol{D(a)} = 0$.
2. $D(bX) = b^2 D(X)$.
3. $D(a+bX) = b^2 D(X)$.

Следствие. Если две случайные величины X, Y независимы, то $D(X+Y) = D(X) + D(Y)$.

Заметим, что $M(X)$ и $D(X)$ — это числовые характеристики генеральной совокупности (числа), а не функции.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №5

Тема: Коэффициент корреляции и правила его вычисления

Задача №1. Рассчитать урожайность культур в зависимости от балла качественной оценки земли и дозы вносимых удобрений на различных полях хозяйства. Фактическая урожайность зерновых культур по полям, дозы вносимых удобрений и баллы качества почвы приведены в таблице.

2) Вычислить коэффициенты парной и множественной корреляции между указанными показателями и определить достоверность полученных взаимосвязей.

Таблица №1

№ поля	Урожайность, ц./га, **у**								Балл качест-венной	Доза удоб-рений, ц
	озимая пщени-	озимая ячмень	хло-пок	яровая пшеница	ози-мая рож	яч-мень	горох	гречиха		
1	24,1	23,1	25,1	28,1	20,5	30,5	14,3	12,1	30	8
2	24,2	23,4	25,2	28,3	20,7	30,8	14,5	12,2	35	7

3	28,5	27,	29,	32,5	24,5	34,	16,5	14,5	40	10
4	25,0	23,	27,	29,1	21,0	31,	15,1	13,0	45	6
5	24,9	23,	25	29,9	20,9	30,	15,3	12,5	50	5
6	25,2	24,	27,	29,9	21,3	31,	15,6	12,6	55	4
7	31,0	29,	31,	35,0	27,0	38,	17,2	15,5	60	9
8	33,3	32,	34,	37,3	29,3	39,	19,5	17,0	65	10
9	32,1	31,	33,	36,5	28,1	38,	18,3	16,7	70	9
10	33,4	32,	35,	37,4	28,5	38,	18,7	16,9	75	8
11	35,7	34,	37,	39,7	31,7	41,	20,0	17,8	80	9
12	38,0	37,	40,	42,1	34,2	44,	21,1	19,0	85	3
13	34,3	33,	38,	38,5	30,5	40,	19,5	17,1	90	6
14	31,1	31,	35,	35,5	27,8	39,	18,2	15,6	95	2
15	39,0	38,	41,	44,0	35,0	46,	22,0	20,0	100	3

Теоретическая часть

В качестве меры для степени линейной связи двух переменных используется коэффициент их корреляции:

$$r(x,y)=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (1)$$

По формуле коэффициента корреляции видно, что он будет положителен, если отклонения переменных от своих средних значений имеют, как правило, одинаковый знак, и отрицательным – если разные знаки. Коэффициент корреляции является безразмерной величиной (так как размерности числителя и знаменателя есть размерности произведения); его величина не зависит от выбора единиц измерения обеих переменных. Величина коэффициента корреляции меняется от – 1 в случае строгой линейной отрицательной связи до +1 в случае строгой линейной положительной связи. Близкая к нулю величина коэффициента корреляции говорит об отсутствии линейной связи переменных, но не об отсутствии связи между ними вообще. Последнее вытекает из того, что каждой паре одинаковых отклонений переменной от ее среднего значения соответствуют равные по абсолютной величине положительное и отрицательное отклонения переменной от ее среднего.

Далее, в анализе коэффициента корреляции возникает следующий вопрос. Если он равен нулю для генеральной совокупности, это вовсе не значит, что он в точности будет равен нулю для выборки. Наоборот, он обязательно будет отклоняться от истинного значения, но чем больше такое отклонение, тем менее оно вероятно при данном объеме выборки. Таким образом, при каждом конкретном значении коэффициента корреляции величин для генеральной совокупности выборочный коэффициент корреляции является случайной величиной. Следовательно, случайной величиной является также любая его функция, и требуется указать такую функцию, которая имела бы одно из известных распределений, удобное для табличного анализа. Для выборочного коэффициента корреляции такой функцией является статистика, рассчитываемая по $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$.

Методическое указание

Представим исходные данные и расчетные показатели в виде следующей расчетной таблице:

Таблица №2

№	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})*(y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	2	1	4	2	8	16	4
2	6	2	36	12	432	1296	144
3	10	4	100	40	4000	10000	1600
4	14	11	196	154	30184	38416	23716

5	18	12	324	216	69984	104976	46656
Итого	50	30	660	424	279840	435600	179776
Среднее	10	6	132	84,8	11193,6	17424	7191,04

Подставив найденные значения в формулу коэффициента корреляции, находим **R=11193/корень(17424*7191,04)=1** . Значить степень зависимости польная.

Задача №2. Покажем, что $r_{\hat{y},y} = r_{x,y}$ в случае парной регрессии $\hat{y} = y + bx$. Действительно, из отношений

$$cov\left(\hat{y}, y\right) = cov\left(a + bx, y\right) = bcov\left(x, y\right),$$

$$var\left(\hat{y}\right) = var\left(a + bx\right) = b^2 var(x)$$

имеем

$$r_{\hat{y},y} = \frac{cov(\hat{y}, y)}{\sqrt{var(\hat{y})var(y)}} = \frac{cov(x, y)}{\sqrt{var(x)var(y)}} = r_{x,y}$$

Вывод. В случае парной регрессии коэффициент детерминации есть квадрат коэффициента корреляции переменных x и y, т.е. $R^2 = r_{x,y}^2$.

Используя данные таблицы №1 докажите правильность этих соотношений.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №6

Тема: Парная регрессионная модель

Задача №1 Используя данные задачи №1 в практическом задании №5 выпольнить следующее:

1. Определить зависимости между результативными показателями y (урожайностью) отдельно с x_1 (баллом качественной оценки земли) и x_2 (количеством вносимых удобрений).
2. Установить точность параметров полученных уравнений регрессии a_{ij} (коэффициентов регрессии).

Теоретическая часть

Рассмотрим более подробно те случаи, когда одна переменная зависит от другой $Y = \alpha + \beta x + u$. Здесь $\alpha + \beta x$ - неслучайная составляющая, где x выступает как объясняющая переменная, u – случайный член.

Задача регрессионного анализа состоит в получении оценок α и β и следовательно, в определении положения прямой по точкам. Очевидно, что чем меньше значение *u,* тем легче эта задача.

Оценка параметров по методу наименьших квадратов.

