

Задача 1.

По исходным данным за 16 месяцев, представленным в таблице 15, постройте уравнение зависимости объема предложения некоторого блага Y для функционирующей в условиях конкуренции фирмы от цены X_1 этого блага и заработной платы X_2 сотрудников этой фирмы.

Таблица 15.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Y	20	35	30	45	60	69	75	90	105	110	120	130	130	130	135	140
X_1	10	15	20	25	40	37	43	35	38	55	50	35	40	55	45	65
X_2	12	10	9	9	8	8	6	4	4	5	3	1	2	3	1	2

Задание:

1. Для заданного набора данных построить линейную модель множественной регрессии. Оценить точность и адекватность построенного уравнения регрессии.
2. Дать экономическую интерпретацию параметров модели.
3. Для полученной модели проверить выполнение условия гомоскедастичности остатков, применив тест Голдфелда-Квандта.
4. Проверить полученную модель на наличие автокорреляции остатков с помощью теста Дарбина-Уотсона.
5. Проверьте, адекватно ли предположение об однородности исходных данных в регрессионном смысле. Можно ли объединить две выборки (по первым 8 и остальным 8 наблюдениям) в одну и рассматривать единую модель регрессии Y по X ?

Решение:

1. В меню Сервис выбираем строку Анализ данных. Задаём параметры:

Получим результаты:

Контрольная работа выполнена на сайте www.MatBuro.ru
©МатБюро. Решение задач по математике, статистике, теории вероятностей

ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>	
Множественный R	0,984694332
R-квадрат	0,969622927
Нормированный R-квадрат	0,964949532
Стандартная ошибка	7,786996234
Наблюдения	16

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	25161,71497	12580,85748	207,4771689	1,36946E-10
Остаток	13	788,2850346	60,63731035		
Итого	15	25950			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	114,7767307	11,83332445	9,699449313	2,56066E-07	89,21239236	140,341069	89,21239236	140,341069
X1	0,672521327	0,202619002	3,319142433	0,005539908	0,23478967	1,110252983	0,23478967	1,110252983
X2	-9,440467325	0,87048122	-10,84511315	6,9644E-08	-11,32102731	7,559907343	-11,32102731	-7,559907343

R-квадрат – это R^2 . В нашем случае значение $R^2 = 0,97$ свидетельствует о том, что изменения зависимой переменной Y (некоторого блага) в основном (на 97%) можно объяснить изменениями включенных в модель объясняющих переменных – X_1, X_2 . Такое значение свидетельствует об адекватности модели.

В нашем случае расчетное значение F -критерия Фишера составляет 207,48. Значимость $F = 1,37E-10$, что меньше 0,05. Таким образом, полученное уравнение в целом значимо.

Анализ таблицы для рассматриваемого примера позволяет сделать вывод о том, что на уровне значимости $\alpha = 0.05$ значимым оказывается лишь коэффициенты при факторе X_2 , так как только для него Р-значение меньше 0,05. Таким образом, фактор X_1 не существен, и его включение в модель нецелесообразно.

2. Коэффициент $b_1 = 0,67$ означает, что при увеличении цены блага на 1 объём предложения увеличивается на 0,67. Коэффициент $b_2 = -9,44$ означает, что при увеличении заработной платы сотрудников на 1 объём предложения уменьшается на 9,44. Коэффициент $a = 114,78$ означает объём предложения блага при нулевой цене блага и заработной плате сотрудников.

3. Гипотеза о равенстве дисперсий двух нормально распределенных совокупностей, как проверяется с помощью критерия Фишера—Снедекора. Нулевая гипотеза о равенстве дисперсий двух наборов по m наблюдений (т. е. гипотеза об отсутствии гетероскедастичности) отвергается, если:

$$F = \frac{\sum_{i=1}^m e_i^2}{\sum_{i=n-m+1}^m e_i^2} > F_{\alpha, m-p, m-p}, \text{ где}$$

p – число регрессоров.

$$m = n/3 = 16/3 \approx 5$$

Построим модель по первым 5 наблюдениям, упорядочив данные по величине фактора X_2 :

ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>	
Множественный R	0,925937478
R-квадрат	0,857360213
Нормированный R-квадрат	0,714720426
Стандартная ошибка	3,961107998
Наблюдения	5

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	188,6192469	94,30962343	6,01066667	0,142639787
Остаток	2	31,38075314	15,69037657		
Итого	4	220			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	120,9832636	8,314419226	14,55101797	0,00468975	85,20918012	156,7573471	85,20918012	156,7573471
Переменная X 1	0,527196653	0,191739569	2,749545417	0,11073346	-0,297792701	1,352186006	-0,297792701	1,352186006
Переменная X 2	-8,20083682	2,63797494	-3,108762216	0,08975929	-19,5511348	3,149461158	-19,5511348	3,149461158

