Содержание

Практическое занятие № 1-2.ПОСТРОЕНИЕ МНОЖЕСТВЕННОЙ РЕГРЕССИОННОЙ МОДЕЛИ

Постановка задачи: известны статистические данные наблюдений за некоторым количеством однородных экономических объектов.

Требуется:

1. Рассчитать парные коэффициенты корреляции между факторами и результатом. Оценить значимость полученных коэффициентов.
2. Осуществить выбор факторных признаков для построения двухфакторной (или трехфакторной) регрессионной модели.
3. Построить линейное уравнение множественной регрессии, описывающее зависимость между факторами и результатом.
4. Рассчитать коэффициент множественной корреляции, частные коэффициенты корреляции, коэффициенты эластичности. Дать экономическую интерпретацию.

Пример. Известны статистические данные 16 наблюдений некоторых экономических показателей (табл.).

Таблица

Объем реализации товара, млн. руб.	Расходы на оформление товара, руб.	Расходы на рекламу, руб.	Цена товара, руб.	Индекс потребительских расходов, %
y	x_1	x_2	x_3	x_4
44	54	829	37567	100
48	77	943	34767	98
51	47	800	38033	101
y	$x1$	$x2$	$x3$	$x4$
66	90	1500	38733	104
93	115	1429	38733	104
125	91	1643	37567	107
146	193	2357	44800	107
150	212	2929	32900	109
124	251	2732	39433	108
125	134	1771	42000	109
109	161	1486	40600	110
104	87	1186	40133	111
112	165	2057	38500	110
117	70	1186	39200	112
123	76	1214	39900	112
130	131	1071	37800	113

Порядок выполнения работы.

Шаг 0. Подготовительная работа.

Заносим статистические данные в электронную таблицу Excel по столбцам. Определяем, какие из показателей будут являться факторами, а какой из них – результатом. Объем реализации товара в нашем случае выберем в качестве зависимой переменной y. Факторами, или независимыми переменными, объясняющими результат, являются x_1 – расходы на оформление товара, x_2 – расходы на рекламу, x_3 – цена товара, x_4 – индекс потребительских расходов.

Шаг 1. Осуществить выбор факторных признаков для построения двухфакторной (или трехфакторной) регрессионной модели.

Для подбора факторных признаков используем инструмент КОРРЕЛЯЦИЯ. Вызываем команду СЕРВИС – АНАЛИЗ ДАННЫХ (рис. 1). В меню АНАЛИЗ ДАННЫХ выбираем инструмент КОРРЕЛЯЦИЯ.

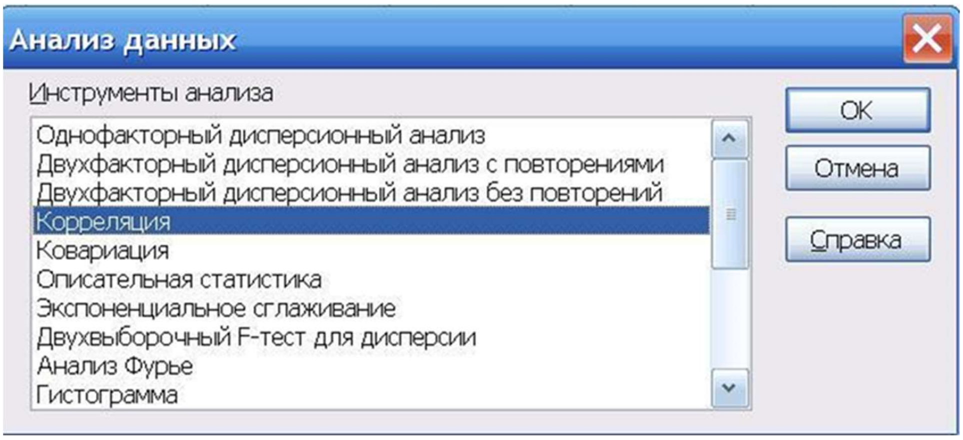
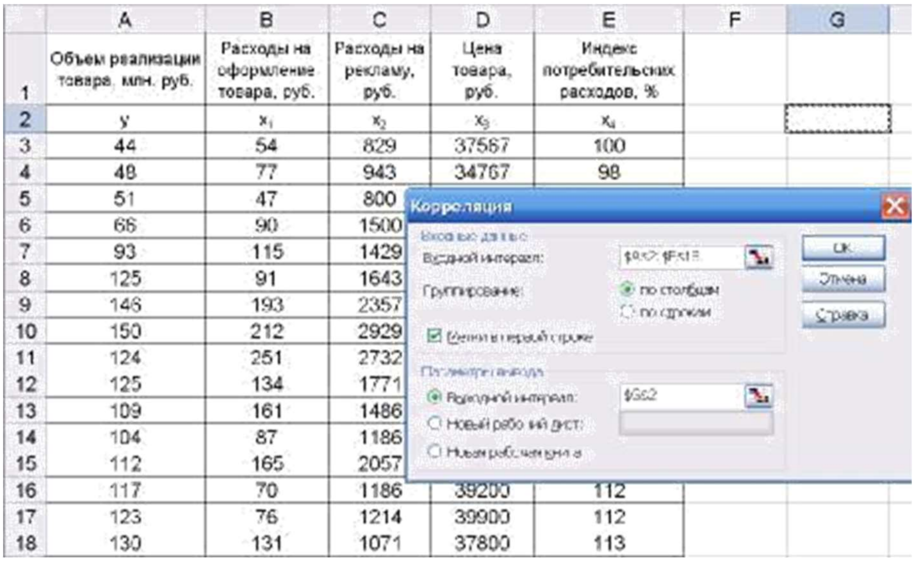