Допустим мы имеем четыре наблюдения для *X* и *Y*. Здесь $\alpha + \beta x$ - неслучайная составляющая, где *x* выступает как объясняющая переменная, *u* – случайный член. Отрезок, отсекаемый прямой на оси Y представляет собой оценку α и обозначен *a*, а угловой коэффициент прямой представляет собой оценку β, и обозначен *b*. Первым шагом является определение остатков для каждого наблюдения. Рассмотрим парную зависимость между стоимостью валовой продукции и баллом бонитета земли. Результат зависимости описывается следующим выражением: $y = \alpha + \beta x + \varepsilon$ и оценена регрессия $\hat{y} = -340{,}05 + 13{,}46x$.

Полученный результат можно истолковать следующим образом. Коэффициент при *x* (коэффициент наклона) показывает, что если *x* увеличивается на одну единицу, то *y* возрастает на 13,4 единиц. *x* измеряется в баллах, *y* измеряется у.е./га; таким образом, коэффициент наклона показывает, что если балл бонитета земли увеличивается на 1 балл, то стоимость валовой продукции возрастает на 13,46 у.е.

Что можно сказать о постоянном члене в уравнении? Формально говоря, он показывает прогнозируемый уровень *y*, когда *x* = 0. Иногда это имеет ясный смысл, иногда – нет. В данном случае константа выполняет единственную функцию: она позволяет определить положение линии регрессии на графике.

Методическое указание

Построим регрессионные зависимости: а) расходов на питание *y* и личного дохода *x*; б) расходов на питание *y* и времени t –по следующим данным (усл.ед.)

Год	1990	1991	1992	1993	1994
x	2	6	10	14	18
y	1	2	4	11	12

Оценим качество подгонки.

а) Пусть истинная модель описывается выражением $y = \alpha + \beta x + \varepsilon$.

По выборочным наблюдениям определяем оценки (*a*,b). Исходные данные и расчетные показатели удобно представить в виде следующей таблицы:

Годы	*x*	*y*	x^2	*xy*	$\hat{y}$	$(y-\bar{y})^2$	$(\hat{y}-\bar{y})^2$	$(y-\hat{y})^2$
1990	*2*	*1*	*4*	*2*	*-0,2*	*25*	*38,44*	*1,44*
1991	*6*	*2*	*36*	*12*	*2,9*	*16*	*9,61*	*0,81*
1992	*10*	*11*	*100*	*40*	*6*	*4*	*0*	*4*
1993	*14*	*12*	*196*	*154*	*9,1*	*25*	*9,61*	*3,61*
1994	*18*	*30*	*324*	*216*	*12,2*	*36*	*38,44*	*0,04*
Итого	50	30	660	424	30	106	96,1	9,9
Среднее	10	6	132	84,8	6	21,2	19,22	1,98

	$\bar{x}$	$\bar{y}$	$\overline{x^2}$	$\overline{xy}$	$\bar{\hat{y}}$	*Var(y)*	*Var($\hat{y}$)*	*Var(e)*

Зависимость переменной в регрессии $y = \alpha + \beta x + \varepsilon$ разбивается на две компоненты: $y = y_1 + y_2$. Рассмотрим две регрессии для компонент: $y_1 = \alpha_1 + \beta_1 x + \varepsilon_1, \quad y_2 = \alpha_2 + \beta_2 x + \varepsilon_2.$

Докажем следующие соотношения для МНК-оценок параметров двух регрессии: $a = a_1 + a_2, \quad b = b_1 + b_2$.

Действительно,

$$b = \frac{cov(x, y)}{var(x)} = \frac{cov(x, y_1 + y_2)}{var(x)} = \frac{cov(x, y_1) + cov(x, y_2)}{var(x)} = b_1 + b_2,$$

$$a = \bar{y} - b\bar{x} = \left(\overline{y_1 + y_2}\right) - \bar{x}\left(b_1 + b_2\right) = a_1 + a_2.$$

Задача №2. Покажем, что если все значения переменных изменить на одно и то же число или в одно и то же число раз, то величина коэффициента b в парной регрессии не изменится.

Пусть $x' = x + c, \quad y' = y + c,$ тогда

$$b' = \frac{\mathrm{cov}(x', y')}{var(x')} = \frac{cov(x + c, y + c)}{var(x + c)} = \frac{cov(x, y)}{var(x)} = b$$

$x' = kx, \quad y' = ky,$ тогда

$$b' = \frac{\mathrm{cov}(x', y')}{var(x')} = \frac{cov(kx, ky)}{var(kx)} = \frac{k^2 cov(x, y)}{k^2 var(x)} = b$$

Используя данные таблицы №1 докажите правильность этих соотношений.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №7

Тема: Метод наименьших квадратов. Построение уравнения регрессии

Задача №1. Используя данные задачи №1 в практическом задании №5 выпольнить следующее: определить регрессионные зависимости между а) y – урожайностью какой-либо культуры и x- баллом качественной оценки земли; б) y - урожайностью какой-либо культуры и x –дозой удобрений. Установить точность параметров полученных уравнений регрессии.

а) Вид модели: $y = \alpha + \beta x + \varepsilon$.

Теоретическая часть

Допустим, мы имеем четыре наблюдения для X и Y. Здесь $\alpha + \beta x$ - неслучайная составляющая, где x выступает как объясняющая переменная, u – случайный член.

Отрезок, отсекаемый прямой на оси Y представляет собой оценку α и обозначен a, а угловой коэффициент прямой представляет собой оценку β, и обозначен b. Первым шагом является определение остатков для каждого наблюдения. $e_1 = y_1 - \hat{y}_1$, $e_2 = y_2 - \hat{y}_2$, $e_3 = y_3 - \hat{y}_3$, $e_4 = y_4 - \hat{y}_4$.

Мы хотим построить линии регрессии таким образом, чтобы эти остатки были минимальными. Одним из способов решения поставленной проблемы состоит в минимизации суммы квадратов остатков S.

$$S = e_1^2 + e_2^2 + e_3^2 + e_4^2 \to \min$$

$$S(ab) = \sum_i e_i^2 \to \min$$

$$S(ab) = \sum_i (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \to \min$$

$$\frac{\partial S}{\partial a} = 0. \quad \begin{cases} \frac{\partial S}{\partial a} = -2 \quad \sum_{i=2}^{n} (y_i - a - bx_i) = 0 \\ \frac{\partial S}{\partial b} = -2 \quad \sum_{i=1}^{n} (y_i - a - bx_i)x_i = 0 \end{cases}$$

$$\frac{\partial S}{\partial b} = 0$$

Отсюда находим:

$$\begin{cases} \sum y_i = na + b\sum x_i \\ \sum x_i y_i = a \sum x_i + b\sum x_i^2 \end{cases}$$ Находя $\sum y_i$, $\sum x_i$, $\sum x_i y_i$, $\sum x_i^2$ можно определить a и b

$$b = \frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)}; \; a = \bar{y} - b\bar{x} \qquad \operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\operatorname{var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Методическое указание

Задача №2. Построим регрессионные зависимости: а) расходов на питание *y* и личного дохода *x*; б) расходов на питание *y* и времени t –по следующим данным (усл.ед.)

