

**Московский государственный университет
имени М. В. Ломоносова
МОСКОВСКАЯ ШКОЛА ЭКОНОМИКИ**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Эконометрика - II»

**Направление 080100 Экономика
для подготовки студентов — магистров очного отделения**

**Авторы — составители программы:
профессор, д.ф.-м.н. С. А. Айвазян**

Рабочая программа утверждена
решением Ученого совета МШЭ МГУ
Протокол № от « ___ » _____ 2011 г.

**Москва
2011**

ВВЕДЕНИЕ

Учебная программа по курсу «Эконометрика - II» разработана в соответствии с требованиями Образовательного стандарта Московского государственного университета им. М.В. Ломоносова.

Рабочая программа соответствует учебному плану подготовки магистров по направлению 080100 «Экономика».

Курс посвящен дополнительным главам математико-статистического инструментария эконометрики и предназначен для студентов первого курса магистратуры.

Для усвоения курса необходимы знания (в рамках стандартных учебных программ) по линейной алгебре, математическому анализу, методам оптимизации, теории вероятностей, математической статистике и эконометрике.

Излагаемые в курсе методы и модели используются, в частности, при эконометрическом моделировании социально-экономических процессов (в частности, в ситуациях, когда исследуемые статистические зависимости строятся в условиях мультиколлинеарности или *по регрессионно-неоднородным исходным данным, наконец при реализации двухшагового метода наименьших квадратов в системах регрессионных уравнений большой размерности*), при решении различных задач типологизации объектов, а также при построении интегральных показателей, отборе наиболее информативных переменных и снижении размерностей анализируемых моделей. Они могут оказаться полезными также при организации и анализе результатов работы экспертов.

Изучение курса «Эконометрика - II» предназначено для формирования и усвоения знаний, умений, навыков в области экономической теории и практики, которые необходимы для работы в государственных и частных структурах, а также развития профессиональных качеств, компетенций, необходимых для выполнения функциональных обязанностей в сфере экономики.

Основные задачи преподавания дисциплины:

- ознакомление студентов с основами теории и методологии использования методов многомерного статистического анализа (МСА) в эконометрическом моделировании;
- раскрытие роли методов МСА в реалистическом моделировании механизмов социально-экономических процессов;
- привитие практических навыков эконометрического моделирования с использованием методов МСА;
- формирование общего представления о роли и месте методов МСА в эконометрическом моделировании.

В результате изучения курса студенты должны:

Знать:

- основные определения и понятия теории, методологии и практики применения МСА;
- методику применения МСА для изучения функционирования и прогноза социально-экономических систем;
- теорию МСА в той мере, в какой она нужна для успешного применения его методов в эконометрическом моделировании.

Уметь:

- анализировать различные социально-экономические процессы с применением МСА;
- интерпретировать результаты применения методов МСА к исследованию социально-экономических процессов;
- рассчитывать основные параметры функционирования реальных социально-экономических процессов и/или систем при их исследовании с помощью методов МСА.

Владеть:

- приемами МСА в процессе эконометрического моделирования реальных систем;
- методиками эконометрического моделирования, использующими методами МСА.

Программа предусматривает решение типовых задач в ходе лекций, упражнений и в форме домашних (зачетных) и бонусных заданий. В самостоятельную работу студента входит запись и штудирование лекций, изучение рекомендованной лектором литературы и выполнение домашних заданий.

Контроль и оценка полученных студентами знаний осуществляется в форме проверки домашних заданий и финального письменного экзамена.

Итоговая оценка студента складывается из оценки их активности на упражнениях (включая выполнение домашних заданий) с весом 20% и оценки за финальный письменный экзамен (с весом 80%). При этом, учитываются бонусные очки.

Дисциплина изучается в течение одного семестра при общем объеме учебной нагрузки 78 часов, в том числе, **44** аудиторных часов (**30** часа лекций и **14** часов семинаров) и **32** часа самостоятельных занятий.

