Министерство образования и науки РФ

Федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования

ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ

УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)

Кафедра экономической математики, информатики и статистики

ЭКОНОМЕТРИКА

Методические указания для выполнения практических и самостоятельных работ
для студентов специальности 080105 - Финансы и кредит
и для направлений: 080100 - Экономика, 080500 – Менеджмент

Зав.кафедрой ЭМИС,
д.ф.-м.н., профессор И.Г.Боровской

Составил: доц. каф. ЭМИС Д.Д. Даммер

2012

АННОТАЦИЯ

Методические указания для выполнения практических и самостоятельных работ для студентов специальности 080105 - Финансы и кредит и для направлений: 080100 - Экономика, 080500 – Менеджмент

В методических указаниях содержатся основные понятия и определения, используемые в эконометрике, приведены примеры решения задач и вопросы для самостоятельной работы студентов.

СОДЕРЖАНИЕ

Введение ……………………………………………………………………………………3
1. Парная линейная регрессия………………………………………………………………4
2. Проверка качества эмпирического уравнения парной линейной регрессии……………………..6
3. Множественная линейная регрессия……………………………………………………..11
4. Анализ качества эмпирического уравнения множественной линейной регрессии……………...13
5.Нелинейная регрессия……………………………………………………………………...17
6.Гетероскедастичность……………………………………………………………………...19
7.Автокорреляция……………………………………………………………………………23
8. Фиктивные переменные в регрессионных моделях ………………………………………24
9.Динамические модели. Лаги в экономических моделях………………………………...26
Задания для самостоятельной работы………...……………………………………………..30
Приложения……………………………………………………………………………………..34
Список рекомендуемой литературы…………………………………………………….…….38

Введение

Последние десятилетия эконометрика как научная дисциплина стремительно развивается. Формально «эконометрика» означает «измерения в экономике». Однако область исследований этой дисциплины гораздо шире. Эконометрика – это наука, в которой на базе реальных статистических данных строятся, анализируются и совершенствуются математические модели реальных экономических процессов. Эконометрика позволяет найти количественное подтверждение либо опровержение того или иного экономического закона или гипотезы. При этом одним из важнейших направлений эконометрики является построение прогнозов по различным экономическим показателям.

Эконометрика как научная дисциплина зародилась на основе слияния экономической теории, математики и статистики. Экономическая составляющая эконометрики, безусловно, является первичной. Экономика определяет постановку задачи и исходные предпосылки. Результат, формируемый на математическом языке, представляет интерес лишь в том случае, если удается его экономическая интерпретация.

Настоящее учебно-методическое пособие ориентировано на студентов экономических специальностей. Также оно может быть полезно всем интересующимся статистическими методами анализа экономических процессов. Предполагается, что студенты изучающие эконометрику, уже прослушали базовые курсы по высшей математике, теории вероятностей и математической статистике, микро- и макроэкономике.

1. Парная линейная регрессия

Цель занятия: по эмпирическим данным научиться строить модели парной регрессии, находить оценки коэффициентов уравнения.

Методические указания:

Если функция регрессии линейна, то говорят о *линейной регрессии. Линейная регрессия (теоретическое линейное уравнение регрессии)* представляет собой линейную функцию между условным математическим ожиданием $M(Y|X=x_i)$ зависимой переменной Y и одной объясняющей переменной X (x_i - значения независимой переменной в i-ом наблюдении, $i=1,2,\ldots,n$).

$$M(Y|X=x_i)=\beta_0+\beta_1 x_i.$$

Для отражения того факта, что каждое индивидуальное значение y_i отклоняется от соответствующего условного математического ожидания, необходимо ввести в последнее соотношение случайное слагаемое ε_i :

$$y_i=M(Y|X=x_i)+\varepsilon_i=\beta_0+\beta_1 x_i+\varepsilon_i.$$

Это соотношение называется *теоретической линейной регрессионной моделью;* β_0 и β_1 — *теоретическими параметрами (теоретическими коэффициентами) регрессии;* ε_i — *случайным отклонением.*

Для определения значений β_0 и β_1 необходимо знать и использовать все значения переменных X и Y генеральной совокупности, что практически невозможно.

Задача линейного регрессионного анализа состоит в том, чтобы по имеющимся статистическим данным (x_i, y_i), $i=1,2,\ldots,n$ для переменных X и Y:

а) получить наилучшие оценки неизвестных параметров β_0 и β_1;

б) проверить статистические гипотезы о параметрах модели;

в) проверить, достаточно ли хорошо модель согласуется со статистическими данными (адекватность модели данным наблюдений).

По выборке ограниченного объема мы сможем построить так называемое *эмпирическое уравнение регрессии*

$$\hat{y}_i=b_0+b_1 x_i,$$

где $\hat{y}_i$ – оценка условного математического ожидания $M(Y|X=x_i)$; b_0 и b_1 – оценки неизвестных параметров β_0 и β_1, называемые *эмпирическими коэффициентами регрессии.* Тогда :

$$y_i=b_0+b_1 x_i+e_i,$$

где *отклонение* e_i — оценка теоретического случайного отклонения ε_i.

Используя МНК можно получить формулы для определения b_0 и b_1:

$$\begin{cases} b_1=\dfrac{\overline{xy}-\bar{x}\cdot\bar{y}}{\overline{x^2}-\bar{x}^2} \\ b_0=\bar{y}-b_1\bar{x} \end{cases}, \text{ где } \quad \begin{matrix} \bar{x}=\dfrac{1}{n}\sum x_i,\ \overline{x^2}=\dfrac{1}{n}\sum x_i^2 \\ \bar{y}=\dfrac{1}{n}\sum y_i \end{matrix}$$

Пример.

Изучается зависимость себестоимости единицы продукции (y, тыс. руб.), от величины выпуска продукции (x, тыс. шт.). Экономист обследовал n=5 предприятий и получил следующие результаты:

Номер	x	y
1	2	1,9
2	3	1,7

3	4	1,8
4	5	1,6
5	6	1,4
Сумма	20	8,4

Полагая, что между переменными имеет место линейная зависимость, определить эмпирическое уравнение регрессии.

Решение:

Воспользуемся формулой для определения оценок коэффициентов уравнения регрессии

$$\begin{cases} b_1 = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}, \\ b_0 = \bar{y} - b_1\bar{x}. \end{cases}$$

Имеем

$$\bar{x} = \frac{2+3+4+5+6}{5} = 4,$$

$$\bar{y} = \frac{1{,}9+1{,}7+1{,}8+1{,}6+1{,}4}{5} = 1{,}68,$$

$$\overline{xy} = \frac{2 \cdot 1{,}9 + 3 \cdot 1{,}7 + 4 \cdot 1{,}8 + 5 \cdot 1{,}6 + 6 \cdot 1{,}4}{5} = \frac{32{,}5}{5} = 6{,}5,$$

$$\overline{x^2} = \frac{4+9+16+25+36}{5} = 18.$$

Тогда

$$\begin{cases} b_1 = \dfrac{6{,}5 - 4 \cdot 1{,}68}{18 - 4^2} = -0{,}11 \\ b_0 = 1{,}68 + 0{,}11 \cdot 4 = 2{,}12, \end{cases}$$

и эмпирическое уравнение регрессии будет иметь вид:

$$\hat{y} = 2{,}12 - 0{,}11x.$$

Задачи

1. Фирма провела рекламную компанию. Через 10 недель фирма решила проанализировать эффективность этого вида рекламы, сопоставив недельные объемы продаж (y, тыс. руб) с расходами на рекламу (x, тыс. руб):

x	5	8	6	5	3	9	12	4	3	10
y	72	76	78	70	68	80	82	65	62	90

Полагая, что между переменными имеет место линейная зависимость, определить эмпирическое уравнение линейной регрессии.

2. Постройте уравнение регрессии между данными о темпах прироста численности занятых и производительности труда сначала по всей совокупности данных, а затем исключив наблюдение для Японии. Дайте экономическую интерпретацию.

Страна	Занятость (прирост, %)	Производительность (прирост, %)
Австрия	2	4,2
Бельгия	1,5	3,9
Канада	2,3	1,3
Дания	2,5	3,2
Франция	1,9	3,8
Италия	4,4	4,2
Япония	5,8	7,8
Нидерланды	1,9	4,1

Норвегия	0,5	4,4
ФРГ	2,7	4,5
Великобритания	0,6	2,8
США	0,8	2,6

3. Для 13 клиентов спортивного магазина зафиксирована сумма покупки x_t (в у.е.) и время разговора с продавцом y_t (в мин.):

x_t	40	50	60	80	100	110	120	130	150	160	180	200	310
y_t	14	14	17	19	17	20	24	22	25	24	18	20	26

Требуется: оценить с помощью МНК параметры линейного регрессионного уравнения, предположив, что переменная «длительность разговора с продавцом» объясняется переменной «величина покупки»; оценить с помощью методов МНК параметры линейного регрессионного уравнения, предположив, что переменная «величина покупки» объясняется переменной «длительность разговора с продавцом»; нарисовать диаграмму рассеяния величин (x_t, y_t), обе линии регрессии и объяснить, почему получаются различные уравнения регрессии.

4. Исследуется зависимость затрат на рекламу y от годового оборота x в некоторой отрасли. Для этого собрана информация по 20 случайно выбранным предприятиям этой отрасли о годовом обороте x_i и соответствующих расходах на рекламу y_i. Из выборки получены следующие данные: $\overline{x} = 17{,}3; \overline{y} = 1{,}2; \sum x_i y_i = 944{,}3; \sum x_i^2 = 9250; \sum y_i^2 = 127{,}2$. Предполагается, что зависимость y_i от x_i описывается уравнением: $y_i = a + bx_i + u_i (i = 1, \dots 20)$. Требуется оценить параметры a и b с помощью МНК.

5. Исследователь считает, что $y = \beta x + u$. Выведите формулу МНК для расчета определения оценки b регрессионного параметра β.

6. Исследователь имеет ежегодные данные о временных рядах для совокупной заработной платы (*W*), совокупной прибыли (*П*) и совокупного дохода (*Y*) для страны за период в *n* лет. По определению $Y = W + П$. Используя МНК, получены уравнения регрессии: $\hat{W} = a_0 + a_1 Y$ и $\hat{П} = b_0 + b_1 Y$. Покажите, что коэффициенты регрессии будут автоматически удовлетворять следующим уравнениям: $a_1 + b_1 = 1$ и $a_0 + b_0 = 0$.

2. Проверка качества эмпирического уравнения парной линейной регрессии.

Цель занятия: научиться проверять качество уравнения регрессии оцениванием значения коэффициента детерминации, проверять гипотезы относительно коэффициентов уравнения регрессии, строить интервальные оценки коэффициентов и доверительные интервалы для зависимой переменной.

Методические указания.

После определения оценок теоретических коэффициентов регрессии нужно решить еще некоторые важные задачи: насколько точно эмпирическое уравнение регрессии соответствует уравнению для всей генеральной совокупности, насколько близки оценки b_0 и b_1 коэффициентов к своим теоретическим прототипам β_0 и β_1, как близко оцененное значение $\hat{y}_i$ к условному математическому ожиданию $M(Y|X = x_i)$.

Доказано, что для получения по МНК наилучших результатов необходимо, чтобы выполнялся ряд предпосылок относительно случайного отклонения (предпосылки Гаусса-Маркова):

1. Математическое ожидание случайного отклонения ε_i равно нулю: $M(\varepsilon_i)=0$ для всех наблюдений. Данное условие означает, что случайное отклонение в среднем не оказывает влияния на зависимую переменную. В каждом конкретном наблюдении может быть ибо положительным, либо отрицательным, но он не должен иметь систематического смещения. Выполнимость $M(\varepsilon_i)=0$ влечет выполнимость $M(Y|X=x_i)=\beta_0+\beta_1 x_i$
2. Дисперсия случайных отклонений ε_i постоянна: $D(\varepsilon_i)=D(\varepsilon_j)=\sigma^2$ для любых наблюдений i и j. Данное условие подразумевает, что несмотря на то, что при каждом конкретном наблюдении случайное отклонение может быть либо большим, либо меньшим, не должно быть некой причины, вызывающей большую ошибку (отклонение).
3. Случайные отклонения ε_i и ε_j являются независимыми друг от друга для $i \neq j$. Выполнимость данной предпосылки предполагает, что отсутствует систематическая связь между любыми случайными отклонениями.
4. Случайное отклонение должно быть независимо от объясняющих переменных.
5. Модель является линейной относительно параметров.

Теорема Гаусса-Маркова. Если предпосылки 1-5 выполнены, то оценки, полученные по МНК, обладают следующими свойствами:

1. Оценки являются несмещенными, т.е. $M(b_0)=\beta_0$, $M(b_1)=\beta_1$. Это вытекает из того, что $M(\varepsilon_i)=0$, и говорит об отсутствии систематической ошибки в определении положения линии регрессии.
2. Оценки состоятельны, так как дисперсия оценок параметров при возрастании числа наблюдений n стремится к нулю: $D(b_0)\xrightarrow[n\to\infty]{}0$, $D(b_1)\xrightarrow[n\to\infty]{}0$. Другими словами, при увеличении объема выборки надежность оценок увеличивается (b_0 близко к β_0, а b_1 близко к β_1.
3. Оценки эффективны, т.е. они имеют наименьшую дисперсию по сравнению с любыми другими оценками данных параметров, линейными относительно величин y_i.