Год	1990	1991	1992	1993	1994
x	2	6	10	14	18
y	1	2	4	11	12

И оценим качество подгонки.

а) Пусть истинная модель описывается выражением $y = \alpha + \beta x + \varepsilon$.

По выборочным наблюдениям определяем оценки (*a*,b).

Исходные данные и расчетные показатели удобно представить в виде следующей таблицы:

Годы	*x*	*y*	x^2	*xy*	$\hat{y}$	$(y-\bar{y})^2$	$(\hat{y}-\bar{y})^2$	$(y-\hat{y})^2$
1990	*2*	*1*	*4*	*2*	*-0,2*	*25*	*38,44*	*1,44*
1991	*6*	*2*	*36*	*12*	*2,9*	*16*	*9,61*	*0,81*
1992	*10*	*11*	*100*	*40*	*6*	*4*	*0*	*4*
1993	*14*	*12*	*196*	*154*	*9,1*	*25*	*9,61*	*3,61*
1994	*18*	*30*	*324*	*216*	*12,2*	*36*	*38,44*	*0,04*
Всего	50	30	660	424	30	106	96,1	9,9

В среднем	10	6	132	84,8	6	21,2	19,22	1,98
	$\bar{x}$	$\bar{y}$	$\overline{x^2}$	$\overline{xy}$	$\bar{\hat{y}}$	*Var(y)*	*Var($\hat{y}$)*	*Var(e)*

Окончательно имеем

$$cov(x, y) = \overline{xy} - \bar{x}\bar{y} = 84{,}8\text{-}60\text{=}24{,}8,$$

$$var(x) = \overline{x^2} - (\bar{x})^2 = 132 - 100 = 32,$$

$$b = \frac{cov(x, y)}{var(y)} = \frac{24{,}8}{32} = 0{,}775, \quad a = \bar{y} - b\bar{x} = 6 - 0{,}775*10 = -1{,}75\,.$$

Следовательно, $\hat{y} = -1{,}75 + 0{,}775x\,.$

Коэффициент b=0,775 показывает, что при увеличении дохода на 1 усл. ед. расходы на питание увеличиваются в среднем на 0,775 усл.ед.

Замечание. В Excel оценки (a,b) можно также определить с помощью функций:

$a = ОТРЕЗОК(массив\ y_1\ массив\ x)$,

$b = НАКЛОН(массив\ y_1\ \ массив\ x)$.

Условие $var(y) = var(\bar{y}) + var(e)$ выполняется.

Качество подгонки оценим коэффициентом детерминации:

$$R^2 = \frac{var(\hat{y})}{var(y)} = \frac{19{,}22}{21{,}2} = 0{,}907,$$

т.е. 90,7% вариации зависимой переменной (расходы на питание) объясняется регрессией.

Задача №3 Покажем, что в модели регрессии *без свободного члена* $\Upsilon = \beta X + \varepsilon$ оценка МНК для β есть $b = \dfrac{\sum x_i y_i}{\sum x_i^2} = \dfrac{\overline{xy}}{\overline{x^2}}.$

Выборочная регрессия для этой модели $\hat{y} = bx\,.$ Наблюдаемые значения зависимой переменной связаны с расчетным уравнением $y_i = \hat{y}_i + e_i\,.$ Оценку b найдем из минимизации величины

$$Q = \sum e_i^2 = \sum (y_i - bx_i)^2 = \sum y_i^2 - 2b\sum x_i y_i + b^2 \sum x_i^2 .$$

Получим $Q'_b = -2\sum x_i y_i + 2b\sum x_i^2 = 0$,

Откуда

$$b = \frac{\sum x_i y_i}{\sum x_i^2} \quad .$$

Вычисление R^2 при отсуствии свободного члена некорректно.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №8

Тема: Интерпретация уравнения регрессии. Качество оценки. Коэффициент детерминации

Задача №1. Используя данные задачи №1 практического задания №5 выполнить следующее: интерпретировать уравнение регрессии, качество оценки, коэффициент детерминации.

Теоретическая часть

При интерпретации уравнения регрессии чрезвычайно важно помнить о трех вещах. Во-первых, a является лишь оценкой α и b – оценкой β. Поэтому вся интерпретация в действительности представляет собой лишь оценку. Во-вторых, уравнение регрессии отражает только общую тенденцию для выборки. При этом каждое отдельное наблюдение подвержено воздействию случайностей. В-третьих, верность интерпретации зависит от правильности спецификации уравнения.

Качество оценки.

Качество оценки уравнения регрессии проводится с определением показателя коэффициента детерминации

$$D = r^2 \textit{ или } \qquad R^2 = 1 - \frac{\mathrm{var}(e)}{\mathrm{var}(y)}$$

Максимальное значение коэффициента R^2 равно единице. Это происходит в том случае, когда линия регрессии точно соответствует всем наблюдениям, т.к.

$\hat{y}_i = y_i$ *для всех i и все остатки равны нулю.*

Тогда $\operatorname{var}(\hat{y}) = \operatorname{var}(y)$, $\operatorname{var}(e) = 0$ *и* $R^2 = 1$.

Если в выборке отсутствует видимая связь между *y* и *x*, то коэффициент R^2 будет близок к нулю.

Методическое указание

1. Покажем, что в модели регрессии $Y = \alpha + \varepsilon$ оценка МНК для α есть $a = \bar{y}$.

Выборочная регрессия для заданной модели есть $\hat{y}_i = a$. Наблюдаемые значения зависимой переменной связаны с расчетными значениями уравнением $y_i = \hat{y}_i + e_i = a + e_i$. Оценку найдем из минимизации величины

$$Q = \sum e_i^2 = \sum (y_i - a)^2 = \sum y_i^2 - 2a \sum y_i + na^2 .$$

Получим

$$Q'_a = -2 \sum y_i + 2an = 0 ,$$

Откуда $a = \dfrac{1}{n} \sum y_i = \bar{y}$.

Выборочная регрессия $\hat{y} = \bar{y}$.

2. По данным примера №1 практического задания №5 покажите, что зависимость урожайности *y* от балла качественной оценки земли *x* для модели регрессии *без свободного члена* есть $\hat{y} = 0{,}197x$, при этом $\hat{y} \neq \bar{y}$ и $var(y) \neq var(\hat{y}) + var(e)$.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №9

Тема: Свойства коэффициентов регрессии и проверка гипотез

Задача №1. Используя данные задачи №1 практического задания №5 в модели регрессии $\hat{y} = 17{,}8 + 0{,}197x$ вычислить стандартные ошибки коэффициентов.

