МИНИСТЕРСТВО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
КРАСНОЯРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ
КАФЕДРА СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО ПЛАНИРОВАНИЯ

ЭКОНОМЕТРИКА

Учебно-методический комплекс

Для экономических специальностей
дневной и заочной формы обучения

Красноярск 2003

Одобрено решением методического Совета экономического факультета Декан экономического факультета Е.Б.Бухарова_________________ «_____»________________2003 г.	Программа составлена в соответствии с государственными образовательными стандартами высшего профессионального образования

ББК 65в6

Автор-составитель: Е.В.Зандер

Эконометрика: Учебно-методический комплекс. Красноярск: РИО КрасГУ, 2003. 36 с.

Предназначена для специальностей: «Государственное муниципальное управление», «Финансы и кредит», «Налоги и налогообложение», «Антикризисное управление», «Экономика труда», «Финансы и кредит», «Бухгалтерский учет и аудит», «Мировая экономика», «Менеджмент» дневной и заочной форм обучения, бакалавриата по направлениям «Экономика» и «Менеджмент» дневной формы обучения.

1. ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКИЙ РАЗДЕЛ

Цель курса

Конечной целью изучения дисциплины является формирование у будущих специалистов теоретических знаний и практических навыков по применению статистических методов для исследования и обобщения эмпирических зависимостей экономических переменных, а также построения надежных прогнозов в банковском деле, финансах, различных сферах предпринимательской деятельности с целью обоснования принимаемых решений.

Задачи курса

Основной задачей изучения дисциплины "Эконометрика" является реализация требований, установленных Государственным стандартом высшего профессионального образования к подготовке специалистов в области экономических и бизнес-дисциплин. Данный курс рассчитан на студентов дневного отделения, обучающихся по специальностям «Финансы и кредит», «Антикризисное управление», «Государственное муниципальное управление», «Налоги и налогообложение», «Бухгалтерский учет и статистика», «Мировая экономика», «Государственное и муниципальное управление», «Менеджмент», «Маркетинг», а также для студентов, обучающихся по направлениям «Экономика» и «Менеджмент» (бакалавриат).

Место курса в системе дисциплин экономического образования

Эконометрика объединяет совокупность методов и моделей, позволяющих на базе **экономической теории, экономической статистики и математико-статистического инструментария** исследовать количественные выражения качественных зависимостей. При изучении дисциплины «Эконометрика» предполагается, что студент владеет основами теории вероятностей, математической статистики и матричной алгебры в объеме курса высшей математики для экономических специальностей.

Требования к уровню освоения содержания курса

В результате изучения курса «Эконометрика» дипломированный специалист в области экономики должен знать и уметь:

- формировать концепцию эконометрической модели на основе качественного анализа объекта исследования;
- собирать и проводить статистическую обработку экономической информации с целью выявления основных характеристик числовой совокупности;
- проводить оценку взаимосвязей экономических показателей с помощью статистических методов, интерпретировать полученные результаты по оценке взаимосвязей с точки зрения экономической сущности явлений;

- строить эконометрические модели с использованием процедур регрессионного анализа и анализа временных рядов;
- оценивать качество построенных эконометрических моделей с точки зрения их адекватности фактическим данным;
- применять эконометрические модели в практике хозяйственного управления.

2. СОДЕРЖАНИЕ КУРСА

Курс «Эконометрика» на дневном отделении экономического факультета Красноярского государственного университета представлен двумя модулями: «Основы эконометрики» и «Эконометрические модели», на заочном отделении проводится обучение по единому модулю «Эконометрика». Изучение курса "Эконометрика" предусматривает проведение лекционных и семинарских занятий (большая часть которых проводится в компьютерном классе), а также выполнение двух контрольных и двух лабораторных работ, сопровождающееся последующей защитой полученных результатов, а также самостоятельную проработку специальной литературы.

Распределение часов по темам и разделам курса

Тема	Лекции	Семинары
Тема 1	4 часа	6 часов
Тема 2	4 часа	8 часов
Тема 3	6 часов	12 часов
Тема 4	4 часа	8 часов
Тема 5	8 часов	9 часов
Тема 6	4 часа	4 часа
Тема 7	4 часа	4 часа
Итого	34 часа	51 час

Тема 1. МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ КУРСА (4 часа)

Предмет эконометрики. Примеры применения методов анализа данных. Основные математические предпосылки эконометрического моделирования. Закон больших чисел, теоремы Чебышева, Бернулли, Ляпунова.

Этапы и проблемы эконометрического моделирования. Эконометрическая модель и экспериментальные данные. Пространственная выборка, временной (динамический) ряд, пространственно-временная выборка.

Основные этапы предварительной обработки данных. Основные описательные статистики и их анализ. Проверка выборочного распределения на стационарность и однородность. Выявление аномальных наблюдений. Отсев грубых погрешностей. Проверка распределения на нормальность. Преобразование распределения к нормальному.

Тема 2. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ (4 часа)

Понятия функциональной, статистической и корреляционной зависимости. Типы связи экономических переменных: линейные и нелинейные связи. Меры тесноты линейной связи переменных: парный, частный и множественный коэффициенты корреляции. Проверка статистических гипотез для оценки значимости корреляции. Свойства основных корреляционных коэффициентов. Корреляционное отношение как оценка нелинейной связи. Оценка тесноты связи между ординальными (порядковыми) переменными – коэффициент ранговой корреляции Спирмена.

Тема 3. МОДЕЛИ И МЕТОДЫ РЕГРЕССИОННОГО АНАЛИЗА (6 часов)

Задачи регрессионного анализа. Понятия регрессионного анализа: результирующая (зависимая, эндогенная) переменная **y** и объясняющие (предикторные, экзогенные) переменные **X**, функция регрессии **y** по **X**, возмущения.

Основные предпосылки регрессионного анализа (теорема Гаусса-Маркова). Уравнение регрессионной связи между **y** и **X**. Геометрическая интерпретация регрессии. Классическая линейная модель множественной регрессии.

Метод наименьших квадратов для оценки параметров регрессионной модели. Статистические свойства оценок параметров. Стандартизованные коэффициенты регрессии и коэффициенты эластичности.

Нелинейные модели регрессии и линеаризующие преобразования.

Анализ вариации результирующего показателя и выборочный коэффициент детерминации. Проверка значимости уравнения регрессии и коэффициентов уравнения регрессии. Оценка качества регрессионной зависимости.

Построение точечных и интервальных прогнозов, основанных на моделях линейной регрессии. Построение доверительного интервала для параметров регрессионной модели.

Тема 4. ПРОБЛЕМЫ ПРАКТИЧЕСКОГО ИСПОЛЬЗОВАНИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ (4 часа)

Понятие мультиколлинеарности. Признаки и причины мультиколлинеарности. Методы устранения мультиколлинеарности. Отбор наиболее существенных переменных в классической линейной модели множественной регрессии. Скорректированный коэффициент детерминации.

Линейные регрессионные модели с переменной структурой. Фиктивные переменные.

Тема 5. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ (8 часов)

Понятие временного ряда и его отличия от случайной выборки. Составляющие временного ряда. Проверка гипотезы о неизменности среднего значения временного ряда как процедура проверки наличия тренда. Процедуры аналитического выравнивания (сглаживания) временного ряда. Подбор порядка аппроксимирующего полинома с помощью метода последовательных разностей.

Стационарные временные ряды и их характеристики. Понятия автокорреляции, автокорреляционной функции, временного лага, коэффициента автокорреляции, коррелограммы. Интерпретация коррелограмм.

Гетероскедастичность пространственной выборки. Искажение характеристик точности МНК-оценок, обусловленное игнорированием автокоррелированности остатков. Проверка гипотезы о наличии/отсутствии автокоррелированности регрессионных остатков. Положительная и отрицательная автокорреляция.