Если предпосылки 2 и 3 нарушены, то свойства несмещенности и состоятельности сохраняются, но свойство эффективности – нет.

Анализ точности определения оценок коэффициентов регрессии

В силу случайного отбора элементов в выборку являются также оценки b_0 и b_1 коэффициентов β_0 и β_1 теоретического уравнения регрессии. Их математические ожидания при выполнении предпосылок об отклонениях ε_i равны соответственно $M(b_0)=\beta_0$, $M(b_1)=\beta_1$. При этом оцени тем надежнее, чем меньше их разброс вокруг β_0 и β_1, т.е. чем меньше дисперсии $D(b_0)$ и $D(b_1)$ оценок, рассчитываемые по формулам:

$$D(b_0)=\frac{\sigma^2\sum x_i^2}{n\sum(x_i-\bar{x})^2},\quad D(b_1)=\frac{\sigma^2}{\sum(x_i-\bar{x})^2}.$$

Так как ε_i неизвестны, они заменяются величинами $e_i=y_i-b_0-b_1x_i$. Дисперсия случайных отклонений $D(\varepsilon_i)=\sigma^2$ заменяется ее несмещенной оценкой

$$S^2=\frac{1}{n-2}\sum(y_i-b_0-b_1x_i)^2=\frac{\sum e_i^2}{n-2}.$$

Тогда:

$$D(b_0)\approx S_{b0}^2=\frac{S^2\sum x_i^2}{n\sum(x_i-\bar{x})^2}=\overline{x^2}S_{b1}^2,$$

$$D(b_1)\approx S_{b1}^2=\frac{S^2}{\sum(x_i-\bar{x})^2},$$

где $S^2 = \frac{\sum e_i^2}{n-2}$ – необъясненная дисперсия (мера разброса зависимой переменной вокруг линии регрессии), тогда $S = \sqrt{\frac{\sum e_i^2}{n-2}}$ – стандартная ошибка оценки (стандартная ошибка регрессии), $S_{b0} = \sqrt{S_{b0}^2}$ и $S_{b1} = \sqrt{S_{b1}^2}$ – стандартные отклонения случайных величин b_0 и b_1, называемые стандартными ошибками коэффициентов регрессии.

Проверка гипотез относительно коэффициентов линейного уравнения регрессии

При проведении статистического анализа перед исследователем зачастую возникает необходимость сравнения эмпирических коэффициентов регрессии b_0 и b_1 с некоторыми теоретически ожидаемыми значениями β_0 и β_1 этих коэффициентов. Данный анализ осуществляется при помощи статистической проверкой гипотез.

Для проверки гипотезы $b_1 = \beta_1$ используется статистика:

$$t = \frac{b_1 - \beta_1}{S_{b1}},$$

которая при справедливости $b_1 = \beta_1$ имеет распределение Стьюдента с числом степеней свободы $v = n-2$, где n – объем выборки. Следовательно, гипотеза $b_1 = \beta_1$ отклоняется на основании данного критерия, если:

$$|T_{набл}| = \left|\frac{b_1 - \beta_1}{S_{b1}}\right| \geq t_{\frac{\alpha}{2},n-2},$$

где α - требуемый уровень значимости. При невыполнении этого условия считается, что нет оснований для отклонения $b_1 = \beta_1$.

Задача установления значимости коэффициентов решается при помощи отношений, называемых t-статистикой:

$$t = \frac{b_1}{S_{b1}} \quad \text{и} \quad t = \frac{b_0}{S_{b0}}.$$

В случае, если $|t| > t_{\frac{\alpha}{2},n-2}$, то статистическая значимость соответствующего коэффициента регрессии подтверждается. Значения $t_{\frac{\alpha}{2},n-2}$ в зависимости от уровня значимости и числа степеней свободы v ($v = n-2$) приведены в приложении 1.

Интервальные оценки коэффициентов линейного уравнения регрессии

Доверительные интервалы коэффициентов b_0 и b_1, которые с надежностью $(1-\alpha)$ накрывают определяемые параметры β_0 и β_1, определяются по формулам:

$$\left(b_0 - t_{\frac{\alpha}{2},n-2}S(b_0);\ b_0 + t_{\frac{\alpha}{2},n-2}S(b_0)\right) \quad \text{и} \quad \left(b_1 - t_{\frac{\alpha}{2},n-2}S(b_1);\ b_1 + t_{\frac{\alpha}{2},n-2}S(b_1)\right).$$

Доверительные интервалы для зависимой переменной

Интервал, определяющий границы, за пределами которых могут оказаться не более $100\alpha\%$ точек наблюдений при $X = x_p$, рассчитывается следующим образом:

$$\left(b_0 + b_1 x_p \pm t_{\frac{\alpha}{2},n-2} S\sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum (x_i - \bar{x})^2}}\right).$$

Проверка общего качества уравнения регрессии.

Общее качество уравнения регрессии оценивается по тому, как хорошо эмпирическое уравнение регрессии согласуется со статистическими данными. Мерой общего качества уравнения регрессии является *коэффициент детерминации* R^2. В случае парной регрессии коэффициент детерминации будет совпадать с квадратом коэффициента корреляции. В общем случае коэффициент детерминации рассчитывается по формуле:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}$$

Справедливо соотношение $0 \le R^2 \le 1$. Чем теснее линейная связь между X и Y, тем ближе коэффициент детерминации R^2 к единице.

Пример.

Для анализа зависимости объема потребления Y (у.е.) домохозяйства от располагаемого дохода X (у.е.) отобрана выборка n=12 (помесячно в течение года) необходимо провести регрессионный анализ.

Данные и расчеты представлены в таблице:

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$\hat{y}_i$	e_i	e_i^2
1	107	102	11449	10914	10104	103,63	-1,63	2,66
2	109	105	11881	11445	11025	105,49	-0,49	0,24
3	110	108	12100	11880	11664	106,43	1,57	2,46
4	113	110	12769	12430	12100	109,23	0,77	0,59
5	120	115	14400	13800	13225	115,77	-0,77	0,59
6	122	117	14884	14274	13689	117,63	-0,63	0,40
7	123	119	15129	14637	14161	118,57	0,43	0,18
8	128	125	16384	16000	15625	123,24	1,76	3,10
9	136	132	18496	17952	17424	130,71	1,29	1,66
10	140	130	19600	18200	16900	134,45	-4,45	19,8
11	145	141	21025	20445	19881	139,11	1,89	3,57
12	150	144	22500	21600	20736	143,78	0,22	0,05
Сумма	1503	1448	190617	183577	176834	-	0	35,3
Среднее	125,25	120,67	15884,75	15298,08	14736,17	-	-	-

Решение:

Найдем сначала оценки коэффициентов:

$$\begin{cases} b_1 = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \dfrac{15298{,}08 - 125{,}25 \cdot 120{,}67}{15884{,}75 - (125{,}25)^2} = 0{,}9339, \\ b_0 = \bar{y} - b_1\bar{x} = 120{,}67 - 0{,}9339 \cdot 125{,}25 = 3{,}699. \end{cases}$$

Таким образом, уравнение парной регрессии имеет вид: $\hat{Y} = 3{,}699 + 0{,}9339X$. По этому уравнению рассчитаем $\hat{y}_i$, а также $e_i = y_i - \hat{y}_i$.

В нашем примере коэффициент b_1 может трактоваться как предельная склонность к потреблению. Фактически он показывает, на какую величину изменится объем потребления, если располагаемый доход возрастает на одну единицу. Свободный член b_0 определяет прогнозируемое значение Y при величине располагаемого дохода X, равной нулю (т.е. автономное потребление).

Рассчитаем другие показатели.

$$S^2 = \frac{\sum e_i^2}{n-2} = \frac{35{,}3}{12-2} = 3{,}53; \quad S = \sqrt{S^2} = 1{,}88,$$

$$S_{b1}^2 = \frac{S^2}{\sum (x_i - \bar{x})^2} = \frac{3{,}53}{2366.25} = 0{,}0015\,; \quad S_{b1} = \sqrt{S_{b1}^2} = 0{,}039\,,$$

$$S_{b0}^2 = \overline{x^2} S_{b1}^2 = 15884{,}75 \cdot 0{,}0015 = 23{,}83\,; \qquad S_{b0} = \sqrt{S_{b0}^2} = 4{,}88\,,$$

Проверим статистическую значимость коэффициентов b_0 и b_1:

$$t_1 = \frac{b_1}{S_{b1}} = \frac{0{,}9339}{0{,}039} = 23{,}946 \quad \text{и} \quad t_0 = \frac{b_0}{S_{b0}} = \frac{3{,}699}{4{,}88} = 0{,}76\,.$$

Критическое значение при уровне значимости $\alpha = 0{,}05$ равно (см. приложение 1) $t_{крит} = t_{\frac{\alpha}{2}, n-2} = t_{0{,}025;10} = 2{,}228$. Так как $|t_1| = 23{,}946 > 2{,}228$, то это подтверждает статистическую значимость коэффициента регрессии b_1. Аналогично для другого коэффициента: так как $|t_0| = 0{,}76 < 2{,}228$, то гипотеза о статистической значимости коэффициента b_0 отклоняется. Это означает, что в данном случае свободным членом уравнения регрессии можно пренебречь, рассматривая регрессию как $Y = b_1 X$.

Определим доверительные интервалы коэффициентов регрессии (по формуле 2.20), которые с надежность 95% $(\alpha = 0.05)$ будут следующими:

для b_0 $(3{,}699 - 2{,}228 \cdot 4{,}88;\ 3{,}699 + 2{,}228 \cdot 4{,}88) = (-7{,}173;\ 14{,}572)$,

для b_1 $(0{,}9339 - 2{,}228 \cdot 0{,}039;\ 0{,}9339 + 2{,}228 \cdot 0{,}039) = (0{,}8470;\ 1{,}021)$.

Рассчитаем границы интервала, в котором будет сосредоточено 95% возможных объемов потребления при неограниченно большом числе наблюдений и уровне дохода $X = 160$:

$$3{,}699 + 0{,}9339 \cdot 160 \pm 2{,}228 \cdot 1{,}88 \cdot \sqrt{1 + \frac{1}{12} + \frac{(125{,}25 - 160)^2}{2102{,}1875}}\,.$$

Таким образом, этот интервал имеет вид: (147,4898; 158,7082).

Рассчитаем коэффициент детерминации:

$$R^2 = 1 - \frac{35{,}3}{2108{,}6668} = 0{,}983\,.$$

Столь высокое значение коэффициента детерминации свидетельствует о высоком общем качестве построенного уравнения регрессии.

Задачи

1. В условиях задачи № 3 из предыдущего раздела для модели, в которой переменная «величина покупки» объясняется переменной «длительность разговора с продавцом»:

1) проверить статистическую значимость коэффициентов регрессии с уровнем значимости 5%;

2) определить доверительные интервалы для коэффициентов регрессии с уровнем значимости 1%;

3) определить доверительные интервалы для зависимой переменной при $y^* = 22$ для уровня значимости 10%;

4) проверить качество уравнения регрессии и статистическую значимость коэффициента детерминации (уровень значимости 10%).

2. Исследователь изучает зависимость между совокупным спросом на услуги (y) и совокупным располагаемым личным доходом (x) по данным для американской экономики (обе величины измеряются в млрд.дол.), используя ежегодные данные временных рядов и модель: $y = \alpha + \beta x + u$. Исследователь получает уравнение, проводя регрессионный анализ при помощи МНК. Предполагая, что обе величины x и y могут быть существенно занижены в системе национальных счетов из-за стремления людей уклониться от уплаты налогов, исследователь принимает два альтернативных метода уточнения заниженных оценок: 1) он добавляет к каждому году 90 млрд.дол. к показателю y и 200 млрд.дол. к показателю x; 2) он увеличивает значения как для x, так и для y на 10% за каждый год. Оцените влияние этих корректировок на результаты оценивания регрессии.

3. Регрессионная зависимость расходов на питание y от времени t задана уравнением: $\hat{y} = 95{,}3 + 2{,}53t$. Стандартная ошибка коэффициента при t составила 0,08. Проверьте нулевую гипотезу о том, что истинное значение коэффициента равно нулю при 5%-ном и 1%-ном уровнях значимости. Сделайте выводы.

3. Множественная линейная регрессия

Цель занятия: научиться строить модели множественной линейной регрессии, находить оценки коэффициентов уравнения.

Методические указания.