Теоретическая часть

В случае парного регрессионного анализа, коэффициенты регрессии должны рассматриваться как случайные переменные специального вида, случайные компоненты которых обусловлены наличием в модели случайного члена. Каждый коэффициент регрессии вычисляется как функция значений y и независимых переменных в выборке, а y в свою очередь определяется независимыми переменными и случайным членом. Отсюда следует, что коэффициенты регрессии определяются значениями переменных и случайным членом, а их свойства существенно зависят от свойств последнего.

Считаем, что для регрессионного анализа выполняются следующие условия Гаусса-Маркова: 1) математическое ожидание u в любом наблюдении равно нулю; 2) теоретическая дисперсия его распределения одинакова для всех наблюдений; 3) теоретическая ковариация его значений в любых двух наблюдениях равняется нулю.

Существуют еще два практических требования. Во-первых, нужно иметь достаточное количество данных для проведения линии регрессии, что означает наличие стольких (независимых) наблюдений, сколько параметров необходимо оценить. В общем случае можно сказать, что коэффициенты регрессии, скорее всего, являются более точными:

1) чем больше число наблюдений в выборке.
2) чем больше дисперсия выборки объясняющих переменных.
3) чем меньше теоретическая дисперсия случайного члена.

Методическое указание

Для полученной в методическом указании практического задания №7 зависимости расходов на питание от личного дохода $\hat{y} = -1,75 + 0,775x$ рассчитаем стандартные ошибки коэффициентов регрессии.

Исходные данные:

$n = 5, \ \ var(x) = 32, \quad \bar{x}^2 = 132, \ var(e) = 198.$

Остаточная дисперсия S^2 и стандартная ошибка регрессии S равны соответственно

$$S^2 = \frac{n}{n-2} var(e) = \frac{5}{3} {*} 1,98 = 3,3, \qquad S = \sqrt{3,3} = 1,816\,.$$

Для расчета стандартной ошибки можно также воспользоваться функцией Excel.

S=СТОШYX(массив *y*; массив *x*)

Стандартные ошибки коэффициентов регрессии

$$\boldsymbol{S_a = S\sqrt{\frac{\bar{x}^2}{nvar(x)}}} = 1,816\sqrt{\frac{132}{5{*}32}} = 1,65,$$

$$\boldsymbol{S_b = \frac{S}{\sqrt{nvar(x)}}} = \frac{1,816}{\sqrt{5832}} = 0,143,$$

Задача №2. Покажем, что в выборочной регрессии *без свободного члена* $\hat{y} = bx$ стандартная ошибка оценки b.

$$\boldsymbol{S_b = \frac{S}{\sqrt{n{*}\bar{x}^2}}}\,.$$

где,

$$\boldsymbol{S^2 = \frac{1}{n-1}\sum (y_i - bx_i)^2}$$

Подставим в оценку b выражение $\boldsymbol{y_i = \beta x_i + \varepsilon_i}$,

$$\boldsymbol{b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(\beta x_i + \varepsilon_i)}{\sum x_i^2} = \beta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}}$$

Оценка b является *несмещенной*, так как M(b)= $\boldsymbol{\beta}$,

Дисперсия оценка b

$$\boldsymbol{D}(\boldsymbol{b})=\frac{\sum \boldsymbol{x}_i^2 \boldsymbol{D}(\boldsymbol{\varepsilon}_i)}{(\sum \boldsymbol{x}_i^2)^2}=\frac{\boldsymbol{\sigma}^2}{\sum \boldsymbol{x}_i^2}=\frac{\boldsymbol{\sigma}^2}{\boldsymbol{n}\bar{\boldsymbol{x}}^2}$$

В исходной модели оценивается один параметр, поэтому оценкой $\boldsymbol{\sigma}^2$ является

$$\boldsymbol{S}^2=\frac{1}{\boldsymbol{n}-1}\sum \boldsymbol{e}_i^2=\frac{1}{\boldsymbol{n}-1}\sum (\boldsymbol{y}_i-\boldsymbol{b}\boldsymbol{x}_i)^2$$

Следовательно, $\boldsymbol{S}_b=\dfrac{\boldsymbol{S}}{\sqrt{\boldsymbol{n}\bar{\boldsymbol{x}}^2}}$.

Задача №3. Покажем, что в выборочной регрессии $\hat{y}=a$ стандартная ошибка оценки a

$$S_a=\frac{S}{\sqrt{n}}\text{ . Где } S^2=\frac{1}{n-1}\sum (y_i-\bar{y})^2 \quad .$$

Подставим в оценку для a выражение $y_i=\alpha+\varepsilon_i$

$$a=\frac{\sum y_i}{n}=\frac{\sum (\alpha+\varepsilon_i)}{n}=\alpha+\frac{\sum \varepsilon_i}{n} \quad .$$

Оценка является *несмещенной*, так как $\mathrm{M}(a)=\alpha$.

Дисперсия оценки a

$$\mathrm{D}(a)=\frac{\sum D(\varepsilon_i)}{n^2}=\frac{\sigma^2}{n} \quad .$$

В исходной модели оценивается один параметр, поэтому оценкой σ^2 является $S^2=\dfrac{1}{n-1}\sum e_i^2=\dfrac{1}{n-1}\sum (y_i-\bar{y})^2$.

Следовательно, $S_a=\dfrac{S}{\sqrt{n}}$.

Задача №4. По данным приведенным в методическом указании практического задания №6 построим зависимость расходов на питание y от личного дохода *x* для модели регрессии *без свободного члена* и рассчитаем стандартную ошибку коэффициента регрессии.

Исходные данные и расчетные показатели представим в виде таблицы:

Год	x	y	x^2	xy	$\hat{y}$	$(y-\bar{y})^2$	$(\hat{y}-\bar{y})^2$	$(y-\hat{y})^2$
1990	2	1	4	2	1,28	25	26,378	0,0806
1991	6	2	36	12	3,85	16	6,594	3,429
1992	10	4	100	40	6,42	4	0	5,856
1993	14	11	196	154	8,99	25	6,594	4,048
1994	18	12	324	216	11,56	36	26,378	0,193
итого	50	30	660	424	32,1	106	65,946	13,608
Сред-	10	6	132	84,8	6,42	21,2	13,189	2,721
нее	$\bar{x}$	$\bar{y}$	$\overline{x^2}$	$\overline{xy}$	$\bar{\hat{y}}$	*Var(y)*	*Var*$(\hat{y})$	*Var(e)*

Коэффициент b определяется выражением

$$b = \frac{\overline{xy}}{\overline{x^2}} = \frac{84{,}8}{132} = 0{,}642\ .$$

Следовательно, $\hat{y} = 0{,}642x$. Заметим, что в отсутствие свободного члена $\bar{\hat{y}} \neq \bar{y},\ \ var(y) \neq var\left(\hat{y}\right) + var(e)$.

