

Модель линейной множественной регрессии. Фиктивные переменные. Спецификация модели

В качестве основной модели воспользуемся уравнением линейной множественной регрессией.

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p,$$

где y – результативная переменная,

x_i – факторы.

В нашем случае результативной, т.е. моделируемой переменной будет рыночная стоимость автомобиля, измерять ее будем в денежных единицах. Количество факторов и их набор будем определять в ходе исследования. Сначала на основе качественных рассуждений необходимо сформировать список всех факторов, которые влияют на результативную переменную – стоимость автомобиля. Затем методами корреляционного анализа выделить те, которые статистически значимо влияют на моделируемую переменную.

В качестве источника информации воспользуемся рыночной информацией – объявлениями о продаже автомобилей. В них учитывается следующая информация:

- год выпуска автомобиля;
- пробег, км;
- тип кузова (седан, хетчбэк, универсал и т.д.);
- коробка передач (автомат, механика);
- объем двигателя, л;
- тип двигателя (бензиновый, дизельный, гибридный);
- привод (передний, задний, полный);
- руль (левый, правый);
- цвет;
- состояние (не битый, битый).

Часть перечисленных факторов являются количественными (год выпуска, пробег, объем двигателя), но большинство - качественные. Стоит

отметить, что показатель «объем двигателя» имеет низкую изменчивость, т.е. по сути дела тоже является качественным показателем. Характеристики, измеренные в номинативной или порядковой шкале, могут быть включены в регрессионную модель как фиктивные переменные.

Рассмотрим объявления о продаже автомобилей с пробегом в интернете, например, на сайте Drom.ru. Здесь в объявлениях о продаже указаны значения не всех факторов, перечисленных выше. Поэтому для первого численного эксперимента остановимся на модели следующего вида:

$$y = b_0 + b_1x_1 + b_2x_2 + \varepsilon,$$

где y – цена автомобиля, тыс. руб.;

x_1 – возраст автомобиля, лет;

x_2 – пробег, тыс.км.

Фактор x_1 будем измерять по формуле: $2014 - \text{год выпуска}$. В качестве фильтра отбора объявлений укажем наш регион – Краснодарский край. Для первого численного эксперимента рассмотрим автомобиль Hyundai Accent, для которого разброс года выпуска достаточный, а соответственно и пробег и цена. Это позволит построить модель, у которой вариация переменных достаточная для выявления статистических закономерностей. Выборка составляет 50 наблюдений.

Сначала проведем корреляционный анализ – рассчитаем матрицу коэффициентов парной корреляции. Воспользуемся инструментом *Корреляция* надстройки *Анализ данных* в электронных таблицах MS Excel. Корреляционная матрица представлена в таблице 2.1.

Таблица 2.1.

	Цена	Возраст	Пробег
Цена	1		
Возраст	-0,73669342	1	
Пробег	-0,39680497	0,093565312	1

Получили, что коэффициент линейной парной корреляции между результативной переменной – ценой и первым фактором – возрастом

машины $r_{yx_1} = -0,74$. Это говорит о том, что между ценой и возрастом сильная и обратная связь, т.к. $|r_{yx_1}| > 0,7$. Наличие обратной связи говорит о том, что с увеличением возраста машины, ее стоимость снижается и этот вывод соответствует общепринятым представлениям.

Коэффициент линейной парной корреляции между результативной переменной – ценой и вторым фактором – пробегом $r_{yx_2} = -0,40$. Это говорит о том, что между ценой и пробегом связь умеренная и обратная, т.к. $0,3 < |r_{yx_2}| < 0,7$. Наличие обратной связи говорит о том, что с увеличением пробега стоимость машины снижается, это тоже соответствует здравому смыслу. Стоит сказать, что связь возраста с ценой сильнее, чем связь пробега с ценой. Но наличие умеренной корреляции дает нам хороший повод включить фактор x_2 в модель.