На любой экономический показатель чаще всего оказывает влияние не один, а несколько факторов. В этом случае вместо парной регрессии $M(Y|X) = f(x)$ рассматривается *множественная регрессия*

$$M(Y|x_1, x_2, \ldots, x_m) = f(x_1, x_2, \ldots, x_m)$$

Уравнение множественной регрессии может быть представлено в виде:

$$Y = f(\beta, X) + \varepsilon,$$

где $X = (X_1, X_2, \ldots, X_m)$ – вектор независимых (объясняющих) переменных; β - вектор параметров (подлежащих определению); ε - случайная ошибка (отклонение); Y - зависимая (объясняемая) переменная.

Теоретическое линейное уравнение регрессии имеет вид:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon,$$

или для индивидуальных наблюдений $i, i = 1,2,\ldots,n$:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im} + \varepsilon_i.$$

Здесь $\beta = (\beta_0, \beta_1, \ldots, \beta_m)$ – вектор размерности $(m+1)$ неизвестных параметров. $\beta_j, j = 1,2,\ldots,m$ называется j-тым теоретическим коэффициентом регрессии (частичным коэффициентом регрессии). β_0 - свободный член, определяющий Y в случае, когда все объясняющие переменные X_j равны нулю.

После выбора линейной функции в качестве модели зависимости необходимо оценить параметры регрессии. Пусть имеется n наблюдений вектора объясняющих переменных $X = (X_1, X_2, \ldots, X_m)$ и зависимой переменной Y:

$$(x_{i1}, x_{i2}, \ldots, x_{im}, y_i), i = 1,2,\ldots,n.$$

Для того, чтобы однозначно можно было бы решить задачу отыскания параметров $\beta_0, \beta_1, \ldots, \beta_m$ (т.е. найти некоторый наилучший вектор β), должно выполняться неравенство $n \geq m+1$.

Как и в случае парной регрессии, истинные значения параметров β_j по выборке получить невозможно. В этом случае вместо теоретического уравнения регрессии оценивается эмпирическое уравнение регрессии:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_m X_m + e.$$

Здесь $b_0, b_1, \ldots, b_m$ - оценки теоретических значений $\beta_0, \beta_1, \ldots, \beta_m$ коэффициентов регрессии (эмпирические коэффициенты регрессии); e - оценка отклонения ε. Для индивидуальных наблюдений имеем:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_m x_{im} + e_i.$$

При выполнении предпосылок МНК относительно ошибок ε_i оценки $b_0, b_1, \ldots, b_m$ параметров $\beta_0, \beta_1, \ldots, \beta_m$ множественной линейной регрессии по МНК являются несмещенными, эффективными и состоятельными.

Обозначим

$$Y=\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix},\quad X=\begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix},\quad B=\begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_m \end{pmatrix},\quad e=\begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}.$$

Здесь $Y - n$-мерный вектор-столбец наблюдений зависимой переменной X; X - матрица размерности $n \times (m+1)$, в которой i-тая строка $(i = 1,2,\dots,n)$ представляет наблюдение вектора значений независимых переменных $X_1, X_2, \dots, X_m$; единица соответствует переменной при свободном члене b_0; B - вектор-столбец размерности $(m+1)$ параметров уравнения регрессии; e - вектор-столбец размерности n отклонений выборочных (реальных) значений y_i зависимой переменной Y от значений $\hat{y}_i$, получаемых по уравнению регрессии

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_m x_{im},\ i = 1,2\dots n.$$

Тогда МНК-оценки параметров $\beta_0, \beta_1, \dots, \beta_m$ будут определяться следующей формулой:

$$B = (X^T X)^{-1} X^T Y.$$

Здесь $(X^T X)^{-1}$ – матрица, обратная к $X^T X$.

Полученные общие соотношения справедливы для уравнений регрессии с произвольным количеством m объясняющих переменных. Например, для $m = 2$ получим:

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2,$$

$$b_1 = \frac{\sum(x_{i1}-\bar{x}_1)(y_i-\bar{y})\cdot\sum(x_{i2}-\bar{x}_2)^2 - \sum(x_{i2}-\bar{x}_2)(y_i-\bar{y})\cdot\sum(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)}{\sum(x_{i1}-\bar{x}_1)^2\sum(x_{i2}-\bar{x}_2)^2 - (\sum(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2))^2},$$

$$b_2 = \frac{\sum(x_{i2}-\bar{x}_2)(y_i-\bar{y})\cdot\sum(x_{i1}-\bar{x}_1)^2 - \sum(x_{i1}-\bar{x}_1)(y_i-\bar{y})\cdot\sum(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)}{\sum(x_{i1}-\bar{x}_1)^2\sum(x_{i2}-\bar{x}_2)^2 - (\sum(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2))^2}.$$

Пример.

Анализируется объем s сбережений домохозяйства за 10 лет. Предполагается, что его размер s_t в текущем году t зависит от величины y_t располагаемого дохода y и от величины z_t реальной процентной ставки z. Статистические данные представлены в таблице:

Год	80	81	82	83	84	85	86	87	88	89	90
y, тыс.у.е.	100	110	140	150	160	160	180	200	230	250	260
z, %	2	2	3	2	3	4	4	3	4	5	5
s, тыс.у.е.	20	25	30	30	35	38	40	38	44	50	55

Необходимо рассчитать оценки коэффициентов уравнения регрессии.

Решение:

Средние значения исходных данных равны: $\bar{y} = 176{,}3636$, $\bar{z} = 3{,}3636$, $\bar{s} = 36{,}8182$.

Представим требующиеся для построения модели множественной регрессии и проведения дальнейшего анализа промежуточные вычисления в таблице:

Год	$(y_i-\bar{y})^2$	$(z_i-\bar{z})^2$	$(s_i-\bar{s})^2$	$(y_i-\bar{y})\cdot(z_i-\bar{z})$	$(y_i-\bar{y})\cdot(s_i-\bar{s})$	$(z_i-\bar{z})\cdot(s_i-\bar{s})$
80	5831,4050	1,8595	282,8512	104,1322	1284,2975	22,9339
81	4404,1322	1,8595	139,6694	90,4959	784,2975	16,1157
82	1322,3140	0,1322	46,4876	13,2231	247,9339	2,4793
83	695,0413	1,8595	46,4876	35,9504	179,7521	9,2975
84	267,7686	0,1322	3,3058	5,9504	29,7521	0,6612
85	267,7686	0,4050	1,3967	-10,4132	-19,3388	0,7521

86	13,2231	0,4050	10,1240	2,3140	11,5702	2,0248
87	558,6777	0,1322	1,3967	-8,5950	27,9339	-0,4298
88	2876,8595	0,4050	51,5785	34,1322	385,2066	4,5702
89	5422,3140	2,6777	173,7603	120,4959	970,6612	21,5702
90	6995,0413	2,6777	330,5785	136,8595	1520,6612	29,7521
$\sum$	28654,5455	12,5455	1087,6364	524,5455	5422,7273	109,7273

Теперь рассчитаем коэффициенты уравнения регрессии:

$$b_0 = 36{,}8182 - 0{,}124189 \cdot 176{,}3636 - 3{,}553796 \cdot 3{,}3636 = 2{,}962233,$$

$$b_1 = \frac{5422{,}7273 \cdot 12{,}5455 - 109{,}7273 \cdot 524{,}5455}{28654{,}5455 \cdot 12{,}5455 - (524{,}5455)^2} = \frac{10473{,}8639}{84337{,}619} = 0{,}124189,$$

$$b_2 = \frac{109{,}7273 \cdot 28654{,}5455 - 5422{,}7273 \cdot 524{,}5455}{28654{,}5455 \cdot 12{,}5455 - (524{,}5455)^2} = \frac{299718{,}7075}{84337{,}619} = 3{,}553796$$

Таким образом, эмпирическое уравнение регрессии имеет вид:

$$s_t = 2{,}962233 + 0{,}124189 y_t + 3{,}553796 z_t .$$

Задачи

1. Предполагается, что объем предложения товара y линейно зависит от цены товара x_1 и зарплаты сотрудников x_2 : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. Статистические данные собраны за десять месяцев. Оценить по МНК коэффициенты уравнения регрессии для двух вариантов:

1)

y, руб	20	35	30	45	60	70	75	90	105	110
x_1, руб	10	15	20	25	40	37	43	35	40	55
x_2, руб	12	10	9	9	8	8	6	4	4	5

2)

y, руб	75	90	105	110	120	130	130	130	135	140
x_1, руб	43	35	38	55	50	35	40	55	45	65
x_2, руб	6	4	4	5	3	1	2	3	1	2

2. Торговое предприятие имеет несколько филиалов. Найти коэффициенты эмпирического уравнения множественной регрессии, если предполагается, что зависимая переменная y – это годовой товарооборот филиала, а независимые переменные x_1, x_2 – размер торговой площади и среднедневная интенсивность потока соответственно. Зависимость y от x_1, x_2 предполагается линейная. Данные приведены в следующей таблице:

y, млн руб	2,93	5,27	6,85	7,01	7,02	8,35	4,33	5,77	7,68	3,16	1,52	3,15
x_1, тыс м^2	0,31	0,98	1,21	1,29	1,12	1,49	0,78	0,94	1,29	0,48	0,24	0,55
x_2, тыс чел в день	10,24	7,51	10,81	9,89	13,72	13,92	8,54	12,36	12,27	11,01	8,25	9,31

4.Анализ качества эмпирического уравнения множественной линейной регрессии.

Цель занятия: научиться проверять качество уравнения множественной регрессии оцениванием значения коэффициента детерминации, проверять гипотезы относительно коэффициентов уравнения регрессии, строить интервальные оценки коэффициентов и доверительные интервалы для зависимой переменной.

Методические указания.

Прежде чем проводить анализ качества уравнения регрессии, необходимо определить дисперсии и стандартные ошибки коэффициентов, а также интервальные оценки коэффициентов.

Выборочные дисперсии эмпирических коэффициентов регрессии можно определить следующим образом:

$$S_{bj}^2 = S^2 z'_{jj} = \frac{\sum e_i^2}{n-m-1} z'_{jj}, \; j = 1,2,\ldots,m .$$

Здесь z'_{jj} - j-тый диагональный элемент матрицы $Z^{-1} = (X^T X)^{-1}$.

При этом:

$$S^2 = \frac{\sum e_i^2}{n-m-1},$$

где m - количество объясняющих переменных модели.

В частности, для уравнения $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$ с двумя объясняющими переменными используются следующие формулы:

$$S_{b0}^2 = \left[\frac{1}{n} + \frac{\bar{x}_1^2 \sum (x_{i2} - \bar{x}_2)^2 + \bar{x}_2^2 \sum (x_{i1} - \bar{x}_1)^2 - 2\bar{x}_1\bar{x}_2 \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sum (x_{i1} - \bar{x}_1)^2 \sum (x_{i2} - \bar{x}_2)^2 - (\sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2))^2}\right] \cdot S^2,$$

$$S_{b1}^2 = \frac{\sum (x_{i2} - \bar{x}_2)^2}{\sum (x_{i1} - \bar{x}_1)^2 \sum (x_{i2} - \bar{x}_2)^2 - (\sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2))^2} \cdot S^2 \Rightarrow$$

$$S_{b1}^2 = \frac{S^2}{\sum (x_{i1} - \bar{x}_1)^2 \cdot (1 - r_{12}^2)},$$

$$S_{b2}^2 = \frac{\sum (x_{i1} - \bar{x}_1)^2}{\sum (x_{i1} - \bar{x}_1)^2 \sum (x_{i2} - \bar{x}_2)^2 - (\sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2))^2} \cdot S^2 \Rightarrow$$

$$S_{b2}^2 = \frac{S^2}{\sum (x_{i2} - \bar{x}_2)^2 \cdot (1 - r_{12}^2)},$$

$$S_{b0} = \sqrt{S_{b0}^2}, \quad S_{b1} = \sqrt{S_{b1}^2}, \quad S_{b2} = \sqrt{S_{b2}^2} .$$

Здесь r_{12} - выборочный коэффициент корреляции между объясняющими переменными X_1 и X_2; S_{bj} - стандартная ошибка коэффициента регрессии; S - стандартная ошибка регрессии (несмещенная оценка).

По аналогии с парной регрессией после определения точечных оценок b_j коэффициентов β_j ($j = 1,2,\ldots,m$) теоретического уравнения регрессии могут быть рассчитаны интервальные оценки указанных коэффициентов. Доверительный интервал, накрывающий с надежностью $(1-\alpha)$ неизвестное значение параметра β_j, определяется:

$$\left(b_j - t_{\frac{\alpha}{2}, n-m-1} S(b_j); \; b_j + t_{\frac{\alpha}{2}, n-m-1} S(b_j) \right)$$

Проверка статистической значимости коэффициентов уравнения регрессии.

Как и в случае парной регрессии, статистическая значимость коэффициентов множественной линейной регрессии с m объясняющими переменными проверяется на основе t-статистики:

$$t = \frac{b_j}{S_{bj}} \ ,$$

имеющей в данном случае распределение Стьюдента с числом степеней свободы $v = n - m - 1$. При требуемом уровне значимости наблюдаемое значение t-статистики сравнивается с критической точной $t_{\frac{\alpha}{2}, n-m-1}$ распределения Стьюдента.

