

УДК 37.016:51

Мурай С.В., Павлова Т.О.

## ИЗУЧЕНИЕ МНОЖЕСТВЕННОЙ РЕГРЕССИИ С УЧЕТОМ КАЧЕСТВЕННЫХ ФАКТОРОВ С ИСПОЛЬЗОВАНИЕМ ТАБЛИЧНОГО ПРОЦЕССОРА EXCEL

**Постановка проблемы.** Множественная регрессия позволяет выяснить наличие взаимосвязей, определить ее форму, тесноту связи между переменными с учетом качественных факторов. Построенное регрессионное уравнение позволяет произвести анализ взаимосвязи и определить прогноз на будущее. При изучении регрессии на практическом занятии в вузе требуется проводить значительный объем вычислений, поэтому целесообразно использовать программное обеспечение Microsoft EXCEL.

**Анализ последних исследований и публикаций.** Издано большое количество научно-технической и учебной литературы, в которой обстоятельно освещены теоретические вопросы и показано решение широкого круга задач с использованием множественной регрессии [1-5]. Однако в учебной литературе недостаточно освещены вопросы проведения практических занятий при изучении множественной регрессии с помощью современного компьютерного обеспечения.

Нерешенными ранее частями общей проблемы является недостаточное использование программного обеспечения при изучении множественной регрессии в вузе.

**Формулировка целей статьи:** разработать материалы для проведения учебных занятий при изучении множественной регрессии с использованием EXCEL.

**Изложение основного материала исследования.** При решении задач регрессионного анализа в большинстве ситуаций значения независимых переменных точно определить не удастся. Рассмотрим способы включения качественных факторов, например сезонных колебаний, в анализ методом множественной регрессии на примере прогнозирования поквартальных продаж автомобилей в США. Исходные данные для решения данной задачи показаны на рисунке 1. (Значения продаж указаны в тысячах автомобилей.)

Квартал года – это *качественная независимая переменная* (qualitative independent variable). Чтобы смоделировать такую переменную, создаем независимую переменную (называемую *фиктивной*) для всех возможных значений качественной переменной, за исключением одного. (В данном примере это *Квартал 4*). Фиктивные переменные указывают, какое значение количественной переменной наблюдается. Таким образом, фиктивные

переменные для *Квартала 1*, *Квартала 2*, *Квартала 3* содержат следующие свойства:

- фиктивная переменная для *Квартала 1* равна 1, если квартал года – первый, и равна нулю в противном случае;
- фиктивная переменная для *Квартала 2* равна 1, если квартал года – второй, и равна нулю в противном случае;
- фиктивная переменная для *Квартала 3* равна 1, если квартал года – третий, и равняется нулю в противном случае.

	A	B	C	D	E	F
9	Исходные данные					
10	Год	Квартал	Продажи	ВВП	Безраб	Ставка
11	79	1	Продажи	2541	5,9	9,4
12	79	2	2910	2640	5,7	9,4
13	79	3	2562	2595	5,9	9,7
14	79	4	2385	2701	6	11,9
15	80	1	2520	2785	6,2	13,4
16	80	2	2142	2509	7,3	9,6
17	80	3	2130	2570	7,7	9,2
18	80	4	2190	2667	7,4	13,6
19	81	1	2370	2878	7,4	14,4
20	81	2	2208	2835	7,4	15,3
21	81	3	2196	2897	7,4	15,1
22	81	4	1758	2744	8,3	11,8
23	82	1	1944	2582	8,8	12,8
24	82	2	2094	2613	9,4	12,4
25	82	3	1911	2529	10	9,3
26	82	4	2031	2544	10,7	7,9

Рис. 1. Данные о продажах автомобилей

Наблюдение *Квартала 4* будет определяться тем фактом, что фиктивные переменные для *Кварталов 1–3* равны 0. В случае, если включим переменную *Квартала 4*, каждая точка данных будет удовлетворять линейной зависимости:

$$(\text{Фиктивная переменная Квартала 1}) + (\text{Фиктивная переменная Квартала 2}) + (\text{Фиктивная переменная Квартала 3}) + (\text{Фиктивная переменная Квартала 4}) = 1$$

Чтобы создать фиктивную переменную для *Квартала 1*, нужно скопировать из ячейки G12 в диапазон G13:G42 формулу ЕСЛИ (B12=1;1;0). Эта формула помещает 1 в столбец G, если квартал года – первый, и 0, если квартал года не является первым. По аналогии создаем фиктивные переменные для *Квартала 2* (в диапазоне H12:H42) и *Квартала 3* (в диапазоне I12:I42). Результаты применения формул показаны на рис. 2.