Использование авторегрессионных моделей: модель авторегрессии порядка **p**, определение порядка авторегрессионной модели. Методы исключения из временных рядов основной тенденции с целью устранения автокорреляции: метод последовательных или конечных разностей и метод коррелирования отклонений уровней ряда от основной тенденции.

Способы построения множественной регрессионной модели по временным рядам. Модели рядов, содержащих сезонную компоненту. Определение максимального количества тригонометрических составляющих при представлении временного ряда в

виде ряда Фурье. Оценивание параметров периодической функции, проверка их значимости.

Тема 6. СИСТЕМЫ ЛИНЕЙНЫХ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ. (4 часа)

Общий вид системы одновременных уравнений. Модель спроса-предложения как пример системы одновременных уравнений. Условия идентифицируемости уравнений системы. Структурная и приведенная формы эконометрической модели, построенной на базе систем одновременных уравнений. Рекурсивная модель как частный случай модели в структурной форме. Идентификация систем одновременных уравнений (статистическое оценивание неизвестных значений параметров системы): идентификация рекурсивных систем, косвенный метод наименьших квадратов, двухшаговый МНК оценивания структурных параметров отдельного уравнения, трехшаговый МНК одновременного оценивания всех параметров системы. Оценивание параметров системы внешне не связанных уравнений.

Тема 7. ОСНОВНЫЕ ПОНЯТИЯ И МОДЕЛИ ДИСПЕРСИОННОГО АНАЛИЗА (4 часа)

Дисперсионный анализ как метод организации (планирования), статистического анализа и интерпретации результатов эксперимента, где изучается зависимость количественной переменной от сочетания градаций качественных переменных. Сущность и задачи дисперсионного анализа. Полные и неполные, случайные и рандомизированные планы эксперимента. Модели одно- и двухфакторного дисперсионного анализа. Анализ вариации зависимой переменной. Проверка статистических гипотез относительно наличия/отсутствия влияния неколичественного признака.

3. ФОРМЫ КОНТРОЛЯ

Курс «Эконометрика» на дневном отделении рассчитан на 2 семестра, в течение первого семестра студенты выполняют комплексную контрольную работы по проверке теоретических знаний в области корреляционно-регрессионного анализа, а также лабораторную работу «Множественный корреляционно-регрессионный анализ». Во втором семестре студенты выполняют комплексную контрольную работу по теоретическим аспектам моделирования временных рядов, а также лабораторную работу «Анализ временных рядов».

Итоговый контроль для студентов дневного отделения всех специальностей заключается в сдаче письменного зачета по результатам обучения в первом семестре, по результатам обучения за 2 семестра студенты сдают устно-письменный экзамен по программе всего курса. Для студентов заочного отделения всех экономических специальностей предполагается проведение письменной аудиторной работы с ее последующей защитой, по результатам которой выставляется зачет.

Примерные вопросы к экзамену

1. Предмет эконометрики.
2. Этапы эконометрического моделирования.
3. Этап предварительной обработки данных: простые статистики (показатели уровня и меры рассеяния числовой совокупности).
4. Способы отсева грубых погрешностей.

5. Способы проверки распределения на нормальность.
6. Формулы преобразования матрицы исходных данных в случае невыполнения гипотезы о нормальности распределения.
7. Выборочный парный коэффициент корреляции (формула для расчета, интерпретация).
8. Процедура проверки на значимость парных коэффициентов корреляции (t-статистика).
9. Доверительный интервал коэффициента корреляции (формула для расчета, интерпретация).
10. Выборочное корреляционное отношение (формула для расчета, интерпретация).
11. Проверка значимости корреляционного отношения (F-критерий).
12. Выборочный множественный коэффициент корреляции (формула для расчета, интерпретация).
13. Процедура проверки на значимость множественного коэффициента корреляции.
14. Коэффициент детерминации (формула для расчета, интерпретация).
15. Выборочный частный коэффициент корреляции (формула для расчета, интерпретация).
16. Процедура проверки на значимость выборочного частного коэффициента корреляции.
17. Коэффициент ранговой корреляции Спирмена (формула для расчета, интерпретация).
18. Процедура проверки на значимость коэффициента ранговой корреляции.
19. Задачи регрессионного анализа, основные предпосылки регрессионного анализа.
20. Использование МНК для расчета оценок параметров регрессионного уравнения.
21. Упрощенные формулы для расчета оценок параметров в случае парной линейной регрессии.
22. Свойства оценок параметров, полученных по МНК.
23. Стандартизованные коэффициенты уравнения регрессии, коэффициенты эластичности (формулы для расчета, интерпретация).
24. Линеаризующие преобразования (для функций, нелинейных по факторам и для функций, нелинейных по параметрам).
25. Характеристики качества уравнения регрессии: стандартная ошибка уравнения и множественный коэффициент детерминации (формулы для расчета и интерпретация).
26. Процедура проверки значимости уравнения регрессии.
27. Процедура проверки значимости параметров уравнения регрессии.
28. Формула для расчета стандартных ошибок параметров уравнения регрессии.
29. Доверительный интервал для параметров уравнения регрессии (формула для расчета, интерпретация).
30. Построение точечных прогнозов.
31. Интервальная оценка линии регрессии (формула для расчета, интерпретация).
32. Доверительный интервал для индивидуального прогнозного значения зависимой переменной.
33. Понятие мультиколлинеарности, причины ее возникновения.
34. Следствия мультиколлинеарности.
35. Признаки наличия мультиколлинеарности.
36. Формальные критерии проверки наличия мультиколлинеарности.
37. Методы устранения мультиколлинеарности.
38. Критерии качества уравнения регрессии с целью сравнения подмножеств факторов.
39. Понятие временного ряда, его характерные особенности.
40. Понятие тенденции временного ряда (тренд).
41. Тенденции среднего уровня, дисперсии и автокорреляции временного ряда.
42. Процедура проверки наличия тренда.
43. Процедуры сглаживания временных рядов.

44. Формулы для аналитического выравнивания временных рядов.
45. Понятие автокорреляции, автокорреляционной функции.
46. Коэффициент автокорреляции (формула для расчета, интерпретация).
47. Примеры интерпретации коррелограмм.
48. Процедура проверки на наличие автокорреляции (критерий Дарбина-Уотсона).
49. Процедура построения авторегрессионных уравнений.
50. Коэффициент множественной автокорреляции.
51. Методы устранения автокорреляции: метод последовательных разностей.
52. Методы устранения автокорреляции: метод коррелирования отклонений уровня ряда от основной тенденции.
53. Коэффициент лаговой корреляции (формула для расчета, интерпретация).
54. Понятия периода колебаний временного ряда, частоты, фазы, амплитуды.
55. Определение количества гармоник, входящих в разложение детерминированной составляющей временного ряда (для рядов с четным и нечетным периодом колебаний).
56. Разложение временного ряда в ряд Фурье.
57. Понятие дисперсионного анализа, его сущность и задачи.
58. Формирование планов эксперимента: полные и неполные, случайные и рандомизированные планы эксперимента.
59. Разложение общей суммы квадратов в однофакторном дисперсионном анализе. Оценки дисперсий.
60. Разложение общей суммы квадратов в двухфакторном дисперсионном анализе. Оценки дисперсий.
61. Понятие системы одновременных регрессионных уравнений: общий вид, модель спроса-предложения.
62. Структурная и приведенная формы эконометрической модели, построенной на базе систем одновременных уравнений. Рекурсивная модель.
63. Идентификация систем одновременных уравнений (статистическое оценивание неизвестных значений параметров системы): идентификация рекурсивных систем, косвенный метод наименьших квадратов.