В случае, если $|t| > t_{\frac{\alpha}{2}, n-m-1}$, то статистическая значимость соответствующего коэффициента регрессии подтверждается. Это означает, что фактор X_j линейно связан с зависимой переменной Y. Если же установлен факт незначимости коэффициента b_j, то рекомендуется исключить из уравнения переменную X_j. Это не приведет к существенной потере качества модели, но сделает ее более конкретной.

Проверка общего качества уравнения регрессии.

Для этой цели, как и в случае парной регрессии, используется *коэффициент детерминации* R^2:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} \ .$$

Справедливо соотношение $0 \le R^2 \le 1$. Чем ближе этот коэффициент к единице, тем больше уравнение регрессии объясняет поведение Y.

Рекомендуется после проверки общего качества уравнения регрессии провести анализ его статистической значимости. Для этого используется F-статистика:

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

Если $F = 0$, то $R^2 = 0$, следовательно, величина Y линейно не зависит от $X_1, X_2, \ldots, X_m$.

Расчетное значение F сравнивается с критическим $F_{кр}$, которое определяется на основе распределения Фишера (приложение 2) исходя из требуемого уровня значимости α и числами степеней свободы $v_1 = m$ и $v_2 = n - m - 1$. Если $F > F_{кр}$, то R^2 признается статистически значимым.

Пример.

В условиях задачи из примера предыдущего раздела получить следующее

1) дисперсию регрессии,

2) дисперсии и стандартные ошибки коэффициентов,

3) соответствующие t-статистики,

4) проверить статистическую значимость коэффициентов на основе распределения Стьюдента с уровнем значимости $\alpha = 0{,}05$,

5) определить 95%-е интервальные оценки коэффициентов,

6) рассчитать коэффициент детерминации,

7) проанализировать статистическую значимость коэффициента детерминации с уровнем значимости $\alpha = 0{,}05$.

Решение:

Подставляя соответствующие значения y_t и z_t в эмпирическое уравнение регрессии:

$$\hat{s}_t = 2{,}962233 + 0{,}124189 y_t + 3{,}553796 z_t$$

получаем значения $\hat{s}_t$. Расчет отклонений e_i реальных значений от модельных представлен в таблице:

Год	s	$\hat{s}$	e_i	e_i^2	$e_i - e_{i-1}$	$(e_i - e_{i-1})^2$
80	20	22,489	-2,48873	6,19375	-	-

81	25	23,731	1,26939	1,61134	3,75811	14,1234
82	30	31,01	-1,01008	1,02026	-2,27947	5,19597
83	30	28,698	1,30183	1,69475	2,31191	5,34491
84	35	33,494	1,50614	2,26845	0,20431	0,04174
85	38	37,048	0,95234	0,90696	-0,55380	0,30669
86	40	39,531	0,46856	0,21955	-0,48378	0,23404
87	38	38,461	-0,46142	0,21291	-0,92998	0,86487
88	44	45,741	-1,74089	3,03069	-1,27947	1,63703
89	50	51,778	-1,77846	3,16293	-0,03758	0,00141
90	55	53,02	1,97965	3,91900	3,75811	14,1234
сумма	405	405	≈ 0	24,24060	4,46837	41,8734
	36,8182	36,8182				

Рассчитаем дисперсию регрессии по формуле :

$$S^2 = \frac{\sum e_i^2}{n-m-1} = \frac{24,2406}{11-2-1} = 3,03.$$

Определим дисперсии и стандартные ошибки коэффициентов:

$$S_{b0}^2 = \left[\frac{1}{11} + \frac{(176,3636)^2 \cdot 12,5455 + (3,3636)^2 \cdot 28654,5455 - 2 \cdot 176,3636 \cdot 3,3636 \cdot 524,5455}{28654,5455 \cdot 12,5455 - (524,5455)^2}\right] \cdot 3,03,$$

$$S_{b0} = \sqrt{S_{b0}^2} = \sqrt{3,5832} = 1,8929,$$

$$S_{b1}^2 = \frac{12,5455}{28654,5455 \cdot 12,5455 - (524,5455)^2} \cdot 3,03 = 0,00054, \quad S_{b1} = \sqrt{S_{b1}^2} = \sqrt{0,00054} = 0,0212,$$

$$S_{b2}^2 = \frac{28654,5455}{28654,5455 \cdot 12,5455 - (524,5455)^2} \cdot 3,03 = 1,0294, \quad S_{b2} = \sqrt{S_{b2}^2} = \sqrt{1,0294} = 1,0146,$$

Рассчитаем соответствующие t-статистики:

$t_{b0} = 1,565$, $t_{b1} = 5,858$, $t_{b2} = 3,503$.

Проверим статистическую значимость коэффициентов на основе распределения Стьюдента. По таблице, приведенной в приложении 1, определим критические значения с уровнем значимости $\alpha = 0,05$: $t_{кр} = t_{\frac{\alpha}{2};n-m-1} = t_{0,025;\ 8} = 2,306$. Таким образом, $|t_{b0}| < t_{кр}$, $|t_{b1}| > t_{кр}$, $|t_{b2}| > t_{кр}$.

Определим 95%-е интервальные оценки коэффициентов:

для β_0: (2,962233-2,306*1,8929; 2,962233+2,306*1,8929), т.е. (-1,4028; 7,3273),

для β_1: (0,124189-2,306*0,0212; 0,124189+2,306*0,0212), т.е. (0,0753; 0,1731),

для β_2: (3,553796-2,306*1,0146; 3,553796+2,306*1,0146), т.е. (1,2141; 5,8935),

Рассчитаем коэффициент детерминации:

$$R^2 = 1 - \frac{24,2406}{1087,6364} = 0,9777.$$

Анализ статистической значимости коэффициента детерминации осуществляется на основе F-статистики :

$$F = \frac{0,9777}{1-0,9777} \cdot \frac{11-2-1}{2} = 175,3722$$

Определим по приложению 2 критическую точку распределения Фишера: $F_{кр} = F_{0,05;2;8} = 4,46$ с 95%-ой вероятностью. Очевидно, что 175,3722>4,46, следовательно, коэффициент детерминации статистически значим, т.е. совокупное влияние переменных y и z на переменную s существенно.

На основе проведенных рассуждений и вычислений можно заключить, что построенное уравнение регрессии объясняет 97,77% разброса зависимой переменной S.

По всем статистическим показателям модель может быть признана удовлетворительной.

Задачи

1. Найти стандартную ошибку регрессии и стандартную ошибку коэффициентов в условиях задачи 1 и задачи 2 из предыдущего раздела.

2. В результате решения задачи 1 и задачи 2 из предыдущего раздела проверить статистическую значимость коэффициентов уравнения с уровнем значимости $\alpha = 0{,}05$, определить 95%-е интервальные оценки коэффициентов, рассчитать коэффициент детерминации, проанализировать статистическую значимость коэффициента детерминации с уровнем значимости $\alpha = 0{,}05$. Сделать вывод о качестве модели.

5. Нелинейная регрессия

Цель занятия: по эмпирическим данным научиться строить нелинейные регрессионные модели, оценивать коэффициенты таких моделей.

Методические указания.

Построение и анализ нелинейных моделей имеют свою специфику. Рассмотрим нелинейные модели, допускающими сведение их к линейным. Такие модели называют *линейные относительно параметров модели*. Будем рассматривать модели парной нелинейной регрессии:

1) Логарифмические модели: $Y = AX^{\beta}$, где A, β – параметры модели (константы, подлежащие определению). Для анализа такой функции используется логарифмирование всего выражения:

$$\ln Y = \ln A + \beta \ln X .$$

С целью статистической оценки коэффициентов добавим в модель случайную погрешность ε и заменим: $\ln A = \beta_0$, $Y^* = \ln Y$ и $X^* = \ln X$, получаем линейную модель:

$$Y^* = \beta_0 + \beta X^* + \varepsilon ,$$

и при большем числе переменных:

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \ldots + \beta_m \ln X_m + \varepsilon .$$

2) Полулогарифмические модели: $\ln Y = \beta_0 + \beta X + \varepsilon$, $Y = \beta_0 + \beta \ln X + \varepsilon$. После замены $Y^* = \ln Y$ и $X^* = \ln X$, получаем линейную модель.

3) Обратная модель: $Y = \beta_0 + \beta_1 \cdot \dfrac{1}{X} + \varepsilon$. Сводится к линейной путем замены $X^* = \dfrac{1}{X}$.

4) Показательная модель $Y = \beta_0 e^{\beta x}$. Сначала сводится к лог-линейной $\ln Y = \ln \beta_0 + \beta X$, а потом к линейной модели.

Пример.

Анализируется индекс потребительских цен Y по объему денежной массы X на основании приведенных в таблице данных. Необходимо построить логарифмическую модель**.**

Год	Y	X	Год	Y	X
81	65	110	89	95	235
82	68	125	90	100	240
83	72,5	132	91	106,5	245
84	77,5	137	92	112	250
85	82	160	93	115,5	275
86	85,5	177	94	118,5	285
87	88,5	192	95	120	295
88	91	215	96	120,5	320
			97	121	344

Решение:

Логарифмическая модель имеет вид: $Y = AX^{\beta}$. Данная модель сводится к линейной следующим образом: $\ln Y = b_0 + b \ln X$. Для определения коэффициентов в этой модели определим логарифмы переменных Y и X, $(\ln X)^2$, $(\ln X)\cdot(\ln Y)$ и представим их в таблице.

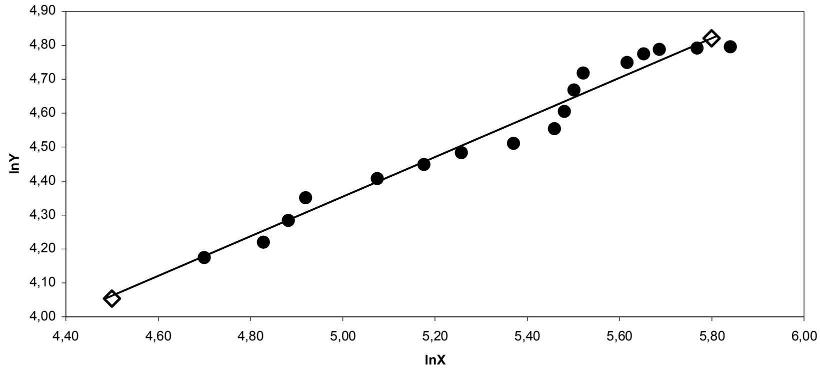
Год	Y	X	$\ln Y$	$\ln X$	$(\ln X)^2$	$(\ln X)\cdot(\ln Y)$
81	65	110	4,1744	4,7005	22,0947	19,6218
82	68	125	4,2195	4,8283	23,3125	20,3730
83	72,5	132	4,2836	4,8828	23,8417	20,9160
84	77,5	137	4,3503	4,9200	24,2064	21,4035
85	82	160	4,4067	5,0752	25,7577	22,3649
86	85,5	177	4,4485	5,1761	26,7920	23,0259
87	88,5	192	4,4830	5,2575	27,6413	23,5694
88	91	215	4,5109	5,3706	28,8433	24,2262
89	95	235	4,5539	5,4596	29,8072	24,8625
90	100	240	4,6052	5,4806	30,0370	25,2393
91	106,5	245	4,6681	5,5013	30,2643	25,6806
92	112	250	4,7185	5,5215	30,4870	26,0532
93	115,5	275	4,7493	5,6168	31,5484	26,6759
94	118,5	285	4,7749	5,6525	31,9508	26,9901
95	120	295	4,7875	5,6870	32,3420	27,2265
96	120,5	320	4,7916	5,7683	33,2733	27,6394
97	121	344	4,7958	5,8406	34,1126	28,0103
Сумма	1639	3737	77,3217	90,7392	486,3122	413,8784
Среднее	96,4118	219,8235	4,5483	5,3376	28,6066	24,3458

Затем, по аналогии с примером, приведенным в разделе 1, рассчитываются коэффициенты для этой модели следующим образом:

$$\begin{cases} b = \dfrac{\overline{(\ln X)\cdot\ln(Y)} - \overline{\ln X}\cdot\overline{\ln Y}}{\overline{(\ln X)^2} - \left(\overline{\ln X}\right)^2} = \dfrac{24{,}3458 - 5{,}3376\cdot 4{,}5483}{28{,}6066 - (5{,}3376)^2} = \dfrac{0{,}0688}{0{,}1166} = 0{,}5901, \\ b_0 = \overline{\ln Y} - b\cdot\overline{\ln X} = 4{,}5483 - 0{,}5901\cdot 5{,}3376 = 1{,}3986. \end{cases}$$

Следовательно, модель имеет вид: $\ln Y = 1{,}3986 + 0{,}5901\cdot\ln X$. Если свести данную модель к виду $Y = AX^{\beta}$, то получим: $Y = 4{,}0495\cdot X^{0{,}5901}$ (т.к. $\ln A = b_0 = 1{,}3986$, следовательно, $A = e^{b_0} = 4{,}0495$).