Рис. 1. Меню АНАЛИЗ ДАННЫХ

В диалоговом окне КОРРЕЛЯЦИЯ во *Входном интервале* вводим диапазон ячеек, содержащих исходные данные задачи (ввод данных осуществляем путем выделения необходимого диапазона ячеек). Отметим группирование по столбцам и, если вводим заголовки столбцов, то устанавливаем метку в окошке *Метки* в первой строке. Затем выбираем *Параметры вывода* (необходимо указать, где именно компьютер должен вывести информацию): если мы решили расположить результат на данном листе, то выбираем *Выходной интервал* и указываем адрес ячейки, начиная с которой результат будет представлен (рис. 2); или на *Новый рабочий лист*, тогда необходимо указать его название; либо *Новая рабочая книга*, тогда результаты расчетов будут выведены в новом файле.



	A	B	C	D	E	F	G
1	Объем реализации товара, млн. руб.	Расходы на оформление товара, руб.	Расходы на рекламу, руб.	Цена товара, руб.	Индекс потребительских расходов, %		
2	y	x_1	x_2	x_3	x_4		
3	44	54	829	37567	100		
4	48	77	943	34767	98		
5	51	47	800				
6	66	90	1500				
7	93	115	1429				
8	125	91	1643				
9	146	193	2357				
10	150	212	2929				
11	124	251	2732				
12	125	134	1771				
13	109	161	1486				
14	104	87	1186				
15	112	165	2057				
16	117	70	1186	39200	112		
17	123	76	1214	39900	112		
18	130	131	1071	37800	113		

Рис. 2. Работа в меню КОРРЕЛЯЦИЯ

В результате мы получаем матрицу коэффициентов парной корреляции (см. рис. 3).

	G	H	I	J	K	L
2		*y*	*x1*	*x2*	*x3*	*x4*
3	y	1				
4	x1	**0,663812**	1			
5	x2	**0,68306**	**0,909616**	1		
6	x3	0,278488	0,127788	0,045103	1	
7	x4	**0,809924**	0,354051	0,299836	0,306524	1

Рис. 3. Корреляционная матрица

Проведем анализ коэффициентов парной корреляции. Зависимая переменная y имеет тесную связь с переменной x_4: и достаточно высокое влияние на нее оказывают переменные x_1 и x_2. Однако факторы x_1 и x_2 тесно связаны между собой ($r_{x1x2} = 0{,}91$), а это свидетельствует о наличии мультиколлинеарности. Следовательно, решаем оставить фактор x_2, а исключить фактор x_1, т.к. он слабее влияет на результат ($r_{x2\,y} = 0{,}68 > r_{x1\,y} = 0{,}66$).

Шаг 2. Проверим значимость коэффициентов корреляции. Основы проверки статистических гипотез приведены в Приложении А.

Проверка значимости осуществляется путем сопоставления табличного и расчетного значений t-статистики Стьюдента. Последняя определяется по формуле:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

Величина t следует t-распределению Стьюдента. Найденное по данной формуле значение t^* сопоставляем с табличным значением t_α при $(n-2)$ степенях свободы (приложение Б).