Для проведения зачета по первому модулю курса «Эконометрика» на дневном отделении используются вопросы из данного списка №№1-38.

Примерные варианты заданий аудиторной работы

ВАРИАНТ 1.

Задание 1. По данным n=15 предприятий, каждое из которых характеризуется по трем показателям: x_1 – объем сменной выработки, x_2 – себестоимость продукции и x_3 – фондоотдача; получена матрица парных коэффициентов корреляции:

$$R=\begin{pmatrix} 1 & -0,6 & 0,8 \\ -0,6 & 1 & -0,6 \\ 0,8 & -0,6 & 1 \end{pmatrix}$$

Определите оценку частного коэффициента корреляции $\mathbf{r_{23.1}}$.

Задание 2. По данным задания 1 проверить при α=0,05 значимость частного коэффициента корреляции $\mathbf{r_{23.1.}}$

Задание 3. По данным задания 1 найти точечную оценку множественного коэффициента корреляции, характеризующего тесноту связи между себестоимостью и остальными переменными.

Задание 4. По данным заданий 1 и 3 при α=0,05 проверить значимость множественного коэффициента корреляции $r_{2.13}$.

Задание 5. По данным заданий 1 и 4 определите, какая доля дисперсии x_2 объясняется влиянием показателей x_1 и x_3.

Задание 6. Известно, что факторный признак x_3 усиливает связь между величинами x_1 и x_2.По результатам наблюдений получен частный коэффициент корреляции $r_{12..3}$= -0,45. Какое значение может принять парный коэффициент корреляции r_{12}:
а) –0,4; б) 0,344; в) – 0,8; г)1,2.

Задание 7. Предположим, Вы обследовали два динамических ряда А и В, каждый из которых содержит по 60 наблюдений. Данные ряды были исследованы на автокорреляцию уровней, в результате чего были получены следующие автокорреляционные функции:

Лаг (τ)	1	2	3	4	5	6	7
r_τ(A)	0,68	0,23	-0,055	-0,145	-0,106	-0,052	-0,135

Лаг (τ)	1	2	3	4	5	6	7
r_τ(B)	-0,03	0,15	0,17	-0,08	-0,05	0,20	-0,01

Сделайте заключение о внутренней структуре временных рядов.
Для описания какого из этих рядов лучше подойдет авторегрессионная модель?

Задание 8. При исследовании зависимости себестоимости продукции y от объема выпуска x_1 и производительности труда x_2 по данным n=20 предприятий получено уравнение регрессии $\hat{y}$=2,88-0,72x_1-1,51x_2. С доверительной вероятностью p=0,99 определите, на какую величину максимально может измениться себестоимость продукции y, если объем производства x_1 увеличить на единицу (известно, что S_{b1}= 0,052; S_{b2}=0,5):
а) –0,6; б) 0,72; в) –1,5; г) –0,83.

Задание 9. Для уравнения регрессии $\hat{y}$=2,88-0,72x_1-1,51x_2 рассчитан множественный коэффициент корреляции $r_{y.x1x2}$=0,84. Какая доля вариации результативного показателя $\boldsymbol{y}$ (в %) объясняется входящими в уравнение регрессии переменными x_1 и x_2:
а) 70,6; б) 16,0; в) 84,0; г) 29,4.

Задание 10. Предположим, в результате Вашего исследования было получено два вида трендовых моделей, каждая из которых содержит по четыре объясняющих переменных. При этом было обследовано 35 объектов. Построенные модели имеют следующие характеристики:
Модель 1. R^2=0.95; F=0.5; DW=3,5;
Модель 2. R^2=0.76; F=1.85; DW=2,1.
Какая модель является более адекватной и почему?

ВАРИАНТ 2

Задание 1. На основании данных n=5 стран о зависимости темпов роста ВНП (y) и темпов прироста промышленного производства (x) получена оценка уравнения регрессии $\hat{y}$= 1,56+0,21x и оценка остаточной дисперсии, равная 0,04. Матрица значений аргументов X имеет вид:

$$X=\begin{pmatrix} 1\ 1\ 1\ 1\ 1 \\ 1\ 3\ 4\ 6\ 8 \end{pmatrix}^T$$

Проверить значимость коэффициента регрессии H_0: β_1=0 при α=0,05

Задание 2. По данным задания 1 при доверительной вероятности p=0,95 определите интервальную оценку для β_0.

Задание 3. По данным задания 1, приняв $\bar{y}=2,5$ и $\bar{x}=3,5$ определите, на сколько процентов в среднем изменится y, если x увеличится на 1%.

Задание 4. По данным задания 1 определите с доверительной вероятностью p=0,95 интервальную оценку для $\beta_{1.}$

Задание 5. Множественный коэффициент корреляции $r_{1.23}$ =0.8. Определите, какой процент дисперсии величины x_1 объясняется влиянием величин x_2 и x_3:
а) 28%; б) 32%; в) 64%; г)80%.

Задание 6. По результатам n=25 наблюдений получен парный коэффициент корреляции r_{12}=0,6. Известно, что x_3 занижает связь между x_1 и x_2. Какое значение может принять частный коэффициент корреляции $r_{12\cdot3}$:
а) –0,8; б) –0,2; в) 0,5; г) 0,8.

Задание 7. Предположим, Вы обследовали 66 наблюдений над некоторыми экономическими показателями X_1, X_2, X_3 и X_4. В результате получилась следующая матрица парных коэффициентов корреляции: $R=\begin{pmatrix} 1 & 0,78 & 0,25 & -0,11 \\ 0,78 & 1 & 0,98 & 0,02 \\ 0,25 & 0,98 & 1 & 0,53 \\ -0,11 & 0,02 & 0,53 & 1 \end{pmatrix}$

Проверьте полученные парные коэффициенты корреляции на значимость (α=0,05) и определите, считая показатель X_1 откликом, а X_2, X_3 и X_4 – объясняющими переменными, какие из факторов, скорее всего, войдут в уравнение регрессии, а какие нет и почему?

Задание 8. Какие требования в модели регрессионного анализа предъявляются к распределению ошибок наблюдения ε_i, а именно к их математическому ожиданию и дисперсии $D\varepsilon_i$:
а) $M\varepsilon_i=1$; $D\varepsilon_i=\sigma^2$; б) $M\varepsilon_i=0$; $D\varepsilon_i=1$;
в) $M\varepsilon_i=0$; $D\varepsilon_i=\sigma^2$; г) $M\varepsilon_i=1$; $D\varepsilon_i=0$.

Задание 9. В результате Вашего исследования была построена следующая регрессионная модель: $Уt = -43 + 0{,}39Y_{t-1} + 0{,}92X_{1t} - 0{,}35X_{1t-1} + \varepsilon_t$.
Перечислите: а) объясняющие переменные; б) авторегрессионные составляющие; в) лаговые и трендовые компоненты.

Задание 10. Какое значение может принимать коэффициент детерминации:
а) -0,5; б) –0,2; в) 0,4; г)1,2.

ВАРИАНТ 3

Задание 1. Что минимизируется согласно методу наименьших квадратов:

а) $\sum_{i=1}^{n}(yi-\bar{y})^2$; б) $\sum_{i=1}^{n}\left|yi-\hat{y}_i\right|$; в) $\sum_{i=1}^{n}\left(yi-\hat{y}_i\right)$; г) $\sum_{i=1}^{n}\left(yi-\hat{y}_i\right)^2$;

Задание 2. По данным n=15 фирм исследована зависимость прибыли y от числа работающих x вида: $\hat{y}=b_0+b_1x$. Была получена оценка остаточной дисперсии $S^2_{ост}=2{,}2$ и обратная матрица:

$$(X^TX)^{-1} = \begin{pmatrix} 0{,}31 & -0{,}03 \\ -0{,}03 & 0{,}05 \end{pmatrix}$$

Определить, чему равна ошибка оценки параметра b_1:
а) 1,500; б) 0,332; в) 0,682; г) 0,242.