Представим графически корреляционное поле для переменных $\ln Y$ и $\ln X$, а также график рассчитанной модели $\ln Y = 1{,}3986 + 0{,}5901\cdot\ln X$.



Задачи.

1. В условиях задачи из примера проверить значимость коэффициентов уравнения регрессии, определить их интервальные оценки и рассчитать коэффициент детерминации. Расчет проводится аналогично примеру в разделе 1 для модели вида $\ln Y = 1{,}3986 + 0{,}5901 \cdot \ln X$.

2. Определить экспоненциальную функцию вида $y = \alpha \cdot e^{\beta x}$, где y – совокупные личные расходы, x – располагаемый личный доход (по данным из таблицы индивидуальных заданий). Проверить статистическую значимость коэффициентов регрессии и доверительные интервалы коэффициентов с уровнем значимости 5%, проверить качество уравнения регрессии. Определить для этих же данных линейную регрессию вида $y = \alpha + \beta x$.

6. Гетероскедастичность

Цель занятия: Для построенной модели регрессии научиться определять выполнимость второй предпосылки МНК.

Методические указания.

Гетероскедастичность приводит к тому, что выводы, полученные на основе t- и F-статистик, а также интервальные оценки будут ненадежными. Обнаружение гетероскедастичности является довольно сложной задачей. В настоящее время существует ряд методов, позволяющих определить наличие гетероскедастичности.

1. Тест ранговой корреляции Спирмена

Значения x_i и e_i (абсолютные величины) ранжируются (упорядочиваются по величинам). Затем определяется коэффициент ранговой корреляции:

$$r_{x,e} = 1 - 6 \cdot \frac{\sum d_i^2}{n(n^2 - 1)},$$

где d_i - разность между рангами x_i и e_i, $i = 1,2,\ldots,n$; n - число наблюдений.

Например, если x_{20} является 25-ым по величине среди всех наблюдений, а e_{20} является 32-м, то $d_{20} = 25 - 32 = -7$.

Затем рассчитывается статистика:

$$t = \frac{r_{x,e}\sqrt{(n-2)}}{\sqrt{1 - r_{x,e}^2}}.$$

Если это t значение, превышает критическое $t_{kp} = t_{\frac{\alpha}{2}, n-2}$ (определяемое по приложению 1), то необходимо отклонить гипотезу об отсутствии гетероскедастичности. В противном случае гипотеза об отсутствии гетероскедастичности принимается.

Если в модели регрессии больше, чем одна объясняющая переменная, то проверка гипотезы может осуществляться с помощью t-статистики для каждой из них отдельно.

2. Тест Голдфелда-Квандта

В данном случае предполагается, что стандартное отклонение $\sigma_i = \sigma(\varepsilon_i)$ пропорционально значению x_i переменной X в этом наблюдении, т.е. $\sigma_i^2 = \sigma^2 x_i^2$, $i = 1,2,\dots,n$.

Тест Голдфелда-Квандта состоит в следующем:

1. Все n наблюдений упорядочиваются по величине X.
2. Вся упорядоченная выборка после этого разбивается на три подвыборки размерностей $k, (n-2k), k$ соответственно.
3. Оцениваются отдельные регрессии для первой подвыборки (k первых наблюдений) и для третьей подвыборки (k последних наблюдений). Для парной регрессии Голдфелд и Квандт предлагает следующие пропорции: $n = 30, k = 11$; $n = 60, k = 22$. Если предположение о пропорциональности дисперсий отклонений значениям X верно, то дисперсия регрессии по первой подвыборке (рассчитываемая как $S_1 = \sum_{i=1}^{k} e_i^2$) будет существенно меньше дисперсии регрессии по третьей подвыборке (рассчитываемой как $S_3 = \sum_{i=n-k+1}^{n} e_i^2$).
4. Для сравнения соответствующих дисперсий строится соответствующая F-статистика:
$$F = \frac{S_3/(k-m-1)}{S_1/(k-m-1)} = \frac{S_3}{S_1}.$$
Здесь $(k-m-1)$ – число степеней свободы соответствующих выборочных дисперсий (m - количество объясняющих переменных в уравнении регрессии).
Построенная F-статистика имеет распределение Фишера с числом степеней свободы $v_1 = v_2 = n - m - 1$.
5. Если $F_{набл} = \frac{S_3}{S_1} > F_{кр}$ (где $F_{кр} = F_{\alpha, v_1, v_2}$, определяется из приложения 2, α - выбранный уровень значимости), то гипотеза об отсутствии гетероскедастичности отклоняется.

Этот же тест может использоваться при предположении об обратной пропорциональности между σ_i и значениями объясняющей переменной. При этом статистика Фишера имеет вид:
$$F = \frac{S_1}{S_3}.$$

Для множественной регрессии данный тест обычно проводится для той объясняющей переменной, которая в наибольшей степени связана с σ_i. При этом k должно быть больше, чем $(m+1)$. Если нет уверенности относительно выбора переменной X_j, то данный тест может осуществляться для каждой из объясняющих переменных.

Пример.

Исследуем зависимость между доходом (X) домохозяйства и его расходом (Y) на продукты питания. Выборочные данные по 40 домохозяйствам представлены в таблице:

X	Y
25,5	14,5
26,5	11,3
27,2	14,7
29,6	10,2
35,7	13,5
38,6	9,9
39,0	12,4
39,3	8,6
40,0	10,3
41,9	13,9

X	Y
42,5	14,9
44,2	11,6
44,8	21,5
45,5	10,8
45,5	13,8
48,3	16,0
49,5	18,2
52,3	19,1
55,7	16,3
59,0	17,5

X	Y
61,0	10,9
61,7	16,1
62,5	10,5
64,7	10,6
69,7	29,0
71,2	8,2
73,8	14,3
74,7	21,8
75,8	26,1
76,9	20,0

X	Y
79,2	19,8
81,5	21,2
82,4	29,0
82,8	17,3
83,0	23,5
85,9	22,0
86,4	18,3
86,9	13,7
88,3	14,5
89,0	27,3

Построить эмпирическое уравнение регрессии и провести анализ модели на наличие гетероскедастичности.

Решение:

Определим коэффициенты эмпирического уравнения регрессии: $b_0 = 7{,}04$, $b_1 = 0{,}16$. Следовательно, уравнение имеет вид: $\hat{y}_i = 7{,}04 + 0{,}16 \cdot x_i$.

Определим отклонения e_i (где $e_i = y_i - \hat{y}_i$), e_i^2, ранги X и e_i. Рассчитанные величины представим в таблице:

X	Y	$\hat{Y}$	e_i	e_i^2	Ранг X	Ранг e_i (абсол.вел.)	d_i	d_i^2
25,5	14,5	11,120	3,380	11,4244	1	25	-33	1089
26,5	11,3	11,280	0,020	0,0004	2	1	-17	289
27,2	14,7	11,392	3,308	10,9429	3	23	-30	900
29,6	10,2	11,776	-1,576	2,4838	4	15	-11	121
35,7	13,5	12,752	0,748	0,5595	5	7	-19	361
38,6	9,9	13,216	-3,316	10,9959	6	24	-4	16
39,0	12,4	13,280	-0,880	0,7744	7	9	-9	81
39,3	8,6	13,328	-4,728	22,3540	8	29	1	1
40,0	10,3	13,440	-3,140	9,8596	9	20	-2	4
41,9	13,9	13,744	0,156	0,0243	10	3	-11	121
42,5	14,9	13,840	1,060	1,1236	11	11	-15	225
44,2	11,6	14,112	-2,512	6,3101	12	16	-2	4
44,8	21,5	14,208	7,292	53,1733	13	37	-25	625
45,5	10,8	14,320	-3,520	12,3904	14	26	5	25
45,5	13,8	14,320	-0,520	0,2704	15	5	-4	16
48,3	16,0	14,768	1,232	1,5178	16	14	-13	169
49,5	18,2	14,960	3,240	10,4976	17	22	-15	225
52,3	19,1	15,408	3,692	13,6309	18	27	-17	289
55,7	16,3	15,952	0,348	0,1211	19	4	-3	9
59,0	17,5	16,480	1,020	1,0404	20	10	-5	25
61,0	10,9	16,800	-5,900	34,8100	21	30	15	225
61,7	16,1	16,912	-0,812	0,6593	22	8	5	25
62,5	10,5	17,040	-6,540	42,7716	23	32	18	324
64,7	10,6	17,392	-6,792	46,1313	24	34	21	441
69,7	29,0	18,192	10,808	116,8129	25	40	-15	225
71,2	8,2	18,432	-10,232	104,6938	26	39	25	625
73,8	14,3	18,848	-4,548	20,6843	27	28	19	361
74,7	21,8	18,992	2,808	7,8849	28	18	-2	4
75,8	26,1	19,168	6,932	48,0526	29	35	-8	64
76,9	20,0	19,344	0,656	0,4303	30	6	7	49
79,2	19,8	19,712	0,088	0,0077	31	2	11	121
81,5	21,2	20,080	1,120	1,2544	32	12	5	25
82,4	29,0	20,224	8,776	77,0182	33	38	-6	36
82,8	17,3	20,288	-2,988	8,9281	34	19	22	484
83,0	23,5	20,320	3,180	10,1124	35	21	4	16
85,9	22,0	20,784	1,216	1,4787	36	13	8	64
86,4	18,3	20,864	-2,564	6,5741	37	17	24	576
86,9	13,7	20,944	-7,244	52,4755	38	36	36	1296
88,3	14,5	21,168	-6,668	44,4622	39	33	35	1225
89,0	27,3	21,280	6,020	36,2404	40	31	4	16

Применим для обнаружения гетероскедастичности тест ранговой корреляции Спирмена. Для этого рассчитаем коэффициент ранговой корреляции:

$$r_{x,e} = 1 - 6 \cdot \frac{\sum d_i^2}{n(n^2 - 1)} = 1 - 6 \cdot \frac{7595}{40(1600 - 1)} = 0.2875.$$

Рассчитаем t-статистику:

$$t = \frac{r_{x,e}\sqrt{(n-2)}}{\sqrt{1 - r_{x,e}^2}} = \frac{0,2875\sqrt{(40-2)}}{\sqrt{1-(0,2875)^2}} = 1,8504.$$

Из приложения 1 определим критическое значение t-статистики для числа степеней свободы $v = n - 2 = 38$ и уровня значимости $\alpha = 0,10$: $t_{кр} = 1,303$. Так как рассчитанное значение t-статистики превышает критическое, определенное по приложению 1, то гипотеза об отсутствии гетероскедастичности отклоняется с уровнем значимости $\alpha = 0,10$.

Проверим гипотезу об отсутствии гетероскедастичности с помощью теста Голдфелда-Квандта. Для этого разобьем ряд на три подвыборки размерности 14, 12, 14.

Определим дисперсии отклонений для первой и третьей подвыборок:

$$S_1 = \sum_{i=1}^{14} e_i^2 = 142,4165 \text{ и } S_3 = \sum_{i=27}^{40} e_i^2 = 315,6039.$$

Определим значение F-статистики. $F = \dfrac{S_3}{S_1} = \dfrac{315,6039}{142,4165} = 2,2161.$

Из приложения 2 определим критическое значение F-статистики для числа степеней свободы $v_1 = v_2 = n - m - 1 = 40 - 1 - 1 = 38$ и уровня значимости $\alpha = 0,10$: $F_{кр} = 1,51$. Так как рассчитанное значение F-статистики превышает критическое, определенное по приложению 2, то гипотеза об отсутствии гетероскедастичности отклоняется с уровнем значимости $\alpha = 0,10$.

Следовательно, по всем тестам гетероскедастичность в данной модели присутствует.

Задачи

1. По приведенным в таблице данным оценить при помощи МНК регрессионную зависимость расходов на образование от валового внутреннего продукта. Проанализировать полученную модель на наличие гетероскедастичности графически, при помощи теста ранговой корреляции Спирмена или теста Голдфелда-Квандта, сделать выводы.

Страна	гос.расходы на образование	ВВП
Люксембург	0,34	5,67
Уругвай	0,22	10,13
Сингапур	0,32	11,34
Ирландия	1,23	18,88
Израиль	1,81	20,94
Венгрия	1,02	22,16
Новая Зеландия	1,27	23,83
Португалия	1,07	24,67
Гонконг	0,67	27,56
Чили	1,25	27,57
Греция	0,75	40,15
Финляндия	2,80	51,62
Норвегия	4,90	57,71
Югославия	3,50	63,03
Дания	4,45	66,32
Турция	1,60	66,97
Австрия	4,26	76,88
Швейцария	5,31	101,65

Саудовская Аравия	6,40	115,97
Бельгия	7,15	119,49
Швеция	11,22	124,15
Австралия	8,66	140,98
Аргентина	5,56	153,85
Нидерланды	13,41	169,38
Мексика	5,46	186,33
Испания	4,79	211,78
Бразилия	8,92	249,72
Канада	18,90	261,41
Италия	15,95	395,52
Великобритания	29,90	534,97
Франция	33,59	655,29
ФРГ	38,62	815,00
Япония	61,61	1040,45
США	181,30	2586,40

2. Торговое предприятие имеет несколько филиалов. Предполагается, что переменная y – годовой товарооборот линейно зависит от переменной x_1 – размер торговой площади. Проверить модель на наличие гетероскедастичности. Данные приведены в следующей таблице:

y, млн руб	2,93	5,27	6,85	7,01	7,02	8,35	4,33	5,77	7,68	3,16	1,52	3,15
x_1, тыс м^2	0,31	0,98	1,21	1,29	1,12	1,49	0,78	0,94	1,29	0,48	0,24	0,55

7. Автокорреляция

Цель занятия: Для построенной модели регрессии научиться определять выполнимость третьей предпосылки МНК.