Шаг 3. Построить линейное уравнение регрессии, описывающее зависимость между факторами и результатом. Обозначим:

m - количество независимых переменных;

n - объем выборки;

$A = \{a_j\}, j = 0, 1, 2, ..., m$ - вектор оценок параметров регрессии;

$Y = \{y_i\}, i = 0, 1, 2, ..., n$ - вектор значений зависимой переменной;

$X = \{x_{ij}\}, j = 0, 1, 2, ..., m$ - матрица значений независимых переменных.

Вектор оценок параметров определяется по формуле:

$$A = (X^T X)^{-1} \cdot X^T y.$$

Для определения вектора A необходимо по данным наблюдения найти матрицу, обратную матрице $X^T X$**,** и вектор $X^T y$. Матрица X в развернутом виде:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \ldots & x_{2m} \\ . & . & . & \ldots & \ldots & \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nm} \end{bmatrix},$$

Рассчитаем параметры модели с помощью матричных функции *Excel* (Приложение В).

Получена модель, описывающую зависимость между объемом реализации продукта, расходами на рекламу и индексом потребительских расходов:

$$\hat{y} = -470 + 0{,}025x_2 + 4{,}99x_4$$

Шаг 3. Рассчитать коэффициент множественной корреляции, частные коэффициенты корреляции, коэффициенты эластичности.

Расчеты проведите самостоятельно, используя Приложения Г и Д.

Шаг 4. Оформить отчет о проделанной работе.

ВАРИАНТЫ ЗАДАНИЙ

Вариант 1

Таблица 1

№	Валовой продукт, млн. руб.	Балансовая стоимость оборудования, млн. руб.	Объем промышленного производства, млн. руб.	Количество занятых, тыс.чел.
i	y	x_1	x_2	x_3
1	1365	3938	625	161
2	1398	3002	475	186
3	1969	6269	528	267
4	1000	2270	242	129
5	761	1879	197	100
6	1253	4709	407	149
7	1590	3976	702	144
8	1425	2946	521	153
9	1127	3151	369	128
10	1595	3424	447	178
11	1636	4110	645	207
12	2110	3910	717	171
13	1131	3045	299	99
14	2005	4575	464	166
15	768	2126	169	105
16	1682	4692	579	167
17	2146	5873	468	255
18	2865	6906	850	299
19	3133	8678	924	458
20	1706	3988	507	182
21	1456	3840	475	179
22	2616	6368	801	244
23	2657	6396	604	304
24	1538	3632	592	170
25	1249	2681	220	111
26	2960	6675	672	306
27	2255	5298	712	241
28	1423	3185	214	133
29	2488	7364	602	245
30	1426	3812	225	116

Вариант 2

Таблица 2

№	Валовой продукт, млн. руб.	Балансовая стоимость оборудования, млн. руб.	Объем промышленного производства, млн. руб.	Количество занятых, тыс.чел.
i	y	x_1	x_2	x_3
1	4197	17125	1613	163
2	3562	14656	1273	177
3	5514	25701	1380	255
4	2710	10790	632	128
5	1975	10325	514	94
6	3209	13252	1073	142
7	4600	16397	2303	139
8	3478	12580	1395	146
9	2618	12854	907	122
10	3875	14790	1096	174
11	4148	16799	1745	201
12	4923	16228	1861	163
13	2690	11334	889	98
14	4756	18798	1595	160
15	1755	8672	477	103
16	4171	19553	1538	169
17	6059	23146	1260	247
18	6545	29168	2191	295
19	8782	40712	2475	452
20	4153	15059	1315	176
21	3920	17104	1210	174
22	6047	26266	2287	240
23	6811	30628	1598	292
24	3567	15688	1378	162
25	2581	11840	593	110
26	7677	27988	1621	289
27	6923	23448	2225	227
28	3348	12346	475	124
29	6432	31504	1873	237
30	2777	19119	559	123