**Учебно-методический план
(тематическое распределение часов и литературы)**

№№ п.п.	Тема занятий	Всего часов	в том числе		Литература (п.п. [1])
			лекции	упражнения	
1	Содержание и назначение прикладного МСА	1	1	—	гл. 9
2	Способы описания поведения многомерного признака	13	9	4	2.4, 2.5.3, 2.6.1, 2.6.6, 3.1.4, 3.1.5, гл. 11
3	Классификация многомерных наблюдений и статистические методы распознавания образов	12	8	4	гл. 12
4	Снижение размерности исследуемого признакового пространства и отбор наиболее информативных показателей	14	10	4	гл.13
5	Статистический анализ экспертных оценок	4	2	2	11.3, 11.4
ИТОГО:		44	30	14	

Краткое содержание программы

Т е м а 1 . **Содержание и назначение прикладного многомерного статистического анализа** (лекция 1)

Понятия многомерного признака и многомерного наблюдения. Смешанная природа компонент многомерного признака. Определение количественных и категоризованных (ординальных и номинальных) переменных. Основные формы записи массива исходных данных в многомерном статистическом анализе (МСА). Сущность и назначение МСА. Три базовые проблемы прикладного МСА: статистическое исследование зависимостей, классификация объектов (признаков) и снижение размерности исследуемого признакового пространства.

Основные типы социально-экономических задач, решаемых с привлечением аппарата МСА: разведочный анализ и визуализация (формирование рабочих гипотез о природе анализируемых данных); сжатие больших массивов информации; диагностика финансово-экономического состояния фирмы (банка, предприятия); замена многокритериальной оптимизационной схемы ее однокритериальной аппроксимацией; прогнозирование, нормирование, оптимальное управление; анализ экспертных оценок; отбор наиболее информативных переменных. Примеры.

Т е м а 2 . **Способы описания поведения многомерного признака** (лекции 1 ~ 5)

Многомерные (совместные), частные (маргинальные) и условные законы распределения вероятностей многомерного признака. Понятие статистической независимости случайных величин. Пояснение этих понятий на примерах.

Основные числовые характеристики многомерного закона распределения вероятностей (з.р.в.) и их статистические оценки. Многомерный нормальный з.р.в. и его свойства. Оценки параметров многомерного нормального з.р.в. Измерители степени тесноты статистической связи между компонентами многомерного признака (корреляционный анализ многомерной выборки): парные, частные и множественные коэффициенты корреляции, корреляционное отношение; ранговые корреляции; анализ таблиц сопряженности и информационная мера связи.

Т е м а 3 . Классификация многомерных наблюдений и статистические методы распознавания образов (лекции 6 ~ 9)

Общая экстремальная постановка задачи классификации. Две основные формы задания исходной информации в задачах классификации. Основные типы задач классификации.

Основные типы расстояний между объектами и между классами объектов, используемые в процедурах классификации. Обобщенные расстояния Колмогорова.

Параметрические и непараметрические методы классификации **при наличии обучающих выборок** (методы дискриминантного анализа). Задачи социально-экономической диагностики. Связь дискриминантного анализа с логит- и пробит-моделями.

Параметрические и непараметрические методы классификации **без обучающих выборок**: оценка параметров смеси распределений, методы кластер-анализа (в том числе — метод k -средних и иерархические кластер-процедуры). Задачи типологизации объектов.

Т е м а 4 . Снижение размерности исследуемого признакового пространства и отбор наиболее информативных показателей (лекции 10 ~ 14)

Общая экстремальная постановка задачи снижения размерности и два подхода к определению критерия информативности показателей.

Основные методы снижения размерности, нацеленные на оптимизацию критериев автоинформативности: метод главных компонент, факторный анализ, метод экстремальной группировки признаков, методы целенаправленного проецирования исходных многомерных данных. Использование метода главных компонент при построении моделей регрессии, при реализации двухшагового метода наименьших квадратов, в задачах построения интегральных показателей.

Методы снижения размерности, ориентированные на «внешние» критерии информативности; экспертно-статистический метод построения сводного (интегрального) латентного показателя эффективности функционирования (или каче-

ства) системы по набору частных показателей; отбор наиболее информативных показателей в моделях регрессии и дискриминантного анализа.

Построение вспомогательного координатного пространства заданной размерности на базе парных сравнений объектов (*многомерное шкалирование*).