Коэффициент линейной межфакторной корреляции близок к нулю $r_{x_1x_2} = 0,09$. Это означает, что связь между факторами практически отсутствует. Этот факт требует объяснения, потому что с увеличением возраста машины, как правило, увеличивается и пробег. Такая ситуация должна приводить к наличию корреляционной связи между возрастом и пробегом. Возможно, причина низкой корреляции заключается в достаточно сильной изменчивости характера эксплуатации автомобилей. Рассматриваемая модель машины в год может проезжать 10 тыс. км., а может 150 тыс. км. Таким образом, теряется связь между возрастом и пробегом на выборке автомобилей со средним объемом. Для построения регрессионной модели отсутствие межфакторной корреляции является положительным явлением.

Проведем регрессионный анализ с использованием инструмента *Регрессия* надстройки *Анализ данных* электронных таблиц MS Excel. Результаты анализа представлены на рис. 2.1.

	A	B	C	D	E	F	G
1	Вывод ИТОГОВ						
2							
3	<i>Регрессионная статистика</i>						
4	Множественный R	0,806950627					
5	R-квадрат	0,651169315					
6	Нормированный R-квадрат	0,636325456					
7	Стандартная ошибка	36,28327264					
8	Наблюдения	50					
9							
10	Дисперсионный анализ						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
12	Регрессия	2	115502,134	57751,06698	43,8679266	1,78421E-11	
13	Остаток	47	61874,36605	1316,475873			
14	Итого	49	177376,5				
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>
17	Y-пересечение	464,3769493	25,09884895	18,50192215	3,25366E-23	413,8845789	514,8693197
18	Переменная X 1	-16,73067185	2,051320749	-8,156048663	1,49134E-10	-20,85739684	-12,60394686
19	Переменная X 2	-0,674960212	0,17657026	-3,822615504	0,000387393	-1,030173752	-0,319746673

Рис. 2.1. Результаты регрессионного анализа для двухфакторной модели стоимости Hyundai Accent

Полученные результаты позволяют сделать следующие выводы.

Множественный коэффициент корреляции $R = 0,81$. Т.к. его значение больше 0,7, то можно сказать о сильной связи между результативной переменной (ценой) и выбранными факторами. Далее указаны значения R-квадрат и Нормированный R-квадрат. Это значения коэффициента детерминации и скорректированного коэффициента детерминации. Как известно, если множественный коэффициент корреляции возвести в квадрат, то получим смещенную оценку коэффициента детерминации. Поэтому рассчитывают скорректированный коэффициент детерминации. Его значение несколько меньше. С увеличением объема выборки разница между этими показателями уменьшается и можно использовать просто квадрат коэффициента корреляции. В нашем случае, $R^2 = 0,64$. Это означает, что 64% вариации результативной переменной – цены объясняется вариацией выбранных факторов – возрастом и пробегом, а оставшиеся 36% другими – неучтенными в модели причинами. Полученное значение является достаточно высоким и приемлемым для принятия решений. Исходя из того, что коэффициент детерминации является характеристикой качества модели, то можно сделать вывод о достаточном уровне качества.

Следующий раздел расчетов – дисперсионный анализ. С помощью критерия Фишера проверяется статистическая значимость уравнения регрессии в целом. Фактическое (наблюдаемое) значение критерия равно 43,87 и соответствующий ему уровень значимости $\alpha = 1,78420604560769E - 11$. Число представлено в электронных таблицах в экспоненциальной форме, традиционная математическая запись:

$$\alpha = 1,78420604560769 \cdot 10^{-11},$$

т.е. первые отличные от нуля цифры находятся на 11 месте после запятой. Это число очень близко к нулю, а т.к. оно характеризует вероятность ошибки, то для полученной модели это означает, что признать уравнение регрессии статистически значимым можно с очень высокой вероятностью.