Вариант 3

Таблица 3

№	Валовой продукт, млн. руб.	Балансовая стоимость оборудования, млн. руб.	Объем промышленного производства, млн. руб.	Количество занятых, тыс.чел.
i	y	x_1	x_2	x_3
1	4987	45778	1932	162
2	4279	38461	1630	175
3	6721	70406	1853	253
4	3219	26535	818	129
5	2315	22163	642	93
6	4276	40832	1408	143
7	5021	41566	2649	138
8	4137	33209	1367	143
9	3355	36655	1149	122
10	4791	42714	1437	170
11	5115	50006	1987	202
12	6046	48984	2250	165
13	2989	30206	903	96
14	6104	49745	1856	158
15	2500	19844	594	105
16	5943	58668	1936	172
17	7355	61742	1475	253
18	9644	73772	3042	296
19	10452	95564	3337	442
20	6033	42046	1830	182
21	5018	45102	1517	175
22	8610	71929	3019	234
23	8262	75885	2118	288
24	4929	42592	1708	158
25	3161	18392	769	113
26	10624	76039	2285	279
27	9513	64083	3052	233
28	3860	32793	690	123
29	8329	74201	2412	233
30	4178	45298	726	118

Вариант 4

По данным 30 наблюдений постройте модель множественной регрессии удовлетворительного качества (табл. 3.4). Рассчитайте прогноз результата, если прогнозные значения независимых факторов будут составлять 89% от их среднего уровня.

Таблица 4

№	Валовой продукт, млн. руб.	Балансовая стоимость оборудования, млн. руб.	Объем промышленного производства, млн. руб.	Количество занятых, тыс.чел.
i	y	x_1	x_2	x_3
1	6429	44502	2572	153
2	5265	38390	1960	174
3	8027	68857	2162	241
4	3590	31650	959	122
5	2962	26353	894	88
6	5601	43109	1831	143
7	5603	40878	2886	132
8	4778	35776	1558	133
9	4014	37539	1432	116
10	6008	54230	1675	159
11	6190	50117	2312	189
12	7515	55941	2404	159
13	3796	29658	1180	85
14	7632	52303	1904	144
15	2907	19397	677	105
16	7484	61771	2476	168
17	9743	71384	1671	241
18	10652	72981	3242	263
19	13424	98853	3118	417
20	6656	48372	2350	175
21	5657	50514	1725	170
22	9651	67027	3206	227
23	9803	73278	2518	281
24	5425	39549	1959	150
25	3784	28495	770	98
26	11913	88151	2256	258
27	9867	63565	1918	229
28	4359	31248	651	112
29	10506	52097	2726	226
30	4911	45076	756	112

Вариант 5

Таблица 5

№	Валовой продукт, млн. руб.	Балансовая стоимость оборудования, млн. руб.	Объем промышленного производства, млн. руб.	Количество занятых, тыс.чел.
i	y	x_1	x_2	x_3
1	11269	49834	4406	169
2	8740	39381	3518	180
3	12992	67977	3471	268
4	5744	31643	1777	127
5	4769	26884	1353	91
6	7957	39625	2843	153
7	10625	40718	5610	145
8	7654	32934	2209	137
9	7192	38551	2743	122
10	9425	53544	2726	166
11	10012	50181	3998	194
12	12092	56124	4776	163
13	6746	29025	2647	89
14	12284	52154	3893	149
15	5416	22855	1510	105
16	13727	61445	5280	174
17	13581	64618	2780	246
18	15887	69583	5723	278
19	22837	95113	5340	431
20	12328	50083	4749	185
21	9084	49066	3324	186
22	18397	59871	5820	243
23	15962	73762	4235	289
24	8505	38833	3033	156
25	5666	28139	1249	104
26	18306	85022	3537	272
27	12422	60576	2695	236
28	7061	28439	1015	114
29	18285	53924	4975	231
30	7021	44981	997	112

Вариант 6

Таблица 6

№	Валовой продукт, млн. руб.	Балансовая стоимость оборудования, млн. руб.	Объем промышленного производства, млн. руб.	Количество занятых, тыс.чел.
i	y	x_1	x_2	x_3
1	14816	48598	5927	168
2	11749	43093	5153	181
3	17755	70635	4742	265
4	8554	35263	3173	126
5	5923	27908	1910	91
6	10819	41487	3739	155
7	16007	43040	8758	145
8	10604	36795	3263	137
9	9968	41751	3871	122
10	12719	58613	4086	167
11	14635	52183	6442	195
12	15521	61816	6514	167
13	9430	30269	3624	92
14	20856	66116	6126	151
15	8194	25238	2068	108
16	19613	74340	8145	175
17	19160	68529	4068	248
18	23128	69608	8151	268
19	31436	99719	8205	437
20	18597	60060	7838	190
21	12706	49344	4703	182
22	26952	78010	9110	248
23	22773	89659	5992	286
24	10966	41983	4146	156
25	7232	30569	1662	104
26	25505	100766	5221	274
27	16161	63420	4080	238
28	10060	35084	1340	117
29	21332	55414	5812	231
30	8861	47219	1277	113

