

Отчет по 5 лабораторной работе «Множественная линейная регрессия»

Выполнила работу: студент группы 15.11Д-БИЦТ09/216 Крабу Кира Сергеевна

Для выполнения данной работы мы выбрали задачу для Варианта №2, с файлом Excel «Вариант 2.xlsx».

	A	B	C	D
1	Цена товара	Затраты на рекламу	Число конкурирующих организаций в регионе	Объем продаж
2	X1	X2	X3	Y
3	20	37	3	112
4	20	38	4	132
5	18	36	3	129
6	17	42	5	134
7	17	47	4	132
8	19	55	3	137
9	18	53	2	139
10	16	54	3	139
11	16	49	2	138
12	16	50	2	143
13	15	52	1	141
14	15	52	2	146
15	14	51	1	148
16	14	54	3	150
17				

Рисунок 1. Изначальные параметры

Далее, перейдя во вкладку «Данные» и выбрав пакет «Анализ данных», выбрали Регрессию, настроив ее так:

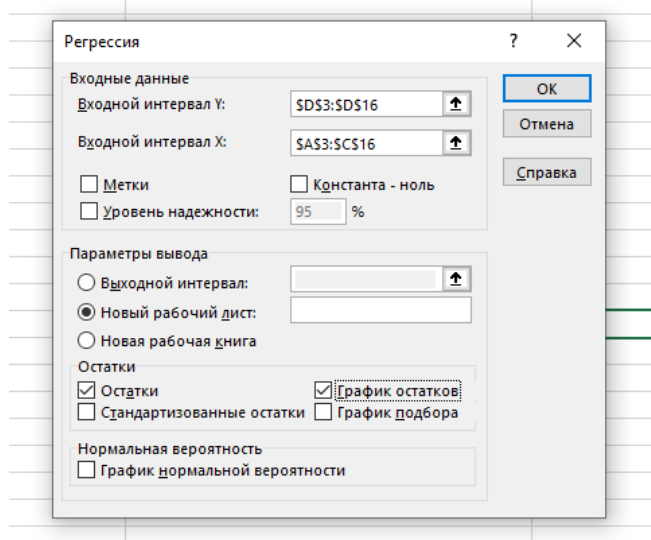


Рисунок 2. Данные для построения регрессии

Полученные данные вывелись на новый лист и выглядели так:

	A	B	C	D	E	F	G	H	I
1	Вывод ИТОГОВ								
2									
3	Регрессионная статистика								
4	Множественный R	0,883787904							
5	R-квадрат	0,781081059							
6	Нормированный R-кв	0,715405376							
7	Стандартная ошибка	5,081946501							
8	Наблюдения	14							
9									
10	Дисперсионный анализ								
11		df	SS	MS	F	Значимость F			
12	Регрессия	3	921,4524833	307,1508278	11,89300256	0,001230541			
13	Остаток	10	258,2618024	25,82618024					
14	Итого	13	1179,714286						
15									
16		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
17	У-пересечение	151,2030285	25,03426551	6,03984281	0,000125286	95,42320892	206,9828481	95,42320892	206,9828481
18	Переменная X 1	-2,718666292	0,933462354	-2,912454134	0,01549945	-4,79855003	-0,638782554	-4,79855003	-0,638782554
19	Переменная X 2	0,629647089	0,274565619	2,29324812	0,044765655	0,017876767	1,241417411	0,017876767	1,241417411
20	Переменная X 3	0,531069209	1,493457424	0,355597154	0,729528943	-2,796561301	3,858699719	-2,796561301	3,858699719
21									

Рисунок 3. Вывод итогов

23			
24	Вывод ОСТАТКА		
25			
26	Наблюдение	Предсказанное Y	Остатки
27	1	121,7198526	-9,719852589
28	2	122,8805689	9,119431114
29	3	126,5275381	2,472461916
30	4	134,0862253	-0,086225325
31	5	136,7033916	-4,703391559
32	6	135,7721665	1,227833525
33	7	136,7004694	2,29953062
34	8	143,2985183	-4,298518262
35	9	139,6192136	-1,61921361
36	10	140,2488607	2,751139302
37	11	143,695752	-2,695751958
38	12	144,2268212	1,773178833
39	13	145,7847712	2,215228839
40	14	148,7358508	1,264149155
41			
42			

