



Множественная регрессия

Лекция 5

2. Применение множественного корреляционно-регрессионного анализа

- Экономические явления часто определяются большим числом одновременно и совокупно действующих факторов.
- Задача состоит в исследовании зависимости переменной y от объясняющих переменных x_1, x_2, x_3, \dots

3. Задачи множественного корреляционно-регрессионного анализа

Измерение тесноты связи между признаками

Отбор факторных признаков в модель

Установление неизвестных причин связей

Определение вида уравнения регрессии

Построение регрессионной модели и оценка ее параметров

Проверка значимости параметров связи

Интервальное оценивание параметров связи

4. Вид уравнения множественной регрессии

- Из-за особенности метода наименьших квадратов во множественной регрессии применяются только линейные уравнения и уравнения, приводимые к линейному виду с помощью преобразования переменных.
- Из-за трудности обоснования формы связи чаще всего используют уравнение вида

$$y_x = a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_k x_{ik} + \varepsilon_i$$

a – параметры модели (коэффициенты регрессии)

ε_i – случайная величина (величина остатка)

5. Смысл коэффициентов уравнения

- Коэффициент регрессии a_i показывает, на какую величину в среднем изменится результативный признак y , если переменную x_i увеличить на 1 при фиксированном (постоянном) значении других факторов, входящих в уравнение регрессии

6. Пример интерпретации коэффициентов уравнения регрессии

Пример 1. Предположим, что зависимость расходов на продукты питания по совокупности семей характеризуется следующим уравнением:

$$Y_{\text{теор}}(x_1, x_2) = 0,6 + 0,27 \cdot x_1 + 0,72 \cdot x_2,$$

где y – расходы семьи за месяц на продукты питания, тыс. руб.;

x_1 – месячный доход на одного члена семьи, тыс. руб.;

x_2 – размер семьи, чел.

Анализ данного уравнения позволяет сделать выводы: – с ростом дохода на одного члена семьи на 1 тыс. руб. расходы на питание возрастут в среднем на 270 руб. при том же среднем размере семьи. Увеличение размера семьи при тех же её доходах предполагает дополнительный рост расходов на питание на 720 руб. Параметр $a = 0,6$ не имеет экономической интерпретации.

7. Пример преобразования к линейному виду

При нелинейной зависимости признаков, приводимой к линейному виду, значения коэффициентов множественной регрессии также определяются также с помощью метода наименьших квадратов лишь с той разницей, что он применяется не к исходной информации, а к преобразованным данным. Так, рассматривая степенную функцию

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_p^{b_p} \cdot \varepsilon,$$

мы преобразовываем её в линейный вид:

$$\ln(y) = \ln(a) + b_1 \cdot \ln(x_1) + b_2 \cdot \ln(x_2) + \dots + b_p \cdot \ln(x_p) + \ln(\varepsilon),$$

где переменные выражены в логарифмах.

Поскольку параметры степенной функции представляют собой коэффициенты эластичности, то они сравнимы по разным факторам.

8. Интерпретация коэффициентов степенной функции при ее линеаризации

Пример 2. При исследовании спроса на некоторый продукт получено следующее уравнение

$$\ln(y) = -1,28 - 0,888 \cdot \ln(x_1) + 1,126 \cdot \ln(x_2) + \varepsilon,$$

где y – количество продукта на душу населения (кг); x_1 – цена (руб.); x_2 – доход на душу населения (тыс. руб.)

Из этого уравнения видно, что с ростом цены на 1% при том же доходе спрос снижается в среднем на 0,888%, а увеличение дохода на 1% при неизменных ценах вызывает увеличение спроса на 1,126%.