Т е м а 5 . Статистический анализ экспертных оценок (лекция 15)

Основные типы организации работы экспертов (коллегиальный, частично-коллегиальный, индивидуально-автономный). Основные виды экспертных оценок (баллы, ранжировки, парные сравнения). Дуализм в интерпретации эксперта и оцениваемого объекта как многомерного наблюдения. Задачи статистического анализа экспертных мнений: исследования структуры совокупности мнений; анализ взаимной согласованности экспертных мнений и оценка компетентности экспертов; построение единого (группового) мнения.

Семинары №№ 1 и 2 по теме 2 (4 часа после лекции 5)

Свойства многомерного нормального закона распределения вероятностей и корреляционный анализ многомерной выборки.

Семинары №№ 3 и 4 по теме 3 (4 часа после лекции 9)

Методы классификации многомерных наблюдений в условиях наличия и отсутствия обучающих выборок.

Семинары №№ 5 и 6 по теме 4 (по 4 часа после лекции 14)

Методы снижения размерности. Пояснения к домашнему заданию.

Семинар № 7 по теме 5 (2 часа после лекции 15)

Статистический анализ экспертных оценок.

Литература

А. Основная:

- [1] Айвазян С.А., Мхитарян В.С. (2001). Прикладная статистика и основы эконометрики. Издание 2-е. Том 1: Теория вероятностей и прикладная статистика. М.: Юнити. — 656 с.

- [2] Айвазян С.А., Мхитарян В.С. (2001). Прикладная статистика в задачах и упражнениях. — М.: Юнити. — 270 с.

Б. Дополнительная:

- [1*] Андерсон Т. (1963). Введение в многомерный статистический анализ. М.: Физматгиз.

Контрольные вопросы для оценки качества освоения дисциплины

1. Сформулировать три основные проблемы многомерного статистического анализа. Краткая характеристика информационной базы и математического аппарата, привлекаемых для их решения. Пояснить на примерах решения социально-экономических задач.

2. Что понимается под многомерным признаком? Как проявляется смешанная природа его компонент? Дать определение и примеры количественных, порядковых (ординальных) и классификационных (номинальных) компонент многомерного признака.

3. Многомерный признак как многомерная случайная величина. Когда многомерная случайная величина считается заданной? Пояснить понятия многомерного, частного (маргинального) и условного распределений на примере двумерной нормальной плотности. Ковариационная матрица и ее свойства. Записать функцию многомерной плотности и дать вероятностную интерпретацию участвующих в этой записи параметров. Как оценить эти параметры по исходным статистическим данным?

4. Проанализировать понятия совместного, частного и условного распределений на примере распределения двумерной дискретной случайной величины $X = (x^{(1)}, x^{(2)})^T$, где $x^{(1)}$ — среднедушевой доход наугад выбранной семьи (признак имеет градации: «низкий доход», «средний доход» и «высокий доход»), а $x^{(2)}$ — качество жилищных условий у той же семьи (признак имеет четыре градации: «низкое», «удовлетворительное», «хорошее» и «очень хорошее»), см. таблицу:

Градации признака $x^{(1)}$	Градации признака $x^{(2)}$			
	низкое	удовлетворит.	хорошее	очень хорошее

низкий	0,06	0,03	0,01	0,00
средний	0,05	0,25	0,35	0,05
высокий	0,01	0,02	0,07	0,10

5. Основные характеристики взаимосвязи *количественных* компонент многомерного признака: ковариации, парные и частные коэффициенты корреляции, множественный коэффициент корреляции в терминах парных и частных коэффициентов корреляции; их вероятностная интерпретация. Правила проверки вышеперечисленных характеристик на их статистически значимое отличие от нуля.

6. Понятие корреляционного отношения, его вычисление, основные свойства и проверка на статистически значимое отличие от нуля.

7. Основные характеристики взаимосвязи *порядковых (ординальных)* компонент многомерного признака — ранговый коэффициент корреляции Спирмэна, коэффициент конкордации (согласованности) Кендалла. Коэффициент согласованности (конкордации) *нескольких* (не менее двух) упорядочений. Правила проверки вышеперечисленных характеристик на их статистически значимое отличие от нуля.