Задание 3. Предположим, Вы обследовали два динамических ряда А и В, каждый из которых содержит по 40 наблюдений. Данные ряды были исследованы на автокорреляцию уровней, в результате чего были получены следующие автокорреляционные функции:

Лаг (τ)	1	2	3	4	5	6	7
$r_\tau(A)$	0,67	-0,21	-0,28	0,73	-0,05	-0,13	0,78
Лаг (τ)	1	2	3	4	5	6	7
$r_\tau(B)$	0,80	0,75	0,70	0,61	0,53	0,44	0,35

Сделайте заключение о внутренней структуре временных рядов.
Для описания какого из этих рядов лучше подойдет трендовая модель?

Задание 4. Какое значение может принять множественный коэффициент корреляции:
а) –1; б) –0,5; в) 0; г) 1,2.

Задание 5. Дайте определение и укажите различие парного и частного коэффициентов корреляции.

Задание 6. Известно, что при фиксированном значении X_3 между величинами X_1 и X_2 существует положительная связь. Какое значение может принять частный коэффициент корреляции $r_{12.3}$:
а) –0,8; б) 0; в) 0,4; г) 1,3.

Задание 7. На основании данных о динамике процента хронических больных на 1000 жителей, приведенных в таблице, а также предположения, что генеральное уравнение регрессии имеет вид: $\breve{y}= \beta_0 + \beta_1x$ определить оценки b_0 и b_1 параметров уравнения регрессии

Годы (Х)	0	1	2	3	4
Доля хронических больных на 1000 жителей в % (У)	10	8	5	3	4

Задание 8. По данным задания 7 определить величину остаточной дисперсии($S^2_{ост}$), объясненной дисперсии ($S^2_{объясн}$) и общей дисперсии ($S^2_{общ}$). Проверить значимость уравнения регрессии при α=0,05.

Задание 9. Статистический анализ экономических показателей производственного объединения показал, что зависимость объема выпускаемой продукции (У) от производительности труда (Х) оценивается двумя уравнениями регрессии:
1) У=2 + 4Х; 2) У=2,5 + 4,5lnX.
Какое уравнение регрессии точнее описывает зависимость объема выпускаемой продукции (У) от производительности труда (Х), если известно, что в первом уравнении 80% общей дисперсии объема выпускаемой продукции определяет влияние неучтенных факторов, а во втором уравнении 60% общей дисперсии объема выпускаемой продукции обусловлено влиянием производительности труда?

Задание 10. По результатам n=20 наблюдений найден множественный коэффициент корреляции $r_{1.23}=0.8$. Проверьте значимость множественного коэффициента корреляции H0: $r_{1\text{-}23}=0$ при α=0,05.

ВАРИАНТ 4

Задание 1. На основании данных о темпе прироста (%) ВНП (У) и промышленного производства (Х) десяти развитых стран мира за 1992 г., приведенных в таблице, и предположения, что генеральное уравнение регрессии имеет вид $\breve{y}=\beta_0+\beta_1x$, определить оценки параметров уравнения регрессии.

СТРАНЫ	У	Х
Япония	3,5	4,3
США	3,1	4,6
Германия	2,2	2,0
Франция	2,7	3,1
Италия	2,7	3,0
Великобритания	1,6	1,4
Канада	3,1	3,4
Австралия	1,8	2,6
Бельгия	2,3	2,6
Нидерланды	2,3	2,4

Задание 2. По данным задания 1 определить величины остаточной, объясненной и общей дисперсии.

Задание 3. По данным задания 1 и результатам расчетов в задании 2 при α=0,05 проверить значимость уравнения регрессии.

Задание 4. По данным задания 1 и результатам расчетов в задании 2 при α=0,05 проверить значимость коэффициентов уравнения регрессии.

Задание 5. Предположим, Вы исследовали экономическую природу некоторого показателя У. В результате на основании n=25 наблюдений было построено уравнение регрессии от двух факторов X_1 и X_2 следующего вида:
$\hat{y}= -18{,}53+2{,}38X_1-0{,}76X_2$. Величина остаточной дисперсии составляет 6,79; величина объясненной дисперсии равна 15,75.

Определите стандартную ошибку оценки по регрессии (среднеквадратическое отклонение от линии регрессии).

Задание 6. По данным задания 5 определите коэффициент множественной корреляции $r_{Y.X1X2}$ и коэффициент детерминации.

Задание 7. По данным задания 5 определите, является ли уравнение регрессии значимым по критерию Фишера при α=0,05.

Задание 8. В результате исследования экономической природы выпуска некоторого продукта было построено уравнение регрессии от двух факторов L (труд) и K(капитал) на основе обследования n=20 предприятий некоторой отрасли. Полученное уравнение регрессии имеет следующий вид:
$У=5{,}03K^{0,3}L^{0,7}$. Остаточная дисперсия составляет 9,18; объясненная дисперсия равна 15,32. Определите стандартную ошибку оценки по регрессии (среднеквадратическое отклонение от линии регрессии).

Задание 9. По данным задания 8 определите коэффициент множественной корреляции $r_{Y.KL}$ и коэффициент детерминации.

Задание 10. По данным задания 8 определите, является ли уравнение регрессии значимым по критерию Фишера при α=0,05.

4. КРАТКИЙ КОНСПЕКТ ЛЕКЦИЙ

Тема 1. МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ КУРСА (4 часа)

Предмет эконометрики - **исследование и установление количественных закономерностей и количественных взаимозависимостей в экономической жизни при помощи математических и математико-статистических методов.**

Этапы эконометрического моделирования.

- изучение объекта
- сбор и предварительная обработка информации
- построение модели
- статистический анализ модели
- проверка модели на адекватность
- практическое использование модели

Наиболее часто употребляемые характеристики случайной величины (и соответствующего распределения вероятностей) - *моменты и квантили*.

Среднее значение наблюдаемого признака $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} xi$

Дисперсия или второй центральный момент эмпирического распределения

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{причем} \quad S^2 = m_2$$

В случае одномерного эмпирического распределения произвольным моментом порядка k называется сумма k-ых степеней отклонений результатов наблюдений от произвольного числа c, деленная на объем выборки n:

$m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^k$, где k может принимать любые значения натурального ряда чисел.

Если $c = 0$ - то момент *начальный*. Начальным моментом первого порядка является выборочное среднее $\bar{x}$ При $c = \bar{x}$ момент называют *центральным.*

Первый центральный момент

$$m_1 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^1 = 0$$

Второй центральный момент $m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ представляет собой дисперсию S^2 эмпирического распределения.

Выборочное среднеквадратическое отклонение $S = \sqrt{S^2}$

Выборочный коэффициент вариации $v = \frac{S}{\bar{x}}$ или $v = \frac{S}{\bar{x}} \cdot 100\,\%$

Выборочная квантиль - решение уравнения $F_n(x) = p$, выборочная медиана есть решение уравнения $F_n(x) = 0.5$

Для умеренно асимметричных распределений существует соотношение

$$X_{Мода} = X_{Сред.арифм.} - 3(X_{Сред.арифм.} - X_{Медиана})$$

Все виды средних характеризуют **уровень** числовой совокупности.

К характеристикам **меры рассеяния** (амплитуды рассеяния) относятся дисперсия, среднеквадратическое отклонение, коэффициент вариации и вариационный размах R.