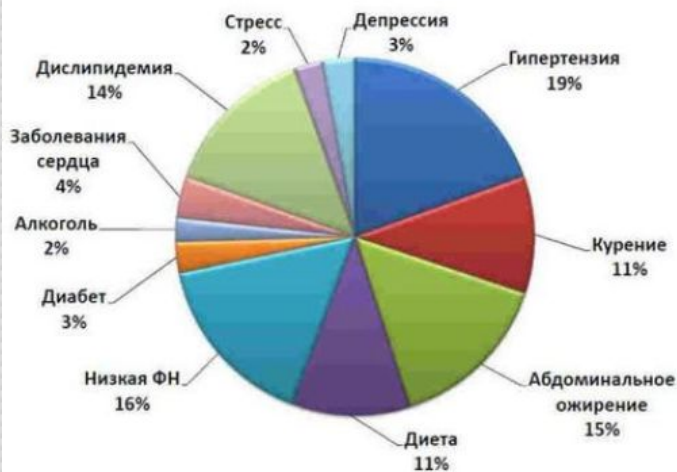
9. Рекомендации по выбору факторов, входящих в модель

I правило Признаки-факторы должны находиться в причинной связи с результивным признаком (следствием).

ПРИМЕР:

- Недопустимо в модель себестоимости у вводить в качестве одного из факторов x_j коэффициент рентабельности, хотя включение такого «фактора» значительно повышает коэффициент детерминации.

Факторы риска инсульта



10. Рекомендации по выбору факторов, входящих в модель

2 правило

- Признаки-факторы не должны быть составными частями результативного признака или его функциями

3 правило

- Признаки-факторы не должны дублировать друг друга, т. е. быть коллинеарными (с коэффициентом корреляции более 0,8).

ПРИМЕР

- Не следует в модель производительности труда включать и энерговооруженность рабочих, и их фондовооруженность, так как эти факторы тесно связаны друг с другом в большинстве объектов.

II. Рекомендации по выбору факторов, входящих в модель

4 правило

- Не следует включать в модель факторы разных уровней иерархии, т. е. фактор ближайшего порядка и его субфакторы.

ПРИМЕР:

- В моделях себестоимости зерна не следует включать и урожайность зерновых культур, и дозу удобрений под них или затраты на обработку гектара, показатели качества семян, плодородия почвы, т. е. субфакторы самой урожайности.

12. Рекомендации по выбору факторов, входящих в модель

- **5 правило (желательное)**
- Желательно, чтобы между результативным признаком и факторами соблюдалось единство единицы совокупности, к которой они отнесены.

ПРИМЕР

Если y - валовой доход предприятия, то и все факторы должны относиться к предприятию: стоимость производственных фондов, уровень специализации, численность работников и т. д.

Если y - средняя зарплата рабочего на предприятии, то факторы должны относиться к рабочему: разряд или классность, стаж работы, возраст, уровень образования, энерговооруженность и т. д. Правило это не категорическое, в модель зарплаты рабочего можно включить, например и уровень специализации предприятия.

13. Рекомендации по выбору факторов, входящих в модель

6 правило

- Математическая форма уравнения регрессии должна соответствовать логике связи факторов с результатом в реальном объекте.

ПРИМЕР

- Такие факторы урожайности, как дозы разных удобрений, уровень плодородия, число прополок и т. п., создают прибавки величины урожайности, мало зависящие друг от друга; урожайность может существовать и без любого из этих факторов. Такому характеру связей отвечает аддитивное уравнение регрессии.
- Наоборот, если y - объем валовой продукции завода, x_1 - число работников, x_2 - стоимость основных производственных фондов, x_3 - затраты на энергию, топливо, сырье, материалы, то результат без любого из факторов не существует, поэтому большинство экономистов-статистиков строят КРМ (называемую производственной функцией, что весьма не удачно терминологически) в мультипликативной форме.

14. Рекомендации по выбору факторов, входящих в модель

7 правило

- Принцип простоты: предпочтительнее модель с меньшим числом факторов при том же коэффициенте детерминации или даже при несущественно меньшем коэффициенте.

