

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Севастопольский государственный университет»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ
к проведению практических занятий по
дисциплине «Эконометрия»
студентов экономических специальностей
всех форм обучения

Севастополь

2019

УДК 658

Методические указания к проведению практических занятий по дисциплине «Эконометрия» для студентов экономических специальностей всех форм обучения // сост. И.А. Гребешкова – 2019 – 62с.

Целью методических указаний является систематизация теоретического материала и практических рекомендаций для выполнения заданий по дисциплине «Эконометрия» для студентов всех форм обучения экономических специальностей

Утверждена на заседании кафедры Менеджмента и бизнес-аналитики от «30» июня 2019., протокол №10

Содержание

Практическое занятие № 1 Срез остаточных знаний по дисциплине «Бизнес-статистика», «Численные методы в управлении», «Высшая математика».....	4
Практическое занятие № 2 Проверка гипотез	7
Практическое занятие № 3 Анализ модели парной регрессии	12
Практическое занятие № 4 Анализ модели множественной регрессии	16
Практическое занятие № 5 Анализ модели с гетероскедастичной случайной составляющей	28
Практическое занятие № 6 Анализ методом главных компонент	37
Практическое занятие № 7 Анализ временных рядов	44
Практическое занятие № 8 Оценка параметров систем линейных уравнений	53

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ № 1
**СРЕЗ ОСТАТОЧНЫХ ЗНАНИЙ ПО ДИСЦИПЛИНЕ «БИЗНЕС-
СТАТИСТИКА», «ЧИСЛЕННЫЕ МЕТОДЫ В УПРАВЛЕНИИ», «ВЫСШАЯ
МАТЕМАТИКА»**

Задача 1.

Из партии готовой продукции в порядке механической выборки проверено 50 лампочек на продолжительность горения. Последняя оказалась равна 840 часов при среднем квадратичном отклонении 60 ч.

Определить:

- 1) среднюю ошибку (μ) выборочной средней продолжительности горения лампочки;
- 2) с вероятностью 0,95 доверительные пределы продолжительности горения лампочки в генеральной совокупности.

Задача 2.

На городской телефонной станции в порядке собственно случайной выборки проведено 100 наблюдений и установлено, что средняя продолжительность телефонного разговора составляет 10 мин при среднем квадратичном отклонении 5 мин.

1. С вероятностью 0,997 определить доверительные пределы для генеральной средней.
2. Можно ли считать данную выборку репрезентативной ?
3. С какой вероятностью можно утверждать, что при определении средней продолжительности разговора допущена ошибка, не превышающая 1 мин.

Задача 3.

Из партии готовой продукции в порядке механической бесповторной выборки проверено 400 изделий и установлено, что 80% из них соответствует первому сорту.

С вероятностью 0,9545 определить долю продукции первого сорта во всей партии.

Задачу решить в двух вариантах:

- 1) численность изделий в партии готовой продукции неизвестна;
- 2) в партии готовой продукции 2000 изделий.

Задача 4.

В порядке случайной выборки обследован дневной надой молока 50 коров. Результаты обследования приведены в таблице.

Дневная удойность, кг	Количество коров
10-14	5
14-18	15
18-22	20
Свыше 22	10
Итого	50

Определить:

- 1) по выборочным данным средний дневной надой молока от одной коровы;
- 2) среднюю ошибку выборки;
- 3) вероятность того, что при определении выборочного среднего надоя молока допущена ошибка, не превышающая 1 кг.

Задача 5.

Намечается провести выборочное обследование покупателей в одном из крупных универмагов города в целях определения доли покупателей из других городов. Каким должен быть объем выборки, чтобы с вероятностью 0,9545 можно было бы гарантировать точность результата до 5%.

Задача 6.

На предприятии выборочно проверен стаж у 12 мужчин и 8 женщин. Результаты наблюдения следующие:

Группа рабочих	n_i	Средний стаж работы, лет, $\tilde{x}_i$	Средне квадратичное отклонение стажа, лет, σ_i
Мужчины	12	14	3
Женщины	8	11	2

1. Рассчитать общий средний стаж работы для рабочих по выборочным данным.
2. С вероятностью 0,954 определить доверительные пределы среднего стажа работы рабочих в генеральной совокупности.

3. Определить можно ли считать расхождения выборочной средней стажа работы у мужчин и женщин случайным при уровне значимости $\alpha = 0,05$.

Задача 7.

Для определения среднего процента выполнения норм выработки проведена 5%-ная типическая выборка из трех групп рабочих с разным стажем. Результаты следующие:

Группы со стажем	Объем выборки, чел., n_i	Средний процент выполнения норм $\tilde{x}_i$	Среднее квадратичное отклонение, %, σ_i
1-2 года	15	98	3
3-5 лет	20	102	2
Более 5 лет	65	104	4

Определить:

- 1) средний % выполнения норм для всех рабочих в выборочной совокупности;
- 2) вероятность того, что выборочная средняя (процент выполнения норм) отличается от генеральной не более, чем на 1%.

Задача 8.

Имеются следующие данные выборочного обследования размера заработной платы у 5 мужчин и 5 женщин в одном НИИ:

Заработная плата, грн	
Женщины	Мужчины
500	660
520	700
660	730
670	840
700	920

Определить:

- 1) среднюю зарплату отдельно для мужчин и для женщин, а так же общую среднюю, то есть для всей выборочной совокупности;
- 2) среднюю ошибку выборки для общей средней заработной платы;
- 3) вероятность того, что общая выборочная средняя зарплата отличается от генеральной не более, чем на 50 грн.
- 4) Можно ли случайными расхождения между средней зарплатой мужчин и женщин?

Задача 9.

Выборочное обследование 10 электрических лампочек для определения их средней продолжительности горения дало следующие результаты:

Продолжительность горения, час	Количество лампочек
900-920	2
920-940	3
940-960	4
960-980	1

Определить:

- 1) среднюю продолжительность горения лампочек по выборочным данным;
- 2) с вероятностью 0,97 оценить доверительные пределы для генеральной средней.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ № 2

ПРОВЕРКА ГИПОТЕЗ

Статистическая гипотеза обычно представляет собой некоторое предположение об одном или нескольких параметрах функции распределения случайной величины.

В теории статистического вывода обычно проверяются гипотезы на основе выборочной информации. В практике статистической работы чаще всего имеют дело с двумя конкурирующими гипотезами: нулевой гипотезой, обозначаемой H_0 ; и альтернативной гипотезой, обозначаемой H_1 . Нулевая гипотеза используется при статистической проверке гипотез об отсутствии существующих различий между несколькими выборочными совокупностями, для суждения о близости фактического распределения к теоретическому (нормальному), об отсутствии зависимости между признаками. Суть нулевой гипотезы состоит в признании того, что выборки взяты из одной совокупности, фактическое распределение укладывается в теоретическое, зависимость между признаками отсутствует и т. д. Следовательно, нулевая гипотеза — это гипотеза, подлежащая проверке. И если отвергается нулевая гипотеза как неподходящая в каком-то статистическом смысле, то принимается альтернативная гипотеза.

Так как мы имеем дело с неизвестной генеральной совокупностью и выносим суждения о ней на основе выборочной информации, то мы можем и не прийти к правильному выводу. Мы сделаем неправильный вывод, если отвергнем нулевую гипотезу, когда она справедлива (ошибка I рода), или примем нулевую гипотезу, когда она ошибочна (ошибка II рода).

В большинстве случаев при проведении проверки гипотез в экономике задается некоторый допустимый уровень вероятности совершения ошибки I рода (α) и осуществляется проверка на основе выборочной информации. В классическом статистическом выводе существует два общих правила для определения величины α :

- 1) чем больше степень уверенности в нулевой гипотезе, тем меньше должно быть значение α .
- 2) чем больше цена отбрасывания справедливой нулевой гипотезы, тем меньше значение должно иметь α .

Сформулируем общий алгоритм проверки статистических гипотез. Процедуру проверки можно описать следующими шагами:

1. Формулирование нулевой и альтернативной гипотез.
2. Выбор уровня значимости.
3. Выбор теста.
4. Определение критической области.
5. Выполнение необходимых вычислений.
6. Принятие статистического решения.

Уровень значимости α - это такое малое значение вероятности попадания критерия в критическую область при условии справедливости гипотезы, что появление этого события может расцениваться как следствие существующего расхождения выдвинутой гипотезы и результатов выборки.

Допустим, рассчитанное по эмпирическим данным значение критерия попало в критическую область. Тогда при условии верности проверяемой гипотезы H_0 вероятность этого события будет не больше уровня значимости α . Поскольку α выбирается достаточно малым, то такое событие является маловероятным, и, следовательно, проверяемая гипотеза может быть отвергнута. Если же наблюдаемое значение характеристики не принадлежит к критической области, и, следовательно, находится в области допустимых значений, то проверяемая гипотеза H_0 не отвергается. Вероятность попадания критерия в область допустимых значений при справедливости проверяемой гипотезы H_0 равна $1 - \alpha$. Чем меньше уровень значимости, тем меньше вероятность забраковать проверяемую гипотезу, когда она верна, то есть меньше вероятность совершить ошибку первого рода. Но при этом расширяется область допустимых значений и, значит, увеличивается вероятность совершения ошибки II рода.

Если альтернативная гипотеза $X_1 \neq X_2$, то гипотеза называется двухсторонней. Если гипотеза имеет вид $X_1 > X_2$ и $X_1 < X_2$, то такую гипотезу называют односторонней. При проверке двухсторонней гипотезы с уровнем значимости α критическое значение будет определено с уровнем значимости $\alpha/2$ и с соответствующим числом степеней свободы. При проверке односторонней гипотезы критическое значение будет определено с соответствующим числом степеней свободы и уровнем значимости α .

ПРИМЕР ПРОВЕРКИ СТАТИСТИЧЕСКОЙ ГИПОТЕЗЫ

Задание. Плановый объем налогов поступлений от предприятий города за день в среднем должен составлять 14 тыс. руб.. Поступления за шесть дней в бюджет составили соответственно: 14.15 13.85 13.95 14.20 14.30 14.35. Полагая, что уровень значимости принят 1%, определить:

- 1) выполнила ли налоговая администрация план по поступлениям средств в бюджет;
- 2) перевыполнила ли налоговая администрация план по поступлениям средств в бюджет.

Решение.

Часть I.

♦ Формулирование нулевой и альтернативной гипотез.

$H_0: \mu = 14$

$H_1: \mu \neq 14$.

♦ Выбор уровня значимости.

$\alpha = 0.01$.

♦ Выбор теста.

t критерий Стьюдента.

♦ Определение критической области.

Так как альтернативная гипотеза имеет следующий вид: $H_1: \mu \neq 14$, то необходимо применять двусторонний тест. Для определения критической области рассчитаем $t(5, 0.005) = 4.032 [t(n-1, \alpha/2)]$.

Правило принятия (отвержения) гипотезы: если $TR < -4.032$ и $TR > 4.032$, то отвергается H_0 и принимается H_1 .

♦ Выполнение необходимых вычислений.

Вычисление t расчетного (TR) производится по следующей формуле:

$$TR = \frac{\bar{x} - \mu_{H_0}}{s / \sqrt{n}}, \text{ где } s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}.$$

Таблица 3.1 Вспомогательные расчеты

X	$x - \bar{x}$	$(x - \bar{x})^2$
14.15	0.0167	0.00028
13.85	-0.2833	0.08026
13.95	-0.1833	0.03360
14.20	0.0667	0.00445
14.30	0.1667	0.02779
14.35	0.2167	0.04696
		0.19334

Таким образом имеем:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{0.19334}{5}} = 0.1966$$

$$TR = \frac{\bar{x} - \mu_{H_0}}{s / \sqrt{n}} = \frac{14.1333 - 14}{0.1966 / \sqrt{6}} = 1.66$$

♦ Принятие статистического решения.

Так как $-4.032 < TR < 4.032$, тогда с $\alpha = 0.01$ гипотеза H_0 о равенстве среднего значения налогов поступлений за день в размере 14 тыс.руб. принимается при $\alpha = 0.01$. Следовательно, налоговая администрация выполнила план по поступлениям средств в бюджет.

♦ При ответе на второй вопрос задания нулевая и альтернативная гипотеза будут иметь вид:

$$H_0: \mu = 14$$

$$H_1: \mu > 14.$$

Так как альтернативная гипотеза имеет следующий вид: $H_1: \mu > 14$, то необходимо применять односторонний тест. Для определения критической области рассчитаем $t(5, 0.01) = 3.365 [t(n-1, \alpha)]$.

$|TR| < 3.365$, то гипотеза H_0 о равенстве среднего значения налогов поступлений за день в размере 14 тыс.руб. принимается при $\alpha = 0.01$. Следовательно, нельзя говорить о перевыполнении налоговой администрацией плана по поступлениям средств в бюджет.

Проверка гипотез о равенстве двух средних для зависимых выборок

Инженеры разработали новую технологию сборки воздушных компрессоров. Необходимо проверить результативность этой технологии. Исходные данные приведены в таблице 1.

Рабочий	Производительность после внедрения новой технологии, x_1	Производительность до внедрения новой технологии, x_2	Разница $x_1 - x_2$ d	$(d - \bar{d})$	$(d - \bar{d})^2$
1	85	80	5	3	9
2	84	88	-4	-6	36
3	80	76	4	2	4
4	93	90	3	1	1
5	83	74	9	7	49
6	71	70	1	-1	1
7	79	81	-2	-4	16
8	83	83	0	-2	4

Для того чтобы сравнить две зависимых выборки применяется тот же подход, что и в случае одной выборки. Но в этом случае тест выполняется для разности между парами значений.

Шаг 1. *Формирование гипотезы.*

Нулевая гипотеза - $H_0 : \mu_d = 0$.

Альтернативная гипотеза - $H_1 : \mu_d \neq 0$.

Шаг 2. *Выбор уровня значимости.*

$$\alpha = 0.05$$

Шаг 3. *Выбор теста.*

двухсторонний t – критерий

Шаг 4. *Определение значений доверительного интервала.*

число степеней свободы - 7, доверительный интервал - 0.025,

значение критерия 2.365

Шаг 5. *Принятие решения.*

Отклонить H_0 и принять H_1 , если $TR < -2.365$ _or_ $TR > +2.365$

Шаг 6. *Вычисления.*

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}} = 4.14 \text{ и } TR = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{2.00 - 0}{4.14 / \sqrt{8}} = 1.37.$$

Шаг 7. *Выводы.*

Принимается нулевая гипотеза, поэтому нельзя сделать заключение о том, что новый метод однозначно лучше.

ЗАДАНИЕ

Задание 1

1 Сформировать гипотезы двух видов, представленных в теоретическом материале.

2 Выполнить проверку гипотез с учетом уровня значимости 1 %, 5 % и 10 %.

АНАЛИЗ МОДЕЛИ ПАРНОЙ РЕГРЕССИИ

Регрессионный анализ предназначен для установления функциональной связи между одной зависимой переменной и одной или несколькими независимыми переменными.

Рассмотрим уравнение регрессии в случае, если устанавливается функциональная связь между зависимой переменной и одной независимой переменной.

Тогда регрессионное уравнение для линейной модели примет вид:

$Y_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$, где Y_i - зависимая переменная, X_i - независимая переменная, α_0, α_1 - коэффициенты регрессионного уравнения, ε_i - ошибка с нормальным законом распределения, средним равным нулю и стандартным отклонением σ .

Так как в экономике в большинстве практических случаев сбор данных или весьма затруднителен, или связан с большими затратами, поэтому чаще всего каждому значению независимой переменной соответствует только одно наблюдение зависимой переменной. В нашем случае Y_i - это одно наблюдение, соответствующее X_i .

Для оценки коэффициентов регрессионного уравнения будем использовать метод наименьших квадратов (МНК). Регрессионное уравнение, коэффициенты которого оценены при помощи МНК, будет иметь вид:

$\hat{Y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 X_i$, где $\hat{Y}_i$ - оцененное значение, $\hat{\alpha}_0, \hat{\alpha}_1$ - оценки коэффициентов регрессионного уравнения, полученные с помощью МНК, знак $\wedge$ показывает, что это оценка случайной величины.