$R = X_{max} - X_{min}$

Отсев грубых погрешностей при *n≤25:* $\frac{|x_i - \bar{x}|}{S} \le \tau_{1-p}$

Показатель асимметрии $g_1 = \frac{m_3}{m_2^{3/2}}$

Для симметричных распределений $m_3 = 0$ и $g_1 = 0$

Для нормального распределения $\frac{m_4}{m_2^2} = 3$

Показатель эксцесса $g_2 = \frac{m_4}{m_2^2} - 3$

Для распределений, имеющих крутую левую ветвь гистограммы и пологую правую, матрица исходных данных преобразуется по формулам:

$x' = \lg(x \pm a) \cdot 10^b$ или $x' = \frac{1}{x}$ или $x' = \frac{1}{\sqrt{x}}$

Для распределений, смещенных вправо, матрицу исходных данных преобразуют по формуле $x' = x^a$ (при *a=1.5; 2*)

Тема 2. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ (4 часа)

Корреляционный анализ (корреляционная модель) - метод, применяемый тогда, когда данные наблюдений или эксперимента можно считать случайными и выбранными из совокупности, распределенной по многомерному нормальному закону.

Изображение точек выборки $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ на координатной плоскости - **корреляционное поле.**

Мерой линейной статистической связи двух случайных величин, имеющих нормальное распределение, является коэффициент парной корреляции. **Выборочный парный коэффициент корреляции** определяется по формуле:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Проверка на значимость. Гипотезы: $H_0 : r = 0; \quad H_1 : r \neq 0$

Статистический критерий проверки гипотез:

При справедливости гипотезы H_0 статистика критерия

$t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}}$ имеет t-распределение Стьюдента с $(n-2)$ степенями свободы.

Гипотеза H_0 отвергается, т.е. выборочный коэффициент корреляции значимо отличается от нуля, если: $t_{расч} > t_{\alpha,n-2}$, где $t_{\alpha,n-2}$ - табличное значение критерия Стьюдента, определенное на уровне значимости α при числе степеней свободы $(n-2)$

Доверительный интервал (интервальная оценка) значимого коэффициента корреляции

$$p = r \pm u_\alpha \frac{1-r^2}{\sqrt{n}}$$

Здесь u_α - критическая точка стандартного нормального распределения, соответствующая уровню значимости α, n - объем выборки.

В случае нелинейной зависимости тесноту связи между величинами оценивают по величине **корреляционного отношения.**

$\eta = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$, где y_i - наблюдаемые значения, а $\hat{y}_i$ - расчетные значения зависимого признака, которые вычисляют на основе уравнения парной регрессии $\hat{y}_i = f(x_i)$. Интервал изменения корреляционного отношения $0 \le \eta \le 1$

Величина $\eta_{yx}{}^2$, называемая **коэффициентом детерминации**, показывает, какая часть общей вариации Y обусловлена вариацией X.

Проверка значимости корреляционного отношения η основана на том, что статистика

$$F = \frac{\eta^2(n-p)}{(1-\eta^2)(p-1)}$$

p – число факторов

n – количество наблюдений (где p – число факторов)

имеет F- распределение Фишера - Снедекора с $f_1 = p-1$ и $f_2 = n-p$ степенями свободы. Поэтому η значимо отличается от нуля, если $F > F_{\alpha;f1;f2}$ где

$F_{\alpha;f1;f2}$ - табличное значение F - критерия на уровне значимости α при числе степеней свободы $f_1 = p-1$ и $f_2 = n-p$.

МНОЖЕСТВЕННЫЙ КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Пусть имеется совокупность случайных переменных $X_1, X_2, ..., X_i, ..., X_j, ..., X_p$, имеющих совместное нормальное распределение.

Матрица Q

$$Q = \begin{pmatrix} 1 & r_{12} & r_{13} & ...r_{1p} \\ r_{21} & 1 & r_{23} & ...r_{2p} \\ r_{31} & r_{32} & 1 & ...r_{3p} \\ ... & ... & ... & ... \\ r_{p1} & r_{p2} & r_{p3} & ...1 \end{pmatrix}$$

составленная из парных коэффициентов корреляции r_{ij}, где $i, j = 1,2,..., p$, называется **корреляционной**.

Теснота линейной взаимосвязи одной переменной Xi с совокупностью других $(p-1)$ переменных, рассматриваемой в целом, измеряется с помощью **выборочного коэффициента множественной корреляции *Ri*** *(i=1,p):*

$R_i = \sqrt{1 - \frac{Q}{Qii}}$, где Q - определитель корреляционной матрицы, Q_{ii} - алгебраическое дополнение корреляционной матрицы.

0≤R≤1.

Величина ***R²***, называемая **выборочным множественным коэффициентом детерминации**, показывает, какую долю вариации исследуемой переменной объясняет вариация остальных переменных.

Коэффициент множественной корреляции значимо отличается от нуля, если значение статистики $F > F_{\alpha;f1;f2}$, где F рассчитывается по формуле

$$F = \frac{R^2(n-p)}{(1-R^2)(p-1)}$$

n - объем выборки;

p - количество переменных;

$F_{\alpha;f1;f2}$ - табличное значение

F - критерия на уровне значимости α при числе степеней свободы $f_1=p-1$ и $f_2=n-p$.

Выборочным частным коэффициентом корреляции между переменными X_i и X_j при фиксированных значениях остальных *(p-2)* переменных называется выражение:

$$r_{ij,1...p} = \frac{-Qij}{\sqrt{QiiQjj}},$$

где *Qij, Qii* и *Qjj* - алгебраические дополнения соответствующих элементов матрицы парных коэффициентов корреляции.

Проверка значимости частного коэффициента корреляции осуществляется также, как парного коэффициента корреляции, только ***n-2*** при этом заменяется на ***(n-p)***, т.е. вычисляется статистика Стьюдента

$$t = \frac{|r_{ij}|\sqrt{n-p}}{\sqrt{1-r_{ij}{}^2}},$$

где *n* - количество наблюдений; *p* - количество факторов.

Коэффициент частной корреляции считается значимым, если $t > t_{v,\alpha}$, причем значение $t_{v,\alpha}$ определяется по таблицам распределения Стьюдента: α - уровень значимости, $v=n-p$ число степеней свободы.

Коэффициент ранговой корреляции Спирмена $\rho = -\frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n}$,

где r_i и s_i ранги *i*- того объекта по переменным *X* и *Y; n* – число пар наблюдений.
При проверке значимости ρ исходят из того, что в случае справедливости нулевой гипотезы об отсутствии корреляционной связи между переменными при $n > 10$ статистика

$$t_{расч} = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}}$$

имеет *t*-распределение Стьюдента с *(n-2)* степенями свободы.
Коэффициент ранговой корреляции *p* значим на уровне α, если $|t_{расч}| > t_{\alpha;n-2}$, где $t_{\alpha;n-2}$ табличное значение *t*-критерия Стьюдента, определенное на уровне значимости *a* при числе степеней свободы (n-2).

Тема 3. МОДЕЛИ И МЕТОДЫ РЕГРЕССИОННОГО АНАЛИЗА (6 часов)

К задачам регрессионного анализа относятся:

- установление формы зависимости между переменными;
- оценка модельной функции (модельного уравнения) регрессии;
- оценка неизвестных значений (прогноз значений) зависимой переменной.

В регрессионном анализе рассматривается односторонняя зависимость переменной *Y* (ее еще называют **функцией отклика**, **результативным признаком, предсказываемой переменной**) от одной или нескольких независимых переменных *X* (называемых также **объясняющими** или **предсказывающими переменными**, **факторными признаками**).

Парная регрессионная модель $y = f(x) + \varepsilon$ где ε - случайная переменная, характеризующая отклонение от модельной функции регрессии (она также называется **возмущающей** или просто **возмущением)**.