15. Пример построения модели множественной регрессии

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	№ организации	Объем реализованной продукции, млн.руб Y	Расходы на рекламу, тыс.руб X1	Цена единицы продукции, руб X2	Отдел маркетинга в организации (1-есть, 0-нет) X3									
2	1	1,27	138	140	1									
3	2	1,34	134	141	1									
4	3	1,25	116	136	0									
5	4	1,28	137	149	1									
6	5	1,43	127	154	0									
7	6	1,25	125	143	0									
8	7	1,53	116	155	1									
9	8	1,57	134	155	1									
10	9	1,27	145	151	1									
11	10	1,46	135	154	1									
12	11	1,28	164	147	0									
13	12	1,55	109	151	0									
14	13	1,35	145	144	0									
15	14	1,49	144	156	1									
16	15	1,46	132	152	0									
17	16	1,25	122	141	0									
18	17	1,29	163	148	1									
19	18	1,28	139	141	1									
20	19	1,33	134	139	0									
21	20	1,51	136	147	1									
22														

Регрессия
?
✕

Входные данные

Входной интервал Y:

Входной интервал X:

Метки Константа - ноль

Уровень надежности: %

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки График остатков

Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

OK

Отмена

Справка

16. Уравнение модели

Вывод итогов					
<i>Регрессионная статистика</i>					
Множественный R	0,806634				
R-квадрат	0,650659				
Нормированный R-квадрат	0,585158				
Стандартная ошибка	0,073662				
Наблюдения	20				
<i>Дисперсионный анализ</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	3	0,161702	0,053901	9,933510781	0,000614708
Остаток	16	0,086818	0,005426		
Итого	19	0,24852			
	<i>Коэффициент</i>	<i>Стандартная ошибка</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	-0,19995	0,451283	-0,44306	0,663651301	-1,156625345
Расходы на рекламу, тыс.руб X1	-0,00298	0,001269	-2,3459	0,032195827	-0,005666696
Цена единицы продукции, руб X2	0,013347	0,002845	4,691053	0,000245307	0,007315304
Отдел маркетинга в организации (1-есть, 0-нет) X3	0,015308	0,036151	0,423457	0,677598986	-0,061327828

$$\bar{y}(x) = -0.200 - 0.003x_1 + 0.013x_2 + 0.015x_3$$

17. Коэффициент множественной

детерминации

- R^2 – используют для оценки качества множественных регрессионных моделей
- Показывает, какую долю вариации y учтена в модели и обусловлена влиянием факторов x
- Чем ближе R^2 к 1, тем выше качество модели
- Если $R^2 = 1$ то модель применять нельзя

18. Скорректированный R^2

- При добавлении независимых переменных x R^2 увеличивается, поэтому его корректируют с учетом числа независимых переменных – скорректированный R^2

19. Проверка значимости модели

- Для проверки значимости модели используют F-критерий Фишера
- Если расчетное значение критерия больше табличного при заданном уровне значимости, то модель считается значимой

20. Оценивание

достоверности каждого из

параметров модели

- Оценивание достоверности каждого из параметров модели по t -критерию Стьюдента (для всех коэффициентов a)
- Если надежность коэффициента регрессии не подтверждается, то следует вывод о несущественности в модели факторного i -того признака и необходимости его устранения из модели или замены на другой факторный признак

21. Пояснения по оценкам модели

- Значимость $F < 0.05$ уравнение регрессии следует признать адекватным
- Множественный коэффициент корреляции $R=0.807$ – связь тесная
- Множественный коэффициент детерминации $R^2 = 0,651$ показывает, что 65% вариаций y учтено в модели и обусловлено влиянием факторов
- Значимость коэффициентов регрессии a_0 и коэффициент при x_3 не значимы и из модели нужно исключить x_3

	P-Значение
Y-пересечение	0,663651301
Расходы на рекламу, тыс.руб X1	0,032195827
Цена единицы продукции, руб X2	0,000245307
Отдел маркетинга в организации (1-есть, 0-нет) X3	0,677598986

