

Гомельский государственный технический университет имени П.О. Сухого

Кафедра «Маркетинг и отраслевая экономика»

Лабораторная работа № 3

**«Применение многофакторного регрессионного анализа в задачах
принятия решений»**

по дисциплине: “Эконометрика и экономико-математические методы и модели ”

Дата сдачи отчета 29.03.2021

Выполнил студент группы МГ-21
Мельников О.В.

Допуск к защите

Принял к.э.н., доцент Винник О.Г.

Цель работы: моделирование задач с несколькими количеством независимых факторов, определяющих экономическую ситуацию. Для моделирования используется метод регрессионного анализа с использованием инструментов Excel.

Теоретическая часть:

Существует множество задач, в которых каждое значение y определяется целым набором независимых факторов x_1, x_2, \dots, x_p , значения которых определяются не только временными интервалами. В этих случаях при моделировании неизвестных оценок Y определяемого фактора y уже требуется учитывать взаимосвязи фактических данных. Они определяются на основе взятых из наблюдений данных, которые задаются следующей матрицей (1):

y_1	x_{11}	x_{12}	...	x_{1n}
y_2	x_{21}	x_{22}	...	x_{2n}
...
y_m	x_{m1}	x_{m2}	...	x_{mn}

В таких задачах используют метод регрессионного анализа. Регрессия – это статистический метод, позволяющий найти уравнение, которое наилучшим образом описывает множество данных. Уравнение регрессии $Y=f(x_1, x_2, \dots, x_p)$ (2) выбирают исходя из характера взаимосвязей (наблюдаемого в опыте или на графиках). Параметры уравнения (коэффициенты, свободный член) находят по методу наименьших квадратов, находя сумму квадратов отклонений L фактических значений y_i от найденных Y_i по уравнению регрессии (2) при значениях факторов x_{ik} , взятых из матрицы (1):

$$L = \sum_{i=1}^m [y_i - Y(x_{i1}, x_{i2}, \dots, x_{in})]^2$$

и затем минимизируя ее. Регрессионный анализ позволяет исследовать линейные и нелинейные взаимосвязи между задаваемыми факторами x_1, x_2, \dots, x_p и определяемым фактором y . Этот метод применяют как для прогнозирования, так и для оценки значений y при варьировании факторов x_1, x_2, \dots, x_p внутри интервалов их допустимых значений, например, для принятия решений по вопросам финансирования операций, проведения маркетинговых исследований и т. п.

Excel предоставляет следующие возможности для анализа:

- инструменты Пакета анализа (Регрессия и др.);
- функции ЛИНЕЙН, ТЕНДЕНЦИЯ, ЛГРФПРИБЛ для построения уравнений регрессии;
- функции ФРАСП, СТЬЮДРАСП для оценки достоверности уравнения регрессии и его коэффициентов;
- диаграммы и линии тренда для графической иллюстрации взаимосвязей.

Однофакторный линейный регрессионный анализ

Регрессия называется однофакторной (или парной), если она описывает зависимость между функцией и одной переменной. При однофакторном анализе в матрице (1) остаются только первый и второй столбцы данных, а уравнение регрессии (2) выглядит как $Y=f(x_1)$ или просто $Y=f(x)$. Оно может быть как линейным $Y=a \cdot x + b$, так и нелинейным.

Для получения уравнения регрессии необходимо:

- определить значения коэффициентов в уравнении;
- оценить достоверность полученного уравнения.

Регрессия называется множественной, если она описывает зависимость функции от нескольких переменных и имеет вид:

$$Y = f(x_1, x_2, \dots, x_n)$$

Регрессия является линейной в том случае, когда уравнение имеет вид:

$$Y = b + M_1 \cdot x_1 + M_2 \cdot x_2 + \dots + M_n \cdot x_n$$

Задания к практической части:

Задание 7. По данным производительности x_1 (количество операций в час), характеристики качества x_2 (время работы на отказ в днях) и зависящей от них цены бытовой техники y (тыс. руб) определите линейное уравнение регрессии, устанавливающее зависимость цены y от x_1 и x_2 . Проведите анализ достоверности полученного уравнения регрессии.

Задание 8. Постройте ту же самую зависимость цены на бытовую технику от производительности труда и характеристики качества. Рассчитайте прогнозное значение y .

Выполнение:

Выполнение задания 7:

1. Занёс данные в рабочий лист Excel в соответствии с вариантом, определяемом номером в журнале

x1	x2	y
125	420	4600
207	960	7700
320	145	3200
418	221	5510
500	285	5100
890	325	6458

Рисунок 1 – исходные данные

2. Выделил блок свободных ячеек, состоящий из 5 строк и столбцов, количество которых равно $n+1$, где n – количество независимых параметров x . Ввел знак «=» и формулу массива ЛИНЕЙН. После появления диалогового окна скопировал диапазон данных о цене бытовой техники в «Изн_знач_y», а диапазон данных, характеризующих производительность и качество в «Изн_знач_x», состоящий из 2-х столбцов. В окне «Константа» и «Стат» ввёл слово «истина». Результат показан на скриншоте снизу:

5,202784655	3,2271212	2061,9202
1,100797499	1,18758678	807,3069205
0,886803936	672,52114	#Н/Д
11,75134415	3	#Н/Д
10629905,95	1356854,051	#Н/Д

Рисунок 2 – результат выполненных в п.2 действий

Выполнение задания 8:

1. Построил уравнение множественной регрессии, используя Пакет анализа (инструмент Регрессия), методика выполнения которого неоднократно использовалась в задании 2.