Остаточная дисперсия S^2 и стандартная ошибка регрессии S равны соответственно

$$S^2 = \frac{1}{n-1}\sum e_i^2 = \frac{13{,}608}{4} = 3{,}397\ ,\quad S = \sqrt{3{,}397} = 1{,}843\ .$$

Стандартная ошибка коэффициента регрессии $S_b = \dfrac{S}{\sqrt{n\overline{x^2}}} = \dfrac{1{,}843}{\sqrt{5*132}} = 0{,}071$.

Задача №5. Зависимость расходов на питание y от личного дохода x по данным приведенных в методическом указании практического задания №6 имеет вид $\hat{y} = -1{,}75 + 0{,}775\,x$

(1,65) (0,143)

(в скобках указаны стандартные ошибки).

Оценим значимость коэффициента регрессии b=0,775 и построим доверительный интервал для β при 5%-ном уровне значимости.

Наблюдаемое значение критерия $t = \dfrac{b}{S_b} = \dfrac{0{,}775}{0{,}143} = 5{,}4$.

Значимость t=0,0124, соответствующую расчетному значению критерия t=5,4, определяем с помощью функции Значимость =СТЬЮДРАСП(t; v;2), v=3. Поскольку *значимость* t=0,0124<0,05, то коэффициент регрессии b=0,775 значим.

При $\alpha = 0{,}05$ критическое значение критерия $t_{кр} = 3{,}18$ определяется с помощью функции $t_{кр} = СТЬЮДРАСПОБР(\alpha; v)$.

Доверительный интервал для β.

0,775-3,18 *0,143<$\beta < 0{,}775 + 3{,}18 * 0{,}143,\ 0{,}32 < \beta < 1{,}23$.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №10

Тема: Временные ряды и тренды

В следующей таблице приводятся данные о производстве стали в 1995-2009 гг. (млн. т.).

1995	71+k
1996	76+k
1997	80+k
1998	85+k
1999	91+k

2000	97+k
2001	102+k
2002	107+k
2003	110+k
2004	116+k
2005	121+k
2006	126+k
2007	131+k
2008	136+k
2009	141+k

k – порядковый номер студента по журналу.

Методическое указание

Задача №1. Рассчитаем скользящую среднюю по данным об урожайности зерновых культур (ц/га) за 10 лет.

Исходные данные и расчетные показатели представим в следующей таблице:

Год	Фактический уровень	Скользящая средняя		Центрированная скользящая средняя
		трехлетняя	четырехлетняя	
1996	15	-	-	-
1997	13	14,33	14,75	-
1998	15	14,67	15,50	15,125
1999	16	16,33	16, 50	16
2000	18	17,00	16,75	16,625
2001	17	17,00	17,50	17,125
2002	16	17,33	17,25	17,375
2003	19	17,33	18	17,625
2004	17	18,67	-	-
2005	20	-	-	-

Период скольжения может быть четным и нечетным. Для нечетного периода (трехлетнего) первое значение скользящей средней есть (15+13+15)/3=14,33, второе – (13+15+16)/3=14,67 и т.д., причем полученные результаты скользящей средней отнесены к середине периода скольжения.

Для четного периода (четырехлетнего) первое значение скользящей средней есть (15+13+15+16)/4=14,75, второе – (13+15+16+18)/4=15,50 и т.д. Одноко рассчитанные усредненные значения нельзя сопоставить каким-либо определенным значением t, поэтому применяют *процедуру центрирования*

(вычисляют среднее значение из двух последовательных скользящих средних). Первое значение центрированной скользящей средней есть (14,75+15,50)/2=15,125, второе – (15,50+16,50)/2=16 и т.д., причем первая центрированная средняя будет отнесена к третьему году, т.е. к 1998-му.

З а м е ч а н и е. *Используя пакет анализа Excel (программа «Скользящее среднее»), можно также получить результаты сглаживания.*

Задача №2. Имеются данные о розничном товарообороте региона (усл.ед.) за 10 лет:

Год	1	2	3	4	5	6	7	8	9	10
Товарооборот	11	13	22	18,5	20	19	25	23	24,5	35

Постройте следующие трендовые модели товарооборота и выберите из них наиболее подходящую:

Вид уравнения	Уравнение	R^2
Линейное	$y = 1{,}94t + 10{,}43$	0,767
Полином второго порядка	$y = 0{,}07t^2 + 1{,}127t + 12{,}06$	0,774
Полином третьего порядка	$y = 0{,}121t^3 + 1{,}921t^2 + 10{,}33t$	0,886
Логарифмическое	$y = 7{,}70lnt + 9{,}46$	0,709
Степенное	$y = 10{,}88t^{0{,}407}$	0,815
Экспоненциальное	$y = 11{,}84e^{0{,}096t}$	0,781

З а м е ч а н и е. *Для нахождения наиболее адекватного уравнения тренда в Excel используется инструмент «Подбор линии тренда» из Мастера диаграмм.*

Анализ аддитивной модели

Общий вид аддитивной модели таков: $Y = T + S + \varepsilon$.

Построение модели включает в себя следующие шаги:

1. выравнивание исходного ряда методом скользящей средней;
2. расчет значений сезонной компоненты S -
3. устранение сезонной компоненты из исходных уровней ряда (У- S) и получение выравненных данных ($T + \varepsilon$);
4. аналитическое выравнивание уровней ($T + \varepsilon$) и расчет значений T с использованием полученного уравнения тренда;
5. расчет полученных по модели значений (74 S) ;
6. расчет абсолютных ошибок.

Задача №3. Имеются поквартальные данные об объеме потребления электроэнергии y в некотором районе за четыре года (усл.ед.):

Год Квартал	1	2	3	4
	6,0	7,2	8,0	9,0
	4,4	4,8	5,6	6,6
	5,0	6,0	6,4	7,0
	9,0	10,0	11,0	10,8

В качестве зависимой переменной при анализе временного ряда выступает фактические уровни ряда y_1, а в качестве независимой переменной – время (сквозной номер квартала) t=1,2,…16.

По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) *о д и н а к о в о й* амплитуды, поэтому используется *аддитивная модель*.