Приложение А. Статистическая проверка гипотез

Оценку генерального параметра получают на основе выборочного показателя с учетом ошибки репрезентативности. ***Ошибка выборки*** – это разница между значениями показателя, полученного по выборке, и генеральным параметром. В другом случае в отношении свойств генеральной совокупности выдвигается некоторая гипотеза о величине средней, дисперсии, характере распределения, форме и тесноте связи между переменными. Проверка гипотезы осуществляется на основе выявления согласованности эмпирических данных с гипотетическими (теоретическими). *Если расхождение между сравниваемыми величинами не выходит за пределы случайных ошибок, гипотезу принимают.* При этом не делается никаких заключений относительно правильности самой гипотезы, речь идет лишь *о согласованности сравниваемых данных*. Основой проверки статистических гипотез являются данные случайных выборок. При этом безразлично, оцениваются ли гипотезы в отношении реальной или гипотетической генеральной совокупности. Последнее открывает путь применения этого метода за пределами собственно выборки: при анализе результатов эксперимента данных сплошного наблюдения, но малой численности. В этом случае рекомендуется проверить, не вызвана ли установленная закономерность стечением случайных обстоятельств, насколько она характерна для того комплекса условий, в которых находится изучаемая совокупность.

Статистической гипотезой (обозначается H) называется произвольное предположение о свойстве генеральной совокупности, которое проверяется, опираясь на данные выборки. Так может быть выдвинута гипотеза о том, что средняя μ в генеральной совокупности равна некоторой величине а (записывается H: $\mu = a$) или о том, что генеральная средняя больше некоторой величины H: $\mu > b$. Различают *простые* и *сложные* гипотезы. ***Гипотеза называется простой***, если она однозначно характеризуется параметром распределения случайной величины. Например, H: $\mu = a$. ***Гипотеза называется сложной***, если она состоит из конечного или бесконечного числа простых гипотез, при этом указывается некоторая область вероятных значений параметра. Например, H: $\mu > b$. Эта гипотеза состоит из множества простых гипотез H: $\mu = c$, где c – любое число, большее *в*. Гипотезы о параметрах генеральной совокупности называются параметрическими, о распределениях – ***непараметрическими***.

Гипотеза о том, что две совокупности, сравниваемые по одному или нескольким признакам, не отличаются, называется ***нулевой гипотезой***, или ***нуль-гипотезой*** (обозначается H_0). При этом предполагается, что действительное различие сравниваемых величин равно нулю, а выявленное по данным отличие от нуля носит случайный характер. Например, H_0: $\mu 1 = \mu 2$, и т.д.

Нулевая гипотеза отвергается в том случае, если по выборке получается результат, который при истинности выдвинутой нулевой гипотезы *маловероятен*. *Границей невозможного или маловероятного* обычно считают α = 0,05, т.е. 5%, или 0,01, 0,001. Как правило, значения критериев в статистико-математических таблицах рассчитаны для вероятностей ошибки 0,05; 0,01; 0,001.

Статистическим критерием называют правило, устанавливающее условия отклонения проверяемой нулевой гипотезы. Критерии проверки статистических гипотез – это показатели, вычисляемые на основании фактических наблюдений, позволяющие сделать вывод о принятии или опровержении проверяемой гипотезы.

Проверка статистических гипотез состоит из следующих этапов:

- формулируется в виде статистической гипотезы задача исследования;
- выбирается статистическая характеристика гипотезы;
- выбираются испытуемая и альтернативная гипотезы на основе анализа возможных ошибочных явлений и их последствий;
- определяется область допустимых значений, критическая область, а также критическое значение статистического критерия (t; F; χ^2) по соответствующей таблице;
- вычисляется фактическое значение статистического критерия;
- проверяется гипотеза на основе сравнения фактического и критического значений критерия, и в зависимости от результатов проверки гипотеза либо отклоняется, либо нет.

При проверке гипотез по одному из критериев возможны два ошибочных решения:

1) неправильное отклонение H_0: ошибка 1-го рода;

2) неправильное принятие H_0: ошибка 2-го рода.