МНК минимизирует сумму квадратов ошибок (остатков):

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(Y_i - \hat{Y}_i \right)^2, \text{ где } \hat{Y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 X_i.$$

Оценки коэффициентов регрессионного уравнения, полученные при помощи метода наименьших квадратов равны:

$$\hat{\alpha}_0 = \frac{\sum Y_i \sum X_i^2 - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2};$$

$$\hat{\alpha}_1 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2},$$

Произведя преобразования, получаем:

$$\hat{\alpha}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

и

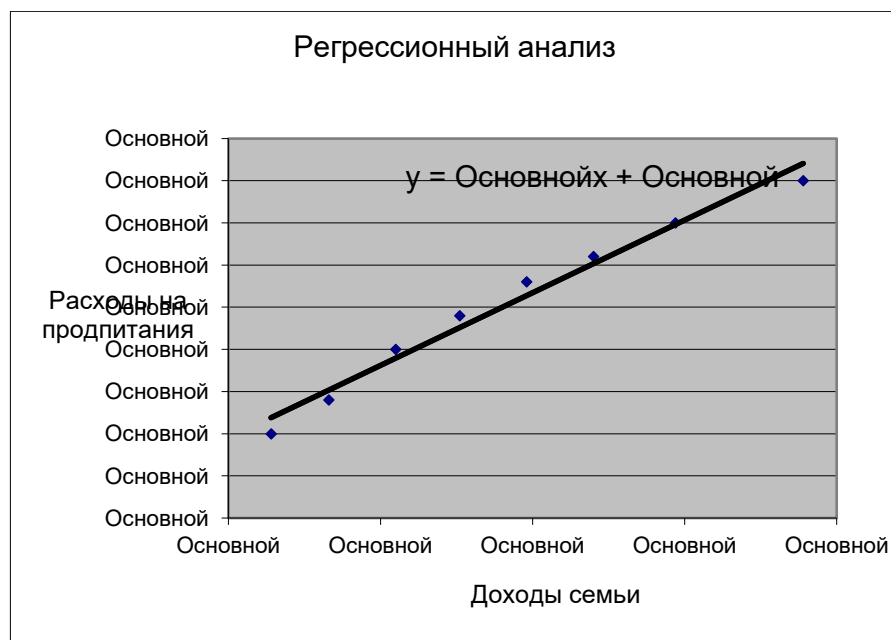
$$\hat{\alpha}_0 = \bar{Y} - \hat{\alpha}_1 \bar{X}, \text{ где } \bar{Y} = \sum y_i / n \text{ и } \bar{X} = \sum x_i / n, \text{ где } x_i = X_i - \bar{X}, y_i = Y_i - \bar{Y}.$$

Так как в процессе определения функциональной связи между исследуемыми переменными регрессионное уравнение строится не на анализе данных генеральной совокупности, а на основе выборки, то следующим шагом анализа будет проверка значимости коэффициентов регрессионного уравнения и проверка адекватности модели.

Рассмотрим пример построения модели парной линейной регрессии

	Доходы семьи (x), тыс.грн.	Расходы на продукты питания (y), тыс.грн.	x-хср	y-уср	(x-хср)kv	(x-хср)(y-уср)
	1,4	1	-7,8	-1,5	60,1	11,8
	3,3	1,4	-5,9	-1,1	34,2	6,6
	5,5	2	-3,7	-0,5	13,3	1,9
	7,6	2,4	-1,6	-0,1	2,4	0,2
	9,8	2,8	0,7	0,3	0,4	0,2
	12	3,1	2,9	0,6	8,1	1,6
	14,7	3,5	5,6	1,0	30,8	5,4
	18,9	4	9,8	1,5	95,1	14,4
среднее	9,2	2,5			244,4	42,1

a1	0,17
a0	0,95



Коэффициент корреляции

Коэффициент корреляции оценивает тесноту связи между переменными.

Линейный коэффициент корреляции (коэффициент корреляции Пирсона) между выборочными величинами X и Y определяется в виде:

$$r = \frac{\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i / n}{\sqrt{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 / n} \cdot \sqrt{\sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2 / n}}$$

Для расчета сформируем таблицу промежуточных расчетов. Последняя строка таблицы представляет собой суммы по столбцам.

Доходы семьи (X), тыс.грн.	Расходы на продукты питания (y), тыс.грн.	$x_i y_i$	x_i^2	y_i^2
1,4	1	1,4	1,96	1
3,3	1,4	4,62	10,89	1,96
5,5	2	11	30,25	4
7,6	2,4	18,24	57,76	5,76
9,8	2,8	27,44	96,04	7,84
12	3,1	37,2	144	9,61
14,7	3,5	51,45	216,09	12,25
18,9	4	75,6	357,21	16
73,2	20,2	226,95	914,2	58,42

Коэффициент корреляции будет иметь вид:

$$r = \frac{226,95 - 73,2 \cdot 20,2 / 8}{\sqrt{914,2 - 73,2^2 / 8} \cdot \sqrt{58,42 - 20,2^2 / 8}} = 0,989.$$

Так как коэффициент корреляции выше 0,6, то это говорит о присутствии сильной прямой связи. Коэффициент имеет положительное значение: с ростом доходов семьи расходы на продукты питания также имеют тенденцию к росту.

Оценка значимости показателя:

$$t_p = \frac{r \cdot \sqrt{k}}{\sqrt{1-r^2}}, \text{ тогда } t_p = \frac{0,989 \cdot \sqrt{8-2}}{\sqrt{1-0,989^2}} = 16,68 \quad \text{Табличное значение при уровне}$$

значимости пять процентов и числе степеней свободы $8-2=6$ равно 2,447. Так как расчетное значение критерия Стьюдента больше табличного, то корреляционная связь между переменными значима.

Кроме данного коэффициента широко используется коэффициент корреляции рангов Спирмена, когда коррелируют не сами значения показателей, а их ранги, то есть номера их мест, занимаемых в каждом ряду значений по возрастанию или убыванию (обозначаются ранги буквой R или N). Коэффициент корреляции рангов Спирмена рассчитывается по формуле (ρ):

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)},$$

где $d = N_x - N_y$, то есть разность каждой пары значений x и y , n - число наблюдений.

ЗАДАНИЕ

Задание 1.

- 1 Сформировать проблему, для ответа на которую необходимо использовать регрессионный однофакторный анализ.
- 2 Для проблемы сформировать статистические данные.
- 3 Построить модель, оценить параметры.
- 4 Сделать выводы.

Задание 2.

При помощи майн-карты представить основные моменты теорического материала по однофакторному регрессионному анализу.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ № 4 АНАЛИЗ МОДЕЛИ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ И ПРИМЕРЫ РАСЧЁТОВ: ЛИНЕЙНАЯ МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ (1МНК)

Термин «регрессия» (термин был впервые использован в работе Пирсона - Pearson, 1908) в переводе означает «отступление», «возврат» и представляет собой количественную меру изменения (уменьшения или увеличения) одного (результативного) признака в зависимости от изменения на определенную величину других коррелирующих с ним (факторных) признаков.

Простая линейная регрессия представляет собой регрессию между двумя переменными y и x , т.е. модель вида $y = f(x) + u = \alpha_0 + \alpha_1 \cdot x + u$, где y — результативный признак; x - признак-фактор, u — случайная ошибка.

В результате построения уравнения регрессии (прямая на рисунке 1) получаются не точные значение (точки y_i), а модельные значения $\tilde{y}_x$, отличающиеся от точных на некоторую величину u_i (ошибку), рисунок 1.

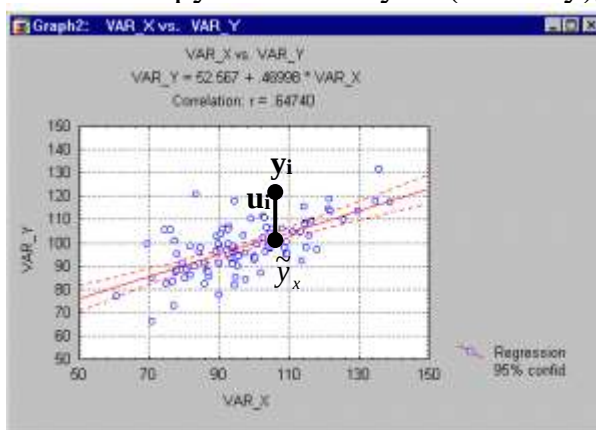


Рисунок 1 – График функции однофакторной линейной регрессии

Множественная регрессия представляет собой регрессию результативного признака с двумя и большим числом факторов, т. е. модель вида $y = \hat{f}(x_1, x_2, \dots, x_n) + u$.

В уравнении регрессии корреляционная по сути связь признаков представляется в виде функциональной связи, выраженной соответствующей функцией (1).

$$y = \tilde{y}_x + u = \alpha_0 + \alpha_1 \cdot x_1 + \dots + \alpha_n x_n + u, \quad (1)$$

где y — фактическое значение результативного признака;

$\tilde{y}_x$ - модельное значение результативного признака;

x_i - признак-фактор;

α_i — параметр регрессионной модели;

u — случайная величина (ошибка, возмущение, остаток), характеризующая отклонения реального значения результативного признака от теоретического

(предсказанные значения “у” минус наблюдаемые). Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения.

Например, специалисты по кадрам обычно используют процедуры множественной регрессии для определения заработной платы, адекватной выполненной работе. Можно определить некоторое количество факторов или параметров, таких, как "размер ответственности" (x_1) или "число подчиненных" (x_2), которые, как ожидается, оказывают влияние на заработную плату (y (стоимость работы)). Кадровый аналитик затем проводит исследование размеров окладов среди сравнимых компаний на рынке, записывая размер жалования и соответствующие характеристики (т.е. значения параметров) по различным должностям. Эта информация может быть использована при анализе с помощью множественной регрессии для построения регрессионного уравнения в следующем виде

$$y = 0.5 + 3.5 \cdot x_1 + 7.3 \cdot x_2 + u,$$

где y – заработная плата, тыс.грн.;

x_1 – размер ответственности;

x_2 – число подчинённых, человек.

Как только эта линия регрессии определена, аналитик оказывается в состоянии построить график ожидаемой оплаты труда и реальных обязательств компании по выплате жалования. Таким образом, аналитик может определить, какие должности недооценены (лежат ниже линии регрессии), какие оплачиваются слишком высоко (лежат выше линии регрессии), а какие оплачены адекватно.

2.1. Оценка параметров модели

Классический подход к оцениванию параметров линейной модели основан на обычном или одношаговом методе наименьших квадратов (МНК).

Этот метод позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y_i от расчетных (теоретических) $\tilde{y}_i$ минимальна (2).

$$\sum_i (y_i - \tilde{y}_i)^2 \rightarrow \min. \quad (2)$$

Чтобы найти минимум функции (5), надо вычислить производные по каждому из параметров и приравнять их к нулю, т.к. равенство нулю производной – необходимое условие экстремума. В результате получается система уравнений (в матричном виде представлена формулой 3, решение которой и позволяет получить оценки параметров регрессии, формула (1).

$$\alpha = [X^T \cdot X]^{-1} \cdot X^T \cdot Y. \quad (3)$$

Пример 1. Составить линейную модель множественной регрессии y от x_1 и x_2 по данным таблицы 1.

Таблица 1 – Исходные данные

Номер предприятия	1	2	3	4
Переменная				
$x_1, (\%)$	1	1	-1	-1
$x_2, (\%)$	1	-1	1	-1
$y, (\text{тыс. грн.})$	0,5	5	7	12

Этап 1. Составим матрицы X (первый столбец всегда состоит из «1», поскольку в модели присутствует α_0 , второй и третий столбцы – значения x_1 и x_2) и Y (значения y)

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \quad Y = \begin{pmatrix} 0,5 \\ 5 \\ 7 \\ 12 \end{pmatrix}$$

$$\text{Этап 2. } X^T \cdot X = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

$$\text{Этап 3. } (X^T \cdot X)^{-1} = \begin{pmatrix} 0,25 & 0 & 0 \\ 0 & 0,25 & 0 \\ 0 & 0 & 0,25 \end{pmatrix}$$

$$\text{Этап 4. } X^T \cdot Y = \begin{pmatrix} 24,5 \\ -13,5 \\ -9,5 \end{pmatrix}$$

$$\text{Этап 5. } \alpha = [X^T \cdot X]^{-1} \cdot X^T \cdot Y = \begin{pmatrix} 6,125 \\ -3,375 \\ -2,375 \end{pmatrix}$$

Получаем модель: $y = 6,125 - 3,375x_1 - 2,375x_2$.

Проверка коэффициентов на значимость

Значимость отдельных коэффициентов уравнения множественной регрессии оценивается с помощью Т-критерия Стьюдента по формуле (4).

$$t_{p_{\alpha_i}} = \frac{|\alpha_i|}{\sqrt{\sigma_{\alpha_i}^2}}, \quad (4)$$

где α_i - оценка i -го коэффициента модели;

$\sigma_{\alpha_i}^2$ - оценка дисперсии параметра α_i - i -й элемент главной диагонали дисперсионно-ковариационной матрицы S_{α_i} , формула (5).

$$S_{\alpha_i} = \sigma_u^2 \cdot [X^T \cdot X]^{-1} = \frac{RSS}{n-k} \cdot [X^T \cdot X]^{-1} = \frac{Y^T \cdot Y - \alpha^T \cdot X^T \cdot Y}{n-k} \cdot [X^T \cdot X]^{-1}, \quad (5)$$

где σ_u^2 - оценка дисперсии ошибки u ;

RSS - сумма квадратов ошибок.

Полученное по формуле (4) значение расчётного критерия Стьюдента $t_{p_{\alpha_i}}$ для каждого коэффициента $\alpha_0, \alpha_1, \dots, \alpha_i$ сравнивается с табличным (приложение Б) при уровне значимости $\alpha = 0,05$ или $0,01$. Если фактическое значение $t_{p_{\alpha_i}}$ превышает табличное, то коэффициент (параметр) модели значим ($\alpha_i \neq 0$) с вероятностью $(1-\alpha) \cdot 100 = 95\%$ или 99% . Следует принимать число степеней свободы $n-k$ (n - число наблюдений, k - число коэффициентов модели).

Порядок работы при проверке значимости коэффициента по t-статистике

1. Формулируем нулевую гипотезу H_0 о не значимости параметра $\alpha_i = 0$, и альтернативную H_1 – о его значимости.

2. Выбираем уровень значимости α (1% или 5%) – максимально допустимой вероятности ошибочного принятия гипотезы H_1 . Вычисляем число степеней свободы $(n-1)$.

3. По таблицам распределения Стьюдента определяем критическое значение критерия (приложение Б)

4. Если расчётный t-критерий по абсолютной величине больше табличного, то коэффициент является значимым при выбранном уровне значимости α . В противном случае коэффициент незначим (на данном уровне α)

Пример 2. Для уравнения регрессии, полученного в Примере 1., вычислить значение Т-критерия Стьюдента и определить статистическую значимость каждого коэффициента модели $\alpha_0, \alpha_1, \alpha_2$ при уровне значимости 5%.

1. Рассчитаем сумму квадратов ошибок:

$$RSS = Y^T \cdot Y - \alpha^T \cdot X^T \cdot Y = 218,25 - 218,1875 = 0,0625.$$

Тогда σ_u^2 - оценка дисперсии ошибки u : $\sigma_u^2 = 0,0625 / (4-3) = 0,0625$.

2. Рассчитаем дисперсионно-ковариационную матрицу

$$\begin{aligned} \frac{RSS}{n-k} \cdot [X^T \cdot X]^{-1} &= (0,0625/1) \cdot \begin{pmatrix} 0,25 & 0 & 0 \\ 0 & 0,25 & 0 \\ 0 & 0 & 0,25 \end{pmatrix} = \\ &= \begin{pmatrix} 0,015625 & 0 & 0 \\ 0 & 0,015625 & 0 \\ 0 & 0 & 0,015625 \end{pmatrix}. \end{aligned}$$

Тогда соответствующие дисперсии коэффициентов

$$\sigma_{\alpha 0}^2 = \sigma_{\alpha 1}^2 = \sigma_{\alpha 2}^2 = 0,015625.$$

Табличное значение (двусторонний) $t_{табл}(0.05; (4 - 3)) = 12.706$.

А расчётные значения t-критериев.

$$t_{p_{\alpha 0}} \frac{|6,125|}{\sqrt{0,015625}} = 49.$$

$$t_{p_{\alpha 1}} \frac{|-3,375|}{\sqrt{0,015625}} = 27.$$

$$t_{p_{\alpha 2}} \frac{|-2,375|}{\sqrt{0,015625}} = 19.$$

Каждый из коэффициентов значим $\alpha_i \neq 0$, т.к. больше табличного значения 12,7 при уровне значимости 5%.

Проверка адекватности уравнения множественной регрессии в целом

Значимость уравнения множественной регрессии в целом ($\alpha_1 \neq \alpha_2 \neq 0$) оценивается с помощью F-критерия Фишера:

$$F_p = \frac{R^2}{1 - R^2} \cdot \frac{n - k}{k - 1}, \quad (6)$$

где R^2 - коэффициент детерминации - индикатор степени подгонки модели к фактическим данным - или часть вариации (дисперсии) зависимой переменной y , которая объясняется уравнением регрессии. Чем меньше разброс значений остатков около линии регрессии по отношению к общему разбросу значений, тем лучше прогноз. Если имеется R-квадрат равный 0.4, то 40% от исходной изменчивости или дисперсии могут быть объяснены, а 60% остаточной изменчивости остаются необъясненными моделью. R-квадрат определяется при помощи формулы (7)

$$R^2 = \frac{\hat{\alpha}^T \cdot X^T \cdot Y - n \cdot \bar{Y}^2}{Y^T \cdot Y - n \cdot \bar{Y}^2}, \quad (7)$$

n - число наблюдений;

k - число коэффициентов факторов;

$\bar{Y}$ - среднее значение зависимой переменной.

Полученное по формуле (6) значение F сравнивается с табличным (приложение Б) при уровне значимости $\alpha = 0,05$ или $0,01$. Если фактическое значение F-критерия Фишера превышает табличное $F_p > F_{табл}$, то уравнение статистически значимо, т.е. модель адекватна ($\alpha_1 \neq \alpha_2 \neq 0$) с вероятностью $(1 - \alpha) * 100 = 95\%$ или 99% . Следует принимать число степеней свободы $k-1$ (по горизонтали), $n-k$ (по вертикали).