Решить следующее упражнение. По результатам ранжирования десяти учащихся в соответствии с их математическими ($x^{(1)}$) и музыкальными ($x^{(2)}$) способностями вычислить ранговый коэффициент корреляции Спирмэна и определить, существует ли между этими ранжировками статистически значимая связь (при уровне значимости критерия $\alpha = 0,05$)

№ учащегося	1	2	3	4	5	6	7	8	9	10
Ранг (место) по математике	2	1	4	3	7	8	5	6	10	9
Ранг (место) по музыке	5	6	4	1	8	10	2	7	9	3

8. Основные характеристики взаимосвязи дискретных (*категоризованных*) компонент многомерного признака — коэффициенты квадратической сопряженности, информационная характеристика связи. Проверка статистических гипотез об отсутствии какой бы то ни было связи между анализируемыми признаками.

Упражнение. В таблице приведены результаты классификации 1725 английских школьников в соответствии с качеством и опрятностью в их одежде ($x^{(1)}$) и в соответствии с их умственными способностями ($x^{(2)}$).

Категории качества одежды (уровни $x^{(1)}$)	Категории способностей (уровни $x^{(2)}$)						Сумма
	Умственно отсталые, тупые	Туповатые	Медл., но не тупые	Дост. умн.	Способные	Очень способные	
Очень Хорошее	33	48	113	209	194	39	636
Хорошее	41	100	202	255	138	15	751
Удовл.	39	58	70	61	33	4	265
Плохое	17	13	22	10	10	1	73
Сумма	130	219	407	535	375	59	1725

Вычислить значение коэффициента квадратической сопряженности X^2 и проверить статистическую значимость исследуемой связи (с уровнем значимости $\alpha = 0.05$).

9. По наблюдениям в течение 20-ти лет в восточных районах Англии за урожайностью кормов трав (y), количеством весенних осадков ($x^{(1)}$) и суммой активных температур весеннего периода ($x^{(2)}$), т.е. по совокупности исходных данных вида $(y_i, x_i^{(1)}; x_i^{(2)})$, $i = 1, 2, \dots, 20$, — были вычислены парные коэффициенты корреляции $r_{yx^{(1)}} = 0,80$; $r_{yx^{(2)}} = -0,40$; $r_{x^{(1)}x^{(2)}} = -0,56$. Вычислить частные коэффициенты корреляции $r_{yx^{(1)}/x^{(2)}}$, $r_{yx^{(2)}/x^{(1)}}$ множественный коэффициент корреляции $R_y(x^{(1)}, x^{(2)})$, дать их содержательный анализ и проверить все эти характеристики на их статистически значимое отличие от нуля.

10. В таблице приведены значения парных коэффициентов корреляции для среднегодовой характеристики качества ткани y (результатирующий показатель), среднегодового числа $x^{(1)}$ профилактических наладок станков автоматической линии, на которых эта ткань производилась (первая объясняющая переменная), и

среднегодового числа $x^{(2)}$ обрывов нити на этих станках (вторая объясняющая переменная), подсчитанные по исходным данным за 18 лет ($n = 18$).

	y	$x^{(1)}$	$x^{(2)}$
y	1	0,105	0,024
$x^{(1)}$	0,105	1	0,996
$x^{(2)}$	0,024	0,996	1

Вычислить частные коэффициенты корреляции $r_{yx^{(1)}/x^{(2)}}$, $r_{yx^{(2)}/x^{(1)}}$ а затем — множественный коэффициент корреляции $R_{y(x^{(1)}, x^{(2)})}$. Проверить их на статистически значимое отличие от нуля и дать содержательную интерпретацию полученных результатов.

11. Параметрические методы классификации при наличии обучающих выборок (методы дискриминантного анализа). Что такое обучающие выборки? Выписать и объяснить дискриминантное правило в случае разбиения **нормальных** наблюдений на два класса, отличающихся только векторами средних значений.