Методические указания.

Автокорреляция (последовательная корреляция) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные ряды). Автокорреляция остатков (отклонений) обычно встречается в регрессионном анализе при использовании данных временных рядов и очень редко при использовании перекрестных данных.

В экономических задачах значительно чаще встречается положительная автокорреляция, нежели отрицательная автокорреляция. В большинстве случаев *положительная автокорреляция* вызывается направленным постоянным воздействием некоторых неучтенных в модели факторов..

Отрицательная автокорреляция фактически означает, что за положительным отклонением следует отрицательное и наоборот.

Обнаружение автокорреляции. Критерий Дарбина-Уотсона.

Рассмотрим популярную в регрессионном анализе статистику Дарбина-Уотсона. При статистическом анализе уравнения регрессии на начальном этапе часто проверяют выполнимость одной предпосылки: условия статистической независимости отклонений между собой (отсутствие автокорреляции). При этом проверяется некоррелированность соседних величин $e_i, i = 1, 2, \ldots, n$. Для анализа коррелированности отклонений используют статистику Дарбина-Уотсона:

$$DW = \frac{\sum (e_i - e_{i-1})^2}{\sum e_i^2},$$

причём $0 \le DW \le 4$.

Критические значения d_1 и d_2 определяются на основе специальных таблиц (приложение 3) для требуемого уровня значимости α, числа наблюдений n и количества объясняющих переменных m.

Выводы об отсутствии автокорреляции остатков осуществляются по следующей схеме:

Пример.

Рассчитаем статистику Дарбина-Уотсона для примера из раздела 4.

Решение:

$$DW = \frac{41{,}8734}{24{,}2406} = 1{,}7274 .$$

По приложению 3 определим критические точки для уровня значимости 0,05 и числа наблюдений 11: $d_1 = 0{,}658; d_2 = 1{,}604$. Таким образом, $1{,}604 < DW < 2{,}396$, т.е. ($d_2 < DW < 4 - d_2$), следовательно, имеются основания считать, что автокорреляция отсутствует. Это является одним из подтверждений высокого качества модели.

Задачи.

1. По таблице индивидуальных заданий определить зависимость личного дохода от времени (линейная зависимость, МНК). Проверить модель на наличие автокорреляции аналитически (критерий Дарбина-Уотсона). Сделать выводы.

2. В условиях задачи из раздела 6 проверить модель на наличие автокорреляции.

8. Фиктивные переменные в регрессионных моделях

Цель занятия: Научиться строить регрессионные модели с количественными и качественными (фиктивными) переменными.

Методические указания.

В моделях влияние качественного фактора выражается в виде фиктивной (искусственной) переменной, которая отражает два противоположных состояния качественного фактора:

$$D = \begin{cases} 0, & \text{фактор не действует} \\ 1, & \text{фактор действует} \end{cases}$$

Переменная D называется *фиктивной (искусственной, двоичной) переменной (индикатором)*.

Пример.

Рассмотрим зависимость между весом новорожденного Y (в граммах), X - количеством сигарет, выкуриваемых в день будущей матерью во время беременности и фиктивной переменной D, которая отражает факт того, является ребенок первенцем или нет. Пусть $D = 0$, если ребенок – первенец и $D = 1$, если ребенок не первенец. Рассмотри выборку из 20 значений:

наблюдение	Y	X	D
1	3520	10	1
2	3460	19	1
3	3000	16	1
4	3320	26	1
5	3540	4	1
6	3310	14	1

наблюдение	Y	X	D
11	3210	29	1
12	3290	15	1
13	3190	3	0
14	3060	12	0
15	3270	17	0
16	3170	14	0

7	3360	21	1	17	3230	18	0
8	3650	10	1	18	3700	11	0
9	3150	22	1	19	3300	14	0
10	3440	8	1	20	3460	9	0

Данная модель содержит одну количественную и одну качественную переменные. В общем виде запишем ее следующим образом: $Y = b_0 + b_1 X + b_2 D$. Коэффициенты b_0, b_1, b_2 определяются как коэффициенты множественной регрессии (раздел 3). Вспомогательная таблица для расчета коэффициентов имеет вид:

№	$(y_i - \bar{y})$	$(x_i - \bar{x})$	$(d_i - \bar{d})$	$(x_i - \bar{x})^2$	$(d_i - \bar{d})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(d_i - \bar{d}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (d_i - \bar{d})$
1	188,5	-4,6	188,5	21,16	0,16	-867,1	75,4	-1,84
2	128,5	4,4	128,5	19,36	0,16	565,4	51,4	1,76
3	-331,5	1,4	-331,5	1,96	0,16	-464,1	-132,6	0,56
4	-11,5	11,4	-11,5	129,96	0,16	-131,1	-4,6	4,56
5	208,5	-10,6	208,5	112,36	0,16	-2210,1	83,4	-4,24
6	-21,5	-0,6	-21,5	0,36	0,16	12,9	-8,6	-0,24
7	28,5	6,4	28,5	40,96	0,16	182,4	11,4	2,56
8	318,5	-4,6	318,5	21,16	0,16	-1465,1	127,4	-1,84
9	-181,5	7,4	-181,5	54,76	0,16	-1343,1	-72,6	2,96
10	108,5	-6,6	108,5	43,56	0,16	-716,1	43,4	-2,64
11	-121,5	14,4	-121,5	207,36	0,16	-1749,6	-48,6	5,76
12	-41,5	0,4	-41,5	0,16	0,16	-16,6	-16,6	0,16
13	-141,5	-11,6	-141,5	134,56	0,36	1641,4	84,9	6,96
14	-271,5	-2,6	-271,5	6,76	0,36	705,9	162,9	1,56
15	-61,5	2,4	-61,5	5,76	0,36	-147,6	36,9	-1,44
16	-161,5	-0,6	-161,5	0,36	0,36	96,9	96,9	0,36
17	-101,5	3,4	-101,5	11,56	0,36	-345,1	60,9	-2,04
18	368,5	-3,6	368,5	12,96	0,36	-1326,6	-221,1	2,16
19	-31,5	-0,6	-31,5	0,36	0,36	18,9	18,9	0,36
20	128,5	-5,6	128,5	31,36	0,36	-719,6	-77,1	3,36
$\sum$				856,8	4,8	-8278	272	18,8

При этом: $\bar{y} = 3331{,}5$, $\bar{x} = 14{,}6$, $\bar{d} = 0{,}6$.

$$b_0 = 3331{,}5 + 11{,}93 \cdot 14{,}6 - 103{,}39 \cdot 0{,}6 = 3443{,}64$$

$$b_1 = \frac{(-8278) \cdot 4{,}8 - 272 \cdot 18{,}8}{856{,}8 \cdot 4{,}8 - (18{,}8)^2} = -11{,}93$$

$$b_2 = \frac{272 \cdot 856{,}8 - (-8278) \cdot 18{,}8}{856{,}8 \cdot 4{,}8 - (18{,}8)^2} = 103{,}39$$

Таким образом, уравнение регрессии с учетом рассчитанных коэффициентов примет вид: $\hat{Y} = 3443{,}64 - 11{,}93X + 103{,}39D$. Затем рассчитывается статистическая значимость коэффициентов. Рассчитанное значение t - статистики для коэффициента b_2 при фиктивной переменной D составляет $t = 1{,}23$.

Из приложения 1 определим для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $v = n - m - 1 = 20 - 2 - 1 = 17$ критическое значение t - статистики: $t_{кр} = t_{\frac{\alpha}{2}, n-m-1} = t_{0{,}025{,}17} = 2{,}110$.

Так как $t < t_{кр}$, то коэффициент b_2 при фиктивной переменной D является статистически незначимым с уровнем значимости 0,05.

Однако можно предположить, что это объясняется малым размером выборки (20 значений). Если рассмотреть большую выборку, то обнаружится статистическая значимость данного коэффициента.

Задачи

1. На основе представленных в таблице данных о доходах Y, поле (мужчина-женщина) D_1 и наличии детей (D_2) необходимо построить модель с фиктивными переменными вида: $\widehat{Y} = \alpha + \gamma_1 \cdot D_1 + \gamma_2 \cdot D_2$. Дайте полную интерпретацию полученной регрессии. Проверить статистическую значимость коэффициентов. Сделать выводы.

Y	91,8	38,7	34,1	30,8	50	34,3	42,6	63,5	19,9	58,9	72,5
дети	есть	есть	нет	есть	нет	есть	нет	есть	нет	нет	нет
пол	м	ж	м	ж	ж	ж	м	м	ж	м	ж

Y	30	93,7	17,8	78,8	39,7	93,9	86,2	26	37	45,8
дети	нет	нет	нет	нет	нет	есть	есть	нет	есть	есть
пол	ж	ж	ж	м	м	ж	м	ж	м	м

2. Оценка регрессионной зависимости объема жилищного строительства (в течение периода 1977-1982гг., в млрд.долл., в ценах 1972г.) от временного тренда и сезонных фиктивных переменных, определенных для II, III и IV кварталов, дала следующий результат (в скобках указаны стандартные ошибки коэффициентов регрессии):

$$\hat{y} = 13{,}69 + 3{,}02D_2 + 4{,}08D_3 + 3{,}00D_4 - 0{,}31t,$$

(0,65) (0,73) (0,73) (0,73) (0,04)

Дайте полную интерпретацию регрессии и проверьте статистическую значимость коэффициентов.

9. Динамические модели. Лаги в экономических моделях

Цель занятия: для анализа экономических явлений научиться использовать методы исследования и анализа временных рядов.

Методические указания.

При анализе многих экономических показателей (особенно в макроэкономике) часто используются ежегодные, ежеквартальные, ежемесячные, ежедневные данные. Для рационального анализа необходимо систематизировать моменты получения соответствующих статистических данных.

В этом случае следует упорядочить данные по времени их получения и построить так называемые *временные ряды*.

Пусть исследуется показатель Y. Его значение в текущий момент (период) времени t обозначают y_t; значения Y в последующие моменты обозначаются $y_{t+1}, y_{t+2}, \ldots, y_{t+k}, \ldots$; значения Y в предыдущие моменты времени обозначаются $y_{t-1}, y_{t-2}, \ldots, y_{t-k}, \ldots$. В качестве объясняющих переменных используются не только текущие значения переменных, но и некоторые предыдущие по времени значения, а также само время T. Модели такого типа называются *динамическими*.

В свою очередь переменные, влияние которых характеризуется определенным запаздыванием, называются *лаговыми переменными*.

Обычно динамические модели подразделяются на два класса:

1. *модели с лагами* (модели с распределенными лагами) – содержат в качестве лаговых переменных лишь независимые (объясняющие) переменные. Примером является модель:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_k x_{t-k} + \varepsilon_t,$$

2. *авторегрессионные модели* – модели, уравнения которых в качестве лаговых объясняющих переменных включают значения зависимых переменных.

$$y_t = \alpha + \beta_0 x_t + \gamma y_{t-1} + \varepsilon_t.$$

Пример.

В таблице приведены данные по располагаемому доходу домохозяйств (X) и затратами домохозяйств на розничные покупки (Y) за 22 года:

t	Y_t	X_t	t	Y_t	X_t
1	5,49	9,098	12	5,905	11,305
2	5,54	9,137	13	6,125	11,43
3	5,305	9,095	14	6,185	11,45
4	5,505	9,28	15	6,225	11,697
5	5,42	9,23	16	6,495	11,87
6	5,32	9,348	17	6,72	12,018
7	5,54	9,525	18	6,92	12,525
8	5,69	9,755	19	6,47	12,055
9	5,87	10,28	20	6,395	12,088
10	6,157	10,665	21	6,555	12,215
11	6,342	11,02	22	6,755	12,495

Необходимо оценить уравнение регрессии вида $\hat{y}_t = b + b_1 x_t + \gamma y_{t-1}$ (принять $y_0 = 5{,}4$), проверить значимость коэффициентов b_0, b_1, γ, оценить качество построенной модели при помощи коэффициента детерминации.