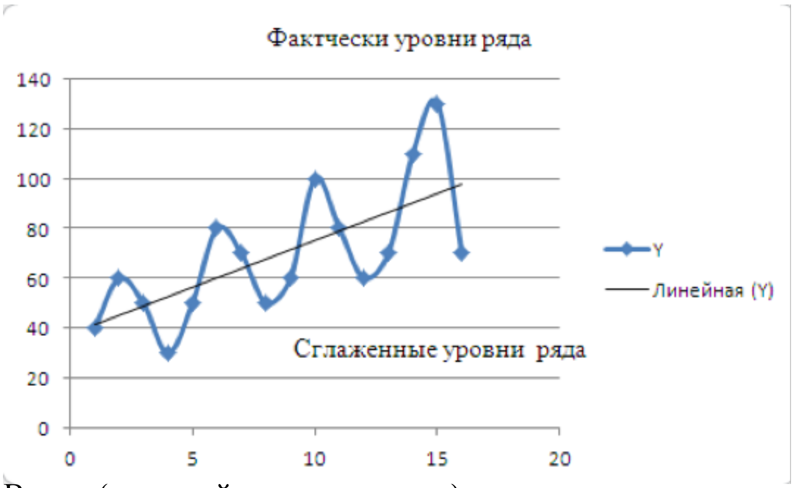
Определим ее компоненты, сезонную и трендовую.

Для исключения влияния сезонной компоненты проведем выравнивание исходного ряда методом скользящей средней за четыре квартала и процедуру центрирования. Результаты расчетов представлены в таблице:

Сквозной номер квартала	Потребление электро-энергии y_1	Скользящая средняя за четыре	Центрированная скользящая средняя	Оценка сезонной вариации
1	6,0	-	-	-
2	4,4	6,10	-	-
3	5,0	6,40	6,250	-1,250
4	9,0	6,50	6,450	2,550
5	7,2	6,75	6,625	0,575
6	4,8	7,00	6,875	-2,075
7	6,0	7,20	7,100	-1,100
8	10,0	7,40	7,300	2,700
9	8,0	7,50	7,450	0,550
10	5,6	7,75	7,625	-2,025
11	6,4	8,0	7,875	-1,475
12	11,0	8,25	8,125	2,875
13	9,0	8,40	8,325	0,675
14	6,6	8,35	8,375	-1,775
15	7,0	-	-	-
16	10,8	-	-	-

На рис. Представлены графики фактических уровней ряда и центрированной скользящей средней (сглаженные уровни).

Пот реб лен ие эле ктр оэн ерг ии



Время (сквозной номер квартала)

Расчет сезонной компоненты выполним в следующей расчетной таблице, в которой оценки сезонной вариации записываются под соответствующим номером квартала в году:

Показатели	Год	Номер квартала в году				
		I	**II**	**III**	**IV**	
	1	-	-	-1,250	2,550	
	2	0,575	-2,075	-1,100	2,700	
	3	0,550	-2,025	-1,475	2,875	
	4	0,675	-1,775	-	-	
Итого		1,800	-5,875	-3,825	8,125	сумма
Среднее		0,600	-1,958	-1,275	2,708	0,075
Скорректированное *Si*		0,581	-1,977	-1,294	2,690	0

Оценки сезонной вариации определяются как разность между фактическими уровнями ряда *y*, и центрированными скользящими средними.

В строке **Среднее** рассчитаны средние сезонной вариации по годам за каждый квартал и их сумма, равная 0,075.

В аддитивной модели предполагается, что сумма всех сезонных компонент по всем кварталам должна быть равна *нулю (условие взаимопогашаемости сезонных воздействий).*

В строке **Скорректированное** S_i рассчитаны значения сезонных компонент S_i как разность между средней сезонной вариацией и корректирующим коэффициентом 0,075/4, при этом $\sum S_i = 0$.

Расчет трендовой компоненты и ошибок выполнено в вышеуказанной таблице:

В столбце *« V - S= T+ε»* исключается влияние сезонной компоненты: вычитая ее значение из каждого уровня исходного ряда, получим только тенденцию и случайную компоненту.

Проводя аналитическое выравнивание ряда *(T + ε)* с помощью линейного тренда, получим следующее уравнение линии тренда: T = 5,706 + 0,187t.

t	Y	S	$Y - S =$ T+ ε	T	Ошибка e
1	6,0	0,581	5,419	5,893	-0,474
2	4,4	-1,977	6,337	6,088	0,256
3	5,0	-1,294	6,294	6,268	0,025
4	9,0	2,690	6,310	6,455	-0,145
5	7,2	0,581	6,619	6,642	-0,023
6	4,8	-1,977	6,777	6,829	-0,052
7	6,0	-1,294	7,294	7,016	0,277
8	10,0	2,690	7,310	7,204	0,106
9	8,0	0,581	7,419	7,391	0,027
10	5,6	-1,977	7,577	7,578	-0,001
11	6,4	-1,294	7,694	7,765	-0,071
t	Y	S	$Y - S =$ T+ ε	T	Ошибка e
12	11,0	2,690	8,310	7,952	0,357
13	9,0	0,581	8,419	8,139	0,279
14	6,6	-1,977	8,577	8,326	0,250
15	7.0	-1,294	8,294	8,514	-0,220
16	10,8	2,690	8,110	8,701	-0.591

Уровни ряда T для каждого $t =$ 1, 2,..., 16 указаны в вышеприведенной таблице. Расчет ошибки в аддитивной модели осуществляется по формуле

$e = Y - (T + S)$.

Дисперсии фактического ряда и ошибки, рассчитанные в Excel с помощью функции **ДИСПР**, составляют: var(y) = 4,196; var(ε) = 0,0684.

Для оценки качества построенной модели по аналогии с моделью рефессии можно использовать выражение

$$1 - \frac{var(e)}{var(y)} = 1 - \frac{0{,}0684}{4{,}196} = 0{,}984,$$

т.е. аддитивная модель объясняет 98,4% общей вариации уровней исходного временного ряда.

Анализ мультипликативной модели

Общий вид мультипликативной модели таков:

Y= TSe.

Построение модели включает в себя следующие шаги:

1) выравнивание исходного ряда методом скользящей средней;

2) расчет значений сезонной компоненты;

3) устранение сезонной компоненты исходных уровней ряда (***Y/S)*** и получение выравненных данных *(T*ε*)]*

4) аналитическое выравнивание уровней *(T*ε*)* и расчет значений *T* с использованием полученного уравнения тренда;

5) расчет полученных по модели значений (US');

6) расчет ошибок.