Основные предпосылки регрессионного анализа:

1. зависимая переменная y_i (или возмущение ε_i) есть величина случайная, а объясняющая переменная x_i есть величина неслучайная;
2. математическое ожидание возмущения ε_i равно нулю: $M(\varepsilon_i)=0;$
3. дисперсия зависимой переменной y_i (или возмущения ε_i) постоянна для любого i: $D(\varepsilon_i) = \sigma_\varepsilon^2$;
4. переменные y_i и y_i (или возмущения ε_i и ε_j) не коррелированы: $M(\varepsilon_i \cdot \varepsilon_j)=0;$

5.зависимая переменная y_i (или возмущение ε_i)есть нормально распределенная случайная величина.

Простейшая модель регрессионного анализа, когда функция *f(x)* линейна как по параметрам, так и по переменным x_i *(i=1,n):*

$y_i=\beta_0 + \beta_1 x_i + \varepsilon_i.$

Оценкой линейной модели по выборке является уравнение регрессии $y_x = b_0 + b_1 x$. Параметры b_0 и b_1 определяются на основе **<u>метода наименьших квадратов</u>**.

Модель множественной регрессии, включающая p объясняющих переменных $x_1,..., x_p$, имеет вид:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ...+ \beta_p x_{ip} + \varepsilon_i$$

Уравнение регрессии с оценками параметров можно записать как:

$$y_x = b_0 + b_1 x_1 + b_2 x_2 + ...+ b_p x_p.$$

Оценки параметров уравнения множественной регрессии $b_1,b_2,...,b_p$ получают по методу наименьших квадратов. Оценки, полученные на основе применения этого метода, обладают следующими свойствами: *несмещенностью* (т.е. они не содержат систематических ошибок при оценивании), *состоятельностью* (т.е. при n→∞ с вероятностью, как угодно близкой к 1, сходятся к оцениваемым параметрам) и *эффективностью* (т.е. обладают наименьшими дисперсиями среди всех возможных несмещенных оценок параметров).

Стандартизованные коэффициенты регрессии b'_j и коэффициенты эластичности $Э_j$ *(j=1,p):*

$$b_j' = b_i \frac{Sx_j}{S_y} \qquad Э_j = b_j \frac{\bar{x}_j}{\bar{y}}$$

Стандартизованный коэффициент регрессии b_j' показывает, на сколько величин *Sy* изменится в среднем зависимая переменная *Y* при увеличении только *j*-ой объясняющей переменной на Sx_j, а коэффициент эластичности $Э_j$ - на сколько процентов (от средней) изменится в среднем *Y* при увеличении только X_j на 1% .

Нелинейная регрессия

С позиции использования МНК различают следующие виды зависимостей:

1) функции, нелинейные по факторам, например:
$y=a_0 + a_1 x^2$ или $y=a_0 + a_1 logx;$
2) функции, нелинейные по параметрам, например:
$y = e^{a_0 + a_1 x}$ или $y=aK^{\alpha}L^{1-\alpha};$
3) функции, не приводимые к линейному виду.

В первом и втором случае МНК для оценки параметров модели используется после проведения линеаризующих преобразований, приводящих функцию к линейному виду. К уравнениям второго вида вначале требуется применить линеаризующее преобразование, а затем уже заменить переменные. В третьем случае, когда функцию невозможно привести к линейному виду, оценивание параметров осуществляют с помощью нелинейного МНК.

Анализ вариации зависимого показателя

Общая сумма квадратов отклонений зависимой переменной:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

Сумма квадратов отклонений фактических значений от расчетных называется **остаточной суммой квадратов** и обозначается как $Q_{ост} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Сумма квадратов отклонений расчетных значений от среднего называется **объясненной суммой квадратов** и обозначается как $Q_{объясн} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

Третий элемент в разложении представляет собой сумму произведений объясненной и остаточной компонент регрессии и равен нулю. В результате общая сумма квадратов раскладывается на остаточную сумму квадратов и объясненную сумму квадратов.

Для получения оценок соответствующих дисперсий все перечисленные выше суммы квадратов делятся на соответствующие значения степеней свободы:

- общую сумму квадратов *Q* на *(n-1)* для получения оценки общей дисперсии (S^2y) зависимой переменной, которая характеризует разброс значений показателя вокруг среднего;
- объясненную сумму квадратов на *p* (количество факторов в уравнении регрессии) для получения оценки объясненной дисперсии переменной у $(S^2 объясн)$, которая характеризует вариацию зависимого показателя, объясненную построенным уравнением регрессии;
- остаточную сумму квадратов на *(n-p-1)* для получения оценки остаточной дисперсии зависимой переменной (S^2ост), которая характеризует разброс значений относительно линии регрессии.

Среднеквадратическое отклонение от линии регрессии есть квадратный корень из значения $S^2ост$ (чаще называется **стандартной ошибкой** регрессии $SE = \sqrt{S^2 ост}$).

Величина $\mathbf{R^2}$ – **множественный коэффициент детерминации,** показывает, какая часть дисперсии функции отклика объясняется вариацией линейной комбинации выбранных факторов $x_1,x_2,..,x_j,..,x_p$.

$$R^2 = 1 - \frac{Q_{ост}}{Q_{общ}} = \frac{Q_{объясн}}{Q_{общ}}$$

Измеряется в долях единицы (от 0 до 1) либо в процентах (от 0 до 100%).

Квадратный корень из коэффициента детерминации есть коэффициент множественной корреляции, характеризует тесноту связи между функцией отклика и совокупности факторов, включенных в уравнение.

Проверка значимости уравнения регрессии:

H_0: все $\beta_j = 0$; H_1: существует хотя бы один $\beta_j \neq 0$.

Уравнение регрессии значимо, если

$$Fp = \frac{S^2_{объясн}}{S^2_{ост}} > Fm_{\alpha,\nu 1,\nu 2}$$

(где $\nu_1=p$, $\nu_2=n-p-1$).

Проверка значимости параметров: H_0: $\beta_j=0$; H_1: $\beta_j \neq 0$.

Расчетное значение t-статистики Стьюдента $t = \frac{|b_j|}{S_{bj}}$

где $|bj|$ - абсолютное значение оценки параметра β_j;

S_{bj} - стандартная ошибка параметра, определяемая по формуле:

$$S_{bj} = \sqrt{S^2 ост * c_{jj}}$$

где c_{jj} - диагональный элемент матрицы, обратной матрице нормальных уравнений $((X^T X)^{-1})$.

Если $t_{расч} > t_{табл,\alpha,\nu}$ (где $\nu=n-p-1$), то данный фактор оказывает существенное (значимое) влияние на результирующую переменную.

Доверительный интервал для коэффициентов регрессии:

$$b_j - t_{табл} * S_{bj} \le \beta_j \le b_j + t_{табл} * S_{bj}.$$

Точечная оценка результирующего признака: $y^* = b_0 + b_1 x_1^* + b_2 x_2^* + ... + b_p x_p^*$.

Доверительный интервал для условного математического ожидания $Mx(Y)$ (или же **для линии регрессии**):

$$Mx(Y) = y^* \pm t_{\nu,\alpha} S_{ост} \sqrt{X^T (X^T X)^{-1} X},$$

где $\nu=n-p-1$, n- объем выборки, p- число факторов в уравнении регрессии.

Для парной линейной регрессии доверительный интервал для линии регрессии :

$$y_{1,2} = y^* \pm t_{\nu,\alpha} S_{ост} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - x)^2}}$$

где x* - прогнозное значение фактора, n-объем выборки, t-статистика Стьюдента с числом степеней свободы ν=n-2 и уровнем значимости α.

$S_{ост} = \sqrt{S^2_{ост}}$ - среднеквадратическое отклонение наблюдений от линии регрессии, произведение $S_{ост} \sqrt{X^T (X^T X)^{-1} X}$ есть погрешность оценки регрессии.