Порядок работы при проверке значимости уравнения по F-критерию

1. Формулируем нулевую гипотезу H_0 о неадекватности модели в целом ($\alpha_0 = \alpha_1 = \alpha_2 = 0$) и альтернативную H_1 – о ее адекватности.

2. Выбираем уровень значимости α (1% или 5%). Вычисляем число степеней свободы: $k-1$, $n-k$.

3. По таблицам F-распределения Фишера определяем критическое значение критерия (приложение Б).

4. Если F-расчётное больше F-табличного, то модель является в целом адекватной при выбранном уровне значимости α . В противном случае – не адекватной.

Пример 3. Для уравнения, полученного в Примере 1., вычислить значение F-критерия Фишера и определить статистическую значимость уравнения (адекватность модели) при уровне значимости 5%.

$$1. \hat{\alpha}^T \cdot X^T \cdot Y - n \cdot \bar{Y}^2 = 218,1875 - 150,0625 = 68,125.$$

$$2. Y^T \cdot Y - n \cdot \bar{Y}^2 = 218,25 - 150,0625 = 68,1875.$$

$$3. R^2 = 68,125 / 68,1875 = 0,999.$$

$$4. F_p = 545.$$

Сравним с $F_{табл}(0,05; (3-1); (4-3)) = 199.5$ для уровня значимости 5% и числа степеней свободы 2 и 1 (приложение Б).

Вывод: модель в целом адекватна, т.к. $F_p > F_{табл}$.

Предпосылки метода наименьших квадратов

Регрессионный анализ функции (1), основанный на обычном или одношаговом методе наименьших квадратов (МНК), должен удовлетворять четырем условиям Гаусса—Маркова:

1. Математическое ожидание случайной составляющей (ошибки), $M(u_i)$ в любом наблюдении должно быть равно нулю.

2. Дисперсия случайной составляющей σ_u^2 должна быть постоянна для всех наблюдений.

3. Отсутствие систематической связи между значениями случайной составляющей u_i в любых двух наблюдениях.

4. Случайная составляющая должна быть распределена независимо от переменных x_i , которые не являются стохастическими, т.е. не имеют случайной составляющей.

После того как проведена оценка параметров модели, рассчитывая разности фактических и теоретических значений $y - \tilde{y}_x$ можно получить оценки случайной составляющей (остатков или ошибок) u . В задачу регрессионного анализа входит не только построение самой модели, но и исследование остаточных величин.

Необходимость этого объясняется тем, что при использовании МНК предполагалось, что остатки представляют собой независимые случайные величины (условие №3, 4) и их среднее значение равно 0 (условие № 1); они имеют одинаковую (постоянную) дисперсию (условие №2).

Таким образом, исследование остатков предполагают проверку наличия перечисленных предпосылок МНК:

Случайных характер остатков

Для проверки строится график зависимости остатков u_i от теоретических значений результативного признака $\tilde{y}_x$. Если на графике получена горизонтальная полоса, то остатки представляют собой случайные величины и 1МНК оправдан, а теоретические значения хорошо аппроксимируют фактические значения y . Пример случайности остатков приведен на рисунке 2.

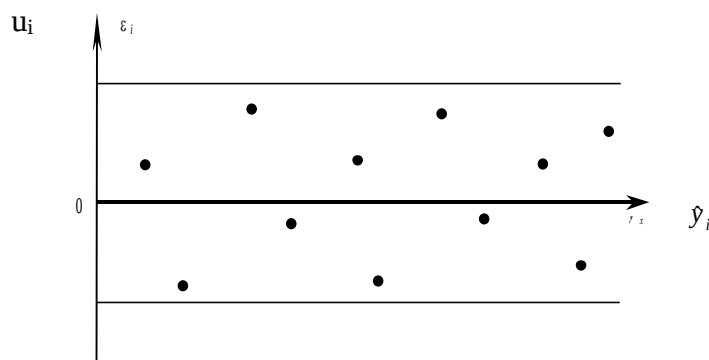


Рисунок 2 - Зависимость u_i от y

Возможны различные случаи зависимости остатков от теоретических

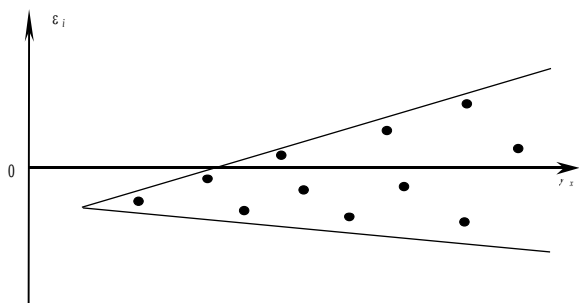
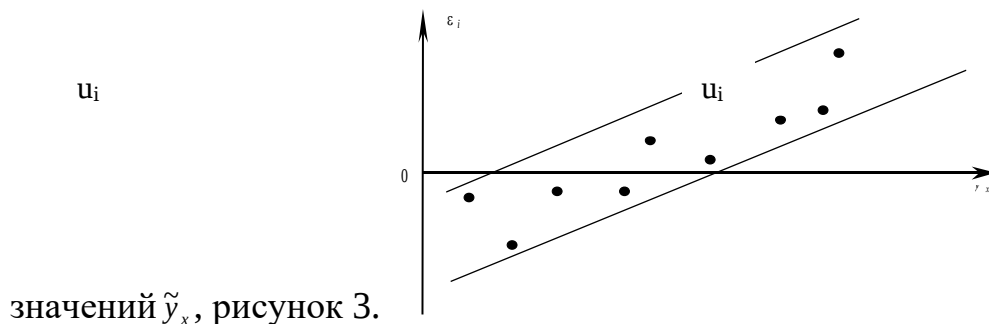


Рисунок 3 - Зависимость u_i от $\tilde{y}_x$

Нулевая средняя величина остатков, не зависящая от x_i

Эта предпосылка означает, что математическое ожидание случайной составляющей (ошибки), $M(u_i)$ в любом наблюдении должно быть равно нулю, т.е. $\sum_{i=1}^n (\tilde{y}_x - y_i) = 0$. Это условие выполнимо для линейных моделей. Для определения независимости величины остатков от x_i , как и в случае определения независимости u_i от y , строится график u_i от x_i . Если остатки на графике расположены в виде горизонтальной полосы, то они независимы от значений x_i . Если же зависимость присутствует, то модель является неадекватной.

Гомоскедастичность

Гомоскедастичность остатков означает, что дисперсия каждого отклонения одинакова для всех значений x . Если это условие не соблюдается, то имеет место гетероскедастичность. Наличие гетероскедастичности можно наглядно видеть из поля корреляции, рисунок 4.

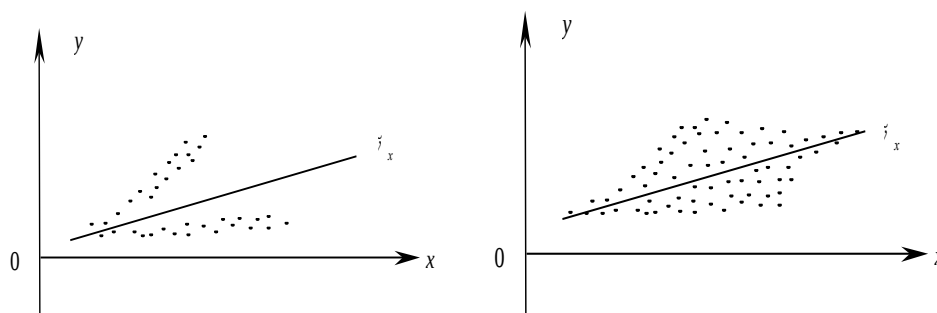


Рисунок 4 - Поле корреляции

Т.к. дисперсия характеризует отклонение, то из рисунков видно, что в первом случае дисперсия остатков растет по мере увеличения x , а во втором – дисперсия остатков достигает максимальной величины при средних значениях величины x и уменьшается при минимальных и максимальных значениях x . Наличие гетероскедастичности будет сказываться на уменьшении эффективности оценок параметров уравнения регрессии. Наличие гомоскедастичности или гетероскедастичности можно определять также по графику зависимости остатков от теоретических значений y .

Отсутствие автокорреляции остатков

Под автокорреляцией остатков понимают зависимость распределения значений остатков u_i друг от друга. Автокорреляция остатков означает наличие корреляции между остатками текущих и предыдущих (последующих) наблюдений. Оценить эту зависимость можно вычислив коэффициент корреляции между этими остатками по формуле (8).

$$r_{u_i u_j} = \frac{\overline{u_i u_j} - \bar{u}_i \cdot \bar{u}_j}{\sigma_{u_i} \sigma_{u_j}}. \quad (8)$$

Если этот коэффициент окажется существенно отличным от нуля, то остатки автокоррелированы.

Пример 4.

По четырем предприятиям региона (таблица 2) изучается зависимость выработки продукции на одного работника y (тыс. грн.) от ввода в действие новых основных фондов x_2 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_1 (%). Составлена модель множественной регрессии $y = 2,38462 + 3,38462x_1 + 2,15385x_2$.

Проверить выполнение предпосылок 1МНК для полученной модели.

Таблица 2 – Анализ остатков (ошибок) уравнения регрессии

Предприятие \ Показатель	№1	№2	№3	№4
x_1 , (%)	1	2	3	5
x_2 , (%)	0	1	3	4
y , (тыс. грн.)	6	11	19	28
$\tilde{y}_x$, (тыс. грн.)	5,76923	11,30769	19	27,92308
u_i , (тыс. грн.)	0,23077	-0,30769	0	0,07692

Вычисляем теоретические значения $\tilde{y}_x$ по уравнению регрессии, а остатки по формуле $u = y - \tilde{y}_x$ и записываем в таблицу 2.

Для проверки первого условия вычислим величину суммарного отклонения:

$$\sum_{k=1}^n (y - \tilde{y}_x) = 0,23077 - 0,30769 + 0 + 0,07692 = 0.$$

По этой величине можно сделать вывод о нулевой средней величине остатков – первое условие соблюдается.

Теперь для проверки случайного характера остатков построим график их зависимости от теоретических значений $\tilde{y}_x$, рисунок 5.

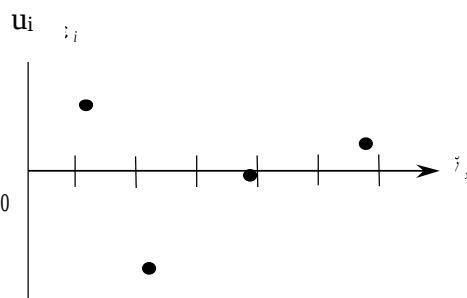


Рисунок 5 - Зависимость u_i от $\tilde{y}_x$

Хотя по четырем точкам судить трудно, но в целом можно сделать вывод, что остатки распределены случайно. Из этого же рисунка можно сделать вывод о гомоскедастичности остатков, т. к. дисперсия каждого отклонения одинакова для всех значений x .

Коэффициент автокорреляции остатков находим по следующим рядам данных:

u_i , (тыс. грн.)	-0,30769	0	0,07692
u_{i-1} , (тыс. грн.)	0,23077	-0,30769	0

$$\bar{u}_i = \frac{-0,307 + 0 + 0,07692}{3} = -0,07692$$

$$\bar{u}_{i-1} = \frac{0,23077 - 0,30769 + 0}{3} = -0,02564$$

$$\overline{u_i u_j} = \frac{(-0,30769) \cdot 0,23077 + 0 \cdot (-0,30769) + 0 \cdot (0,07692)}{3} = -0,02367$$

$$\begin{aligned} \sigma_{u_i} &= \sqrt{\frac{\sum_{k=1}^3 (u_{ik} - \bar{u}_i)^2}{3}} = \\ &= \sqrt{\frac{(-0,30769 + 0,07692)^2 + (0 + 0,07692)^2 + (0,07692 + 0,07692)^2}{3}} = 0,166 \end{aligned}$$

$$\begin{aligned} \sigma_{u_{i-1}} &= \sqrt{\frac{\sum_{k=1}^3 (u_{i-1k} - \bar{u}_{i-1})^2}{3}} = \\ &= \sqrt{\frac{(0,23077 + 0,02564)^2 + (-0,30769 + 0,02564)^2 + (0 + 0,02564)^2}{3}} = 0,2205 \end{aligned}$$

Отсюда находим

$$r_{u_i u_j} = \frac{\overline{u_i u_j} - \bar{u}_i \cdot \bar{u}_j}{\sigma_{u_i} \sigma_{u_j}} = \frac{-0,02367 - (-0,07692) \cdot (-0,02564)}{0,166 \cdot 0,2205} = -0,7$$

Коэффициент автокорреляции показывает существенную связь между значениями случайной составляющей u_i , что является нарушением одной из рассматриваемых предпосылок и некорректности применения 1МНК к исходным данным.

ЗАДАНИЕ

Задание 1.

Даны матрицы $A = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 4 & 1 \\ 1 & -4 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$, $C = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$ и число $\alpha = 2$. Найти $A^T B + \alpha C$, (Приложение А).

Задание 2. Задана квадратная матрица 3-го порядка. Вычислите определитель $\det A$ и обратную матрицу A^{-1} .

$$A = \begin{bmatrix} 1 & 4 & 1 \\ 2 & 3 & 1 \\ 3 & 2 & 0 \end{bmatrix}$$

Задание 3.

По исходным данным таблицы 3 в соответствии с вариантом выполнить:

1. Оценить параметры $\alpha_0, \alpha_1, \alpha_2$ модели $y = y + u = \alpha_0 + \alpha_1 \cdot x_1 + \alpha_2 x_2 + u$
2. Проверить коэффициенты $\alpha_0, \alpha_1, \alpha_2$ на значимость
3. Оценить адекватность (значимость) модели в целом.
4. Проверить были ли все предпосылки к тому, чтобы применять МНК и линейное уравнение регрессии к исходным данным.

Таблица 3 – Исходные данные

Номер предприятия	1	2	3	4
x_1	1	1	-1	-1
x_2	1	-1	1	-1
Вариант	y			
1	6	10	19	29
2	5	11	19	29
3	6	11	16	29
4	8	11	19	29
5	6	11	19	26
6	6	12	19	29
7	10	11	19	29
8	8	13	19	30
9	5	10	19	29
10	6	11	18	31
11	6	11	19	31
12	6	11	17	25
13	6	11	15	26
14	6	11	20	28
15	6	13	18	29
16	6	16	20	29
17	6	11	19	29
18	5	12	19	29
19	9	11	19	29
20	6	11	15	29
21	6	11	18	29
22	6	11	19	30
23	4	11	19	29
24	6	15	19	29
25	6	10	19	29
26	6	11	13	29

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ № 5 «АНАЛИЗ МОДЕЛ С ГЕТЕРОСКЕДАСТИЧНОЙ СЛУЧАЙНОЙ СОСТАВЛЯЮЩЕЙ»

Гомоскедастичность и гетероскедастичность

Регрессионный анализ функции (1), основанный на обычном или одношаговом методе наименьших квадратов (МНК) должен удовлетворять четырем условиям Гаусса—Маркова:

1. Математическое ожидание случайной составляющей, $M(u_i)$ в любом наблюдении должно быть равно нулю.
2. Дисперсия случайной составляющей σ_u^2 должна быть постоянна для всех наблюдений.
3. Отсутствие систематической связи (автокорреляции) между значениями случайной составляющей u_i в любых двух наблюдениях.
4. Случайный характер остатков. Случайная составляющая должна быть распределена независимо от переменных y_i и x_i .

$$y = \alpha_0 + \alpha_1 \cdot x_1 + \dots + \alpha_n x_n + u, \quad (1)$$

где $y, x_1, \dots, x_n$ - зависимая и независимые переменные;

α_i – параметр регрессионной модели

u – аддитивная случайная составляющая (остаток).

Условия 1 и 2 известны как гомоскедастичность, что означает "одинаковый разброс" $\sigma^2_{u_i} = const$, и представлены на рисунке 1 (график №1). Другими словами, вероятность того, что величина u_i примет какое-то данное положительное (или отрицательное) значение, будет одинаковой для всех наблюдений.

Однако, в экономических и социальных исследованиях (во временных и пространственно-временных выборках) условия Гаусса-Маркова часто не выполняются. Если в данных выборках нарушается условие 2, т.е. для различных наблюдений величина u_i распределена по-разному: вероятность того, что в i -ом наблюдении величина примет какое-то конкретное значение отличается от вероятности принятия этого же значения в j -ом наблюдении. Данное свойство называется гетероскедастичностью, что означает "неодинаковый разброс" $\sigma^2_{u_i} \neq const$, и представлено на рисунке 1 (графики №2 и №3).

Следовательно, метод МНК в данном случае не даёт наилучших из всех возможных результатов: $\alpha_1 \dots \alpha_n$ не будут лучшими оценками коэффициентов, неправильно оцененная дисперсионно-ковариационная матрица отрицательно скажется на t - и F -тестах, что приведёт к принятию ошибочных гипотез.