Следующий раздел содержит информацию о значениях коэффициентов регрессии, полученная модель имеет вид:

$$y = 464,38 - 16,73x_1 - 0,67x_2 + \varepsilon.$$

Свободный член уравнения регрессии часто не поддается интерпретации, но в нашем случае его можно рассматривать как некоторую начальную стоимость автомобиля с нулевым возрастом и пробегом.

Коэффициент регрессии перед фактором x_1 говорит о том, что при увеличении возраста машины на 1 год, ее стоимость в среднем снижается на 16,73 тыс. руб. Коэффициент регрессии перед фактором x_2 говорит о том, что при увеличении пробега на 1 тыс. км. ее стоимость снижается на 0,67 тыс.руб. Полученные значения проверяются на статистическую значимость с помощью t-критерия Стьюдента. В столбце p -значение находятся уровни значимости, на которых коэффициенты регрессии признаются статистически значимо отличными от нуля. Для всех коэффициентов получены хорошие оценки, уровни значимости меньше 0,001. Хотя для практики обычно достаточно 0,01.

Рассчитаем характеристику точности модели – среднюю ошибку аппроксимации:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{1}{50} \cdot 4,26 \cdot 100\% = 8,52\%.$$

Считается, что при значении ошибки на уровне 8-10% точность модели хорошая.

В итоге нами получено статистически значимое, качественное уравнение линейной множественной регрессии с хорошими характеристиками точности, которое можно использовать для определения рыночной цены на конкретный автомобиль рассматриваемой модели.

Попробуем включить в модель качественный фактор – «коробка передач» и посмотрим, улучшит ли это качество модели. Эта переменная измерена в шкале наименований и принимает два значения – «механика» и «автомат». Можно дать этим позициям числовые имена:

$$x_3 = \begin{cases} 1, & \text{если автомат,} \\ 2, & \text{если механика.} \end{cases}$$

Дискретные независимые переменные включаются в регрессионную модель с помощью, так называемых, фиктивных переменных. Количество фиктивных переменных на единицу меньше количества значений, которые принимает качественный фактор. В нашем случае качественный фактор принимает два значения, значит достаточно одной фиктивной переменной, которую мы сформируем следующим образом:

$$z = \begin{cases} 1, & \text{если механика,} \\ 0, & \text{в противном случае.} \end{cases}$$

Уравнение регрессии в этом случае имеет вид:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3z + \varepsilon,$$

где y – цена автомобиля, тыс. руб.;

x_1 – возраст автомобиля, лет;

x_2 – пробег, тыс.км.;

z - фиктивная переменная (коробка передач).

Проведем регрессионный анализ теми же инструментальными средствами, т.е. с помощью надстройки *Анализ данных/Регрессия* (рис. 2.2.)

	A	B	C	D	E	F	G
1	ВЫВОД ИТОГОВ						
2							
3	<i>Регрессионная статистика</i>						
4	Множественный R	0,820683508					
5	R-квадрат	0,673521421					
6	Нормированный R-квадрат	0,65222934					
7	Стандартная ошибка	35,48105057					
8	Наблюдения	50					
9							
10	Дисперсионный анализ						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
12	Регрессия	3	119466,8723	39822,29078	31,63248409	3,00231E-11	
13	Остаток	46	57909,62767	1258,904949			
14	Итого	49	177376,5				
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>
17	Y-пересечение	475,9266322	25,39212618	18,74307921	3,60887E-23	424,814934	527,0383305
18	Переменная X 1	-16,69152795	2,006087375	-8,320439156	1,00106E-10	-20,72957233	-12,65348356
19	Переменная X 2	-0,659928225	0,172873937	-3,8173957	0,000401436	-1,007905407	-0,311951043
20	Переменная X 3	-19,45752728	10,96419753	-1,774642169	0,082574819	-41,52731189	2,612257325

Рис. 2.2. Результаты регрессионного анализа для трехфакторной модели стоимости Hyundai Accent

Получили следующие результаты.