Доверительный интервал для индивидуальных значений зависимой переменной:

$$y_{прогноз} = y^* \pm t_{\nu,a} S_{ост} \sqrt{1 + X^T (X^T X)^{-1} X}$$

В случае парной линейной регрессии **доверительный интервал для индивидуального прогнозного значения зависимой переменной**:

$$y_{1,2} = y^* \pm t_{\nu,a} S_{ост} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2} + 1}$$

Тема 4. ПРОБЛЕМЫ ПРАКТИЧЕСКОГО ИСПОЛЬЗОВАНИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ (4 часа)

Мультиколлинеарность- высокая взаимная коррелированность объясняющих переменных.

Следствия мультиколлинеарности:

1.Резко падает точность оценок параметров, получаемых с помощью метода наименьших квадратов. Ошибки некоторых параметров уравнения могут стать очень большими.

2.Выборочные характеристики регрессионной модели становятся крайне неустойчивыми. При добавлении (исключении) некоторого количества наблюдений или факторов к массиву исходной информации может произойти резкое изменение оценок параметров.

3.Из-за неустойчивости модели резко сокращаются возможности содержательной интерпретации модели, а также прогноза значений зависимой переменной у в точках, существенно удаленных от значений объясняющих переменных в выборке в виду ненадежности получаемых результатов.

Признаки наличия мультиколлинеарности.

1) небольшие изменения в данных приводят к широким колебаниям оценок параметров;
2) коэффициенты регрессии имеют высокие стандартные ошибки и высокий уровень значимости, несмотря на тот факт, что совместно они высоко значимы и достаточно высоко значение множественного коэффициента детерминации;
3) коэффициенты могут иметь неверный знак или неправдоподобную величину.

Формальные критерии мультиколлинеарности:

1. **Критерий $\chi 2$**. Высокая коррелированность переменных проявляется в близости к нулю определителя матрицы парных коэффициентов корреляции $R=|rij|$.

Гипотезы процедуры проверки мультиколлинеарности $x_1, x_2, ..., x_p$:

H_0: между объясняющими переменными мультиколлинеарность отсутствует;

H_1: объясняющие переменные высококоррелированы.

Для проверки гипотезы вычисляется определитель корреляционной матрицы R и строится критерий $x^2 = -((n-1) - \frac{(2p+5)}{6}) \ln |R|$, который имеет $\chi 2$- распределение с количеством степеней свободы, равным. $\nu = \frac{p(p-1)}{2}$

Если $\chi 2расч \le \chi 2 \nu, \alpha$, то принимается гипотеза H_0.

2. **Критерий, использующий число обусловленности матрицы $X^T X$.**

Обусловленность матрицы определяется отношением максимального собственного числа (λmax) к минимальному (λmin): $y = \sqrt{\frac{\lambda \max}{\lambda \min}}$

Если $\lambda min \to 0$, то γ бесконечно велико и присутствует мультиколлинеарность. Для $\gamma > 20$ наблюдается приближенная коллинеарность объясняющих переменных. Для $\gamma < 20$ можно считать, что мультиколлинеарность отсутствует.

Сравнение подмножеств факторов по некоторому критерию качества уравнения регрессии:

1. Коэффициент детерминации (или квадрат коэффициента множественной корреляции).

$$R^2 = 1 - \frac{Q_{ост}}{Q_{общ}} = \frac{Q_{объясн}}{Q_{общ}}$$

2. Скорректированный коэффициент детерминации.

$$\bar{R}^2 = \frac{n-1}{n-q-1}(1 - R^2)$$

n-число наблюдений, q- число факторов в уравнении

Тема 5. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ (8 часов)

Временным рядом называют последовательность наблюдений, обычно упорядоченную во времени (возможно упорядочение и по какому-то другому параметру. Д*ва главных отличия* от наблюдений, образующих случайные выборки:

а) образующие временной ряд наблюдения, рассматриваемые как случайные величины, не являются взаимно-независимыми, и, в частности, значение, которое мы получим в момент времени t_k, может существенно зависеть от того, какие значения были зарегистрированы до этого момента времени;

б) наблюдения временного ряда (в отличие от элементов случайной выборки) не образуют стационарной последовательности, т.е. закон распределения вероятностей k-го члена временного ряда не остается одним и тем же при изменении его номера k; в частности от t_k могут зависеть основные числовые характеристики случайной переменной x_k - ее среднее значение и дисперсия. *Иначе говоря, при исследовании временных рядов существенное значение имеет тот порядок, в котором проводились наблюдения над исследуемой величиной.*

Динамика рядов экономических показателей в общем случае складывается из **четырех компонентов**:

1) *тенденции, характеризующей долговременную основную закономерность* развития исследуемого явления;
2) *периодического компонента*, связанного с влиянием сезонности развития изучаемого явления;
3) *циклического компонента*, характеризующего циклические колебания, свойственные любому воспроизводству (например, циклы обновления, связанные с чисто техническими проблемами);
4) *случайного компонента* как результата влияния множества случайных факторов.

Во временных рядах наблюдаются тенденции трех видов: *тенденция среднего уровня, тенденция дисперсии, тенденция автокорреляции.*

Тенденцию среднего уровня наглядно можно представить графиком временного ряда. Аналитически она выражается в виде функции *f(t)*, вокруг которой варьируют фактические значения изучаемого явления. *Тенденция дисперсии* - это изменения отклонений эмпирических значений временного ряда от значений, вычисленных по уровню тренда. *Тенденция автокорреляции* - это тенденция изменения связи между отдельными уровнями временного ряда.

Процедура проверки наличия тренда: временной ряд делится на две примерно равные части, для каждой из которых вычисляются величины средних и дисперсий $(\bar{y}_1, \bar{y}_2, S_1^2, S_2^2)$. После этого проверяется гипотеза о равенстве дисперсий при уровне значимости α, для чего формируются две гипотезы: H_0: $\sigma_1^2 = \sigma_2^2$; H_1: $\sigma_1^2 \neq \sigma_2^2$

Значимость различий проверяется путем вычисления $F_{расч} = \frac{S_2^2}{S_1^2}$ и сравнением ее с критическим значением F при $f_1=n_2-1$ и $f_2=n_1-1$ и уровне значимости α. Если *Fрасч<Fтабл*, то принимается H_0.

После этого проверяется основная гипотеза: H_0: $\bar{y}_1 = \bar{y}_2$ и гипотеза H_1: $\bar{y}_1 \neq \bar{y}_2$, для чего рассчитывается величина

$$T_{расч} = \frac{\bar{y}_1 + \bar{y}_2}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}} * \frac{n_1 n_2 (n_1+n_2-2)}{\sqrt{n_1+n_2}}$$

Если $|T_{расч}| < t_{крит_{\alpha,n-2}}$,то принимается нулевая гипотеза о равенстве средних, расхождение между вычисленными средними незначимо, т.е. тренд отсутствует. В противном случае, когда различие между средними будет значимо, принимается гипотеза H_1 и делается вывод о наличии тренда.

Сглаживающие процедуры

Пусть имеются наблюдения $y_1.....y_T$. Формула линейного фильтра (или сглаженного значения уровня в точке t)

$$y_t* = \sum_{s=-m}^{m} c_s y_{t+s} \qquad t=m+1,T-m \quad \text{причем} \quad \sum_{s=-m}^{m} c_s = 1$$

Если $c_s=const$, то фильтры **симметричные** и результат сглаживания есть вариант *среднего арифметического*:

$$y_t* = \frac{1}{2m+1} \sum_{s=-m}^{m} y_{t+s} \text{ , т.е. } c_S = \frac{1}{(2m+1)}$$

В случае, когда весовые коэффициенты не остаются постоянными, то говорят о *скользящей средней взвешенной*. Если задается $c_s<c_{s+1}$, то фильтр позволяет учесть устаревание данных.