Пример 4.7. Имеются поквартальные данные о выплате доходов компании акционерам в форме дивидендов за последние четыре года (уел. ед.):

Год Квартал	1	2	3	4
I	40	60	50	30
II	50	80	70	50
III	60	100	80	60
IV	70	110	130	70

По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) *в о з р а с т а ю щ е й* амплитуды, поэтому используется *мультипликативная модель.* Определим ее компоненты, сезонную и трендовую.

Для исключения влияния сезонной компоненты проведем выравнивание исходного ряда методом скользящей средней за четыре квартала и процедуру центрирования. Результаты расчетов представлены в таблице:

Сквозной номер квартала	Размер дивидендов *Yt*	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной вариации
1	40	—	—	—
2	60	45	—	—
3	50	47,5	46,25	1,081
4	30	52,5	50	0,600
5	50	57,5	55	0,909
6	80	62,5	60	] ,333
7	70	65	63,75	1,098
8	50	70	67,5	0,741
9	60	72,5	71,25	0,842
10	100	75	73,75	1,356
11	80	77,5	76,25	1,049
12	60	80	78,75	0,762
13	70	92,5	86,25	0,811
14	110	95	93,75	1,173
15	130	—	—	
16	70	-		

Оценки сезонной вариации для мультипликативной модели определяются как частное от деления фактических уровней ряда *y,* на центрированные скользящие средние.

Показате	Год	Номер квартала в году				
		I	II	III	IV	
	1	-	-	1,081	0,600	
	2	0,909	1,333	1,098	0,741	

	3	0,842	1,356	1,049	0,762	
	4	0,811	1,173	-	-	
Итого		2,562	3,862	3,228	2,103	Сумм
Среднее		0,854	1,287	1,076	0,701	3,918
Скорректированно		0,872	1,314	1,099	0,715	4

Расчет сезонной компоненты выполним в следующей расчетной таблице, в которой оценки сезонной вариации записываются под соответствующим номером квартала в году:

В строке **Среднее** рассчитаны средние сезонной вариации по годам за каждый квартал и их сумма, равная 3,918.

В мультипликативной модели предполагается, что сумма всех сезонных компонент по всем кварталам должна быть равна четырем — числу сезонов в году *(условие взаимопогашаемости сезонных воздействий).*

В строке **Скорректированное** S_i рассчитаны значения сезонных компонент S_i как произведение соответствующей средней сезонной вариации на корректирующий коэффициент 4/3,918 = 1,021, при этом $\sum S_i = 4$.

Расчет трендовой компоненты и ошибок выполним в следующей таблице:

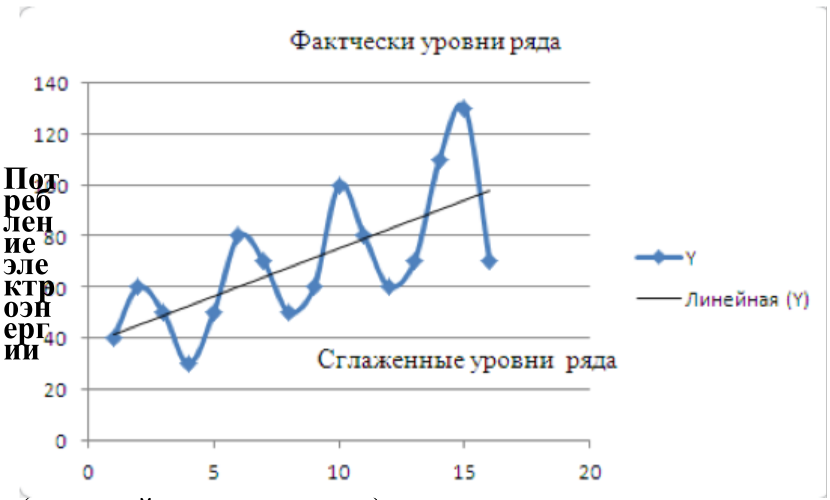
t	Y	S	Y/S=Tε	T	ε=Y/(TS)	e=Y-TS
1	40	0,872	45,871	38,97	1,18	6,02
2	60	1,314	45,662	43,046	1,06	3.44
3	50	1,099	45,496	47,122	0,96	-1,79
4	30	0,715	41,958	51,197	0,82	-6,60
5	50	0,872	57,335	55,273	1,04	1,80
6	80	1,314	60,883	59,349	1,02	2,01
7	70	1,099	63,694	63,425	1,00	0,29
8	50	0,715	69,930	67,501	1,03	1,73
9	60	0,872	68,807	71,577	0,96	-2,41
10	100	1,314	76,103	75,652	1,00	0,59
11	80	1,099	72,793	79,728	0,91	-7,62
12	60	0,715	83,916	83,804	1,00	0,08

13	70	0,872	80,275	87,880	0,91	-6,63
14	110	1,314	83,713	91,956	0,91	-10,83
15	130	1,099	118,289	96,032	1,23	24,46
16	70	0.715	97,902	100,108	0,98	-1,58

Разделив каждый уровень исходного ряда на соответствующие значения сезонной компоненты (столбец *«Y/S= Te»)*, исключим влияние сезонной компоненты и в результате получим только тенденцию и случайную компоненту.

Проводя аналитическое выравнивание ряда *(Te)* с помощью линейного тренда, получим следующее уравнение линии тренда: *T* = 34,89 + 4,087t. Уровни ряда *T* для каждого t= 1, 2 , … , 1 6 указаны в следующей таблице.

Графики исходного ряда и его тренда приведены на рис. 14.



Время (сквозной номер квартала)

Расчет ошибки в мультипликативной модели осуществляется по формуле *e = Y/(TS)*. Чтобы сравнивать мультипликативную модель с другими моделями временного ряда, ошибки в мультипликативной модели определяются как *e = Y- TS.*

Дисперсии фактического ряда и ошибки, рассчитанные в Excel с помощью функции **ДИСПР,** составляют: *var(y)* = 643,36; *var(e)* = 58,18.

Для оценки качества построенной модели можно по аналогии с аддитивной моделью использовать выражение

$$1 - \frac{var(e)}{var(y)} = 1 - \frac{58,18}{643,36} = 0,91 .$$

т.е. мультипликативная модель объясняет 91% общей вариации уровней исходного временного ряда.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №11

Тема: Автокорреляция случайного члена. Проверка наличия автокорреляции

Задача. Используя следующие статистические данные, определить наличие автокорреляции по методу Дарбина-Уатсона.

Таблица №

t	-15	-13	-11	-9	-7	-5	-3	-1	1	3	5	7	9	11	13	15
Год	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
y	47+k	51+k	55+k	59+k	62+k	66+k	70+k	75+k	79+k	82+k	86+k	89+k	92+k	96+k	100+k	103+k

k- порядковый номер студента по журналу.