Рисунок 4. Вывод остатка

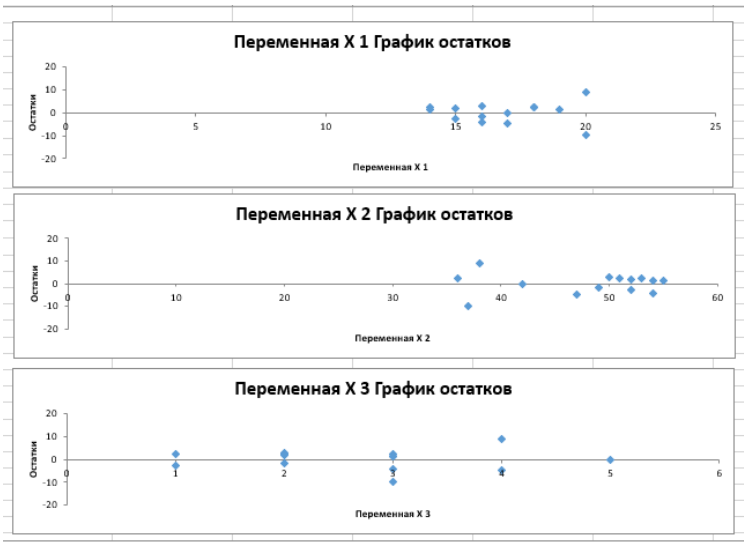


Рисунок 5. График остатков

Имея все эти данные, мы построили уравнение регрессии, которое в округленном виде переменных выглядит так:

$$y = 151,2 - 2,72x_1 + 0,63x_2 + 0,53x_3$$

Коэффициент детерминации равен 78,1%, что означает, что объем продаж на 78,1% зависит от изменений в цене, в затратах на рекламу и в конкуренции. Значимость F равняется 0,001, что меньше уровня значимости $\alpha = 0,05$, что подтверждает значимость коэффициента детерминации. Коэффициент Фишера (Р-значение) меньше уровня значимости $\alpha = 0,05$ в первых трех случаях (Y, x1, x2). Р-значение x3 больше пяти сотых, а значит 3 переменная не является статистически значимой.

Используя тот же самый пакет анализа данных, на новом листе была построена корреляционная матрица:

	A	B	C	D	E	
1		X1	X2	X3	Y	
2	X1	1				
3	X2	-0,61499	1			
4	X3	0,509716	-0,49551	1		
5	Y	-0,81508	0,767452	-0,44978	1	

Рисунок 6. Корреляционная матрица

Исходя из полученных данных, больше всего на объем продаж (Y) влияет переменная x1 – цена. Между переменными тесная корреляционная связь отсутствует, что исключает мультиколлинеальность.

Преобразуем модель, исключая фактор(-ы), не оказывающие заметного влияния на Y. Построим используем пакет «Анализ данных» и регрессию для вывода зависимости между Y и x1 (самым сильным фактором), дальше добавляем другие факторы: Y, x1 и x2, Y, x1 и x3.

Для Y и x1:

Рисунок 7. Входные данные

	А	В
1	ВЫВОД ИТОГОВ	
2		
3	<i>Регрессионная статистика</i>	
4	Множественный R	0,815082938
5	R-квадрат	0,664360195
6	Нормированный R-	0,636390212
7	Стандартная ошибка	5,744265781
8	Наблюдения	14
9		

Рисунок 8. Вывод итогов

Нормированный R-квадрат тут равен 0,636. Коэффициент детерминации здесь – 66,4%.

Для Y, x1, x2:

Рисунок 9. Входные данные

1	ВЫВОД ИТОГОВ	
2		
3	<i>Регрессионная статистика</i>	
4	Множественный R	0,882220405
5	R-квадрат	0,778312843
6	Нормированный R-квадрат	0,738006088
7	Стандартная ошибка	4,875984803
8	Наблюдения	14

Рисунок 10. Вывод итогов

Нормированный R-квадрат тут равен 0,738. Коэффициент детерминации здесь – 77,8%.

Для Y, x1, x3 (предварительно копируя таблицу и удалив столбец с x2, так как входной интервал должен быть непрерывным):

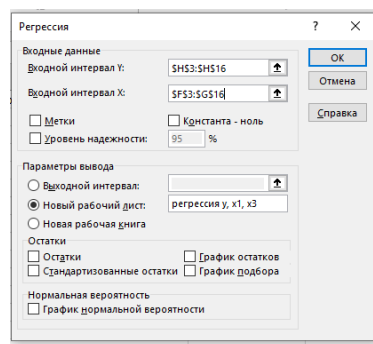


Рисунок 11. Входные данные

	A	B
1	ВЫВОД ИТОГОВ	
2		
3	<i>Регрессионная статистика</i>	
4	Множественный R	0,816058744
5	R-квадрат	0,665951873
6	Нормированный R-квадрат	0,60521585
7	Стандартная ошибка	5,985447101
8	Наблюдения	14

Рисунок 12. Вывод итогов

Нормированный R-квадрат тут равен 0,6. Коэффициент детерминации здесь – 66,6%.

Наиболее высокий результат показал результат добавления к Y и x1 фактора номер 2 (по нормированному R-квадрату и коэффициенту

детерминации). Следовательно, дальше будем работать с таким набором данных.