Множественный коэффициент корреляции $R = 0,82$ - связь между ценой автомобиля и включенными в модель факторами сильная, коэффициент больше чем 0,7.

Скорректированный коэффициент детерминации $R^2 = 0,65$, т.е. 65% вариации стоимости автомобилей с пробегом данной модели объясняется вариацией возраста автомобиля, его пробега и вида коробки передач. Прирост коэффициента детерминации после включения качественного фактора в модель составил всего 1% - это небольшой прирост, нужны дополнительные основания, чтобы считать целесообразным использование трехфакторной модели, а не двухфакторной.

По критерию Фишера уравнение регрессии можно признать статистически значимым с большой доверительной вероятностью, уровень значимости имеет порядок 10^{-11} .

Уравнение регрессии имеет вид:

$$y = 475,93 - 16,70x_1 - 0,66x_2 - 19,46z + \varepsilon.$$

Как видно, коэффициенты регрессии при первых двух факторах фактически не изменились, интерпретация их сохранилась. Коэффициент регрессии при фиктивной переменной позволяет сделать вывод, что стоимость автомобилей с одинаковым возрастом и пробегом дешевле на 19,46 т.р. для механической коробки передач, по сравнению с коробкой передач – автоматом.

Как и в предыдущем варианте модели, коэффициенты регрессии при первом и втором факторах статистически значимы с вероятностью ошибки меньше 0,001. Несколько выше уровень значимости для выводов по критерию Стьюдента для коэффициента регрессии при фиктивной переменной, его можно признать статистически значимо отличающимся от нуля только с вероятностью 91% (уровень значимости $\alpha = 0,082$). Считаем, что с учетом целей нашего исследования такая вероятность является приемлемой.

Рассчитаем коэффициенты парной корреляции для нового фактора:

$$r_{yz} = -0,28; r_{x_1z} = 0,13; r_{x_2z} = 0,19.$$

Связь типа коробки передач с ценой можно считать скорее слабой, чем умеренной, а межфакторная корреляция еще меньше в два раза, но это уже положительный для качества и адекватности модели факт. Поэтому будем считать, что есть основания для включения в модель фиктивной переменной «тип коробки передач».

Рассчитаем характеристику точности модели – среднюю ошибку аппроксимации:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{1}{50} \cdot 0,53 \cdot 100\% = 1,06\%.$$

Получили существенное увеличение точности модели, средняя относительная ошибка снизилась с 8,5% до 1%. Считаем это важным достоинством включения фиктивной переменной в модель.

Таким образом, включение качественного фактора «тип коробки передач» повышает точность модели, позволяет дать полезную качественную интерпретацию коэффициенту регрессии при фиктивной переменной.

Наша выборка имеет объем, позволяющий построить два частных уравнения регрессии – отдельно для каждой коробки передач. Результаты регрессионного анализа представлены на рисунках 2.3 и 2.4.

	A	B	C	D	E	F	G
1	Вывод итогов						
2							
3	<i>Регрессионная статистика</i>						
4	Множественный R	0,799283327					
5	R-квадрат	0,638853837					
6	Нормированный R-квадрат	0,57866281					
7	Стандартная ошибка	51,98006403					
8	Наблюдения	15					
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	57355,27532	28677,63766	10,61377197	0,002218697	
13	Остаток	12	32423,12468	2701,927057			
14	Итого	14	89778,4				
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>
17	Y-пересечение	530,2404121	56,21597096	9,432202327	6,70982E-07	407,7563334	652,7244907
18	Переменная X 1	-21,57011941	8,419392546	-2,561956732	0,024911656	-39,91439989	-3,225838936
19	Переменная X 2	-0,824490157	0,45710516	-1,803720957	0,09641949	-1,820436743	0,17145643

Рис. 2.3. Результаты регрессионного анализа для машин с автоматической коробкой передач и двумя факторами

Получили следующие результаты.