В то время как фактически H_0 верна (1) и H_0 не верна (2), принимают два ошибочных решения:

- H_0 отклоняется и принимается альтернативная гипотеза;
- H_0 не отклоняется.

Вероятности, соответствующие неверным решениям, называются риском 1 и риском 2. Риск 1 равен вероятности ошибки α (уровню значимости), риск 2 равен вероятности ошибки β. Поскольку α всегда больше 0, то всегда есть риск ошибки β. Обычно задают значение α и пытаются сделать β возможно малым. Вероятность 1-β называется *мощностью критерия*: чем она больше, тем меньше вероятность ошибки 2-го рода.

Альтернативная гипотеза H_1 может быть сформулирована по-разному в зависимости от того, какие отклонения от гипотетической величины нас особенно беспокоят: положительные, отрицательные, либо и те, и другие. Соответственно альтернативные гипотезы могут быть записаны:

$$H_1: \mu > a, H_1: \mu < a, H_1: \mu \neq a.$$

Приложение Б. Процентные точки распределения Стьюдента

Степень свободы	Вероятность									
	40 %	25 %	10 %	5 %	2,5 %	1 %	0,5 %	0,25 %	0,1 %	0,05 %
1	0,3249	1,0000	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	318,3088	636,6192
2	2887	0,8561	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	22,3271	31,5991
3	2767	7649	6377	3534	3,1824	4,5407	5,8409	7,4533	10,2145	12,9240
4	2707	7407	5332	1318	2,7764	3,7469	4,6041	5,5976	7,1732	8,6103
5	2672	7267	4759	2,0150	5706	3649	4,0321	4,7733	5,8934	6,8688
6	0,2648	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2076	5,9588
7	2632	7111	4149	8946	3646	2,9980	4995	4,0293	4,7854	4079
8	2619	7064	3968	8595	3060	8965	3554	3,8325	5008	5,0413
9	2610	7027	3830	8331	2622	8214	2498	6897	2968	4,7809
10	2602	6998	3722	8125	2281	7638	1693	5814	1437	5869
11	0,2696	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0247	4,4370
12	2590	6955	3562	7823	1788	6810	0554	4284	3,9296	3178
13	2586	6934	3502	7709	1604	6503	3,0123	3725	8520	2208
14	2582	6924	3450	7613	1448	6245	2,9768	3257	7874	1405
15	2579	6912	3406	7530	1314	6025	9467	2860	7328	0728
16	0,2576	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6862	4,0150
17	2573	6892	3334	7396	1098	5696	8982	2224	6458	3,9651
18	2571	6884	3304	7341	1009	5524	8784	1966	6105	9216
19	2569	6874	3277	7291	0930	5395	8609	1737	5794	8834
20	2567	6870	3253	7247	0860	5280	8453	1534	5518	8495

Приложение В. Матричные функции Excel.

Для работы матричными функциями Excel можно:
обратиться к Мастеру функций и выбрать нужную категорию функций, затем указать имя функции и задать соответствующие диапазоны ячеек;
б) ввести с клавиатуры имя функции задать соответствующие диапазоны ячеек.

Транспонирование матрицы осуществляется с помощью функции **ТРАНСП** (категория функций – Ссылки и массивы). Обращение к функции имеет вид:

ТРАНСП (диапазон ячеек),

где параметр диапазон ячеек задает все элементы транспонируемой матрицы (или вектора).

Умножение матриц осуществляется с помощью функции **МУМНОЖ** (категория функций – Математические). Обращение к функции имеет вид:

МУМНОЖ(*диапазон*_1;*диапазон*_2),

где параметр *диапазон*_1 задает элементы первой из перемножаемых матриц, а параметр *диапазон*_2 – элементы второй матрицы. При этом перемножаемые матрицы должны иметь соответствующие размеры (если первая матрица ($n \times k$), вторая – ($k \times m$) , то результатом будет матрица ($n \times m$)).

Обращение матрицы (вычисление обратной матрицы) осуществляется с помощью функции **МОБР** (категория функций – Математические). Обращение к функции имеет вид:

МОБР (*диапазон ячеек*),

где параметр *диапазон ячеек* задает все элементы обращаемой матрицы, которая должна быть квадратной и невырожденной.