Далее будем рассматривать модель линейной регрессии (1), в которой дисперсии случайных составляющих (остатков) u_i не являются постоянными от наблюдения к наблюдению (гетероскедастичны).

№1 Гомоскедастичность

$$\sigma^2_{u_i} = const$$

№2 Растущая

гетероскедастичность

$$\sigma^2_{u_i} \neq const$$

№3 U-образная

гетероскедастичность

$$\sigma^2_{u_i} \neq const$$

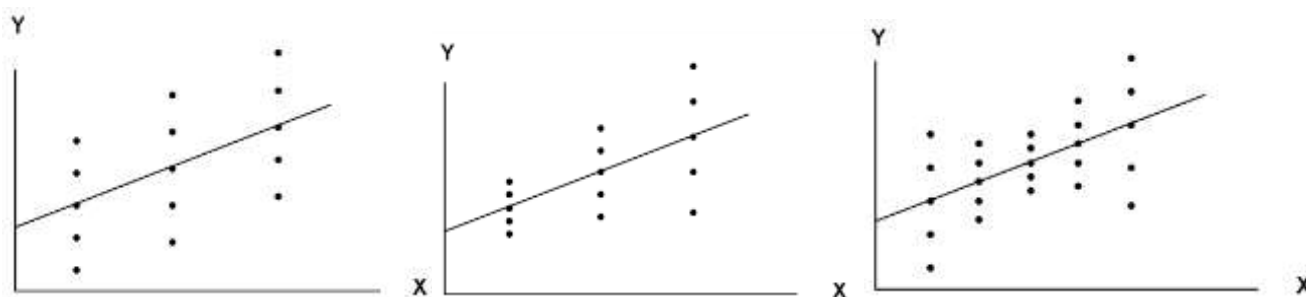


Рисунок 1 – Гомоскедастичные и гетероскедастичные случайные составляющие, u_i

Обнаружение гетероскедастичности

Для формальной проверки существует большое число статистических процедур (тестов), позволяющих обнаружить гетероскедастичность (например, тест Голдфелда-Квандта, тест Бреуша-Погана, тест Уайта, тест Глейзера и т.д.). В каждом тесте пытаются опровергнуть гипотезу о гомоскедастичности, если это удаётся, то можно сделать вывод, что в модели наблюдается гетероскедастичность.

Рассмотрим наиболее известный тест Голдфелда-Квандта (Goldfeld-Quandt, 1965) для обнаружения гетероскедастичности. При проведении проверки по предложенному ими критерию предполагается, что регрессия однофакторная $y = a_0 + a_1 \cdot x + u$ и стандартное отклонение пропорционально значению x . Метод состоит в оценке двух линий регрессии одношаговым методом наименьших квадратов (МНК). Первая линия строится на основе данных с наименьшими значениями регрессионных остатков (и дисперсии ошибки), вторая – на основе данных с наибольшими значениями остатков (и дисперсии ошибки). Если значения остатков (и дисперсий) обеих регрессий примерно одинаковы, то принимается гипотеза о гомоскедастичности. В противном случае можно считать, что в модели присутствует гетероскедастичность.

Тест состоит из следующих этапов:

1. Формулируются гипотезы H_0 = гомоскедастичность, H_1 = гетероскедастичность.
2. Располагаются данные в порядке возрастания величины x , пропорционально которой изменяется стандартное отклонение случайной составляющей.
3. Исключаются средние d наблюдений. d может быть выбрано, например, как $1/5$ всех наблюдений.
4. Оцениваются две регрессии по формуле (2). Первая из них использует наименьшие значения переменной x , вторая – наибольшие значения этой переменной. Каждая из регрессий построена на $(n-d)/2$ наблюдениях с $[(n-$

d)/2]-к степенями свободы (k равно числу коэффициентов, для однофакторной регрессии k=2). Величина d должна быть такой, чтобы гарантировать достаточность степеней свободы для правильной оценки каждой из регрессий.

$$\hat{\alpha} = [X^T \cdot X]^{-1} \cdot X^T \cdot Y, \quad (2)$$

где X - матрица независимых переменных; Y - вектор-столбец зависимой переменной.

5. Вычислить сумму квадратов остатков (ошибок) для каждой из регрессий: RSS_1 для малых значений x , и RSS_2 для больших x , формула (3).

$$RSS = Y^T \cdot Y - \hat{\alpha}^T \cdot X^T \cdot Y. \quad (3)$$

6. В предположении, что ошибки имеют нормальное распределение, статистика RSS_2/RSS_1 будет иметь F-распределение с $(n-d-4)/2$ степенями свободы. Опровергается гипотеза о гомоскедастичности в выбранном уровне значимости, если вычисленная статистика превышает соответствующее критическое значение F-распределения.

Тест Голдфелда-Квандта может применяться не только в случае парной линейной регрессии. В случае множественной модели число степеней свободы F-распределения будет $(n-d-2k)/2$.

Пример 1. Проверим однофакторную регрессионную модель $y = -70,4 + 2,52x$, построенную по данным таблицы 1 методом 1МНК, на гетероскедастичность (тест Голдфелда-Квандта, $\alpha = 1\%$ и 5%), исключив средние $d=N/5$ наблюдений, ($N=20$).

Таблица 1 – Исходные и сортированные данные X и Y

№	X	Y	X- сорт.	Y- сорт.
1	100	500	5	1
2	15	3	10	2
3	10	2	15	3
4	20	4	20	4
5	25	6	25	6
6	5	1	30	8
7	30	8	35	10
8	35	10	40	11
9	75	110	45	9
10	45	9	50	100
11	50	100	55	15
12	95	200	60	2
13	65	7	65	7
14	40	11	70	15

15	70	15	75	110
16	90	115	80	35
17	60	2	85	90
18	55	15	90	115
19	85	90	95	200
20	80	35	100	500

1. Сформулируем гипотезы H_0 = гомоскедастичность, H_1 = гетероскедастичность.
2. Отсортируем данные X и Y по возрастанию X и поместим результаты в таблицу 1 как X -сорт и Y -сорт.
3. Определим $d=4$ и исключим соответствующие данные из середины отсортированной выборки, получим две новые выборки.
4. Оценим параметры и RSS двух соответствующих регрессий по формулам (2 и 3).

Для первой регрессии рассчитаем коэффициенты

$$[X^T \cdot X]^{-1} = \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 10 & 15 & 20 & 25 & 30 & 35 & 40 \end{bmatrix} \bullet \begin{bmatrix} 1 & 5 \\ 1 & 10 \\ 1 & 15 \\ 1 & 20 \\ 1 & 25 \\ 1 & 30 \\ 1 & 35 \\ 1 & 40 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 8 & 180 \\ 180 & 5100 \end{bmatrix}^{-1} = \begin{bmatrix} 0,607143 & -0,021429 \\ -0,021429 & 0,000952 \end{bmatrix}.$$

$$X^T \cdot Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 10 & 15 & 20 & 25 & 30 & 35 & 40 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 6 \\ 8 \\ 10 \\ 11 \end{bmatrix} = \begin{bmatrix} 45 \\ 1330 \end{bmatrix}.$$

$$\hat{\alpha} = \begin{bmatrix} 0,607143 & -0,021429 \\ -0,021429 & 0,000952 \end{bmatrix} \bullet \begin{bmatrix} 45 \\ 1330 \end{bmatrix} = \begin{bmatrix} -1,1786 \\ 0,3024 \end{bmatrix}.$$

Рассчитаем RSS_1

$$Y^T \cdot Y = 351.$$

$$\hat{\alpha}^T \cdot X^T \cdot Y = [-1,1786 \quad 0,3024] \bullet \begin{bmatrix} 45 \\ 1330 \end{bmatrix} = 349,131.$$

$$RSS_1 = 351 - 349,131 = 1,869.$$

Для второй регрессии рассчитаем коэффициенты

$$[X^T \cdot X]^{-1} = \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 65 & 70 & 75 & 80 & 85 & 90 & 95 & 100 \end{bmatrix} \bullet \begin{bmatrix} 1 & 65 \\ 1 & 70 \\ 1 & 75 \\ 1 & 80 \\ 1 & 85 \\ 1 & 90 \\ 1 & 95 \\ 1 & 100 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 8 & 660 \\ 660 & 55500 \end{bmatrix}^{-1} = \begin{bmatrix} 6,607143 & -0,078571 \\ -0,078571 & 0,000952 \end{bmatrix}.$$

$$X^T \cdot Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 65 & 70 & 75 & 80 & 85 & 90 & 95 & 100 \end{bmatrix} \bullet \begin{bmatrix} 7 \\ 15 \\ 110 \\ 35 \\ 90 \\ 115 \\ 200 \\ 500 \end{bmatrix} = \begin{bmatrix} 1072 \\ 99555 \end{bmatrix}.$$

$$\hat{\alpha} = \begin{bmatrix} 6,607143 & -0,078571 \\ -0,078571 & 0,000952 \end{bmatrix} \bullet \begin{bmatrix} 1072 \\ 99555 \end{bmatrix} = \begin{bmatrix} -739,3214 \\ 10,5857 \end{bmatrix}.$$

Рассчитаем RSS_2

$$Y^T \cdot Y = 324924.$$

$$\hat{\alpha}^T \cdot X^T \cdot Y = [-739,3214 \quad 10,5857] \bullet \begin{bmatrix} 1072 \\ 99555 \end{bmatrix} = 261308,2.$$

$$RSS_2 = 324924 - 261308,2 = 63615,79.$$

5. Рассчитаем $\frac{RSS_2}{RSS_1} = \frac{63615,79}{1,869} = 34037$, а также число степеней свободы

$$(n-d-4)/2 = (20-4-4)/2 = 6$$

Для уровня значимости $\alpha = 1\%$

$$F_{\text{табл}} = 8,47$$

Для уровня значимости $\alpha = 5\%$

$$F_{\text{табл}} = 4,28$$

Поскольку вычисленное значение 34037 больше табличных значений F – критериев при выбранных уровнях значимости 1% и 5%, то в обоих случаях принимается гипотеза H_1 о наличии гетероскедастичности. Данный вывод подтверждается графически (рисунок 2).

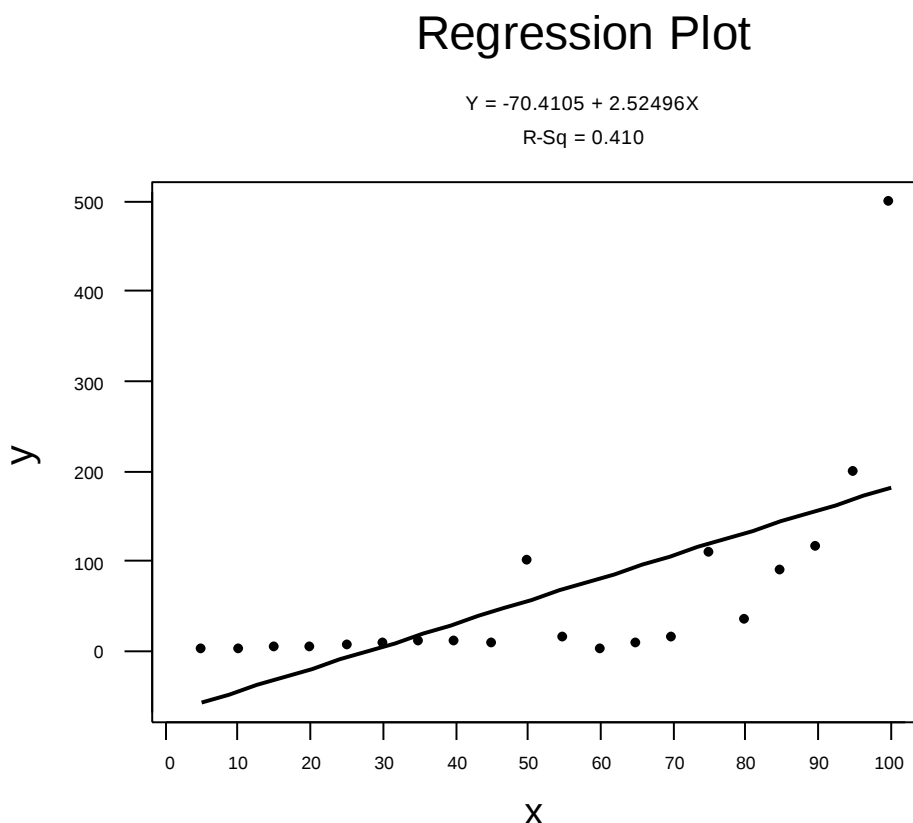


Рисунок 2 – Иллюстрация модели с гетероскедастичностью остатков

Использование взвешенного метода наименьших квадратов (ВМНК) для оценки моделей с гетероскедастичностью

Для улучшения качества оценки коэффициентов регрессионных моделей с неодинаковой дисперсией (гетероскедастичной случайной составляющей) используется взвешенный метод наименьших квадратов (ВМНК). На данную модель как и на классическую распространяется правило об отсутствии зависимости между случайными составляющими любых последовательных наблюдений. Метод ВМНК, как и МНК, применим к однофакторной и множественной линейной регрессии и использует то же правило минимизации суммы квадратов остатков (ошибок), RSS , однако, во ВМНК вместо одинаковых весов для каждого наблюдения им приписываются значения, обратные соответствующим дисперсиям ошибки, формула (4).

$$RSS = \sum_{i=1}^n w_i (y_i - \tilde{y}_{x_i})^2 = \sum_{i=1}^n \frac{1}{\sigma_{u_i}^2} (y_i - \tilde{y}_{x_i})^2, \quad (4)$$

где $y_i, \tilde{y}_{x_i}$ - фактическое и модельное i -е значение зависимой переменной;

$\sigma_{u_i}^2$ - дисперсия i -й случайной составляющей (ошибки);

w_i – вес i -го наблюдения, определяемый по формуле (5).

$$w_i = \frac{1}{\sigma_{u_i}^2}, \quad (5)$$

Тогда коэффициенты линейной регрессии найдём по формуле (6).

$$\hat{\alpha} = [X^T \cdot W \cdot X]^{-1} \cdot X^T \cdot W \cdot Y, \quad (6)$$

где $W = \text{diag}\{w_1, \dots, w_n\}$ – диагональная матрица весов;
 n – число наблюдений.

В редких случаях когда известны $\sigma_{u_i}^2$ данная процедура оценки предполагает составление матрицы весов W по известным значениям $\sigma_{u_i}^2$ и используется формула (7) для расчёта коэффициентов.

В более частом случае, дисперсия ошибки $\sigma_{u_i}^2$ неизвестна, но между ней и значением какой-либо объясняющей переменной x_k в регрессионной модели существует линейная зависимость вида (7).

$$\sigma_{u_i}^2 = c \cdot x_{1i}^2, \quad (7)$$

где c – ненулевая константа;

x_{1i} – значение объясняющей переменной x_1 (в данном примере) в i -ом наблюдении.

В случае подобного соотношения $\sigma_{u_i}^2$ можно считать известными, т.к. постоянная величина « c » не влияет на взвешенную процедуру (сокращается в процессе вычисления коэффициентов). Тогда значения весов w_i определяем по формуле (8) для регрессионной модели (9).

$$w_i = \frac{1}{c \cdot x_{1i}^2}, \quad (8)$$

$$y_i = \alpha_0 + \alpha_1 x_{1i} + \dots + \alpha_p x_{pi} + u_i. \quad (9)$$

Составив таким образом диагональную матрицу весов W , также используем формулу (6) для расчёта коэффициентов.

Пример 2.

Рассмотрим пример расчёта коэффициентов линейной однофакторной регрессии $y_i = \alpha_0 + \alpha_1 x_{1i} + u_i$ с гетероскедастичностью по данным таблицы 2, $\sigma_{u_i}^2 = c \cdot x_{1i}^2$.

Таблица 2 – Величины X и Y

X	Y
10	1

20	3
30	4
80	11

Составим матрицу весов W .

$$W = \begin{bmatrix} \frac{1}{c \cdot 100} & 0 & 0 & 0 \\ 0 & \frac{1}{c \cdot 400} & 0 & 0 \\ 0 & 0 & \frac{1}{c \cdot 900} & 0 \\ 0 & 0 & 0 & \frac{1}{c \cdot 6400} \end{bmatrix}.$$

В дальнейшем опустим константу «с», поскольку она сокращается в процессе расчётов. Оценим коэффициенты по формуле (6).

$$\begin{aligned} [X^T \cdot W \cdot X]^{-1} &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 10 & 20 & 30 & 80 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{100} & 0 & 0 & 0 \\ 0 & \frac{1}{400} & 0 & 0 \\ 0 & 0 & \frac{1}{900} & 0 \\ 0 & 0 & 0 & \frac{1}{6400} \end{bmatrix} \cdot \begin{bmatrix} 1 & 10 \\ 1 & 20 \\ 1 & 30 \\ 1 & 80 \end{bmatrix}^{-1} = \begin{bmatrix} 239,2523364 & -11,71339564 \\ -11,71339564 & 0,823468328 \end{bmatrix}. \\ [X^T \cdot W \cdot Y] &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 10 & 20 & 30 & 80 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{100} & 0 & 0 & 0 \\ 0 & \frac{1}{400} & 0 & 0 \\ 0 & 0 & \frac{1}{900} & 0 \\ 0 & 0 & 0 & \frac{1}{6400} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 4 \\ 11 \end{bmatrix} = \begin{bmatrix} 0,023663194 \\ 0,520833333 \end{bmatrix}. \end{aligned}$$

$$\hat{\alpha} = [X^T \cdot W \cdot X]^{-1} \cdot X^T \cdot W \cdot Y = \begin{bmatrix} 239,2523 & -11,71334 \\ -11,7134 & 0,823468 \end{bmatrix} \cdot \begin{bmatrix} 0,023663 \\ 0,520833 \end{bmatrix} = \begin{bmatrix} -0,43925 \\ 0,151713 \end{bmatrix}.$$

Полученная модель: $y = -0,43925 + 0,151713x$.