Автоматическая коробка передач:

$$y = 530,24 - 21,57x_1 - 0,82x_2 + \varepsilon, \quad R^2 = 57,8\%.$$

Механическая коробка передач:

$$y = 415,00 - 15,13x_1 - 0,41x_2 + \varepsilon, \quad R^2 = 73\%.$$

В первом случае автомобиль в среднем дороже, но и быстрее дешевеет с возрастом и пробегом. Второе частное уравнение имеет наивысший коэффициент детерминации, что позволяет использовать именно его для моделирования стоимости автомобиля с механической коробкой, а для автоматической коробки можно использовать уравнение с фиктивной переменной.

	A	B	C	D	E	F	G
1	ВЫВОД ИТОГОВ						
2							
3	<i>Регрессионная статистика</i>						
4	Множественный R	0,863833796					
5	R-квадрат	0,746208827					
6	Нормированный R-квадрат	0,730346879					
7	Стандартная ошибка	25,50854115					
8	Наблюдения	35					
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	61221,65851	30610,82925	47,04395777	2,96227E-10	
13	Остаток	32	20821,94149	650,6856716			
14	Итого	34	82043,6				
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>
17	Y-пересечение	415,0034902	24,10197791	17,21864868	9,66225E-18	365,909368	464,0976125
18	Переменная X 1	-15,12975423	1,583975549	-9,551759958	6,87817E-11	-18,35620682	-11,90330163
19	Переменная X 2	-0,408784839	0,165104939	-2,47590921	0,018768684	-0,745092593	-0,072477085

Рис. 2.4. Результаты регрессионного анализа для машин с механической коробкой передач и двумя факторами

Рассмотрим фактор «объем двигателя». С одной стороны он является количественным, т.е. принимает числовые значения и измеряется в метрической шкале. Но с другой стороны, он стандартизирован настолько, что встречается всего три его значения в нашей выборке: 1400; 1500 и 1600 (куб.см). Поэтому будем включать в модель этот фактор разными способами – как количественный и как качественный показатель.

Включаем в модель фактор x_4 - объем двигателя (куб. см.). Корреляционная матрица представлена в таблице 2.2.

Таблица 2.2.

	Цена	Возраст	Пробег	Коробка передач	Объем двигателя
Цена	1				
Возраст	-0,737	1			
Пробег	-0,397	0,094	1		
Коробка передач	-0,177	0,018	0,050	1	
Объем двигателя	-0,108	-0,049	-0,024	0,028	1

Как видно, корреляция показателя «объем двигателя» с другими факторами фактически отсутствует, по модулю меньше 0,1. Но с другой стороны, связь переменной «объем двигателя» с результативной переменной

тоже очень слабая $r_{yx_4} = -0,108$. Посмотрим, что получится в результате регрессионного анализа, расчеты представлены на рис. 2.5.

	A	B	C	D	E	F	G
1	ВЫВОД ИТОГОВ						
2							
3	<i>Регрессионная статистика</i>						
4	Множественный R	0,833657206					
5	R-квадрат	0,694984337					
6	Нормированный R-квадрат	0,667871834					
7	Стандартная ошибка	34,67391294					
8	Наблюдения	50					
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	4	123273,8893	30818,47232	25,6333518	4,15267E-11	
13	Остаток	45	54102,61073	1202,280238			
14	Итого	49	177376,5				
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>
17	Y-пересечение	766,744171	165,3028539	4,638420649	3,03392E-05	433,8071376	1099,681204
18	Переменная X 1	-16,85847918	1,962695772	-8,589451006	4,8587E-11	-20,81155132	-12,90540703
19	Переменная X 2	-0,666284537	0,168979088	-3,942999955	0,000278402	-1,006625886	-0,325943188
20	Переменная X 3	-18,87988009	10,71969567	-1,761232844	0,084992749	-40,47045515	2,710694964
21	Переменная X 4	-0,19004891	0,106801127	-1,779465409	0,081917693	-0,405157418	0,025059597

Рис. 2.5. Результаты регрессионного анализа для автомобилей Hyundai Accent с четырьмя факторами

Получили следующее уравнение регрессии:

$$y = 766,74 - 16,86x_1 - 0,67x_2 - 18,88z - 0,19x_4 + \varepsilon,$$

где y – цена автомобиля, тыс. руб.;

x_1 – возраст автомобиля, лет;

x_2 – пробег, тыс.км.;

$z = \begin{cases} 1, & \text{если механика,} \\ 0, & \text{в противном случае,} \end{cases}$

x_4 - объем двигателя (куб. см.).