Теоретическая часть

Корреляционная зависимость между последовательными уровнями временного ряда называется автокорреляцией уровней ряда.

Коэффициент автокорреляции порядка *x* определяется как коэффициент корреляции между рядами y_t и $y_{t-\tau}$;

$$R_t = \frac{cov(y_t, y_{t-\tau})}{\sqrt{var(y_t)var(y_{t-\tau})}}$$. Число периодов τ, по которым рассчитывается коэффициент автокорреляции, называется **лагом**.

С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается.

- если автокорреляции нет, то R=0
- если автокорреляция полная, то R=+-1

Критерий Дарбина-Уотсона показывает тогда:

- если автокорреляции нет, то d=2
- если автокоррелляция полная, то R=0 или 4

Оценки R и d являются интервальными. Существует таблица критерия Д-У при уровне существенности 5%

Таблица 4

Число наблюдений (n)	m=1		m=2		m=3		m=4		m=5	
	d	d	d	d	d	d	d	D	d	d
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,47
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

Методическое указание

Задача №1. Имеются данные об объеме предложения товара y, его цены x_1 и зарплаты сотрудников x_2 за 10 месяцев. Выявим на уровне значимости 0,05 наличие автокорреляции остатков в модели регрессии

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon .$$

Исходные данные и результаты промежуточных расчетов (усл. ед.) представлены в следующей таблице:

t	x_1	x_2	y	e_t	e_{t-1}
1	10	12	20	8,30	—
2	15	10	35	4,26	8,30
3	20	9	30	-12,46	4,26
4	25	9	45	-1,86	-12,46
5	40	8	60	-7,38	-1,86
6	37	8	70	5,26	-7,38
7	43	6	75	-9,66	5,26
8	35	4	90	-2,26	—9,66
9	40	4	105	8,34	-2,26
10	55	5	110	7,46	8,34

Выборочная регрессия для этой модели:

$\hat{y} = 90,72 + 0,88x_1 - 7,32x_2$.

Коэффициент автокорреляции остатков первого порядка r = -0,02512, следовательно, значение критерия Дарбина-Уотсона для этой модели составляет DW= 2,05. По таблице распределения Дарбина-Уотсона (см. Приложение) находим $d_x = 0,70$ и $d_2 = 1,64$. Поскольку $d_2 < DW < 4 - d_2$, то нет оснований отклонять гипотезу H_0 об отсутствии автокорреляции в остатках.

Замечание. Тест Дарбина-Уотсона построен в предположении, что объясняющие переменные некоррелированы со случайным членом. Поэтому этот тест неприменим к моделям, включающим в качестве объясняющих переменных лаговые значения зависимой переменной y.

Задача №. Пусть изучается зависимость среднедушевых расходов на конечное потребление y от среднедушевого дохода x; по данным некоторой страны за 16 лет.

Пусть исходная модель имеет вид $y_t = \alpha + \beta x_t + \varepsilon_t$.

Исходные (y_t, x_t) и расчетные $(e_t,\ y'_t\ x'_t)$ данные (уел. ед.) представлены в следующей таблице:

t	y_t,	x_t	e_t	$y_t^{;'}$	x'_t
1	70	73	0,18	-	-
2	73	76	0,76	37, 51	38,99
3	78	83	0,12	40,99	44,47
4	83	89	0,28	43,45	46,92
5	86	95	-1,55	43,92	49,88
6	89	100	-2,58	45,40	51,83
7	96	107	-1,22	50,88	56,30
8	96	108	-2,03	47,33	53,75
9	103	113	0,94	54,33	58,24
10	109	119	2,10	56,78	61,15
11	112	121	3,49	56,74	60,66
12	114	122	4,69	57,22	60,65

13	115	131	-1,56	57,20	69,14
14	118	135	-1,79	59,70	68,58
15	122	139	-1,01	62,17	70,55
16	123	140	-0,82	61,15	69,53

По исходным данным с использованием МНК получено следующее оцененное уравнение регрессии:

$$\hat{y}_t = \underset{(2,8)}{11,0} + \underset{(0,025)}{0,80}x_t, \qquad R^2 = 0,986$$

(в скобках указаны стандартные ошибки).

Коэффициент автокорреляции остатков первого порядка составляет ***r*** = 0,507, следовательно, ***DW***= 2(1 - ***r)*** = 0,986. При уровне значимости 5% табличное значение d_1 **-** 1,10 и d_2 = 1,37. Поскольку 0< ***DW***< d_1 то имеется положительная автокорреляция остатков. Применяя МНК к преобразованным данным

$$y'_t = y_t - 0,507y_{t-1},$$
$$x'_t = yx_t - 0,507x_{t-1} \qquad (t \geq 2),$$

получим оценку преобразованного уравнения

$$(\hat{y}_t)' = \underset{(3,0)}{6,2} + \underset{(0,05)}{0,79}x'_t, \qquad R^2 = 0,95$$

(в скобках указаны стандартные ошибки).

Коэффициент автокорреляции остатков первого порядка составляет ***r*** = 0,145, следовательно, ***DW***- 2(1 - ***r)*** = 1,71. Поскольку d_2 < ***DW***< ***4 -*** d_2, то автокорреляция остатков отсутствует.

Пересчитывая оценку α = **6**,2/(1 - 0,507) = 12,62, получим следующую оценку исходной модели:

$$\hat{y}_t = 12,62 + 0,79x_t, \qquad R^2 = 0,993$$

Это уравнение отличается от полученного ранее уравнения, оцененного обычным МНК.

Литература

1. Замков О.О. и др. Математические методы в экономике. М., изд-во ДИС, 1998.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. М., Высшая школа, 1991.
3. Кубонива М. Математическая экономика на ПК. М., Финансы и статистика, 1991.
4. Р.Г.Кравченко «Мтематическое моделирование экономических процессов в сельском хозяйстве. М: 1978г.
5. К. Доугерти «Введение в эконометрику», перевод с английского, М. 2001г.
6. Полунин И.Ф. Курс математического программирования. Минск, 1975.
7. Кади Дж. Количественные методы в экономике. М,1977.
8. Экономика и математические методы. www.bigmax.ru
9. Математические методы анализа экономики. http://www.shop4.ru
10. Журнал «Экономика и математические методы». adeptis.ru
11. Полтерович В.М. Экономика и математические методы. www.management.edu.ru

Содержание

Шадманова Гулчера

Мирзаев Сайибджан Сабитович

«Прикладная эконометрика»

Редактор

Подписано в печать________2022 г. Формат бумаги 60x84. Объем ___ п.л. Тираж 15. Заказ №____. Отпечатано в типографии НИУ «ТИИИМСХ».

Ташкент-100000, улица Кары-Ниязова, 39.