ЗАДАНИЕ

Задание 1. Проверить модель $y = -83 + 2,52x$, построенную по данным таблицы 3 с использованием МНК, на гетероскедастичность (тест Голдфелда-Квандта $\alpha = 5\%$). Исключить средние $d=N/5$ наблюдений. Построить график.

Таблица 3. – Исходные и сортированные данные X и Y.

№	X	Y	X- сорт.	Y- сорт.
1	105	500		
2	20	3		
3	15	2		
4	25	4		
5	30	6		
6	10	1		
7	35	8		
8	40	10		
9	80	110		
10	50	9		
11	55	100		
12	100	200		
13	70	7		
14	45	11		
15	75	15		
16	95	115		
17	65	2		
18	60	15		
19	90	90		
20	85	35		

Задание 2.

Оценить коэффициенты линейной однофакторной регрессии $y_i = \alpha_0 + \alpha_1 x_i + u_i$ с гетероскедастичностью методом ВМНК по данным таблицы 4, $\sigma_{u_i}^2 = c \cdot x_{ii}^2$.

Таблица 4 – Величины X и Y

X	Y
30	1
40	3
50	4
100	11

Анализ главных компонент как метод редукции данных

Предположим, что необходимо оценить удовлетворенность людей жизнью, для чего составляется вопросник с различными пунктами, среди которых: удовлетворены ли люди своим хобби (п.1) и как интенсивно они им занимаются (п. 2). Две переменные -исходные признаки- x_1 и x_2 (ответы на данные два пункта) коррелированы между собой. Из высокой коррелированности двух этих переменных можно сделать вывод об избыточности двух пунктов вопросника и возможности объединения их в один фактор- новую переменную y_1 , которая будет являться линейной комбинацией x_1 и x_2 . Зависимость между переменными x_1 и x_2 можно обнаружить путем подгонки линии регрессии (МНК), на основе которой можно определить y_1 , которая будет включать наиболее существенные черты обеих переменных.

Если приведённый выше пример распространить на большее число переменных $x_1, \dots, x_p$, то он будет иллюстрировать основную идею метода главных компонент – представление двух или более коррелированных (зависимых между собой) переменных одним фактором или главной компонентой y_1 . Метод главных компонент (РСА) был предложен Пирсоном в 1901 году и затем вновь открыт и детально разработан Хоттелингом (1933). Данный метод снижения размерности позволяет путём анализа меньшего набора показателей более просто объяснить многомерные структуры с минимальной потерей информации.

Представим графически исходное пространство переменных x_1, x_2, x_3 (зависимость между ними) как диаграмму рассеяния (рисунок 1). Процедура выделения главных компонент y_i подобна вращению новой оси y , максимизирующему дисперсию исходного пространства переменных. Этот тип вращения называется вращением, максимизирующим дисперсию, так как цель вращения заключается в максимизации дисперсии "новой" переменной (фактора) y_i и минимизации разброса вокруг нее.

Задача анализа метода главных компонент состоит в максимизации дисперсии, связанной с первой главной компонентой. Геометрически это выглядит как ориентация новой координатной оси y_1 (ось совпадает с линией регрессии) вдоль направления наибольшей вытянутости эллипсоида рассеивания объектов исследуемой выборки в пространстве признаков $x_1, \dots, x_p$, (рисунок 1).

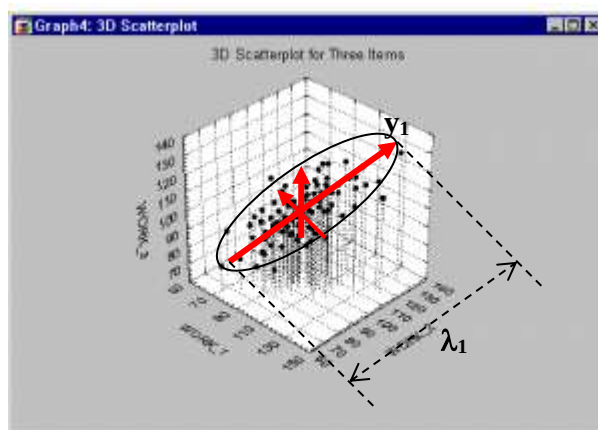


Рисунок 1 - Диаграмма рассеяния x_1, x_2, x_3

После того, как найдена линия (первый фактор y_1 выделен), для которой дисперсия максимальна, вокруг нее остается некоторый разброс данных. Процедура повторяется и определяется следующая линия (фактор y_2), максимизирующая остаточную вариацию (разброс данных вокруг первой прямой), и т.д. Таким образом, последовательно выделяются факторы y_i . (i равно числу начальных переменных x). Так как каждый последующий фактор определяется так, чтобы максимизировать изменчивость, оставшуюся от предыдущих, то факторы оказываются независимыми друг от друга (некоррелированными или ортогональными). Для случая более трех переменных, становится невозможным представить точки на диаграмме рассеяния, однако логика вращения осей с целью максимизации дисперсии нового фактора остается прежней.

Т.о. метод главных компонент осуществляет переход к новой системе координат $y_1, \dots, y_p$ в исходном пространстве признаков $x_1, \dots, x_p$, которая является системой ортонормированных линейных комбинаций (1).

$$\left\{ \begin{array}{l} y_j(x) = w_{1j} \left(\frac{x_1 - \bar{x}_1}{\sigma_1} \right) + \dots + w_{pj} \left(\frac{x_p - \bar{x}_p}{\sigma_p} \right); \\ \sum_{i=1}^p w_{ij}^2 = 1 \quad (j = 1 \dots p); \\ \sum_{i=1}^p w_{ij} w_{ik} = 0 \quad (j, k = 1 \dots p, j \neq k); \end{array} \right. \quad (1)$$

где $\bar{x}_j$ и σ_j — среднее арифметическое и среднеквадратическое отклонение признака x_i ;

w_{ij} - коэффициенты главных компонент, максимизирующие дисперсию y_j , находятся из уравнения $(S - \lambda E) \cdot \vec{w} = 0$, которое имеет решение, если $|S - \lambda E| = 0$, где S - ковариационная матрица, λ_i - соответствующие собственные числа этой матрицы - равны дисперсиям проекций множества объектов на оси главных компонент (или диагоналям эллипса рассеяния, рисунок 1).

Векторы $w_1 = (w_{11}, \dots, w_{p1})'$, ... , $w_p = (w_{1p}, \dots, w_{pp})'$ единичной длины называются собственными (характеристическими) векторами матрицы S .

Этапы метода главных компонент

1. Подготовительный этап.

Центрирование и нормирование переменных – переход к $\frac{x_{ij} - \bar{x}_j}{\sigma_j}$ (j - номер исходной переменной, i -номер наблюдения или реализации j -й переменной, а $\bar{x}_j$ и σ_j - её

среднее арифметическое и среднеквадратическое отклонение). Данный этап важен, если показатели измеряются в различных единицах или имеется большой разброс значений. В результате x_j – центрированное и нормированное, имеет нулевое среднее и единичную дисперсию.

Пример 1. Пусть заданы средние месячные доходы x_1 (тыс.грн.) и средние месячные сбережения x_2 (тыс.грн.) для 4-х групп опрошенных.

исходные показатели:

x_1	x_2
8	1,79
18	4,00
16	6,11
20	8,42

преобразованные показатели:

x_1	x_2
-1,426	-1,158
0,475	-0,380
0,095	0,363
0,856	1,176

$$\bar{x}_1 = 15,5; \bar{x}_2 = 5,08,$$

$$\sigma^2_1 = 27,66667; \sigma^2_2 = 8,069.$$

2. Вычисление матрицы ковариаций (S) или корреляций (R).

Диагональные элементы матрицы ковариаций S являются оценками дисперсий случайных величин x_j , остальные элементы – соответствующие коэффициенты ковариации, рассчитываемые по формуле (2). Диагональные элементы матрицы корреляций R равны «1», а остальные элементы – соответствующие коэффициенты корреляции, рассчитываемые по формуле (3).

$$\text{cov}_{kj} = \frac{1}{n-1} \sum_{i=1}^n (x_{ik} - \bar{x}_k) (x_{ij} - \bar{x}_j), \quad (2)$$

$$r_{kj} = \frac{\text{cov}_{kj}}{\sigma_k \cdot \sigma_j}. \quad (3)$$

Тогда матрица ковариаций рассматриваемого примера:

$$S = \begin{pmatrix} \sigma^2_{x1} & \text{cov}_{1,2} \\ \text{cov}_{2,1} & \sigma^2_{x2} \end{pmatrix} = \begin{pmatrix} 1 & 0,837 \\ 0,837 & 1 \end{pmatrix}.$$

3. Нахождение собственных чисел $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ симметричной положительно определенной матрицы S (или R) как вектора решений характеристического уравнения $|S - \lambda E| = 0$ (если используется матрица ковариаций S) или $|R - \lambda E| = 0$ (если используется матрица корреляций R). E – единичная матрица (квадратная матриц, элементы главной диагонали которой равны единице, а остальные равны нулю)

Для рассматриваемого примера:

$$\left| \begin{pmatrix} 1 & 0,837 \\ 0,837 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = \begin{vmatrix} 1-\lambda & 0,837 \\ 0,837 & 1-\lambda \end{vmatrix} = 0,$$

$$\begin{aligned}(1-\lambda)^2 - 0,837^2 &= 0, \\ (1-\lambda-0,837) \cdot (1-\lambda+0,837) &= 0, \\ \lambda_2 &= 0,163 \quad \cup \quad \lambda_1 = 1,837.\end{aligned}$$

4. Нахождение для каждого корня характеристического уравнения λ_i собственного вектора. Собственный вектор – это решение системы уравнений (4), если используется матрица S и решение системы (5), если используется матрица R.

$$(S - \lambda E) \cdot \vec{w} = 0, \quad (4)$$

$$(R - \lambda E) \cdot \vec{w} = 0, \quad (5)$$

где $\vec{w}$ - собственный вектор.

Для рассматриваемого примера воспользуемся системой (4):

А) для $\lambda_1 = 1,837$

$$\begin{pmatrix} 1-1,837 & 0,837 \\ 0,837 & 1-1,837 \end{pmatrix} \begin{pmatrix} w_{11} \\ w_{21} \end{pmatrix} = 0.$$

$$\begin{cases} -0,837 * w_{11} + 0,837 * w_{21} = 0; \\ 0,837 * w_{11} - 0,837 * w_{21} = 0. \\ (w_{11})^2 + (w_{21})^2 = 1 \end{cases}$$

$$w_{11} = \frac{1}{\sqrt{2}} \quad w_{21} = \frac{1}{\sqrt{2}} \quad (\text{используются арифметические корни})$$

Б) для $\lambda_2 = 0,163$

$$\begin{pmatrix} 1-0,163 & 0,837 \\ 0,837 & 1-0,163 \end{pmatrix} \begin{pmatrix} w_{12} \\ w_{22} \end{pmatrix} = 0$$

$$\begin{cases} 0,837 * w_{12} + 0,837 * w_{22} = 0 \\ 0,837 * w_{12} + 0,837 * w_{22} = 0 \\ (w_{12})^2 + (w_{22})^2 = 1 \end{cases} \quad (\text{используются арифметические корни})$$

$$w_{12} = -\frac{1}{\sqrt{2}} \quad w_{22} = \frac{1}{\sqrt{2}}$$

5. Нахождение линейных комбинаций для всех главных компонент y_j , система уравнений (1).

Для рассматриваемого примера:

$$y_1 = \frac{1}{\sqrt{2}} x_1 + \frac{1}{\sqrt{2}} x_2, \quad y_2 = -\frac{1}{\sqrt{2}} x_1 + \frac{1}{\sqrt{2}} x_2.$$

6. Анализ вклада каждой главной компоненты и их ранжирование по убыванию (таблица 1).

Таблица 1 – Анализ вклада главных компонент y_1, y_2

Главная компонента y_j	Y_1	y_2
Собственное число λ_j	1,837	0,163
Вклад j -главной компоненты, I_j %	91,85	8,15

Суммарный вклад, $I_2\%$	91,85	100
--------------------------	-------	-----

$$I_1 = \frac{1,837}{1,837 + 0,163} * 100 = 91,85\%, \quad I_2 = 8,15\%.$$

$$I = I_1 + I_2 = 100\%.$$

Пример 2.

Пусть в результате анализа трёх исходных переменных x_1, x_2, x_3 получена корреляционная матрица:

$$R = \begin{pmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{2,1} & 1 & r_{2,3} \\ r_{3,1} & r_{3,2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0,8 & 0 \\ 0,8 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

где r_{ij} -коэффициент корреляции переменных x_i и x_j , формула (3). Найдём собственный вектор из системы (5), решив характеристическое уравнение $|R - \lambda E| = 0$.

$$\left| \begin{pmatrix} 1 & 0,8 & 0 \\ 0,8 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right| = \left| \begin{pmatrix} 1-\lambda & 0,8 & 0 \\ 0,8 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{pmatrix} \right| = 0;$$

$$\text{Det} = (1-\lambda) \cdot (1-\lambda)^2 - 0,8 \cdot (0,8 \cdot (1-\lambda)) = 0;$$

$$(1-\lambda) \cdot ((1-\lambda)^2 - 0,8^2) = 0;$$

$$\lambda_1 = 1,8; \lambda_2 = 1; \lambda_3 = 0,2;$$

$$1) \lambda_1 = 1,8;$$

$$\begin{pmatrix} -0,8 & 0,8 & 0 \\ 0,8 & -0,8 & 0 \\ 0 & 0 & -0,8 \end{pmatrix} \begin{pmatrix} w_{11} \\ w_{21} \\ w_{31} \end{pmatrix} = 0 \Leftrightarrow \begin{cases} -0,8 \cdot w_{11} + 0,8 \cdot w_{21} = 0; \\ 0,8 \cdot w_{11} - 0,8 \cdot w_{21} = 0; \\ -0,8 \cdot w_{31} = 0; \\ (w_{11})^2 + (w_{21})^2 + (w_{31})^2 = 1. \end{cases}$$

$$\begin{cases} w_{11} = w_{21}; \\ w_{31} = 0; \\ 2(w_{11})^2 = 1. \end{cases}$$

$$w_{11} = w_{21} = \frac{1}{\sqrt{2}}, w_{31} = 0.$$

$$y_1 = \frac{1}{\sqrt{2}} x_1 + \frac{1}{\sqrt{2}} x_2$$

$$2) \lambda_2 = 1;$$

$$\begin{pmatrix} 0 & 0,8 & 0 \\ 0,8 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} w_{12} \\ w_{22} \\ w_{32} \end{pmatrix} = 0$$

$$\begin{cases} 0,8 \cdot w_{22} = 0; \\ 0,8 \cdot w_{12} = 0; \\ (w_{12})^2 + (w_{22})^2 + (w_{32})^2 = 1. \end{cases}$$

$$w_{12}=w_{22}=0, w_{32}=1.$$

$$y_2 = x_3$$

$$3)\lambda_3 = 0,2;$$

$$\begin{pmatrix} 0,8 & 0,8 & 0 \\ 0,8 & 0,8 & 0 \\ 0 & 0 & 0,8 \end{pmatrix} \begin{pmatrix} w_{13} \\ w_{23} \\ w_{33} \end{pmatrix} = 0 \quad \Leftrightarrow \quad \begin{cases} 0,8*w_{13} + 0,8*w_{23} = 0; \\ 0,8*w_{13} + 0,8*w_{23} = 0; \\ 0,8*w_{33} = 0; \\ (w_{13})^2 + (w_{23})^2 + (w_{33})^2 = 1. \end{cases}$$

$$w_{13} = \frac{1}{\sqrt{2}}, w_{23} = -\frac{1}{\sqrt{2}}, w_{33} = 0.$$

$$y_3 = \frac{1}{\sqrt{2}} x_1 - \frac{1}{\sqrt{2}} x_2.$$

Таблица 2 – Анализ вклада главных компонент y_1, y_2, y_3

главная компонента y_i	y_1	y_2	y_3
собственное число λ_j	1,8	1	0,2
вклад j -главной компоненты, I_j %	60	33,33	6,67
суммарный вклад, I %	60	93,33	100

ЗАДАНИЕ

Задание 1.

Дана матрица значений трёх переменных x_1, x_2, x_3 в таблице 3 и таблице 4.