Регрессия						
Регрессионная статистика						
Линейная						
Необработанный вывод						
ЛИНЕЙН						
5,202784655	3,2271212	2061,9202				
1,100797499	1,18758678	807,3069205				
0,886803936	672,52114	#Н/Д				
11,75134415	3	#Н/Д				
10629905,95	1356854,051	#Н/Д				
Статистика регрессии						
R^2	0,886803936					
Среднеквадратичная ошибка	672,52114					
Число переменных	2					
Наблюдения	6					
Скорректированный коэффициент	0,811339893					
Дисперсионный анализ (ANOVA)						
	df	SS	MS	F	Значимость F	
Регрессия	2	10629905,95	5314952,974	11,75134415	0,038084389	
Остатки	3	1356854,051	452284,6838			
Итого	5	11986760				
Уровень доверия к статистическим выводам						
	0,95					
	Коэффициент	Среднеквадратичная ошибка	t-статистика	P-значение	Нижний 95%	Верхний 95%
Пересечение	2061,9202	807,3069205	2,554072247	0,083645179	-507,290726	4631,131127
X1	3,2271212	1,18758678	2,717377166	0,072710095	-0,55230996	7,006552361
X2	5,202784655	1,100797499	4,726377614	0,017946368	1,699555723	8,706013587
X1	X2	Прогнозное значение Y	Остатки			
125	420	4650,479905	4600	-50,4799053		
207	960	7724,607557	7700	-24,6075573		
320	145	3849,002759	3200	-649,002759		
418	221	4560,672271	5510	949,3277295		
500	285	5158,274427	5100	-58,2744268		
890	325	6624,963081	6458	-166,963081		

Рисунок 3 – Уравнение множественной регрессии

2. Скопировал на новый рабочий лист исходные данные этого задания.

Используя функцию ТЕНДЕНЦИЯ, введя все необходимые данные, я получил то же самое прогнозное значение, что и подсчитанное по уравнению регрессии.

3. Ввёл 10 произвольных дополнительных значений x_1 и x_2 . Для получения прогнозного значения y выделил столбец, состоящий из 10 ячеек и ввёл формулу массива «=ТЕНДЕНЦИЯ», и, пройдя этап введения данных, результат выполнения этих заданий выглядел следующим образом:

C11			\sum	=	{=ТЕНДЕНЦИЯ(C2:C7;A2:B7;A9:B17;1)}							
	A	B	C	D	E	F	G	H	I	J	K	L
1	x1	x2	y									
2		125	420	4600		Регрессия						
3		207	960	7700		Регрессионная	Линейная					
4		320	145	3200								
5		418	221	5510		Необработанный вывод ЛИНЕЙН						
6		500	285	5100		5,202784655	3,2271212	2061,9202				
7		890	325	6458		0,394547779	0,170392632	203,9783515				
8		1000	400	7370,155262		0,989170222	323,0687056	#Н/Д				
9		1100	350	7432,728149		593,6969709	13	#Н/Д				
10		1200	450	8275,718735		123932329,3	1356854,051	#Н/Д				
11		1300	643	9602,568293								
12		1400	567	9529,868779		Статистика регрессии						
13		1500	784	10981,58517		R^2	0,989170222					
14		1600	678	10752,80212		Среднеквадратическая ошибка	323,0687056					
15		1700	801	11715,45675		Число переменных	2					
16		1800	712	11575,12103		Наблюдения	16					
17		1900	950	13136,0959		Скорректированные наблюдения	0,987504102					
18												
19						Дисперсионный анализ (ANOVA)						
20						df	SS	MS	F	Значимость F		
21						Регрессия	2	123932329,3	61966164,63	593,6969709	1,67891E-13	
22						Остатки	13	1356854,051	104373,3886			
23						Итого	15	125289183,3				
24												
25						Уровень доверия к статистике	0,95					
26												
27						Коэффициент	Среднеквадратическая ошибка	t-статистика	P-значение	Нижний 95%	Верхний 95%	
28						Пересечение	2061,9202	203,9783515	10,10852468	1,58722E-07	1621,251763	2502,588637
29						X1	3,2271212	0,170392632	18,93932366	7,54583E-11	2,859010299	3,595232101
30						X2	5,202784655	0,394547779	13,18670369	6,69074E-09	4,350415999	6,05515331
31												
32						X1	X2	Прогнозное Y	Остатки			
33							125	420	4650,479905	4600	-50,47990526	
34							207	960	7724,607557	7700	-24,60755726	
35							320	145	3849,002759	3200	-649,0027592	
36							418	221	4560,672271	5510	949,3277295	
37							500	285	5158,274427	5100	-58,27442682	
38							890	325	6624,963081	6458	-166,963081	
39							1000	400	7370,155262	7370,155262	0	
40							1100	350	7432,728149	7432,728149	0	
41							1200	450	8275,718735	8275,718735	0	
42							1300	643	9602,568293	9602,568293	0	
43							1400	567	9529,868779	9529,868779	0	
44							1500	784	10981,58517	10981,58517	0	

Рисунок 4 – Итог выполнения задания 8, п.2,3.

Таким образом, мной было освоено моделирование задач с двумя независимыми факторами, определяющими экономическую ситуацию. Для моделирования использовался метод регрессионного анализа с использованием инструментов Libreoffice.