Решение:

Для расчета коэффициентов составим вспомогательную таблицу (при этом рассчитанные средние значения равны $\bar{y}_t = 6{,}04223$, $\bar{y}_{t-1} = 5{,}98067$, $\bar{x} = 10{,}79914$):

t	Y_t	X_t	Y_{t-1}	$(y_t-\bar{y}_t)$	$(x_t-\bar{x})$	$(y_{t-1}-\bar{y}_{t-1})$	$(y_t-\bar{y}_t)^2$	$(x_t-\bar{x})^2$	$(y_{t-1}-\bar{y}_{t-1})^2$	$(x_t-\bar{x})\cdot(y_t-\bar{y}_t)$	$(y_{t-1}-\bar{y}_{t-1})\cdot(y_i-\bar{y}_t)$	$(x_t-\bar{x})\cdot(y_{t-1}-\bar{y}_{t-1})$
1	5,49	9,098	5,4	-0,5522	-1,7011	-0,581	0,3050	2,8939	0,3371	0,9394	0,3206	0,9877
2	5,54	9,137	5,49	-0,5022	-1,6621	-0,491	0,2522	2,7627	0,2407	0,8348	0,2464	0,8155
3	5,305	9,095	5,54	-0,7372	-1,7041	-0,441	0,5435	2,9041	0,1942	1,2563	0,3248	0,7509
4	5,505	9,28	5,305	-0,5372	-1,5191	-0,676	0,2886	2,3078	0,4565	0,8161	0,3630	1,0264
5	5,42	9,23	5,505	-0,6222	-1,5691	-0,476	0,3872	2,4622	0,2262	0,9764	0,2960	0,7463
6	5,32	9,348	5,42	-0,7222	-1,4511	-0,561	0,5216	2,1058	0,3143	1,0481	0,4049	0,8136
7	5,54	9,525	5,32	-0,5022	-1,2741	-0,661	0,2522	1,6234	0,4364	0,6399	0,3318	0,8417
8	5,69	9,755	5,54	-0,3522	-1,0441	-0,441	0,1241	1,0902	0,1942	0,3678	0,1552	0,4601
9	5,87	10,28	5,69	-0,1722	-0,5191	-0,291	0,0297	0,2695	0,0845	0,0894	0,0501	0,1509
10	6,157	10,665	5,87	0,1148	-0,1341	-0,111	0,0132	0,0180	0,0122	-0,0154	-0,0127	0,0148
11	6,342	11,02	6,157	0,2998	0,2209	0,176	0,0899	0,0488	0,0311	0,0662	0,0529	0,0390
12	5,905	11,305	6,342	-0,1372	0,5059	0,361	0,0188	0,2559	0,1306	-0,0694	-0,0496	0,1828
13	6,125	11,43	5,905	0,0828	0,6309	-0,076	0,0069	0,3980	0,0057	0,0522	-0,0063	-0,0477
14	6,185	11,45	6,125	0,1428	0,6509	0,144	0,0204	0,4236	0,0208	0,0929	0,0206	0,0940
15	6,225	11,697	6,185	0,1828	0,8979	0,204	0,0334	0,8062	0,0418	0,1641	0,0374	0,1835
16	6,495	11,87	6,225	0,4528	1,0709	0,244	0,2050	1,1467	0,0597	0,4849	0,1106	0,2617
17	6,72	12,018	6,495	0,6778	1,2189	0,514	0,4594	1,4856	0,2646	0,8261	0,3486	0,6269
18	6,92	12,525	6,72	0,8778	1,7259	0,739	0,7705	2,9786	0,5467	1,5149	0,6490	1,2760
19	6,47	12,055	6,92	0,4278	1,2559	0,939	0,1830	1,5772	0,8824	0,5372	0,4018	1,1797
20	6,395	12,088	6,47	0,3528	1,2889	0,489	0,1244	1,6612	0,2395	0,4547	0,1726	0,6307
21	6,555	12,215	6,395	0,5128	1,4159	0,414	0,2629	2,0047	0,1717	0,7260	0,2125	0,5867
22	6,755	12,495	6,555	0,7128	1,6959	0,574	0,5080	2,8760	0,3299	1,2088	0,4094	0,9740
Σ							5,3998	34,1000	5,2208	13,0114	4,8397	12,5953

Имеем:

$$b_1 = \frac{(13{,}0114)\cdot 5{,}2208 - 4{,}8397\cdot 12{,}5953}{34{,}100\cdot 5{,}2208 - (12{,}5953)^2} = \frac{6{,}9724}{19{,}3877} = 0{,}36,$$

$$\gamma = \frac{4{,}8397\cdot 34{,}10 - 13{,}0114\cdot 12{,}5953}{34{,}100\cdot 5{,}2208 - (12{,}5953)^2} = \frac{1{,}1513}{19{,}3877} = 0{,}06,$$

$$b_0 = 6{,}04223 - 0{,}36\cdot 10{,}79914 - 0{,}06\cdot 5{,}98067 = 1{,}80.$$

Таким образом, уравнение регрессии с учетом рассчитанных коэффициентов примет вид: $\hat{y}_t = 1{,}8 + 0{,}36x_t + 0{,}06y_{t-1}$.

Для определения статистической значимости коэффициентов и оценки качества уравнения регрессии составим следующую вспомогательную таблицу:

t	Y_t	$\hat{Y}_t$	e_i	e_i^2
1	5,49	5,3960	-0,0940	0,008843
2	5,54	5,4153	-0,1247	0,015542
3	5,305	5,4032	0,0982	0,009642
4	5,505	5,4558	-0,0492	0,002422
5	5,42	5,4497	0,0297	0,00088
6	5,32	5,4871	0,1671	0,02791
7	5,54	5,5448	0,0048	2,29E-05
8	5,69	5,6406	-0,0494	0,002444
9	5,87	5,8383	-0,0317	0,001006
10	6,157	5,9874	-0,1696	0,028757
11	6,342	6,1321	-0,2099	0,044047
12	5,905	6,2456	0,3406	0,11601
13	6,125	6,2646	0,1396	0,019496
14	6,185	6,2849	0,0999	0,009975
15	6,225	6,3773	0,1523	0,023186
16	6,495	6,4419	-0,0531	0,002824
17	6,72	6,5111	-0,2089	0,043635
18	6,92	6,7068	-0,2132	0,045453
19	6,47	6,5496	0,0796	0,006342
20	6,395	6,5348	0,1398	0,019545
21	6,555	6,5760	0,0210	0,000442
22	6,755	6,6862	-0,0688	0,00473
сумма			≈ 0	0,433155

Рассчитаем необъясненную дисперсию и стандартные отклонения случайных величин:

$$S^2 = \frac{\sum e_i^2}{n-m-1} = \frac{0{,}43155}{22-2-1} = 0{,}0227 .$$

$$S_{b0}^2 = \left[\frac{1}{22} + \frac{(10{,}79914)^2 \cdot 5{,}2208 + (5{,}98067)^2 \cdot 34{,}1 - 2 \cdot 10{,}79914 \cdot 5{,}98067 \cdot 12{,}5953}{34{,}1 \cdot 5{,}2208 - (12{,}5953)^2}\right] \cdot 0{,}0227,\ S_{b0} = \sqrt{S_{b0}^2} = \sqrt{0{,}238} = 0{,}4879 ,$$

$$S_{b1}^2 = \frac{5{,}2208}{34{,}1 \cdot 5{,}2208 - (12{,}5953)^2} \cdot 0{,}0227 = 0{,}0061 ,\ S_{b1} = \sqrt{S_{b1}^2} = \sqrt{0{,}0061} = 0{,}0781 ,$$

$$S_{\gamma}^2 = \frac{34{,}10}{34{,}1 \cdot 5{,}2208 - (12{,}5953)^2} \cdot 0{,}0227 = 0{,}0399 ,\ S_{\gamma} = \sqrt{S_{\gamma}^2} = \sqrt{0{,}0399} = 0{,}1997 .$$

Определим значение t-статистики для каждого из коэффициентов:

$$t_{b0} = \frac{1{,}8}{0{,}4879} = 3{,}689 ,\ t_{b1} = \frac{0{,}36}{0{,}0781} = 4{,}609 ,\ t_{\gamma} = \frac{0{,}06}{0{,}1997} = 0{,}300 .$$

Критическое значение определим из приложения 1 для уровня значимости 0,1 и числа степеней свободы $\nu = 22\text{-}2\text{-}1\text{=}19$: $t_{кр} = t_{\frac{0{,}1}{2},19} = 1{,}729$.

Очевидно, что коэффициенты b_0, b_1 являются статистически значимыми, а коэффициент γ является статистически незначимым с уровнем значимости 0,1.

Определим для рассчитанного уравнения коэффициент детерминации $R^2 = 1 - \dfrac{0,433155}{5,3998} = 0,92$. Столь высокое значение коэффициента детерминации свидетельствует о высоком качестве модели. Поэтому не будем удалять переменную y_{t-1} из уравнения.

Задачи

1. Оценена следующая авторегрессионная модель:

$y_t = 3,5 + 0,5x_t + 0,9y_{t-1}$, и $R^2 = 0,97$, $DW = 2,15$,

S (0,5) (0,06)

Проанализировать качество модели.

2. Анализируется среднедушевой расход на развлечения людей до 25 лет. По 35 годовым данным по МНК построено следующее уравнение регрессии:

$y_t = 43,5 + 0,251x_t + 0,545y_{t-1}$, и $DW = 1,9$,

S (0,105) (0,135)

где y_t – среднедушевой расход на развлечения молодых людей в момент времени t, x_t – среднедушевой располагаемый доход в момент времени t.

1) Построить 95% -й доверительный интервал для теоретического коэффициента регрессии при переменной x_t, пояснить экономический смысл этого коэффициента;

2) Проверить гипотезу об отсутствии автокорреляции остатков.

Задания для самостоятельной работы

Задание 1.

Изучается зависимость доходности акций предприятия y (%) от темпа роста валового внутреннего продукта x (%). Полученные результаты отражены в таблице:

год	x	y
2000	5,5	14,1
2001	6,2	18,7
2002	7,7	23,1
2003	7,2	18,1
2004	4,8	8,7

Определить, есть ли между переменными линейная зависимость.

Задание 2.

По данным n=12 угольных шахт провести регрессионный анализ зависимости полной себестоимости добычи 1т.угля y (тыс. руб.) от средней суточный добычи угля на шахте x_1 и удельного веса комбайновой проходки выработки x_2 (%).

№ п/п	y	x_1	x_2
1	12,2	4795	69
2	7,6	6062	82
3	10,0	6571	87
4	49,9	4249	92
5	15,7	9540	23
6	14,0	3488	31
7	12,7	4888	55
8	10,5	6237	81
9	15,1	2997	65
10	10,6	2990	98
11	15,2	1748	100
12	17,2	2128	69

1) проверить статистическую значимость коэффициентов регрессии с уровнем значимости 5%;

2) определить доверительные интервалы для коэффициентов регрессии с уровнем значимости 1%;

4) проверить качество уравнения регрессии и статистическую значимость коэффициента детерминации (уровень значимости 5%, 10%).

Задание 3.

Данные о прибыли предприятия y (млн долл.) и расходах на рекламу x за 9 лет представлены в таблице.

y	5	7	12	16	23	21	19	18	16
x	0,8	1,1	1,8	2,5	4,1	5,5	7,3	8,1	8,9

Требуется:

1) построить корреляционное поле и выдвинуть гипотезу о форме зависимости между рассматриваемыми показателями;

2) оценить по МНК коэффициенты линейного уравнения регрессии $\hat{y} = \beta_0 + \beta_1 x$ и сделать вывод о качестве уравнения регрессии;

3) оценить по МНК коэффициенты параболического уравнения регрессии $\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$ и сделать вывод о качестве уравнения регрессии;

4) построить обратную регрессионную модель и оценить её коэффициенты;

5)обосновать выбор лучшей модели.

Задание 4.

По данным за 15 лет построены два уравнения регрессии:

$$\hat{y} = 3{,}45 - 0{,}55x\,, \quad R^2 = 0{,}68$$

$$t = (20{,}5)\ (-4{,}3)$$

$$\ln \hat{y} = 0{,}85 - 0{,}25x\,, \quad R^2 = 0{,}78$$

$$t = (44{,}9)\ (-5{,}3)$$

где y - ежедневное среднедушевое потребление кофе (в чашках по 100г); x – среднегодовая цена кофе (в руб./кг).

Требуется:

1) проинтерпретировать коэффициенты каждой из модели;

2) обосновать выбор лучшей модели;

3) ответить на вопрос, можно ли о качестве модели судить по коэффициенту детерминации.

Задание 5.

Исследуется эффективность лекарств y в зависимости от x (возраста пациента). При этом сравнивается эффективность лекарств a и b .

Лекарство	y	x	D
a	54	69	0
b	30	48	1
a	58	73	0
b	66	64	1
b	67	60	1
a	64	62	0
a	67	70	0
a	33	52	0

a	33	63	0
b	42	48	1
b	33	46	1
a	28	55	0
b	30	40	1
b	23	41	1
a	21	55	0
b	43	45	1
a	38	58	0
b	43	58	1
a	43	64	0
b	45	55	1
b	48	57	1
a	48	63	0
a	53	60	0
b	58	62	1

Фиктивная переменная $D=0$, если лекарство *a*; $D=1$, если лекарство *b*. Оценить коэффициенты регрессии: $y=\beta_0+\beta_1 x+\gamma D$.