1. Согласно варианту центрировать и нормировать исходные данные.
2. Вычислить корреляционную матрицу R .
3. Составить характеристическое уравнение и найти собственные числа λ_i корреляционной матрицы R .
4. Найти w_{ij} - коэффициенты главных компонент и составить линейные комбинации для y_i .
5. Рассчитать индивидуальный и суммарный вклад каждой главной компоненты

Таблица 3 - Исходные данные x_2 , x_3

x_2	x_3
-1	-1
-1	1
1	-1
1	1

Таблица 4 - Исходные данные x_1 согласно варианту

Вариант	x_1			
1	1	2	3	4
2	2	13	4	5
3	3	4	15	6
4	14	5	6	7
5	5	6	7	18
6	6	17	8	9
7	7	8	19	10
8	18	9	10	11
9	9	10	1	12
10	10	11	2	13
11	11	12	3	14
12	12	3	14	15
13	13	4	15	16
14	14	5	16	17
15	15	6	17	18
16	16	7	18	19
17	17	8	19	20
18	18	9	20	21
19	19	20	1	22
20	20	21	2	23
21	21	22	23	4
22	22	23	24	5
23	23	24	25	6
24	24	25	26	7
25	25	26	27	8
26	26	7	28	29

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ № 7

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Временной ряд представляет собой совокупность измерений некоторой переменной, производимых по мере возрастания времени. Существенным отличием временного ряда от выборки данных является то, что для временного ряда известен порядок поступления наблюдений.

Одним из методов анализа временных рядов является спектральный (Фурье) анализ, цель которого - разложить комплексные стационарные временные ряды с циклическими компонентами на несколько основных синусоидальных функций с определенной длиной волн. Стационарные временные ряды имеют постоянные по времени основные характеристики, такие как среднее, дисперсия и

автокорреляции; комплексные временные ряды подгоняются к нескольким моделям. Для проведения спектрального анализа нестационарного ряда необходимо предварительное исключение из него трендовой составляющей.

Термин "спектральный" заимствован из оптики: после преломления в призме пучок белого света («белый шум»), составленный хаотически из света с различными длинами волн, разлагается на волны различных частот, т.о. распознавая существенные основные периодические компоненты (рисунок 1).

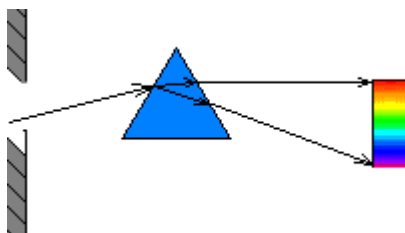


Рисунок 1 - Дисперсия белого видимого света при прохождении через стеклянную прозрачную призму

Аналогично и применение спектрального анализа к временным рядам: возможно изучить определённое явление, обнаружив всего несколько повторяющихся циклов различной длины в интересующих временных рядах, которые, на первый взгляд, выглядят как случайный шум. Спектральный анализ даёт возможность разложить временной ряд на функции синусов и косинусов различных частот, для определения тех, появление которых особенно существенно и значимо. Один из возможных способов сделать это - решить задачу линейной множественной регрессии (оценить коэффициенты функции Фурье, представляющей собой тригонометрический функциональный ряд), где зависимая переменная y_t - наблюдаемый временной ряд, а независимые переменные или регрессоры x_{ts} и x_{tc} - функции синусов и косинусов всех возможных (дискретных) частот, формула (1).

$$\begin{aligned}
 y_t &= a_0 + \sum_{i=1}^q [a_i \cos(2\pi f_i t) + b_i \sin(2\pi f_i t)] = \\
 &= a_0 + a_1 \cos(2\pi f_1 t) + b_1 \sin(2\pi f_1 t) + a_2 \cos(2\pi f_2 t) + b_2 \sin(2\pi f_2 t) + \dots \\
 &+ a_q \cos(2\pi f_q t) + b_q \sin(2\pi f_q t) = a_0 + a_1 x_{1c} + b_1 x_{1s} + a_2 x_{2c} + b_2 x_{2s} + \dots + a_q x_{qc} + b_q x_{qs}
 \end{aligned} \tag{1}$$

где a_0 - константа;

a_i, b_i – амплитуды соответствующих функций -отклонение от среднего значения до пика или впадины (рисунок 2). Величины амплитуд также являются коэффициентами регрессии, полученными в результате подгонки функций синусов и косинусов разных длин к фактическим данным наблюдений. Коэффициенты показывают степень, с которой соответствующие функции коррелируют с фактическими данными. Сами же синусы и косинусы на различных частотах не коррелированы (ортогональны).

$2\pi f_i$ - круговая частота соответствующей функции, радиан;

f_i – частота Фурье в данном частном случае, в общем случае – угловая частота – это число периодов (циклов), содержащихся в единичном интервале, а также величина обратная периоду T (рисунок 2), формула (2);

$$f_i = \frac{i}{N} = \frac{1}{T}, \quad (2)$$

T – период колебаний (длина волны) – время, которое проходит между повторением одинаковых фаз колебаний φ_k ;

φ_k – сдвиг функции относительно начала отсчёта в радианах (от 0 до 2π). В данном частном случае равен нулю, $\cos(2\pi f_i * t + \varphi_k)$;

i – номер соответствующих гармоник (косинуса и синуса) с определённым значением частоты;

N – число наблюдений;

q – для нечётного числа наблюдений $q=(N-1)/2$ (последнее значение опускается) и всего существует q различных синусов и косинусов, а также q значений периодограммы. Для чётного числа наблюдений $q= N/2$ существует q значений косинусов и $q-1$ значений синусов, а также q значений периодограммы.

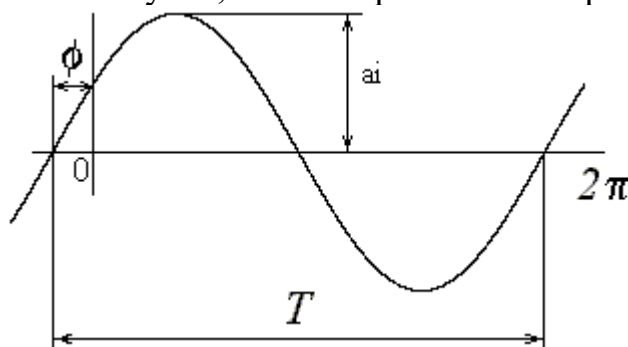


Рисунок 2 - Характеристики периодических функций

Для нечётного числа наблюдений, $q=(N-1)/2$, оценки коэффициентов a_0 , a_i , b_i методом наименьших квадратов будут определяться формулами (3-5)

$$a_0 = \bar{y}, \quad (3)$$

где $\bar{y}$ – среднее значение наблюдений.

$$a_i = \frac{2}{N} \sum_{t=1}^N y_t \cos(2\pi f_i t), \quad (4)$$

где t – номер наблюдения во временном ряду;
 $i=1,2,\dots,q$.

$$b_i = \frac{2}{N} \sum_{t=1}^N y_t \sin(2\pi f_i t), \quad (5)$$

Тогда периодограмма состоит из $q=(N-1)/2$ значений, формула (6).

$$I(f_i) = \frac{N}{2}(a_i^2 + b_i^2), \quad (6)$$

$$i = 1, 2, \dots, q$$

где $I(f_i)$ называется интенсивностью на частоте f_i и представляет собой значения периодограммы;

N - общая длина ряда.

Всё вышеизложенное предполагает наличие нечётного числа наблюдений $N=2q+1$. В случае, когда N – чётно, полагаем $N=2q$, применяя для расчёта значений коэффициентов и периодограммы формулы (4-6) только для $i=1, 2, \dots, q-1$, а для $i=q$ используя формулы (7-9).

$$a_q = \frac{1}{N} \sum_{t=1}^N (-1)^t y_t. \quad (7)$$

$$b_q = 0. \quad (8)$$

$$I(f_q) = I(0,5) = N a_q^2. \quad (9)$$

Таким образом, спектральный анализ определяет корреляцию функций синусов и косинусов различной частоты с наблюдаемыми данными. Если найденная корреляция велика (коэффициент при определенном синусе или косинусе), то можно заключить, что существует строгая периодичность на соответствующей частоте в данных.

Функции синусов и косинусов независимы (или ортогональны); поэтому можно просуммировать квадраты коэффициентов для каждой частоты, чтобы вычислить периодограмму, где $I(f_i)$ - значения периодограммы на частоте f_i .

Значения периодограммы можно интерпретировать как дисперсию (вариацию) данных на соответствующей частоте. Обычно значения периодограммы на графике изображаются в зависимости от частот или периодов.

ПРИМЕР ВЫПОЛНЕНИЯ РАСЧЁТОВ

Пусть существует набор из 12 среднемесячных температур y_t (в градусах Цельсия) для определённого региона в прошедшем году (таблица 1).

По формуле (3) получим $a_0 = 9,466667$.

Поскольку N - чётное, воспользуемся формулами (4-6) для $i=1 \dots q-1$ и формулами (7-9) для $i=q$.

Для вычисления коэффициента a_1 по формуле (4) рассчитаем в третьей строке таблицы величины $\cos(2\pi t/12)$ модельных (оценочных) значений функции косинуса частоты $f_1=1/12$ (формула (2)) в заданные моменты времени $t=1 \dots 12$. Используем таблицы Брадиса (Приложение А), отметив, что $\pi=3,1416$ радиан, а $\pi/6=0,5236$ радиан.

Тогда получим

$$a_1 = (2/12) * (3,4 * \cos(0,5236) + 4,5 * \cos(0,5236 * 2) + 4,3 * \cos(0,5236 * 3) + 8,7 * \cos(0,5236 * 4) + 13,3 * \cos(0,5236 * 5) + 13,8 * \cos(0,5236 * 6) + 16,1 * \cos(0,5236 * 7) + 15,5 * \cos(0,5236 * 8) + 14,1 * \cos(0,5236 * 9) + 8,9 * \cos(0,5236 * 10) + 7,4 * \cos(0,5236 * 11) + 3,6 * \cos(0,5236 * 12)) = -5,3$$

Аналогичным образом в четвёртой строке таблицы 1 рассчитываем величины $\cos(2\pi 2t/12) = \cos(\pi t/3)$ для частоты $f_2 = 2/12 = 1/6$ и получаем

$$a_2 = (2/12) * (3,4 * 0,500 + 4,5 * (-0,500) + \dots + 3,6 * 1) = 0,05;$$

В пятой строке таблицы 1 рассчитываем $\cos(2\pi 3t/12) = \cos(\pi t/2)$, тогда

$$a_3 = (2/12) * (3,4 * 0 + 4,5 * (-1) + \dots + 3,6 * 1) = 0,1;$$

В шестой строке таблицы рассчитываем $\cos(2\pi 4t/12) = \cos(2\pi t/3)$, тогда

$$a_4 = (2/12) * (3,4 * (-0,500) + 4,5 * (-0,5) + \dots + 3,6 * 1) = -0,5166;$$

В седьмой строке таблицы рассчитываем $\cos(2\pi 5t/12) = \cos(\pi 5t/6)$, тогда

$$a_5 = (2/12) * (3,4 * (-0,866) + 4,5 * (0,5) + \dots + 3,6 * 1) = 0,0847;$$

Поскольку N – чётное число, то используем формулу (7) для расчёта a_6 ($q=6$) получим.

$$a_6 = 1/12 (3,4 * (-1)^1 + 4,5 * (-1)^2 + \dots + 3,6 * (-1)^{12}) = -0,3$$

Для вычисления коэффициента b_1 по формуле (5) в восьмой строке рассчитаем $\sin 2\pi 1t/12 = \sin(\pi t/6)$ для частоты $f_1 = 1/12$, тогда

$$b_1 = (2/12) * (3,4 * (0,5) + 4,5 * (0,866) + \dots + 3,6 * 0) = -3,8166;$$

В девятой строке таблицы рассчитаем $\sin 2\pi 2t/12 = \sin(\pi t/3)$, тогда

$$b_2 = (2/12) * (3,4 * (0,866) + 4,5 * (0,866) + \dots + 3,6 * 0) = 0,1732;$$

В десятой строке таблицы рассчитаем $\sin(2\pi 3t/12) = \sin(\pi t/2)$, тогда

$$b_3 = (2/12) * (3,4 * (1) + 4,5 * (0) + \dots + 3,6 * 0) = 0,5;$$

В одиннадцатой строке таблицы рассчитаем $\sin(2\pi 4t/12) = \sin(2\pi t/3)$, тогда

$$b_4 = (2/12) * (3,4 * (0,866) + 4,5 * (-0,866) + \dots + 3,6 * 0) = -0,5196$$

В двенадцатой строке таблицы рассчитаем $\sin(2\pi 5t/12) = \sin(\pi 5t/6)$, тогда

$$b_5 = (2/12) * (3,4 * (0,5) + 4,5 * (-0,866) + \dots + 3,6 * 0) = -0,5834$$

Поскольку N – чётное, используем формулу (8) для расчёта b_6 ($i=q=6$) получим $b_6=0$.

Для $i=1 \dots (q-1)$ вычислим значения периодограммы $I(f_i)$ по формуле (6)

$$I(f_1) = (12/2) * ((-5,2846)^2 + (-3,8166)^2) = 254,9620;$$

$$\begin{aligned}
I(f_2) &= (12/2) * ((0,0500)^2 + (0,1732)^2) = 0,1950; \\
I(f_3) &= (12/2) * ((0,1000)^2 + (0,5000)^2) = 1,5601; \\
I(f_4) &= (12/2) * ((-0,5166)^2 + (-0,5196)^2) = 3,2215; \\
I(f_5) &= (12/2) * ((0,0847)^2 + (-0,5834)^2) = 2,0852.
\end{aligned}$$

Значение периодограммы $I(f_6)$ (для $i=q=6$) определяем по формуле (9)

$$I(f_6) = 12 * (-0,3)^2 = 1,0800.$$

Построим графики периодограммы, на первом из которых ось ОХ – частота Фурье f_i , а на втором ось ОХ – период, рассчитанный как величина обратная f_i . В обоих графиках ОУ- значение периодограммы $I(f_i)$ (таблица 2), (рисунок 3-4).

Вывод по примеру:

В заданном временном ряду основной периодической составляющей является функция $y_t = 9,466667 - 5,2846 \cos(\pi t/6) - 3,8166 \sin(\pi t/6)$ (рисунок 5) на частоте $f_i = 1/12 = 0,083333$, поскольку данная величина частоты соответствует максимальному значению периодограммы $I(f_i) = 254,9620$. Значение периодограммы, в свою очередь, обусловлено максимальными по модулю значениями коэффициентов регрессии a_1 и b_1 , свидетельствующих о корреляции функций синусов и косинусов (независимых переменных x_{ts} и x_{tc}) с наблюдаемыми данными y_t .

Таблица 1 - Набор среднемесячных температур

	t	1	2	3	4	5	6	7	8	9	10	11	12
i	yt	3,4	4,5	4,3	8,7	13,3	13,8	16,1	15,5	14,1	8,9	7,4	3,6
1	$\text{Cos}(2\pi 1t/12)$	0,866	0,500	0,000	-0,500	-0,866	-1,000	-0,866	-0,500	0,000	0,500	0,866	1,000
2	$\text{Cos}(2\pi 2t/12)$	0,500	-0,500	-1,000	-0,500	0,500	1,000	0,500	-0,500	-1,000	-0,500	0,500	1,000
3	$\text{Cos}(2\pi 3t/12)$	0,000	-1,000	0,000	1,000	0,000	-1,000	0,000	1,000	0,000	-1,000	0,000	1,000
4	$\text{Cos}(2\pi 4t/12)$	-0,500	-0,500	1,000	-0,500	-0,500	1,000	-0,500	-0,500	1,000	-0,500	-0,500	1,000
5	$\text{Cos}(2\pi 5t/12)$	-0,866	0,500	0,000	-0,500	0,866	-1,000	0,866	-0,500	0,000	0,500	-0,866	1,000
1	$\text{Sin}(2\pi 1t/12)$	0,500	0,866	1,000	0,866	0,500	0,000	-0,500	-0,866	-1,000	-0,866	-0,500	0,000
2	$\text{Sin}(2\pi 2t/12)$	0,866	0,866	0,000	-0,866	-0,866	0,000	0,866	0,866	0,000	-0,866	-0,866	0,000
3	$\text{Sin}(2\pi 3t/12)$	1,000	0,000	-1,000	0,000	1,000	0,000	-1,000	0,000	1,000	0,000	-1,000	0,000
4	$\text{Sin}(2\pi 4t/12)$	0,866	-0,866	0,000	0,866	-0,866	0,000	0,866	-0,866	0,000	0,866	-0,866	0,000
5	$\text{Sin}(2\pi 5t/12)$	0,500	-0,866	1,000	-0,866	0,500	0,000	-0,500	0,866	-1,000	0,866	-0,500	0,000

Таблица 2 - Значения коэффициентов и периодограммы

i	ai	bi	l(fi)	Частота, fi	Период, T=1/fi
1	-5,2846	-3,8166	254,9620	0,083333	12
2	0,0500	0,1732	0,1950	0,166667	6
3	0,1000	0,5000	1,5601	0,25	4
4	-0,5166	-0,5196	3,2215	0,333333	3
5	0,0847	-0,5834	2,0852	0,416667	2,4
6	-0,3000	0,0000	1,0800	0,5	2