Помимо данных о сезонных колебаниях будем использовать макроэкономические переменные, например валовой национальный продукт (ВВП, млн.), процентные ставки и процент безработицы, для прогнозирования продаж машин. Попытаемся спрогнозировать продажи для второго квартала 1979 г. Поскольку значения ВВП, процентные ставки и процент безработицы на начало второго квартала 1979 г. неизвестны, то нет

возможности использовать их для прогнозирования продаж автомобилей в *Квартале 2* 1979 г. Вместо этого для прогнозирования продаж будем использовать значения ВВП, процентные ставки и процент безработицы с *задержкой на один квартал*. Скопировав из ячейки J12 в диапазон J13:L42 формулу =D11, определим запаздывающие значения ВВП, процентных ставок и процента безработицы. Например, диапазон J13:L42 содержит эти значения для первого квартала 1979 г.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
5	ВВП - валовой национальный продукт; Безраб - процент безработных; Ставка - процентная ставка													
6	ЗапВВП - запаздывающее значение ВВП; ЗапБезраб - запаздывающее значение доли безработных													
7	ЗапСтавка - запаздывающее значение ставки процента													
8														
9	Исходные данные													
10	Год	Квартал	Продажи	ВВП	Безраб	Ставка	K1	K2	K3	ЗапВВП	ЗапБезраб	пСтавка		
11	79	1	2910	2541	5,9	9,4	K1	K2	K3	ЗапВВП	ЗапБезраб	пСтавка		
12	79	2	2910	2640	5,7	9,4	0	1	0	2541	5,9	9,4		
13	79	3	2562	2595	5,9	9,7	0	0	1	2640	5,7	9,4		
14	79	4	2385	2701	6	11,9	0	0	0	2595	5,9	9,7		
15	80	1	2520	2785	6,2	13,4	1	0	0	2701	6	11,9		
16	80	2	2142	2509	7,3	9,6	0	1	0	2785	6,2	13,4		
17	80	3	2130	2570	7,7	9,2	0	0	1	2509	7,3	9,6		
18	80	4	2190	2667	7,4	13,6	0	0	0	2570	7,7	9,2		
19	81	1	2370	2878	7,4	14,4	1	0	0	2667	7,4	13,6		
20	81	2	2208	2835	7,4	15,3	0	1	0	2878	7,4	14,4		
21	81	3	2196	2897	7,4	15,1	0	0	1	2835	7,4	15,3		
22	81	4	1758	2744	8,3	11,8	0	0	0	2897	7,4	15,1		
23	82	1	1944	2582	8,8	12,8	1	0	0	2744	8,3	11,8		
24	82	2	2094	2613	9,4	12,4	0	1	0	2582	8,8	12,8		
25	82	3	1911	2529	10	9,3	0	0	1	2613	9,4	12,4		
26	82	4	2031	2544	10,7	7,9	0	0	0	2529	10	9,3		

Рис. 2. Набор данных для использования фиктивных переменных для выявления квартала, в котором осуществлялись продажи

Теперь можно провести анализ методом множественной регрессии, выбрав в меню *Сервис (Tools)* команду *Анализ данных (Data Analysis)* и затем щелкнув *Регрессия (Regression)* в диалоговом окне *Анализ данных (Data Analysis)*. В качестве *Входного интервала Y (Input Y range)* укажем C11:C42, в качестве *Входного интервала X (Input X range)* – G11:L42, и установим флажок *Метки (Labels)*, а также флажок *Остатки (Residuals)*. Щелкнув *ОК*, получаем вывод, который показан на рисунке 3.