После добавления в модель еще одного фактора коэффициенты регрессии при «старых» факторах практически не изменились. Коэффициент перед четвертым фактором говорит о том, что увеличение объема двигателя на 100 куб. см. приводит к снижению стоимости автомобиля в среднем на 19 тыс. руб. Выходит, что автомобили Hyundai Accent с меньшим объемом двигателя стоят несколько дороже.

Множественный коэффициент корреляции немного увеличился, с 0,82 до 0,83. Соответственно, увеличение коэффициента детерминации тоже незначительно, с 65,2% до 66,8%. Критерий Фишера позволяет сделать вывод о статистической значимости уравнения регрессии в целом, т.к. «Значимость F» < 0,001. По критерию Стьюдента на 10% уровне значимости можно считать не случайными коэффициенты перед z и x_4 , остальные параметры уравнения еще с более меньшей ошибкой.

Таким образом, все характеристики модели приемлемы. Только возникает вопрос о существенности увеличения коэффициента детерминации.

Проверим статистическую гипотезу о равенстве двух коэффициентов детерминации.

$R_1^2 = 0,668$ – коэффициент детерминации для модели с 4-мя факторами.

$R_2^2 = 0,652$ – коэффициент детерминации для модели с 3-мя факторами.

Нулевая гипотеза:

$H_0: R_1^2 = R_2^2$, т.е. качество описания поведения резульативного признака существенно не ухудшилось после удаления фактора x_4 .

$H_1: R_1^2 > R_2^2$, т.е. качество описания поведения резульативного признака ухудшилось существенно.

Рассчитаем статистику $F = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n - m - 1}{k}$, где n – объем выборки, в нашем случае $n = 50$; m – максимальное количество факторов в модели, в нашем случае $m = 4$; k – количество факторов, которые удалили из модели, в нашем случае $k = 1$. Получаем:

$$F = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n - m - 1}{k} = \frac{0,668 - 0,652}{1 - 0,668} \cdot \frac{50 - 4 - 1}{1} = 0,046 \cdot 45 = 2,07.$$

Найдем критическое значение с помощью функции электронных таблиц $\text{FRASПОБР}(\alpha; n_A - 1, n_B - 1)$. Для нашей ситуации:

$$\text{FRASПОБР}(\alpha=0,05; k = 1, n - m - 1 = 45) = 4,057.$$

Для правосторонней проверки, получаем, т.к.

$$F_{\text{факт}} = 2,07 < F_{\text{табл}} = 4,057,$$

то принимаем нулевую гипотезу – $R_1^2 = R_2^2$, т.е. качество описания поведения резульативного признака существенно не ухудшилось после удаления фактора x_4 . Стоит сказать, что рассмотренные нами модели с разным количеством факторов каждый раз имели небольшой прирост коэффициента детерминации. Можно предположить, что такое увеличение тоже не является статистически значимым. Но добавление в модель качественного фактора «коробка передач» позволило нам существенно повысить точность модели, т.е. имеет смысл использовать более точную модель. Рассчитаем среднюю ошибку аппроксимации для модели с четырьмя факторами:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{1}{50} \cdot 0,53 \cdot 100\% = 1,06\%.$$

Получили такое же значение ошибки, как и для модели с тремя факторами, т.е. характеристика точности модели не изменилась после учета в модели характеристики – объем двигателя.