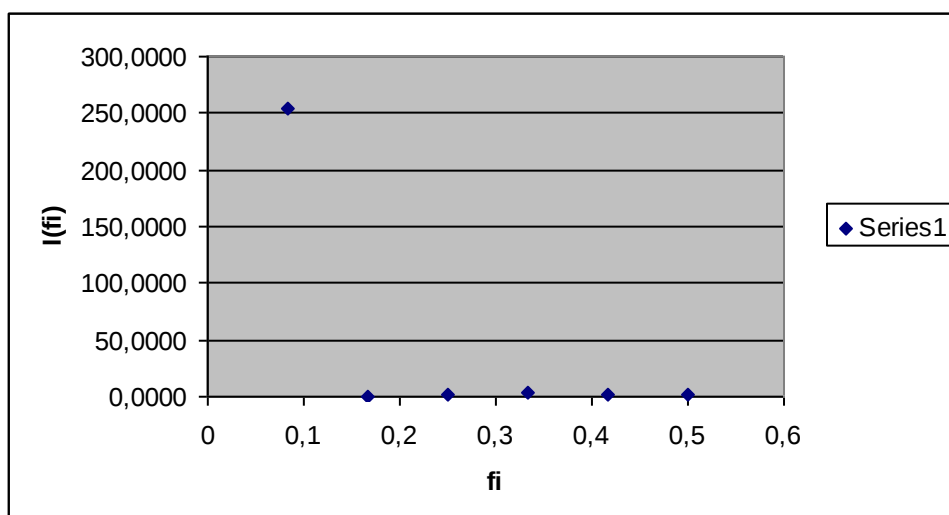


Рисунок 3 – Периодограмма: зависимость интенсивности от частоты

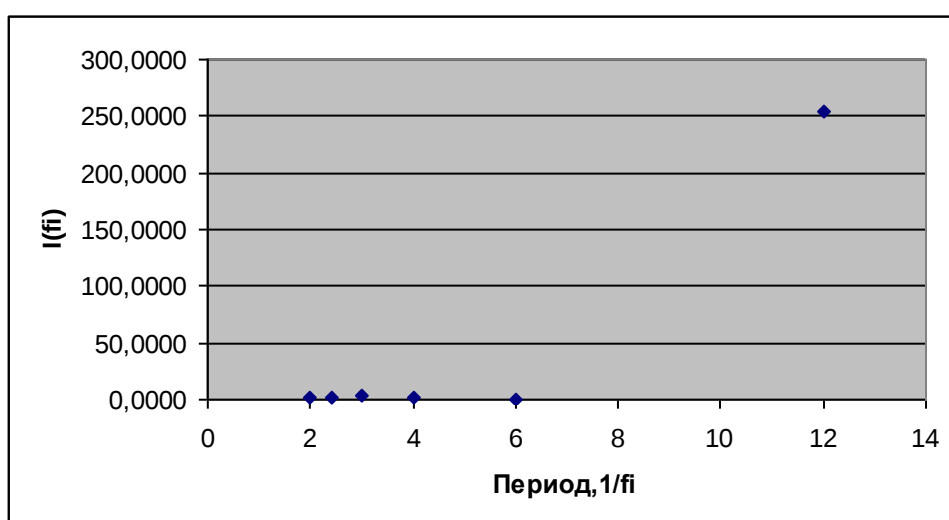


Рисунок 4 – Периодограмма: зависимость интенсивности от периода

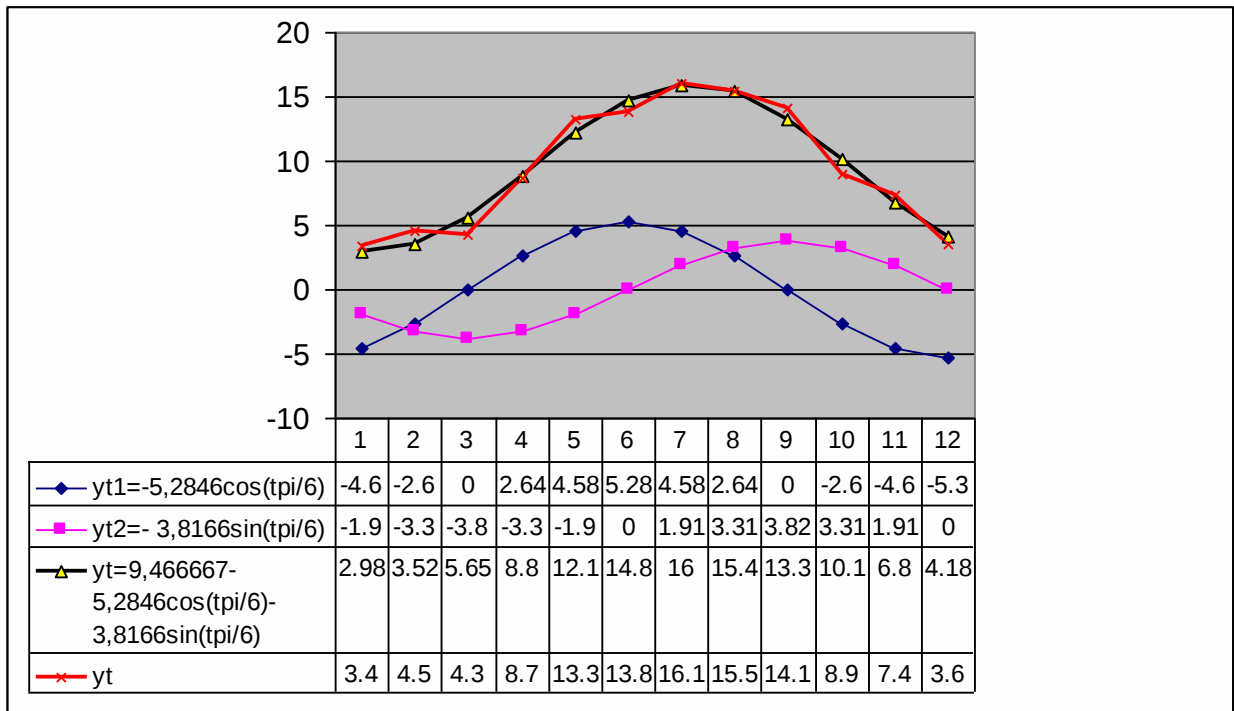


Рисунок 5 - Динамика среднемесячных температур

ИНДИВИДУАЛЬНЫЕ РАСЧЁТНО-ПРАКТИЧЕСКИЕ ЗАДАНИЯ

Дан временной ряд из 5 наблюдений (y_t) среднегодовых оценок студентов отдельного ВУЗа в баллах по 50 бальной шкале (таблица 6). Согласно варианту:

1. Рассчитать оценки функций косинуса и синуса частот f_i в моменты времени $t=1 \dots 5$ и оценить коэффициенты a_0 , a_i и b_i функции Фурье.
2. Рассчитать частоты Фурье f_i , периоды T_i , значения периодограммы $I(f_i)$.
3. Построить график периодограммы, сделать вывод об основных периодических составляющих и их частотах.
4. Составить прогноз уровня временного ряда для $t=6$

Таблица 6 – Временной ряд среднегодовых оценок

Период, t Вариант Yt	1	2	3	4	5
1	34	45	43	87	13
2	34	45	39	38	15
3	43	7	13	45	17
4	40	45	50	34	45
5	45	39	45	38	43
6	8	13	45	39	40
7	45	8	8	13	34
8	40	45	39	38	43
9	40	45	8	13	10
10	8	13	40	45	45
11	39	38	5	45	29
12	45	43	45	7	17
13	34	7	13	38	43
14	40	45	5	36	9
15	8	13	5	43	8
16	23	38	40	45	35
17	45	39	34	6	12
18	33	34	45	49	43
19	34	5	8	13	10
20	40	23	34	33	12
21	6	45	39	34	45
22	45	39	9	45	39
23	18	13	8	13	23
24	39	38	13	40	2
25	45	49	38	43	9
26	10	1	43	11	43

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №8 ОЦЕНКА ПАРАМЕТРОВ СИСТЕМ ЛИНЕЙНЫХ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

Системы эконометрических уравнений. Проблема идентификации

При изучении сложных систем для их описания и объяснения механизмов их функционирования недостаточно построения изолированных уравнений регрессии. При использовании отдельных уравнений регрессии, например, для экономических расчетов, в большинстве случаев предполагается, что аргументы (факторы) можно изменять независимо друг от друга. Однако, это предположение является очень грубым: практически изменение одной переменной, как правило, не может происходить при абсолютной неизменности других. Ее изменение повлечет за собой изменение во всей системе взаимосвязанных признаков. Следовательно, отдельно взятое уравнение множественной регрессии не может характеризовать истинные влияния отдельных признаков на вариацию результирующей переменной.

Этим объясняется необходимость использования не отдельных уравнений, а их систем.

Система уравнений в эконометрических исследованиях может быть построена различными способами:

1. Возможна система независимых уравнений, когда каждая зависимая переменная рассматривается как функция одного и того же набора факторов, модель (1):

$$\begin{cases} y_1 = a_{11} * x_1 + a_{12} * x_2 + \dots + a_{1m} * x_m + e_1 \\ \dots\dots\dots \\ y_n = a_{n1} * x_1 + a_{n2} * x_2 + \dots + a_{nm} * x_m + e_n \end{cases} \quad (1)$$

Примером такой модели может служить модель экономической эффективности сельскохозяйственного производства, где в качестве зависимых переменных выступают показатели эффективности производства (производительность, себестоимость продукции и т.д.), а в качестве факторов – характеристики самого хозяйства (количество голов скота, площадь пашни и т.д.).

Для системы независимых уравнений каждое уравнение может рассматриваться самостоятельно, и его параметры определяются обычным образом по одношаговому методу наименьших квадратов (МНК).

2. Также возможна система рекурсивных уравнений – когда зависимая переменная одного уравнения выступает в виде фактора в другом уравнении, модель (2).

$$\begin{cases} y_1 = a_{11} * x_1 + a_{12} * x_2 + \dots + a_{1m} * x_m + e_1 \\ y_2 = b_{21} * y_1 + a_{21} * x_1 + a_{22} * x_2 + \dots + a_{2m} * x_m + e_2 \\ y_3 = b_{31} * y_1 + b_{32} * y_2 + a_{31} * x_1 + a_{32} * x_2 + \dots + a_{3m} * x_m + e_3 \\ \dots\dots\dots \\ y_n = b_{n1} * y_1 + b_{n2} * y_2 + \dots + b_{nn-1} * y_{n-1} + a_{n1} * x_1 + a_{n2} * x_2 + \dots + a_{nm} * x_m + e_n \end{cases} \quad (2)$$

Для решения этой системы и нахождения ее параметров также используется 1МНК.

3. Наибольшее распространение в эконометрических исследованиях получила система взаимосвязанных (одновременных) уравнений. В ней одни и те же зависимые переменные (у) в одних уравнениях входят в левую часть системы, а в других – в правую часть, модель (3). Тем самым подчеркивается, что в системе одни и те же переменные (у) одновременно рассматриваются как зависимые в одних уравнениях и как независимые в других.

$$\begin{cases} y_1 = b_{12} \cdot y_2 + b_{13} \cdot y_3 + \dots + b_{1n} \cdot y_n + a_{11} \cdot x_1 + a_{12} \cdot x_2 + \dots + a_{1m} \cdot x_m + e_1 \\ y_2 = b_{21} \cdot y_1 + b_{23} \cdot y_3 + \dots + b_{2n} \cdot y_n + a_{21} \cdot x_1 + a_{22} \cdot x_2 + \dots + a_{2m} \cdot x_m + e_2 \\ \dots \\ y_n = b_{n1} \cdot y_1 + b_{n2} \cdot y_2 + \dots + b_{nn-1} \cdot y_{n-1} + a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \dots + a_{nm} \cdot x_m + e_n \end{cases} \quad (3)$$

Данная система уравнений называется структурной формой модели. Эндогенные переменные (у) – взаимосвязанные переменные, которые определяются внутри модели (системы). Экзогенные переменные – независимые переменные, которые определяются вне системы. Предопределенные переменные (х) – экзогенные и лаговые (за предыдущие моменты времени) эндогенные переменные системы. Коэффициенты а и b при переменных – структурные коэффициенты модели.

Запишем данную модель в матричном виде, формула (4).

$$BY + AX = E, \quad (4)$$

где В – матрица порядка n на n коэффициентов при n эндогенных переменных;
А- матрица порядка n на m коэффициентов при m предопределённых переменных (в состав этих коэффициентов включён и свободный член);
Y- вектор-столбец эндогенных переменных;
X- вектор-столбец предопределённых переменных;
Е - вектор-столбец случайных последовательно некоррелированных ошибок с нулевыми средними.

Система взаимосвязанных уравнений получила название системы одновременных (совместных) уравнений. Каждое уравнение такой системы не может рассматриваться самостоятельно, и для нахождения его параметров традиционный 1МНК неприменим. С этой целью используются его модификации: косвенный, двухшаговый и трехшаговый метод наименьших квадратов.

Если матрица В невырожденная (существует обратная матрица B^{-1}), тогда система (3,4) имеет приведённую форму (5). Система линейных функций эндогенных переменных от всех предопределенных переменных системы является приведенной формой модели.

$$\begin{cases} y_1 = p_{11} \cdot x_1 + p_{12} \cdot x_2 + \dots + p_{1m} \cdot x_m + v_1 \\ \dots \\ y_n = p_{n1} \cdot x_1 + p_{n2} \cdot x_2 + \dots + p_{nm} \cdot x_m + v_2, \end{cases} \quad (5)$$

где p_{ij} - коэффициенты приведенной формы модели.

Идентификация модели

Необходимое условие идентификации – выполнение счетного правила:

$D+1=N$ – уравнение точно идентифицируемо;

$D+1<N$ – уравнение неидентифицируемо;

$D+1>N$ – уравнение сверхидентифицируемо.

Где N – число эндогенных переменных в уравнении, D – число предопределенных переменных, отсутствующих в уравнении, но присутствующих в системе.

Уравнение структурной формы называется:

1. Точно идентифицируемым $D+1=N$ – если все участвующие в нем неизвестные коэффициенты однозначно восстанавливаются по коэффициентам приведенной формы. Для решения системы точно идентифицируемых уравнений применяется косвенный МНК (КМНК)
2. Сверхидентифицируемым – если все участвующие в нем неизвестные коэффициенты восстанавливаются по коэффициентам приведенной формы, причем некоторые из них могут принимать одновременно несколько значений. Для решения сверхидентифицируемых уравнений применяется двухшаговый МНК(2МНК)
3. Неидентифицируемым – если хотя бы один из участвующих в нем неизвестных коэффициентов не может быть восстановлен по коэффициентам приведенной формы.

Эконометрическая модель называется точно идентифицируемой, если все уравнения ее структурной формы являются точно идентифицируемыми. Эконометрическая модель называется неидентифицируемой, если хотя бы одно уравнение ее структурной формы является неидентифицируемым.

Запишем приведённую систему в матричной форме (6):

$$Y = P \cdot X + V, \quad (6)$$

где P - матрица коэффициентов порядка n на m приведённой формы;

X - вектор-столбец предопределённых переменных;

V - вектор случайных составляющих (ошибок) приведённой системы.

Имеет место соотношение коэффициентов приведённой и структурной формы, формула (7).

$$P = -B^{-1}A, \quad (7)$$

Все условия Гаусса-Маркова выполняются. Имеет место соотношение ошибок приведённой и структурной формы, формула (8).

$$V = B^{-1}E. \quad (8)$$

Методы решения систем одновременных уравнений: КМНК и 2МНК

Косвенный МНК (КМНК)

КМНК применяется в случае точно идентифицируемой модели для одновременного оценивания коэффициентов всех уравнений системы. Процедура применения КМНК предполагает выполнение следующих этапов:

1. Составляют приведенную форму модели (5,6) и определяют численные значения параметров для каждого ее i -го уравнения обычным 1МНК по формуле (9).

$$\hat{p}_i = (X^T X)^{-1} X^T y_i, \quad (9)$$

где $\hat{p}_i$ - вектор оценок параметров i -го уравнения системы (5), $i=1 \dots n$.

Пример 1. Пусть дана система одновременных уравнений:

$$\begin{cases} y_1 = \chi + \lambda y_2 + \gamma x_2 + \varepsilon_1; \\ y_2 = \beta + \omega y_1 + \nu x_3 + \varepsilon_2. \end{cases}$$

Приведём её к структурному виду (3), введя обозначения:

$$\chi = a_{11}; \lambda = b_{12}; \gamma = a_{12}; x_1 = 1; b_{11} = 1; a_{13} = 0$$

$$\beta = a_{21}; \omega = b_{21}; \nu = a_{23}; a_{22} = 0; b_{22} = 1$$

$$\begin{cases} 1 \cdot y_1 = b_{12} y_2 + a_{11} \cdot 1 + a_{12} x_2 + 0 \cdot x_3 + \varepsilon_1; \\ 1 \cdot y_2 = b_{21} y_1 + a_{21} \cdot 1 + 0 \cdot x_2 + a_{23} \cdot x_3 + \varepsilon_2. \end{cases} \Leftrightarrow \begin{cases} 1 \cdot y_1 - b_{12} y_2 - a_{11} \cdot 1 - a_{12} x_2 - 0 \cdot x_3 = \varepsilon_1; \\ -b_{21} y_1 + 1 \cdot y_2 - a_{21} \cdot 1 - 0 \cdot x_2 - a_{23} \cdot x_3 = \varepsilon_2. \end{cases}$$

тогда

$$B = \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix}, \quad A = \begin{pmatrix} -a_{11} & -a_{12} & 0 \\ -a_{21} & 0 & -a_{23} \end{pmatrix}.$$

$$B \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + A \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}.$$

Отметим, что оба уравнения точно идентифицируемы:

$$1: D+1=N=2,$$

$$2: D+1=N=2.$$

Следовательно, для решения системы применяем КМНК.