22. Новая модель с меньшим числом переменных


A	B	C	D	E	F	G	H	I	J	K
№ организации	Объем реализованной продукции, млн.руб Y	Расходы на рекламу, тыс.руб X1	Цена единицы продукции, руб X2							
					Вывод итогов					
1	1,27	138	140							
2	1,34	134	141		<i>Регрессионная статистика</i>					
3	1,25	116	136		Множественный R	0,804204				
4	1,28	137	149		R-квадрат	0,646744				
5	1,43	127	154		Нормированный R-квадрат	0,605184				
6	1,25	125	143		Стандартная ошибка	0,071862				
7	1,53	116	155		Наблюдения	20				
8	1,57	134	155							
9	1,27	145	151		<i>Дисперсионный анализ</i>					
10	1,46	135	154			df	SS	MS	F	Значимость F
11	1,28	164	147		Регрессия	2	0,160729	0,080364	15,56186516	0,000144132
12	1,55	109	151		Остаток	17	0,087791	0,005164		
13	1,35	145	144		Итого	19	0,24852			
14	1,49	144	156							
15	1,46	132	152		<i>Коэффициенты регрессии</i>					
16	1,25	122	141		Y-пересечение	-0,26411	0,414695	-0,63689	0,532682421	-1,139044587
17	1,29	163	148		Расходы на рекламу, тыс.руб X1	-0,00282	0,001187	-2,37964	0,02930742	-0,005327024
18	1,28	139	141		Цена единицы продукции, руб X2	0,0137	0,002654	5,162156	7,81614E-05	0,008100509
19	1,33	134	139							
20	1,51	136	147							


$$\bar{y}(x) = -0.264 - 0.0003x_1 + 0.014x_2$$

23. Исключение незначимого свободного члена из уравнения регрессии

Регрессия

Входные данные


Входной интервал Y: 

Входной интервал X: 

Метки Константа - ноль

Уровень надежности: %

Параметры вывода

Выходной интервал: 

Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки График остатков

Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

OK
Отмена
Справка

24. Модель без свободного члена

Вывод итогов					
<i>Регрессионная статистика</i>					
Множественный R	0,998813				
R-квадрат	0,997628				
Нормированный R-квадрат	0,941941				
Стандартная ошибка	0,070666				
Наблюдения	20				
<i>Дисперсионный анализ</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	37,80631	18,90316	3785,429749	3,00471E-23
Остаток	18	0,089886	0,004994		
Итого	20	37,8962			
<i>Коэффициентная статистика</i>					
		<i>P-Значение</i>	<i>Нижние 95%</i>		
Y-пересечение	0	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Расходы на рекламу, тыс.руб X1	-0,00308	0,001099	-2,80187	0,011789492	-0,005386263
Цена единицы продукции, руб X2	0,012141	0,00101	12,02195	4,90018E-10	0,010019407

$$\bar{y}(x) = -0.003x_1 + 0.012x_2$$

25. Интерпретация

коэффициентов модели

- Параметр регрессии $a_1 = -0,003$ Повышение расходов на рекламу на 1,0 тыс.рублей при фиксированном (постоянном) значении цены на продукцию приводит к уменьшению объема реализации на 3,0 тыс.рублей.
- Параметр регрессии $a_2 = 0,012$ с ростом цены продукта на 1 руб при фиксированном уровне расходов на рекламу объем реализации продукции увеличивается в среднем на 14,0 тыс. рублей

26. Мера точности модели

- В качестве меры точности модели применяют стандартную ошибку

$$S_{\varepsilon} = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{n - k - 1}}$$

27. Частные коэффициенты эластичности

- Служат для оценки влияния каждого факторного признака на результирующий признак

$$\mathcal{E}_i = a_i \cdot \frac{\bar{x}_i}{\bar{y}}$$

- Он показывает, на сколько % изменится Y при изменении X на 1%. Однако, он не учитывает колеблемости факторов