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	<i>Регрессионная статистика</i>								
4	Множественный R	0,882220405							
5	R-квадрат	0,778312843							
6	Нормированный R-к	0,738006088							
7	Стандартная ошибка	4,875984803							
8	Наблюдения	14							
9									
10	<i>Дисперсионный анализ</i>								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
12	Регрессия	2	918,18678	459,09339	19,30973675	0,0002521			
13	Остаток	11	261,5275057	23,7752278					
14	Итого	13	1179,714286						
15									
16		<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>
17	Y-пересечение	152,230763	23,85906602	6,38041585	5,22008E-05	99,71731274	204,7442132	99,71731274	204,7442132
18	Переменная X 1	-2,619324607	0,854580186	-3,065042519	0,010754689	-4,500242915	-0,738406298	-4,500242915	-0,738406298
19	Переменная X 2	0,603448658	0,253776872	2,377870979	0,03663872	0,04488953	1,162007787	0,04488953	1,162007787
20									

Рисунок 13. Полный вывод итогов для корреляционного анализа Y, x1, x2

Новое уравнение регрессии будет:

$$y = 152,23 - 2,62x_1 + 0,6x_2 \quad (1)$$

Значимость F равна 0,00025, что меньше уровня значимости $\alpha = 0,05$, а значит модель значима. Все P-значения также меньше этого значения, а значит все переменные значимы при заданном условии.

Рассчитаем эластичность каждого фактора по формуле:

$$\varepsilon_i = b_i \frac{\bar{x}_i}{\bar{y}}$$

b_i – коэффициент переменной, числитель дроби – среднее значение переменной, знаменатель – среднее значение объема продаж.

Среднее значение объема продаж: 137,1

Для фактора цены (x1): коэффициент равен -2,62; среднее – 16,8.

Для фактора затрат на рекламу (x2): его коэффициент равен 0,6; среднее значение – 47,9.

Следственно, эластичность первого фактора равна: -0,32105.

Эластичность второго: 0,209628.

Найдем среднюю относительную ошибку аппроксимации. Рассчитаем значения Y по формуле (1). Далее воспользуемся формулой:

$$A = \frac{1}{n} \sum \frac{|Y - Y_1|}{Y} * 100\%$$

В формуле: n – количество значений (ячейка G18), в столбце F – формула под знаком суммы ($Y_1 - Y$ по формуле (прогноз)). В ячейке F18 – сама сумма, а в F19 – средняя ошибка аппроксимации, она равна 2,6%, что говорит о качественности построенной модели.

	A	B	C	D	E	F	G
1	Цена товара	Затраты на рекламу	Число конкурирующих организаций в регионе	Объем продаж			
2	X1	X2	X3	Y	Y формула		
3	20	37	3	112	122,03	0,089553571	
4	20	38	4	132	122,63	0,070984848	
5	18	36	3	129	126,67	0,018062016	
6	17	42	5	134	132,89	0,008283582	
7	17	47	4	132	135,89	0,029469697	
8	19	55	3	137	135,45	0,011313869	
9	18	53	2	139	136,87	0,015323741	
10	16	54	3	139	142,71	0,026690647	
11	16	49	2	138	139,71	0,012391304	
12	16	50	2	143	140,31	0,018811189	
13	15	52	1	141	144,13	0,022198582	
14	15	52	2	146	144,13	0,012808219	
15	14	51	1	148	146,15	0,0125	
16	14	54	3	150	147,95	0,013666667	
17							
18						0,362057932	14
19						2,586128087	

Рисунок 14. Расчет средней ошибки аппроксимации

Дадим экономическое пояснение всему вышесказанному. Формула (1) говорит нам о том, что без воздействия фактора 1 и фактора 2, объем продаж будет равен ~152. Если цену поднять на единицу (фактор 1), то объем продаж снизится на ~2,6 единиц. Если же поднять затраты на рекламу на 1 единицу, то объем продаж увеличится на 0,6. Далее поговорим про эластичность факторов. Эластичность цены равна -0,32105, что означает, что при увеличении цены на 1%, объем продаж снизится на -0,32% (округлили). Эластичность затрат на рекламу равна 0,209628, что увеличивает объем продаж на 0,2% (опять же округлили) при увеличении затрат на рекламу на 1%.

Ответим на вопросы нашего варианта.

Как затраты на рекламу влияют на объем продаж?

Как было сказано ранее, если поднять затраты на рекламу на 1 единицу, то объем продаж увеличиться на 0,6. Эластичность затрат на рекламу равна 0,209628, что увеличивает объем продаж на 0,2% (опять же округлили) при увеличении затрат на рекламу на 1%.

Какой ожидаемый объем продаж, если планируемая цена на товар 15, а бюджет на рекламу 60?

Для этого подставим эти значения в формулу (1):

$$y = 152,23 - 2,62 * 15 + 0,6 * 60 = 148,93$$