Составим приведённую форму данной системы:

$$\begin{cases} y_1 = p_{11} \cdot x_1 + p_{12} \cdot x_2 + p_{13} \cdot x_3 + \nu_1; \\ y_2 = p_{21} \cdot x_1 + p_{22} \cdot x_2 + p_{23} \cdot x_3 + \nu_2. \end{cases}$$

Пусть методом 1МНК, формула (9), были получены оценки параметров приведённой формы системы, тогда она примет вид:

$$\begin{cases} y_1 = 1 + 2x_2 + 3x_3 + \nu_1; \\ y_2 = -2 + 1x_2 + 4x_3 + \nu_2. \end{cases}$$

2. Путем алгебраических преобразований переходим от уравнений приведенной формы к уравнениям структурной формы модели, получая тем самым численные оценки структурных параметров.

Воспользовавшись формулой (7), найдём структурные параметры через приведённые, Р.

$$\begin{pmatrix} 1 & 2 & 3 \\ -2 & 1 & 4 \end{pmatrix} = - \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} -a_{11} & -a_{12} & 0 \\ -a_{21} & 0 & -a_{23} \end{pmatrix} = \begin{pmatrix} \frac{-1}{1-b_{21}b_{12}} & \frac{-b_{12}}{1-b_{21}b_{12}} \\ \frac{-b_{21}}{1-b_{21}b_{12}} & \frac{-1}{1-b_{21}b_{12}} \end{pmatrix} \cdot \begin{pmatrix} -a_{11} & -a_{12} & 0 \\ -a_{21} & 0 & -a_{23} \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{a_{11}+b_{12}a_{21}}{1-b_{21}b_{12}} & \frac{a_{12}}{1-b_{21}b_{12}} & \frac{b_{12}a_{23}}{1-b_{21}b_{12}} \\ \frac{b_{21}a_{11}+a_{21}}{1-b_{21}b_{12}} & \frac{b_{21}a_{12}}{1-b_{21}b_{12}} & \frac{a_{23}}{1-b_{21}b_{12}} \end{pmatrix}$$

Подставим в приведённую форму модели полученные выражения:

$$\begin{cases} y_1 = \frac{a_{11}+b_{12}a_{21}}{1-b_{21}b_{12}} + \frac{a_{12}}{1-b_{21}b_{12}} x_2 + \frac{b_{12}a_{23}}{1-b_{21}b_{12}} x_3 + \nu_1; \\ y_2 = \frac{b_{21}a_{11}+a_{21}}{1-b_{21}b_{12}} + \frac{b_{21}a_{12}}{1-b_{21}b_{12}} x_2 + \frac{a_{23}}{1-b_{21}b_{12}} x_3 + \nu_2. \end{cases}$$

$$\begin{cases} \frac{a_{11}+b_{12}a_{21}}{1-b_{21}b_{12}} = 1; \\ \frac{a_{12}}{1-b_{21}b_{12}} = 2; \\ \frac{b_{12}a_{23}}{1-b_{21}b_{12}} = 3; \\ \frac{b_{21}a_{11}+a_{21}}{1-b_{21}b_{12}} = -2; \\ \frac{b_{21}a_{12}}{1-b_{21}b_{12}} = 1; \\ \frac{a_{23}}{1-b_{21}b_{12}} = 4. \end{cases}$$

$$\begin{aligned} a_{11} &= 2,5; b_{21} = 0,5; a_{21} = -2,5; \\ b_{12} &= 0,75; a_{12} = 1,25; a_{23} = 2,5; \end{aligned}$$

Тогда структурная форма примет вид:

$$\begin{cases} y_1 = 0,75y_2 + 2,5 + 1,25x_2 + \varepsilon_1; \\ y_2 = 0,5y_1 - 2,5 + 2,5x_3 + \varepsilon_2. \end{cases}$$

Воспользовавшись формулой (8), выразим ошибки структурной формы системы через ошибки приведённой:

$$\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{1-b_{21}b_{12}} & \frac{b_{12}}{1-b_{21}b_{12}} \\ \frac{b_{21}}{1-b_{21}b_{12}} & \frac{1}{1-b_{21}b_{12}} \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} \frac{\varepsilon_1 + b_{12}\varepsilon_2}{1-b_{21}b_{12}} \\ \frac{b_{21}\varepsilon_1 + \varepsilon_2}{1-b_{21}b_{12}} \end{pmatrix}.$$

$$\begin{cases} v_1 = \frac{\varepsilon_1 + b_{12}\varepsilon_2}{1 - b_{21}b_{12}}; \\ v_2 = \frac{b_{21}\varepsilon_1 + \varepsilon_2}{1 - b_{21}b_{12}}. \end{cases} \Leftrightarrow \begin{cases} 0,625v_1 = \varepsilon_1 + 0,75\varepsilon_2; \\ 0,625v_2 = 0,5\varepsilon_1 + \varepsilon_2. \end{cases} \Leftrightarrow \begin{cases} \varepsilon_2 = v_2 - 0,5v_1; \\ \varepsilon_1 = v_1 - 0,75v_2. \end{cases}$$

Тогда получим следующие параметры и ошибки системы структурной формы:

$$\begin{cases} y_1 = 0,75y_2 + 2,5 + 1,25x_2 + v_1 - 0,75v_2; \\ y_2 = 0,5y_1 - 2,5 + 2,5x_3 + v_2 - 0,5v_1. \end{cases}$$

Двухшаговый МНК (2МНК)

Данный метод используется для оценивания коэффициентов отдельного уравнения системы. Основная идея 2МНК — на основе приведенной формы модели получить для отдельного уравнения сверхидентифицируемой системы теоретические значения (оценки) эндогенных переменных $\hat{Y}_i$, содержащихся в правой части уравнения.

Запишем каждое уравнение сверхидентифицируемой системы структурной формы (3) в виде, представленном формулой (10).

$$\bar{y}_i = Y_i \bar{\beta}_i + X_i \bar{\alpha}_i + \bar{\varepsilon}_i, \quad (10)$$

$\bar{y}_i$ — вектор значений одной эндогенной переменной в левой части уравнения;

Y_i — матрица значений оставшихся эндогенных переменных в правой части i -го уравнения;

X_i — матрица значений предопределённых переменных i -го уравнения.

Эндогенные регрессоры (Y_i в правой части уравнения) «очищаются» от влияния возмущений $\bar{\varepsilon}_i$, т.е. проводится регрессия каждого столбца матрицы Y_i на все предопределённые переменные — методом 1МНК оцениваются параметры модели (11).

$$Y_i = X \cdot P_i + V_i, \quad (11)$$

где P_i — подматрица матрицы P параметров приведенной формы, соответствующих эндогенным переменным, включённым в правую часть i -го структурного уравнения;

V_i — подматрица матрицы V , соответствующая тем же эндогенным переменным.

X — матрица значений всех предопределённых переменных системы.

Затем рассчитываются значения $\hat{Y}_i$, формула (12).

$$\hat{Y}_i = X \cdot \hat{P}_i. \quad (12)$$

Далее, осуществляется регрессия (10) с заменой в правой части Y_i на $\hat{Y}_i$, т.е. строятся 1МНК оценки структурных параметров $\vec{\beta}_i, \vec{\alpha}_i$ в регрессии (13).

$$\bar{y}_i = \hat{Y}_i \vec{\beta}_i + X_i \vec{\alpha}_i + \bar{\varepsilon}_i \quad (13)$$

Метод получил название двухшагового МНК, поскольку дважды используется 1МНК: на первом шаге при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной $\hat{y}_i = p_{i1}x_1 + p_{i2}x_2 + \dots + p_{im}x_m$ и на втором шаге применительно к структурному сверхидентифицируемому уравнению при определении структурных параметров модели по данным теоретических (расчетных) значений эндогенных переменных.

Сверхидентифицируемая структурная модель может быть двух типов:

- все уравнения системы сверхидентифицируемы;
- система содержит наряду со сверхидентифицируемыми точно идентифицируемые уравнения.

В обоих случаях для оценки структурных параметров к каждому уравнению системы может применяться 2МНК метод.

Пример 2. Опустим свободный член в первом уравнении системы в примере 1, тогда первое уравнение станет сверхидентифицированным ($D=2+1=3 > (H=2)$), а второе останется точно идентифицированным ($D+1=H=2$). Оценим методом 2МНК параметры первого уравнения (параметры второго могут быть найдены аналогичным образом). Также используем другие исходные данные по x_1, x_2, x_3, y_1, y_2 , представленные в таблице 1.

Таблица 1 – Исходные данные для примера 2

x_1	x_2	x_3	y_1	y_2	$\hat{y}_2$
1	1	2	1	1	1,056574448
1	1	3	2	2	1,93472179
1	4	8	8	8	8,011967672
1	12	4	5	9	8,99673609

$$\begin{cases} y_1 = \lambda y_2 + \gamma x_2 + \varepsilon_1; \\ y_2 = \beta + \omega y_1 + \nu x_3 + \varepsilon_2. \end{cases}$$

Представим систему в структурном виде (3):

$$a_{11} = 0; \lambda = b_{12}; \gamma = a_{12}; x_1 = 1; b_{11} = 1; a_{13} = 0.$$

$$\beta = a_{21}; \omega = b_{21}; \nu = a_{23}; a_{22} = 0; b_{22} = 1.$$

$$\begin{cases} 1 \cdot y_1 = b_{12}y_2 + 0 \cdot x_1 + a_{12}x_2 + 0 \cdot x_3 + \varepsilon_1; \\ 1 \cdot y_2 = b_{21}y_1 + a_{21} \cdot 1 + 0 \cdot x_2 + a_{23} \cdot x_3 + \varepsilon_2. \end{cases} \Leftrightarrow \begin{cases} 1 \cdot y_1 - b_{12}y_2 - 0 \cdot x_1 - a_{12}x_2 - 0 \cdot x_3 = \varepsilon_1; \\ -b_{21}y_1 + 1 \cdot y_2 - a_{21} \cdot 1 - 0 \cdot x_2 - a_{23} \cdot x_3 = \varepsilon_2. \end{cases}$$

Также представим систему в приведённой форме (5):

$$\begin{cases} y_1 = p_{11} \cdot x_1 + p_{12} \cdot x_2 + p_{13} \cdot x_3 + v_1; \\ y_2 = p_{21} \cdot x_1 + p_{22} \cdot x_2 + p_{23} \cdot x_3 + v_2. \end{cases}$$

Используя данные таблицы 1 и метод 1МНК для расчёта параметров, формула (9), получим:

$$(X^T X)^{-1} = \begin{pmatrix} 1,218682 & -0,03559 & -0,19024 \\ -0,03559 & 0,0129 & -0,00528 \\ -0,19024 & -0,00528 & 0,050357 \end{pmatrix},$$

для первого уравнения:

$$X^T y_1 = \begin{pmatrix} 16 \\ 95 \\ 92 \end{pmatrix},$$

для второго уравнения:

$$X^T y_2 = \begin{pmatrix} 20 \\ 143 \\ 108 \end{pmatrix},$$

и следующую приведённую форму системы:

$$\begin{cases} y_1 = -1,38436 + 0,169879x_2 + 1,087038x_3 + v_1; \\ y_2 = -1,26189 + 0,56217x_2 + 0,878147x_3 + v_2. \end{cases}$$

Найдём оценки (модельные значения) $\hat{y}_2$ по второму уравнению приведённой формы и запишем их в таблицу 1.

Представим первое уравнение структурной формы $y_1 = \lambda y_2 + \mu x_2 + \varepsilon_1$ в виде (10), заменим y_2 его модельными значениями $\hat{y}_2$, формула (13), и применим 1МНК.

$$(X^{*T} X^*)^{-1} = \begin{pmatrix} 0,04208 & -0,03715 \\ -0,03715 & 0,038961 \end{pmatrix},$$

$$X^{*T} y_1 = \begin{pmatrix} 114,0054 \\ 95 \end{pmatrix}.$$

Получим первое структурное уравнение:

$$y_1 = 1,268618y_2 - 0,53341x_2 + \varepsilon_1,$$

где $b_{12} = 1,268618; a_{11} = 0; a_{12} = -0,53341$.

Параметры $b_{21}; a_{21}; a_{23}$ второго точно идентифицируемого уравнения $y_2 = \beta + \omega y_1 + \nu x_3 + \varepsilon_2$ структурной системы могут быть найдены аналогичным образом методом 2МНК. Получим $y_2 = 3,31y_1 + 3,32 - 2,72x_3 + \varepsilon_2$.

Тогда структурная форма системы примет вид:

$$\begin{cases} y_1 = 1,268618y_2 - 0,53341x_2 + \varepsilon_1; \end{cases}$$

$$y_2 = 3,31y_1 + 3,32 - 2,72x_3 + \varepsilon_2.$$

2МНК является наиболее общим и широко распространенным методом решения системы одновременных уравнений.

ИНДИВИДУАЛЬНЫЕ РАСЧЁТНО-ПРАКТИЧЕСКИЕ ЗАДАНИЯ

Задание 1. Ответьте на вопросы:

А) Какие Вы знаете системы эконометрических уравнений?

Б) Когда используется 2МНК и КМНК для оценки параметров одновременных эконометрических уравнений?

В) Рассматривается следующая модель краткосрочного равновесия типа IS-LM

$$\begin{cases} y_t = c_t + i_t + g_t + nx_t \\ c_t = a_{11} + \beta_{11}y_t + \varepsilon_{1t} \\ i_t = a_{21} + a_{21}r_t + \varepsilon_{2t} \\ nx_t = a_{31} + \beta_{31}y_t + \beta_{32}r_t + \varepsilon_{3t} \\ m_t = a_{40} + \beta_{41}y_t + a_{41}r_t + \varepsilon_{4t} \end{cases},$$

где эндогенными переменными являются валовый доход y , объём личных потребительских расходов c , объём инвестиций I , чистый экспорт nx и ставка процента r . Экзогенные переменные: g - совокупные государственные расходы и m – предложение денег. Опишите процедуру оценивания модели с помощью двухшагового метода наименьших квадратов.

Г) Приведите пример системы одновременных уравнений, к которой можно применить косвенный МНК (с объяснением обозначений).

Д) Приведите пример сверхидентифицируемой системы одновременных уравнений(с объяснением обозначений).

Е) Рассмотрите модель

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + \varepsilon_1 \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + \varepsilon_2 \end{cases},$$

где X – экзогенные переменные, а вектор E – случайные последовательно некоррелированные ошибки с нулевыми средними. Обращая в нуль некоторые коэффициенты, определите три альтернативные структуры, для которых простейшими состоятельными процедурами оценивания являются соответственно обычный 1МНК, косвенный МНК и двухшаговый МНК.

Ж) Имеется следующая макроэкономическая модель:

$$\begin{cases} c = a_{10} + \beta_{11}y + \varepsilon_1 \\ i = a_{20} + \beta_{21}y + \beta_{22}y_{t-1} + \varepsilon_2 \\ y = c + i + g \end{cases},$$

где c , i , y - объём потребления, инвестиции и доход, соответственно, а y_{t-1} - доход предыдущего периода, g - государственные расходы.

1. Классифицируйте типы переменных;
2. Определите типы структурных уравнений;
3. Запишите модель в приведённой форме;
4. Определите метод оценки параметров структурной формы модели.

Задание 2. По исходным данным таблиц 2 и 3 оценить параметры структурной формы следующей системы одновременных уравнений:

$$\begin{cases} y_1 = \chi + \lambda y_2 + \mu x_2 + \varepsilon_1; \\ y_2 = \omega y_1 + \nu x_3 + \varepsilon_2. \end{cases}$$

Таблица 2 – Исходные данные

x1	x2	x3	y1
1	2	2	1
1	1	3	2
1	4	8	8
1	12	4	5

Таблица 3 – Значения y_2 по вариантам

Вариант	y2			
1	1	2	8	9
2	1	1	8	9
3	1	3	8	9
4	1	2	9	9
5	1	2	18	9
6	1	4	8	9
7	2	2	8	9
8	3	2	8	9
9	1	6	8	9
10	1	1	8	19
11	1	2	18	9
12	2	3	8	9
13	1	3	9	9
14	3	2	9	9
15	1	3	8	10
16	2	2	9	9
17	10	20	80	90
18	11	21	80	90
19	2	3	9	10
20	1	7	8	9
21	5	8	8	9
22	5	9	8	9
23	4	3	8	10
24	11	21	30	40
25	12	12	18	19
26	1	2	9	16