12. Параметрические методы классификации без обучающих выборок. Понятие смеси вероятностных распределений, привести примеры. Как данная задача классификации сводится к задаче дискриминантного анализа? Описать подход (метод моментов или метод максимального правдоподобия), с помощью которого можно было бы решить задачу оценивания неизвестных параметров в смеси распределений. Пояснить на примере смеси двух одномерных норм. законов.

13. Две основные формы задания исходной информации в задачах классификации, их преимущества и недостатки. Выписать пример функционала качества для случая заранее заданного числа классов. Объяснить алгоритм классификации объектов по методу « k -средних» (Мак-Куина) в случае заранее заданного числа классов.

14. Методы кластер анализа и иерархической классификации как методы решения общей задачи классификации без обучающих выборок. Привести пример конкретного алгоритма кластер-анализа (с объяснением его действия), описать об-

щую логическую схему действия агломеративных (последовательно соединяющих) иерархических алгоритмов классификации.

15. Основные типы расстояний между объектами и между классами объектов, используемых в процедурах классификации. Доказать, что расстояния «ближайшего» и «дальнего соседа», а также — расстояние «средней связи» между классами являются частными случаями обобщенного расстояния Колмогорова.

16. Три типа основных предпосылок, обуславливающих принципиальную возможность снижения размерности исследуемого факторного пространства. Общая экстремальная постановка задачи снижения размерности. Два подхода к определению критерия информативности показателей: автоинформативность и информативность относительно «внешних» показателей. Привести примеры конкретных методов, реализующих тот и другой подходы (явный вид соответствующих критериев информативности, их интерпретация, общая схема соответствующих вычислительных алгоритмов).

17. Метод главных компонент МГК: определение ГК; описание вычислительной процедуры по их построению. Доказать, что построение первой главной компоненты сводится к вычислению максимального собственного значения и соответствующего ему собственного вектора ковариационной матрицы исходного многомерного признака.

18. Некоторые вычислительные свойства главных компонент: их ковариационная матрица (с выводом), соотношения между простыми и обобщенными дисперсиями главных компонент и соответствующими характеристиками рассеяния исходных показателей.

19. Критерий информативности в общей постановке задачи снижения размерности, приводящий к методу главных компонент; его выражение в терминах собственных чисел ковариационной матрицы исходных показателей (с выводом).

20. Оптимальные свойства главных компонент — наименьшая ошибка автопрогноза и наименьшее искажение геометрической структуры исходных точек наблюдений при их проектировании в пространство меньшей размерности (формулировка с пояснениями, без доказательства).

21. Описать экспертно-статистический метод построения сводного (агрегатного) показателя качества функционирования системы по набору частных критерильных переменных и при наличии дополнительной экспертной информации. Привести примеры его возможного применения.

22. Пусть $(x^{(1)}, x^{(2)})$ — двумерная нормальная случайная величина со средними значениями $(\mu^{(1)}, \mu^{(2)})$ и матрицей ковариаций

$$\Sigma = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

(здесь r — коэффициент корреляции между $x^{(1)}$ и $x^{(2)}$). Можно показать, что главные компоненты $Z^{(1)}$ и $Z^{(2)}$ исходного двумерного признака $\begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix}$ определяются

соотношениями:

$$Z^{(1)} = (x^{(1)} - \mu^{(1)}) \cos \alpha + (x^{(2)} - \mu^{(2)}) \sin \alpha$$

$$Z^{(2)} = -(x^{(1)} - \mu^{(1)}) \sin \alpha + (x^{(2)} - \mu^{(2)}) \cos \alpha$$

где $\operatorname{tg} 2\alpha = \frac{2r\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}$.

Требуется:

1) Доказать (с помощью непосредственного подсчета), что главные компоненты $Z^{(1)}$ и $Z^{(2)}$ взаимно-некоррелированы, т.е. что $\operatorname{cov}(Z^{(1)}, Z^{(2)}) = 0$.

2) Вычислить собственные числа λ_1 и λ_2 ковариационной матрицы Σ и дать им вероятностную интерпретацию в терминах случайных величин $Z^{(1)}$ и $Z^{(2)}$.

3) Определить дисперсии $DZ^{(1)}$ и $DZ^{(2)}$ в частных случаях $|r|=1$ и $r=0$ и дать геометрическую интерпретацию главных компонент в этих случаях.