Построим модель по последним 5 наблюдениям:

ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>	
Множественный R	0,966839011
R-квадрат	0,934777673
Нормированный R-квадрат	0,869555346
Стандартная ошибка	6,734657094
Наблюдения	5

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	1300,088788	650,0443938	14,332173	0,065222327
Остаток	2	90,71121235	45,35560618		
Итого	4	1390,8			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	-24,75069603	62,02147047	-0,399066579	0,72842269	-291,607731	242,1063389	-291,607731	242,1063389
Переменная X 1	2,024044546	0,72667612	2,785346166	0,10834806	1,102592623	5,150681715	-1,102592623	5,150681715
Переменная X 2	2,212098203	4,963372062	0,445684542	0,69942626	19,14358301	23,56777942	-19,14358301	23,56777942

Рассчитаем статистику $F_{расч} = ESS_2/ESS_1$ (т.к. $ESS_2 > ESS_1$). Для нашего примера получаем: $F = 90,71/31,38 = 2,89$.

Для того, чтобы узнать табличное значение, воспользуемся встроенной в EXCEL функцией $FРАСПОБР(0,05;6;6)$ с параметрами 0,05 – заданная вероятность ошибки гипотезы H_0 ; $m - p = 5 - 2 = 6$; $m - p = 5 - 2$ – параметры распределения Фишера.

Статистика $F_{расч}$ меньше табличного значения $F = FРАСПОБР(0,05;3;3) = 9,28$. Следовательно, модель гомоскедастична.

4. Коэффициент автокорреляции рассчитывается по формуле:

$$d(DW) = \frac{\sum (e_t - e_{t-1})^2}{\sum e^2}.$$

В специальных таблицах табулированы значения d_1 и d_2 .

Если $d_{набл} \in (0; d_1)$ – уровни сильно автокоррелированы, модель неадекватна. При $d_{набл} \in (d_2; 2)$ – уровни независимы. Если $d_{набл} \in (d_1; d_2)$, требуются дополнительные исследования значимости коэффициента автокорреляции. В данном случае $d_1 = 0,857$, $d_2 = 1,728$

Приведём расчётную таблицу:

y	\hat{y}	e	$(e_t - e_{t-1})^2$	e^2
130	128,87	1,13		1,27
135	135,60	-0,60	2,98	0,36
130	122,80	7,20	60,89	51,89
140	139,61	0,39	46,42	0,15
120	120,08	-0,08	0,22	0,01
130	123,44	6,56	44,05	42,98
90	100,55	-10,55	292,72	111,37
105	102,57	2,43	168,54	5,90
110	104,56	5,44	9,05	29,56
75	87,05	-12,05	305,87	145,26
60	66,15	-6,15	34,79	37,87
69	64,14	4,86	121,39	23,66
30	43,26	-13,26	328,58	175,91
45	46,63	-1,63	135,43	2,64
35	30,46	4,54	38,02	20,61
20	8,22	11,78	52,47	138,85
1424	1424	0	1641,41	788,29

Тогда:

$$d(DW) = \frac{1641,41}{788,29} = 2,082 - \text{автокорреляция отсутствует, так как } d_{набл} \in (2; 4 - d_2).$$

5. Построим модель по первым 8 наблюдениям:

ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>	
Множественный R	0,954391291
R-квадрат	0,910862736
Нормированный R-квадрат	0,87520783
Стандартная ошибка	5,971172886
Наблюдения	8

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	1821,725472	910,8627359	25,5466315	0,002372183
Остаток	5	178,2745282	35,65490564		
Итого	7	2000			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	113,0165375	11,45178673	9,868899949	0,00018211	83,57883061	142,4542443	83,57883061	142,4542443
Переменная X 1	0,79653047	0,212869613	3,741870234	0,01340564	0,249332603	1,343728337	0,249332603	1,343728337
Переменная X 2	-10,66364301	1,901433319	-5,608212973	0,00249205	15,55142497	-5,77586104	-15,55142497	-5,775861042

Построим модель по последним 8 наблюдениям:

ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>	
Множественный R	0,968930882
R-квадрат	0,938827054
Нормированный R-квадрат	0,914357875
Стандартная ошибка	8,563534777
Наблюдения	8