Требуется:

1) проверить статистическую значимость коэффициентов регрессии с уровнем значимости 5%; решить вопрос о целесообразности введения фиктивной переменной.

2) определить доверительные интервалы для коэффициентов регрессии с уровнем значимости 1%;

Задание №6.

Необходимо исследовать зависимость между результатами письменных вступительных и курсовых (на первом курсе) экзаменов по математике. Получены следующие данные о числе решенных задач на вступительных экзаменах x (задание –10 задач) и курсовых экзаменах y (задание – 7 задач) 12-ти студентов, а также распределение этих студентов по фактору “пол”:

Номер студента	Число решенных задач		Пол студента
i	x_i	y_i	
1	10	6	м
2	6	4	ж
3	8	4	м
4	8	5	ж
5	6	4	ж
6	7	7	м
7	6	3	ж
8	7	4	м
9	9	7	м
10	6	3	ж
11	5	2	м
12	7	3	ж

Построить линейную регрессионную модель с использованием фиктивной переменной по фактору “пол”. Можно ли считать, что эта модель одна и та же для юношей и девушек?

Задание 7.

Данные по Великобритании за 20 лет потребления цыплят (y), среднедушевом доходе (x_1), стоимости 1 фунта цыплят (x_2), стоимости 1 фунта свинины (x_3) и стоимости 1 фунта говядины (x_4), представлены в таблице:

t	y	x_1	x_2	x_3	x_4
1	30,8	459,7	39,5	55,3	79,2
2	31,2	492,9	37,3	54,7	77,4
3	33,3	528,6	38,1	63,7	80,2
4	35,6	560,3	39,3	69,8	80,4
5	36,4	624,6	37,8	65,9	83,9
6	36,7	666,4	38,4	64,5	85,5
7	38,4	717,8	40,1	70,0	93,7
8	40,4	768,2	38,6	73,2	106,1
9	40,3	843,3	39,8	67,8	104,8
10	41,8	911,6	39,7	79,1	114,0
11	40,4	931,1	52,1	95,4	124,1
12	40,7	1021,5	48,9	94,2	127,6
13	40,1	1165,9	58,3	123,5	142,9
14	42,7	1349,6	57,9	129,9	143,6
15	44,1	1449,4	56,5	117,6	139,2
16	46,7	1575,5	63,7	130,9	165,5
17	50,6	1759,1	61,6	129,8	203,3
18	50,1	1994,2	58,9	128,0	219,6
19	51,7	2258,1	66,4	141,0	221,6
20	52,9	2478,7	70,4	168,2	232,6

Требуется построить и сравнить уравнения регрессии вида

а) $\hat{y} = \beta_0 x_2^{\beta_2}$ - функция спроса;

б) $\hat{y} = \beta_0 x_1^{\beta_1}$ - функция потребления;

в) $\hat{y} = \beta_0 x_2^{\beta_2} x_1^{\beta_1}$ - функция спроса и потребления;

г) $\hat{y} = \beta_0 x_2^{\beta_2} x_3^{\beta_3} x_4^{\beta_4}$ - функция спроса с учетом цены на товарозаменители.

Задание 8.

Выявить на уровне значимости 0,05 наличие автокорреляции возмущений для временного ряда y_t по данным таблицы, если получено уравнение тренда $\hat{y}_t = 181{,}32 + 25{,}679t$.

t	1	2	3	4	5	6	7	8
y_t	207,0	232,7	258,4	284,0	309,7	335,4	361,1	386,8

Приложения

Приложение 1

Распределение Стьюдента (t-распределение)

		уровень значимости						
		0.40	0.25	0.10	0.05	0.025	0.01	0.005
число степеней свободы	1	0,325	1,000	3,078	6,314	12,706	31,821	63,657
	2	0,289	0,816	1,886	2,920	4,303	6,965	9,925
	3	0,277	0,765	1,638	2,353	3,182	4,541	5,841
	4	0,271	0,741	1,533	2,132	2,776	3,747	4,604
	5	0,267	0,727	1,476	2,015	2,571	3,365	4,032
	6	0,265	0,718	1,440	1,943	2,447	3,143	3,707
	7	0,263	0,711	1,415	1,895	2,365	2,998	3,499
	8	0,262	0,706	1,397	1,860	2,306	2,896	3,355
	9	0,261	0,703	1,383	1,833	2,262	2,821	3,250
	10	0,260	0,700	1,372	1,812	2,228	2,764	3,169
	11	0,260	0,697	1,363	1,796	2,201	2,718	3,106
	12	0,259	0,695	1,356	1,782	2,179	2,681	3,055
	13	0,259	0,694	1,350	1,771	2,160	2,650	3,012
	14	0,258	0,692	1,345	1,761	2,145	2,624	2,977
	15	0,258	0,691	1,341	1,753	2,131	2,602	2,947
	16	0,258	0,690	1,337	1,746	2,120	2,583	2,921
	17	0,257	0,689	1,333	1,740	2,110	2,567	2,898
	18	0,257	0,688	1,330	1,734	2,101	2,552	2,878
	19	0,257	0,688	1,328	1,729	2,093	2,539	2,861
	20	0,257	0,687	1,325	1,725	2,086	2,528	2,845
	21	0,257	0,686	1,323	1,721	2,080	2,518	2,831
	22	0,256	0,686	1,321	1,717	2,074	2,508	2,819
	23	0,256	0,685	1,319	1,714	2,069	2,500	2,807
	24	0,256	0,685	1,318	1,711	2,064	2,492	2,797
	25	0,256	0,684	1,316	1,708	2,060	2,485	2,787
	26	0,256	0,684	1,315	1,706	2,056	2,479	2,779
	27	0,256	0,684	1,314	1,703	2,052	2,473	2,771
	28	0,256	0,683	1,313	1,701	2,048	2,467	2,763
	29	0,256	0,683	1,311	1,699	2,045	2,462	2,756
	30	0,256	0,683	1,310	1,697	2,042	2,457	2,750
	40	0,255	0,681	1,303	1,684	2,021	2,423	2,704
	50	0,255	0,680	1,296	1,676	2,009	2,403	2,678
	60	0,255	0,679	1,296	1,671	2,000	2,390	2,660
	80	0,254	0,679	1,292	1,664	1,990	2,374	2,639
	100	0,254	0,678	1,290	1,660	1,984	2,365	2,626
	120	0,254	0,677	1,289	1,658	1,980	2,358	2,467
	200	0,254	0,676	1,286	1,653	1,972	2,345	2,601

Приложение 2

Распределение Фишера (F-распределение)

$\alpha = 0{,}10$

число степеней свободы v2 \ число степеней свободы v1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	120
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,50	60,71	61,22	61,74	62,00	62,26	62,53	62,79	63,06
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,40	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,22	5,20	5,18	5,18	5,17	5,16	5,15	5,14
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,91	3,90	3,87	3,84	3,83	3,82	3,80	3,79	3,78
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,28	3,27	3,24	3,21	3,19	3,17	3,16	3,14	3,12
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,92	2,90	2,87	2,84	2,82	2,80	2,78	2,76	2,74
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,68	2,67	2,63	2,59	2,58	2,56	2,54	2,51	2,49
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,52	2,50	2,46	2,42	2,40	2,38	2,36	2,34	2,32
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,40	2,38	2,34	2,30	2,28	2,25	2,23	2,21	2,18
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,30	2,28	2,24	2,20	2,18	2,16	2,13	2,11	2,08
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,23	2,21	2,17	2,12	2,10	2,08	2,05	2,03	2,00
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,17	2,15	2,10	2,06	2,04	2,01	1,99	1,96	1,93
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,12	2,10	2,05	2,01	1,98	1,96	1,93	1,90	1,88
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,08	2,05	2,01	1,96	1,94	1,91	1,89	1,86	1,83
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,04	2,02	1,97	1,92	1,90	1,87	1,85	1,82	1,79
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	2,01	1,99	1,94	1,89	1,87	1,84	1,81	1,78	1,75
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,98	1,96	1,91	1,86	1,84	1,81	1,78	1,75	1,72
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,96	1,93	1,89	1,84	1,81	1,78	1,75	1,72	1,69
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,94	1,91	1,86	1,81	1,79	1,76	1,73	1,70	1,67
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,92	1,89	1,84	1,79	1,77	1,74	1,71	1,68	1,64
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,88	1,86	1,81	1,76	1,73	1,70	1,67	1,64	1,60
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,85	1,83	1,78	1,73	1,70	1,67	1,64	1,61	1,57
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,84	1,81	1,76	1,71	1,68	1,65	1,61	1,58	1,54
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,81	1,79	1,74	1,69	1,66	1,63	1,59	1,56	1,52
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,79	1,77	1,72	1,67	1,64	1,61	1,57	1,54	1,50
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,73	1,71	1,66	1,61	1,57	1,54	1,51	1,47	1,42
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,68	1,66	1,60	1,54	1,51	1,48	1,44	1,40	1,35
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,62	1,60	1,55	1,48	1,45	1,41	1,37	1,32	1,26

Приложение 2 (продолжение)

Распределение Фишера (F-распределение)

$\alpha = 0{,}05$	число степеней свободы v1																		
число степеней свободы v2	1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	120
1	161	200	216	225	230	234	237	239	271	242	243	244	246	248	249	250	251	252	253
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,71	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	3,52	3,37	3,29	3,21	3,13	3,05	2,96
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,21	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,87	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35

Приложение 3

Распределение Дарбина-Уотсона

$\alpha = 0{,}01$ (n – объем выборки, m – число объясняющих переменных в уравнении регрессии)

n	m=1		m=2		m=3		m=4	
	d1	d2	d1	d2	d1	d2	d1	d2
6	0,390	1,142						
7	0,433	1,036	0,294	1,676				
8	0,497	1,003	0,343	1,489	0,229	2,102		
9	0,554	0,998	0,408	1,389	0,279	1,873	0,183	2,433
10	0,604	1,001	0,466	1,333	0,340	1,733	0,230	2,193
11	0,633	1,010	0,319	1,297	0,396	1,640	0,286	2,030
12	0,697	1,023	0,369	1,274	0,449	1,373	0,339	1,913
13	0,738	1,038	0,616	1,261	0,499	1,326	0,391	1,826
14	0,776	1,034	0,660	1,234	0,347	1,490	0,441	1,737
15	0,811	1,070	0,700	1,232	0,391	1,464	0,488	1,704
16	0,844	1,086	0,737	1,232	0,633	1,446	0,332	1,663
17	0,874	1,102	0,772	1,233	0,672	1,432	0,374	1,630
18	0,902	1,118	0,803	1,239	0,708	1,422	0,613	1,604
19	0,928	1,132	0,833	1,263	0,742	1,413	0,630	1,384
20	0,932	1,147	0,863	1,271	0,773	1,411	0,683	1,367
21	0,973	1,161	0,890	1,277	0,803	1,408	0,718	1,334
22	0,997	1,174	0,914	1,284	0,831	1,407	0,748	1,343
23	1,018	1,187	0,938	1,291	0,838	1,407	0,777	1,334
24	1,037	1,199	0,960	1,298	0,882	1,407	0,803	1,328
25	1,033	1,211	0,981	1,303	0,906	1,409	0,831	1,323
26	1,072	1,222	1,001	1,312	0,928	1,411	0,833	1,318
27	1,089	1,233	1,019	1,319	0,949	1,413	0,878	1,313
28	1,104	1,244	1,037	1,323	0,969	1,413	0,900	1,313
29	1,119	1,234	1,034	1,332	0,988	1,418	0,921	1,312
30	1,133	1,263	1,070	1,339	1,006	1,421	0,941	1,311
35	1,193	1,307	1,140	1,370	1,083	1,439	1,028	1,312
40	1,246	1,344	1,198	1,398	1,148	1,437	1,098	1,318
50	1,324	1,403	1,283	1,446	1,243	1,491	1,203	1,338
100	1,322	1,362	1,303	1,383	1,482	1,604	1,462	1,623

Список рекомендуемой литературы

1. Эконометрика: Учебное пособие / Сидоренко М.Г.– Томск: Томский межвузовский центр дистанционного образования, 2004.-119с.
2. Эконометрика: Учебник / Домбровский В.В – М.: Новый учебник, 2004.-342с.
3. Эконометрика: Учебник / Кремер Н.Ш., Путко Б.А. – М.: Юнити, 2008.-311с.
4. Эконометрика: задачи и решения: Учебно-практическое пособие / Просветов Г.И. – М.: Альфа –Пресс, 2008.-192с.
5. Эконометрика: Учебное пособие в схемах и таблицах / Гореева Н.М., Орехов С.А. – М.: Эксмо, 2008.-224с.
6. Эконометрика: Учебник / Мхитарян В.С., Архипова М.Ю. – М.: Проспект, 2009.-384с.