1	Вывод ИТОГОВ				
2	Приведенные ниже пояснения вы найдете в справочной системе Excel				
3	Регрессионная статистика				
4	Множественный R	0,884139126	SS – сумма квадратов		
5	R-квадрат	0,781701994	F – F-статистика, или F-наблюдаемое значение		
6	Нормированный	0,727127492	Sreg – регрессионная сумма квадратов		
7	Стандартная с	190,5240756	Sresid – остаточная сумма квадратов		
8	Наблюдения	31	MS (mean square) – среднеквадратичное значение		
9					
10	Дисперсионный анализ				
11		df	SS	MS	F
12	Регрессия	6	3119625	519938	14,3236
13	Остаток	24	871186	36299,4	6,80E-07
14	Итого	30	3990811		

Рис. 3. Итоги и таблица дисперсионного анализа

На рис. 4 видно, что уравнение, применяемое для прогноза квартальных продаж, выглядит так:

$$\text{Спрогнозированные поквартальные продажи} = 3154,7 + 156,833K1 + 379,784K2 + 203,036K3 + 0,174(\text{ЗАПВНП в млрд.}) - 93,83(\text{ЗАПБЕЗРАБ}) - 73,91(\text{ЗАПСТАВКА}) \quad (1)$$

	A	B	C	D	E	F	G	H	I	J
16	Кoeffициенты									
17	Y-пересечени	3154,700285	462,653	6,81872	4,7E-07	2199,83124	4109,57	2199,83	4109,57	
18	K1	156,833091	98,8711	1,58624	0,12577	-47,2268431	360,893	-47,227	360,893	
19	K2	379,7835116	98,0892	3,95241	0,00059	181,46512	578,102	181,465	578,102	
20	K3	203,035501	95,4089	2,12806	0,0438	6,12117245	399,95	6,12117	399,95	
21	ЗапВНП	0,174156906	0,05842	2,98112	0,00649	0,05358395	0,29473	0,05358	0,29473	
22	ЗапБезраб	-93,83233214	28,3233	-3,3129	0,00292	-152,288723	-35,376	-152,29	-35,376	
23	ЗапСтавка	-73,9167147	17,7885	-4,1553	0,00036	-110,630406	-37,203	-110,63	-37,203	
24										

Рис. 4. Коэффициенты регрессии для продаж автомобилей

На рис. 4 также видно, что значение вероятности каждой независимой переменной (столбец E) меньше или равно 0,15. Из полученных данных можно сделать вывод, что все независимые переменные оказывают большое влияние на квартальные продажи автомобилей. Интерпретируем все коэффициенты в нашем уравнении регрессии при прочих равных условиях.

Проанализируем каждый коэффициент:

- рост ВВП на \$1 млрд. в предыдущем квартале увеличивает квартальные продажи машин на 174 единицы;
- рост безработицы на 1% в предыдущем квартале снижает квартальные продажи машин на 93832 единицы;
- увеличение процентной ставки на 1% в предыдущем квартале снижает квартальные продажи машин на 73917 единиц.

Коэффициенты фиктивных переменных показывают влияние сезонных колебаний относительно значений пропущенных качественных переменных, поэтому:

- продажи машин в *Квартале 1* превышают продажи машин в *Квартале 4* на 156833 единицы;
- продажи машин в *Квартале 2* превышают продажи машин в *Квартале 4* на 379784 единицы;
- продажи машин в *Квартале 3* превышают продажи машин в *Квартале 4* на 203036 единицы.

Видно, что максимальное количество машин продано во втором квартале (апрель – июнь), а минимальное – в четвертом квартале (октябрь – декабрь).

Анализируя данные, представленные на рисунке 3, можно сделать следующие выводы:

- Изменение независимых переменных (макроэкономические факторы и сезонные колебания) объясняет 78% изменений зависимой переменной (квартальные продажи машин).

- Стандартная ошибка нашей регрессии – 190524 машин. Можно ожидать, что примерно 68% наших прогнозов точны в пределах  $\pm 190524$  машин, и около 95% – в пределах  $\pm 381048$  машин ( $2 \times 190524$ ).

- Для вычисления регрессии потребовалось 31 наблюдение.