23. По данным измерений (в млн. руб.) объема произведенной за год продукции $(x^{(1)})$, основных фондов $(x^{(2)})$ и фонда оплаты труда $(x^{(3)})$ 24-х однотипных предприятий ($n = 24$) была оценена ковариационная матрица

$$\hat{\Sigma} = \begin{pmatrix} 451,4; & 271,2; & 168,7 \\ 271,2; & 171,7; & 103,3 \\ 168,7; & 103,3; & 66,7 \end{pmatrix}$$

Необходимые вычисления дали:

- характеристические (собственные) числа $\lambda_1 = 680,4$; $\lambda_2 = 6,5$; $\lambda_3 = 2,9$;
- собственные векторы

$$l_1 = \begin{pmatrix} 0,81 \\ 0,51 \\ 0,31 \end{pmatrix}, \quad l_2 = \begin{pmatrix} -0,55 \\ 0,83 \\ 0,10 \end{pmatrix}, \quad l_3 = \begin{pmatrix} -0,21 \\ -0,25 \\ 0,95 \end{pmatrix}$$

1) В результате каких вычислительных процедур получены числа λ_i и векторы $l_i (i = 1, 2, 3)$?

2) Выписать и проанализировать уравнение трех главных компонент исследуемого признака X .

3) Вычислить последовательные меры информативности главных компонент $I_k(Z) = \frac{\lambda_1 + \dots + \lambda_k}{Dx^{(1)} + \dots + Dx^{(3)}} (k = 1, 2, 3)$, на основании их сделать вывод о целесообразном уровне снижения размерности.

24. Дать математическую формулировку следующих задач статистического анализа экспертных мнений, представленных в виде балльных оценок, упорядочений (рангов оцениваемых объектов) или парных сравнений:

- 1) анализ структуры имеющихся мнений;
- 2) анализ попарной и групповой согласованности экспертных мнений;
- 3) построение единого (группового) варианта экспертного мнения;
- 4) анализ уровня компетентности мнений экспертов.

25. Привести и прокомментировать общую постановку задачи метрического многомерного шкалирования (ММШ). Описать решение задачи ММШ, основанное на анализе матрицы $\mathbf{B} = (b_{ij})$, где элементы b_{ij} определяются как скалярные произведения

не известных нам векторов $X_i - \bar{X}$ и $X_j - \bar{X} (i, j = 1, 2, \dots, n; \bar{X} = \sum_{i=1}^n x_i / n)$,

но могут быть вычислены по совокупности заданных попарных евклидовых расстояний γ_{kl} между искомыми точками X_1, X_2, \dots, X_n ($k, l = 1, 2, \dots, n$).

26. Описать процедуры отбора наиболее информативных переменных в задаче анализа *классической линейной модели множественной регрессии* (КЛММР) и в задаче дискриминантного анализа в рамках общей оптимизационной постановки проблемы снижения размерности.

27. Описать процедуру использования метода главных компонент в анализе КЛММР в условиях мультиколлинеарности.

28. Описать процедуру использования метода главных компонент в двухшаговом методе наименьших квадратов (2МНК) при анализе «большеразмерных» систем одновременных эконометрических уравнений.

29. Описать процедуру построения сводного (интегрального) индикатора качества функционирования системы по набору значений характеризующих это функционирование статистических показателей (частных критериев) *в условиях отсутствия «обучения»*. Объясните, на каком свойстве 1-й главной компоненты основано предложение использовать ее в качестве интегрального индикатора.

30. Опишите связь, существующую между постановками задач дискриминантного анализа и эконометрическими моделями множественного выбора. Выведите соотношение, связывающее взвешенное *отношение правдоподобия* в байесовском оптимальном правиле классификации дискриминантного анализа двух выборок с *апостериорной вероятностью* принадлежности наблюдения к определенному классу в модели бинарного выбора. Приведите обоснование логистической формы параметризации этой апостериорной вероятности в случае различения двух нормальных классов с одинаковыми ковариационными матрицами.

Автор программы: _____ /Айвазян С.А./