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	5627,329361	2813,66468	38,367739	0,000925547
Остаток	5	366,6706394	73,33412788		
Итого	7	5994			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	23,92625764	60,77865998	0,393662145	0,71005745	-132,3100064	180,1625217	-132,3100064	180,1625217
Переменная X 1	1,561377527	0,683500707	2,284383191	0,07114707	0,195614105	3,318369159	-0,195614105	3,318369159
Переменная X 2	-1,93951575	4,827548071	-0,401760008	0,70445788	14,34910285	10,47007135	-14,34910285	10,47007135

Рассчитаем статистику F по формуле:

$$F_{расч} = \frac{\left(\sum_{i=1}^n \varepsilon_i^2 - \sum_{i=1}^{n_1} \varepsilon_i^2 - \sum_{i=n_1+1}^n \varepsilon_i^2 \right) / (m+1)}{\left(\sum_{i=1}^{n_1} \varepsilon_i^2 + \sum_{i=n_1+1}^n \varepsilon_i^2 \right) / (n-2m-2)} = \frac{(788,29 - 178,27 - 366,67) / (2+1)}{(178,27 + 366,67) / (16 - 2 \cdot 2 - 2)} = 1,49.$$

Находим табличное значение $F_{табл} = F_{РАСПОБР}(0,05;3;10) = 3,7$.

Так как, $F_{расч} < F_{табл}$, то справедлива гипотеза H_0 , т.е. надо использовать единую модель по всем наблюдениям.

Задача 2.

Изучается зависимость объема ВВП (Y , млрд. долл.) от уровня прибыли в экономике (X_t , млрд. долл.). Получена следующая модель с распределенными лагами:

$$Y_t = -5 + \underset{(2,2)}{1,5 \cdot X_t} + \underset{(2,3)}{2 \cdot X_{t-1}} + \underset{(2,5)}{4 \cdot X_{t-2}} + \underset{(2,3)}{2,5 \cdot X_{t-3}} + \underset{(2,4)}{2 \cdot X_{t-4}} + \varepsilon_t.$$

В скобках указаны значения t -критерия Стьюдента для коэффициентов регрессии. $R^2 = 0,9$.

Задание:

1. Проанализируйте полученные результаты регрессионного анализа.
2. Дайте интерпретацию параметров модели: определите краткосрочный и долгосрочный мультипликаторы.
3. Определите величину среднего лага и медианного лага.

Решение:

1. Видно, что связь между факторами сильная, на что указывает коэффициент детерминации, ошибки коэффициентов регрессии незначительны. Наибольшее влияние на результат оказывает фактор X_{t-2} .

2. Интерпретировать коэффициенты можно следующим образом: увеличение дохода населения в периоде $t-4$ на 1 млрд. долл. приводит к увеличению товарооборота в текущем периоде на 2 млрд. долл., увеличение дохода населения в периоде $t-3$ на 1 млрд. долл. приводит к увеличению товарооборота в текущем периоде на 2,5 млрд. долл., увеличение дохода населения в периоде $t-2$ на 1 млрд. долл. – к увеличению на 4 млрд. долл., увеличение дохода населения в периоде $t-1$ на 1 млрд. долл. – к увеличению на 2 млрд. долл., увеличение дохода населения в текущем периоде на 1 млрд. долл. – к увеличению на 1,5 млрд. долл.

Коэффициент при значении фактора x в момент времени t будет характеризовать среднее абсолютное изменение результативной переменной y в момент времени t при изменении фактора x в момент времени t на единицу своего измерения в некоторый фиксированный момент времени t , без учета воздействия лаговых значений фактора x . Он называется краткосрочным мультипликатором. В данном случае краткосрочный мультипликатор равен 1,5.

Частные промежуточные суммы, получаемые в результате сложения коэффициентов при различных значениях фактора x в различные моменты времени и выражающие в момент $(t + 1)$ совокупное воздействие факторной переменной x в момент t на результативную переменную y и т.п., называются промежуточными мультипликаторами. Последняя, самая большая из таких сумм имеет тот же смысл, но выделяется особым названием — долгосрочный мультипликатор. Определим долгосрочный мультипликатор:

$$1,5 + 2 + 4 + 2,5 + 2 = 12$$

3. Средний лаг представляет собой средний период, в течение которого будет происходить изменение результата под воздействием изменения фактора в момент времени t . В данном случае, этот показатель равен 2.

Медианный лаг — это период времени, в течение которого с момента времени t будет реализована 1/2 общего воздействия фактора на результат. 1/2 общего воздействия фактора на результат будет равна 6, следовательно, медианный лаг равен 2.