На рисунке 5 для каждого наблюдения указаны спрогнозированные продажа и остаток.

	A	B	C	D
27	Вывод остатка			
28				
29	Наблюдение	Предсказанные Продажи	Остатки	
30	1	2728,688616	181,411	
31	2	2587,848606	-25,849	
32	3	2336,034563	48,9654	
33	4	2339,328281	180,672	
34	5	2447,266343	-305,27	
35	6	2400,118977	-270,12	
36	7	2199,7408	-9,7408	
37	8	2076,383266	293,617	
38	9	2276,947422	-68,947	
39	10	2026,185621	169,814	
40	11	1848,731191	-90,731	
41	12	2138,394335	-194,39	
42	13	2212,298456	-118,3	
43	14	2014,216596	-103,22	
44	15	1969,394332	61,6057	
45	16	2186,640544	-120,64	
46	17	2440,632301	61,3677	
47	18	2290,352403	-52,352	
48	19	2131,386979	262,613	
49	20	2433,681173	152,319	
50	21	2728,688616	181,411	

Рис. 5. Значения остатков при продаже автомобилей

Например, для второго квартала 1979 г. (наблюдение 1) спрогнозированные продажи на основе уравнения (1) составляют 2728,6 тыс. автомобилей, а остаток - 181400 автомобилей (2910-2728,6).

**Выводы.** Применение табличного процессора EXCEL при изучении множественной регрессии с включением качественных факторов позволяет рассматривать на учебных занятиях реальные задачи и способы их решения, получать результаты в наглядном табличном виде. Это повышает интерес студентов к рассмотрению конкретных прикладных задач, позволяет рассмотреть большое количество вариантов в наглядном виде и проанализировать полученные результаты для принятия решения в каждом конкретном случае.

### ЛИТЕРАТУРА

1. Кремер Н.Ш. Теория вероятностей и математическая статистика: учебник для студентов вузов, обучающихся по экономическим специальностям / Н.Ш. Кремер. – 3-е изд., перераб. и доп. – М.: ЮНИТИ-ДАНА, 2007. – 551 с.

2. Єремеев В.С. Теорія ймовірностей та математична статистика / В.С. Єремеев, Д.О.Сосновских, О.В. Тітова. – Мелітополь.: ТОВ «Видавничий будинок ММД», 2009. – 188 с.
3. Вилсон У.Л. Microsoft Excel: анализ данных и построение бизнес-моделей / Уэйн Л. Вилсон. – Пер. с англ. – М.: Издательско- торговый дом «Русская Редакция», 2005. -576 с.
4. Мур Д. Экономическое моделирование в Microsoft EXCEL / Джеффри Мур, Лари Р. Уэдерфорд. – М.: Издательский дом «Вильямс», 2004. – 1024 с.
5. Excel для экономистов и менеджеров / А.Г. Дубина, С.С. Орлова, И.Ю. Шубина, А.В. Хромов. –СПб.: Питер, 2004. – 295 с.

#### **АННОТАЦІЯ**

*Мурай С.В., Павлова Т.О. Вивчення множинної регресії з урахуванням якісних факторів з використанням табличного процесора Excel. Представлені матеріали для вивчення множинної регресії з включенням якісних факторів за допомогою табличного процесора Excel.*

**Ключові слова.** *Множинна регресія, табличний процесор Excel, якісні фактори, коефіцієнти регресії, продаж автомобілів*

#### **АННОТАЦИЯ**

*Murai S.V., Pavlova T.O. Изучение множественной регрессии с учетом качественных факторов с использованием табличного процессора Excel. Представлены материалы для изучения множественной регрессии с включением качественных факторов с помощью табличного процессора Excel.*

**Ключевые слова.** *Множественная регрессия, табличный процессор Excel, качественные факторы, коэффициенты регрессии, продажа автомобилей.*

#### **SUMMARY**

*Murai S.V, Pavlova T.O. The study of multiple regression, taking into account qualitative factors using the Excel spreadsheet. The materials for the study of multiple regression with the inclusion of qualitative factors with the help of a table processor Excel.*

**Key words.** *Multiple regression, spreadsheet Excel, qualitative factors, the regression coefficients, car sales.*