При использовании этих функций необходимо соблюдать следующий порядок действий:

- выделить ячейки, в которые будет занесен результат выполнения матричных функций (при этом надо учитывать размеры исходных матриц);
- ввести арифметическое выражение, содержащее обращение к матричным функциям Excel;
- одновременно нажать клавиши [Ctrl], [Shift], [Enter]. Ес-ли этого не сделать, то вычислится только один элемент результирующей матрицы или вектора.

Приложение Г. Сравнение коэффициентов регрессии

Для сравнения влияния факторов и установления относительной важности каждого из них прибегают к нормированию коэффициентов регрессии.

Нормирование осуществляют следующим образом:

$$\beta_j = a_j\left(\frac{s_{x_j}}{s_y}\right).$$

где β_j — коэффициент регрессии после нормирования (бэта-коэффициент);
s_{xj}— средняя квадратичная ошибка переменной x_j,

$$s_{x_j} = a_j \cdot \sqrt{\frac{\sum_{i=1}^{n}(x_{ji} - \bar{x}_j)^2}{n-1}}.$$

S_y — средняя квадратичная ошибка y:

$$s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n-1}}.$$

Для сравнения коэффициентов регрессии (сравнения силы влияния каждого фактора на отклик) также может быть использован коэффициент эластичности:

$$Э_j = a_j \frac{\overline{x_j}}{\bar{y}}$$

Коэффициент эластичности показывает, на сколько процентов изменяется зависимая переменная при изменении соответствующего фактора на 1%.

Приложение Д. Коэффициенты множественной и частной корреляции

Теснота линейной взаимосвязи переменной X_i с совокупностью других (p-1) переменных X_j, рассматриваемых в целом, измеряется с помощью коэффициента множественной корреляции R, который представляет собой обобщение парного коэффициента корреляции.

В случае трех переменных (p=3) расчет ведется по формуле

$$R_{i.jk} = \sqrt{\frac{r_{ij}^2 + r_{ik}^2 - 2 \cdot r_{ij} \cdot r_{ik} \cdot r_{jk}}{1 - r_{jk}^2}}$$

Коэффициент множественной корреляции служит измерителем тесноты взаимосвязи переменных (но не о направлении этой связи). Чем ближе R к ±1, тем теснее связь.

Величина $\boldsymbol{R}^2$, называемая множественным коэффициентом детерминации, показывает, какую долю вариации исследуемой переменной (i) объясняет вариация остальных переменных (j, k).

Коэффициент частной корреляции

Если x — случайная переменная, то коэффициент корреляции является мерой тесноты линейной взаимосвязи между y н набором переменных x.

Если x — случайная переменная, то коэффициент корреляции является мерой тесноты линейной взаимосвязи между y н набором переменных x.

В качестве соответствующих измерителей приняты *коэффициенты частной корреляции.*

Рассмотрим коэффициенты частной корреляции для случая, когда во взаимосвязи находятся три случайные переменные, например y, x и z.

Для этих переменных можно получить простые коэффициенты парной корреляции r_{yx}, r_{yz} и r_{xz} (индексы указывают у на коррелируемые переменные).

Остановимся на одном из этих коэффициентов, например r_{yx}. Большая величина r_{yx} может возникнуть не только потому, что y и x действительно связаны между собой, но и в силу того, что оба эти признака испытывают влияние третьего фактора — z.

Коэффициент частной корреляции измеряет парную корреляцию соответствующих признаков (допустим, y и x) при условии, что влияние на них третьего фактора (z) устранено.

Можно показать, что при наличии трех переменных:

$$r_{yx\cdot z} = \frac{r_{yx} - r_{yz} \cdot r_{xz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}};$$

$$r_{yz\cdot x} = \frac{r_{yz} - r_{yx} \cdot r_{xz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yx}^2)}}.$$

Здесь $r_{yx.z}$ — коэффициент частной корреляции *y* и *x*, влияние *z* элиминировано, $r_{yz.x}$ — коэффициент частной корреляции *y* и *z* при устранении влияния *x*.

Поскольку все три переменные равноправны, то можно найти и третий коэффициент:

$$r_{xz\cdot y} = \frac{r_{xz} - r_{xy} \cdot r_{yz}}{\sqrt{(1 - r_{yz}^2)(1 - r_{xy}^2)}}.$$

Если один из коэффициентов r_{xy} или r_{yz} приближается к единице, то соответственно частный коэффициент корреляции $r_{xz.y}$ стремится к нулю.