28. Интерпретация коэффициентов эластичности

$$\mathcal{E}_1 = a_1 \cdot \frac{\bar{x}_i}{\bar{y}} = -0,003 \cdot \frac{134,75}{1,372} = -0,3$$

$$\mathcal{E}_2 = a_2 \cdot \frac{\bar{x}_i}{\bar{y}} = 0,012 \cdot \frac{147,2}{1,372} = 1,3$$

- По абсолютному приросту наибольшее влияние на объем реализации продукции оказывает фактор x_2 : повышение цены продукции на 1% приводит к росту объема реализации продукции на 1,3%.
- Снижение расходов на рекламу на 1% вызывает повышение объема реализации продукции только на 0,3%

29. Бета -коэффициенты

- Бета коэффициент показывает, на какую часть своего среднего квадратического отклонения изменится в среднем значение Y при изменении X на величину своего среднего квадратического отклонения при фиксированных значениях остальных независимых переменных

$$\beta_i = a_i \frac{S_{xi}}{S_y}$$

S_{xi} – среднее _квadraticеское

отклонение _фактора

S_y – СКО _результативного _признака _у

30. Расчет бета-коэффициентов

$$\beta_1 = -0,003 \frac{13,91}{0,11} = -0,374$$

$$\beta_2 = 0,012 \frac{6,22}{0,11} = 0,660$$

3 I. Дельта-коэффициенты

- Позволяет проранжировать факторы по степени их влияния на признак.
- Показывает долю влияния фактора в суммарном влиянии всех факторов, включенных в модель

$$\Delta_i = r_{yi} \cdot \frac{\beta_i}{R^2}$$

r_{yi} – коэффициент парной

корреляции между фактором и y

32. Расчет дельта-коэффициентов

- Определяются парные коэффициенты корреляции
- Анализ данных / Корреляция
- Входной интервал все x и y с заголовками

Корреляция

Входные данные

Входной интервал:

Группирование: по столбцам по строкам

Метки в первой строке

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

OK
Отмена
Справка

	Объем реализованной продукции, млн.руб Y	Расходы на рекламу, тыс.руб X1	Цена единицы продукции, руб X2
Объем реализ	1		
Расходы на ре	-0,304971423	1	
Цена единиць	0,727375131	0,051692207	1

33. Интерпретация дельта-коэффициентов

$$\Delta_1 = -0,305 \cdot \frac{-0,374}{0,998} = 0,114$$

$$\Delta_2 = 0,727 \cdot \frac{0,660}{0,998} = 0,034$$

- Анализ бета- и дельта-коэффициентов показывает, что на объем реализации продукции наибольшее влияние из двух исследуемых факторов оказывает фактор x_2 – цена продукции, так как фактору соответствуют наибольшие (по модулю) значения коэффициентов

34. Общие рекомендации по значениям коэффициентов

Критерий Фишера	Характеризует значимость уравнения регрессии. Применяется для выбора моделей	Должен быть больше табличного значения, установленного для различных размеров матрицы и вероятностей
Критерий Стьюдента	Характеризует существенность факторов, входящих в модель. Применяется для выбора модели	Больше 2,0 (при вероятности, равной 0,95)
Среднеквадратическая ошибка коэффициента регрессии	Характеризует точность полученных коэффициентов регрессии. Применяется для оценки коэффициентов регрессии	В два и более раза меньше соответствующего коэффициента регрессии
Ошибка аппроксимации	Характеризует допуск прогноза или степень несоответствия эмпирической зависимости теоретической. Применяется для оценки точности модели	Меньше (точнее) 15 %
Коэффициент эластичности	Показывает, на сколько процентов изменяется функция при изменении соответствующего фактора на 1 %. Применяется для ранжирования факторов по их значимости	Больше 0,01



Лекция окончена