Задача 3.

Структурная форма макроэкономической модели имеет вид:

$$\begin{cases} C_t = a_1 + b_{11}D_t + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{22}Y_{t-1} + \varepsilon_2, \\ Y_t = D_t + T_t, \\ D_t = C_t + I_t + G_t, \end{cases}$$

где: C – расходы на потребление в период ,
 Y – чистый национальный продукт,
 D – чистый национальный доход,
 I – инвестиции,
 T – косвенные налоги
 G – государственные расходы,
 t – текущий период,
 $t-1$ – предыдущий период.

Задание:

1. Проверить каждое уравнение модели на идентифицируемость, применив необходимое и достаточное условия идентифицируемости.
2. Записать приведенную форму модели.
3. Определить метод оценки структурных параметров каждого уравнения.

Решение:

1. Необходимое условие идентификации:

$$D + 1 <\text{знак сравнения}> H$$

При знаке '=' уравнение точно идентифицировано.

Здесь:

D – число предопределенных переменных, отсутствующих в уравнении, но присутствующих в системе;

H – число эндогенных переменных в уравнении.

Достаточное условие идентификации: уравнение идентифицируемо, если определитель матрицы, составленной из коэффициентов при переменных, отсутствующих в исследуемом уравнении, не равен нулю, и ранг этой матрицы не менее числа эндогенных переменных системы без единицы.

Проверим каждое уравнение на идентифицируемость:

- 1) $3 + 1 > 2$ – необходимое условие не выполняется.
- 2) $2 + 1 > 2$ - необходимое условие не выполняется.
- 3) $2 + 1 = 2$ - необходимое условие не выполняется.
- 4) $2 + 1 = 3$ - необходимое условие выполняется.

Достаточное условие: отсутствуют Y_t , T , Y_{t-1} . Построим матрицу:

	Y_t	T	Y_{t-1}
Первое уравнение	0	0	0

Второе уравнение	b_{21}	0	b_{22}
Третье уравнение	1	1	0

Определитель матрицы будет равен:

$$\det = (-1) * \det \begin{vmatrix} 0 & 0 & 0 \\ b_{21} & 0 & b_{22} \\ 1 & 1 & 0 \end{vmatrix} = 0 - \text{достаточное условие не выполняется.}$$

2. Запишем приведённую форму модели:

$$\begin{cases} C_t = (a_1 + \frac{(a_1 + a_2)b_{11}}{1 + b_{11} - b_{21}}) + \frac{b_{22}b_{11}}{1 + b_{11} - b_{21}}Y_{t-1} + \frac{(2 - b_{21})b_{11}}{1 + b_{11} - b_{21}}T + \frac{b_{11}}{1 + b_{11} - b_{21}}G, \\ I_t = (a_2 + \frac{(a_1 + a_2)b_{21}}{1 + b_{11} - b_{21}}) + (\frac{b_{22}b_{21}}{1 + b_{11} - b_{21}} + b_{22})Y_{t-1} + \frac{(1 - b_{11})b_{21}}{1 + b_{11} - b_{21}}T + \frac{b_{21}}{1 + b_{11} - b_{21}}G, \\ Y_t = \frac{a_1 + a_2}{1 + b_{11} - b_{21}} + \frac{b_{22}}{1 + b_{11} - b_{21}}Y_{t-1} + \frac{1 - b_{11}}{1 + b_{11} - b_{21}}T + \frac{1}{1 + b_{11} - b_{21}}G, \\ D_t = \frac{a_1 + a_2}{1 + b_{11} - b_{21}} + \frac{b_{22}}{1 + b_{11} - b_{21}}Y_{t-1} + \frac{2 - b_{21}}{1 + b_{11} - b_{21}}T + \frac{1}{1 + b_{11} - b_{21}}G, \end{cases}$$

3. Данная система является сверхидентифицируемой. Если система сверхидентифицируема, то КМНК не используется, так как он не дает однозначных оценок для параметров структурной модели. В этом случае могут использоваться разные методы оценивания, среди которых наиболее распространенным и простым является ДМНК:

- составляют приведенную форму модели и определяют численные значения параметров каждого его уравнения обычным МНК;
- выявляют эндогенные переменные, находящиеся в правой части структурного уравнения, параметры которого определяют косвенным МНК, и находят расчетные значения по соответствующим уравнениям приведенной формы модели;
- обычным МНК определяют параметры структурного уравнения, используя в качестве исходных данных фактические значения предопределенных переменных и расчетные значения эндогенных переменных, стоящих в правой части данного структурного уравнения.