Методы сглаживания, основанные на вычислении экспоненциальных средних значений уровня ряда по формуле типа $y_t* = \sum_{s=-m}^{m} (1-c_s)^m y_t$

Моделирование тенденции временного ряда с помощью аналитического выравнивания.

Функции типа $\bar{y}_t = a_0 + a_1 t$, где $\bar{y}_t$ - сглаженное (выровненное) значение уровня на момент t; коэффициенты $a_1, a_2, ... a_\tau$ - веса, приписываемые уровню ряда, находящемуся на расстоянии τ от момента t. Зависимости такого типа целесообразно применять для временных рядов с *постоянным абсолютным приростом или снижением показателей (когда уровни ряда увеличиваются в арифметической прогрессии).*

Если анализируемая тенденция характеризуется *постоянным темпом роста (рост уровней ряда идет в геометрической прогрессии),* то целесообразно проводить выравнивание по показательной функции: $\bar{y}_t = a_0 + a_1^t$ или $\bar{y}_t = a_0 a_1^{b_1 t + b_2 t^2}$

При выравнивании временных рядов экономических явлений, характеризующихся *стремлением к некоторой предельной величине*, насыщением, используется модифицированная экспонента $\bar{y}_t = a_0 + a_1 a_2^t$

Процессы с *переменными темпами* роста хорошо моделируются S-образными кривыми. К ним относятся логистические кривые и кривая Гомперца.

Метод последовательных разностей *при подборе кривых, описываемых полиномами.* Сущность этого метода заключается в нахождении первых, вторых и т.д. разностей уровней, т.е. $\Delta^1_t = y_t - y_{t-1}$; $\Delta^2_t = \Delta^1_t - \Delta^1_{t-1}$; $\Delta^3_t = \Delta^2_t - \Delta^2_{t-1}$ и т.д. Расчет этих разностей ведется до тех пор, пока разности не будут приблизительно равными. Порядок этих разностей и принимают за порядок искомого полинома.

Автокорреляция - это корреляционная зависимость между последовательными (соседними) значениями уровней временного ряда y_1 и y_2, y_2 и y_3, y_3 и y_4 и т.д.

Чтобы оценить степень зависимости между соседними уровнями временного ряда (автокорреляцию), рассчитывают коэффициенты автокорреляции между уровнями исходного ряда и того же ряда, но сдвинутого на τ шагов во времени. Величина **τ - шаг**.

Последовательность значений коэффициентов автокорреляции r_τ, вычисленных при τ =1,2,...,l , называют **автокорреляционной функцией** (как правило, $\tau \leq T/4$).

Коэффициент автокорреляции порядка τ: $r_\tau = \dfrac{\sum_{t=1}^{T-\tau} (y_t - \bar{y}_1)(y_{t+\tau} - \bar{y}_2)}{\sum_{t=1}^{T-\tau} (y_t - \bar{y}_1)^2 \sum_{t=1}^{T-\tau} (y_{t+\tau} - \bar{y}_2)^2}$

Если на оси абсцисс отложить значения τ, а на оси ординат - значения коэффициентов автокорреляции r_τ, а затем точки с координатами *(τ, r_τ)* соединить отрезками прямой, то получится **коррелограмма**. Совокупность значений коэффициентов автокорреляции с разными лагами $r_1, r_2, ..., r_p$ образует **корреляционную функцию**.

Ситуация, когда дисперсия остаточной компоненты возрастает, т.е распределение отличается от нормального и существует *автокорреляция в остатках* называется **гетероскедастичностью**.

Критерий Дарбина-Уотсона для проверки гипотезы о наличии автокорреляции:

$$DW = \frac{\sum_{t=1}^{T-1}(\varepsilon_{t+1} - \varepsilon_t)^2}{\sum_{t=1}^{T}\varepsilon_t^2}$$

Здесь ε_{t+1} и ε_t - отклонения от тренда. Возможные значения критерия находятся в интервале 0-4. При отсутствии автокорреляции значение *DW* колеблется около 2. Эмпирическое значение *DW* сравнивают с табличным значением для уровня значимости α, ν - число факторов в уравнении регрессии, n - число членов временного ряда, при этом если: 1)*DWрасч<dн* - автокорреляция присутствует; 2) *DWрасч>dв* - автокорреляция отсутствует; 3)*DWн≤DWрасч≤DWв* - необходимо дальнейшее исследование.

Коэффициент автокорреляции остатков: $\rho = \dfrac{\sum_{t=1}^{T}\varepsilon_{t+1}\varepsilon_t}{\sum_{t=1}^{T}\varepsilon_t^{\,2}}$

Методы устранения автокорреляции:

1) *метод последовательных или конечных разностей*;
2) *метод коррелирования отклонений уровней ряда от основной тенденции.*

Метод последовательных разностей - это метод коррелирования первых, вторых и т.д. разностей уровней временных рядов, при котором учитывается вид тренда. Если аппроксимирующие функции линейные, то коррелируются первые разности и коэффициент корреляции последовательных разностей:

$$r_{xy} = \frac{\sum_{t=1}^{T-1}\Delta_{lxt}\Delta_{lyt}}{\sqrt{\sum_{t=1}^{T-1}\Delta_{lxt}^{\,2}\sum_{t=1}^{T-1}\Delta_{lyt}^{\,2}}}$$

Метод коррелирования отклонений временных рядов – это метод измерения тесноты связи между отклонениями эмпирических значений уровней от выравненных по тренду. Формула коэффициента корреляции по отклонениям от трендов:

$$r_{xy} = \frac{\sum_{t=1}^{T}(x_t - x^*{}_t)(y_t - y^*{}_t)}{\sqrt{\sum_{t=1}^{T}(x_t - x^*{}_t)^2\sum_{t=1}^{T}(y_t - y^*{}_t)^2}} = \frac{\sum_{t=1}^{T}\varepsilon_{xt}\varepsilon_{yt}}{\sqrt{\sum_{t=1}^{T}\varepsilon_{xt}^{\,2}\sum_{t=1}^{T}\varepsilon_{yt}^{\,2}}}$$

Здесь x_t, y_t - фактические значения показателей; x_{t*}, y_{t*} - расчетные значения показателей; $\varepsilon_{xt}, \varepsilon_{yt}$ - отклонения от трендов.

Направление и продолжительность отставания уровней одного из взаимосвязанных рядов от уровней другого ряда называются **временным лагом**.

Коэффициент лаговой корреляции $y_{t+\tau}$ и x_t определяется по формуле:

$$r_\tau = \frac{\sum_{t=1}^{T-\tau}(x_t - \bar{x})(y_{t+\tau} - \bar{y})}{\sqrt{\sum_{t=1}^{T-\tau}(x_t - \bar{x})^2\sum_{t=1}^{T-\tau}(y_{t+\tau} - \bar{y})^2}}$$

где x_t и $y_{t+\tau}$ - уровни временных рядов, образующих пары, $\bar{x}$ и $\bar{y}$ - средние значения **укороченных** рядов, *n* - временной интервал наблюдений.

Построение множественной регрессионной модели по временным рядам:

1. Построение модели по уровням временных рядов: $\bar{y} = a_0 + a_1 y_1 + a_2 y_2 + ... + a_p y_p$
2. Построение модели по отклонениям уровней временных рядов от выравненных по тренду уровней: $y - \bar{y}_t = a_0 + a_1(x_1 - \bar{x}_{1t}) + ... + a_p(x_p - \bar{x}_{pt})$, где $\bar{y}_t$, $\bar{x}_{it}$ (i=1,p)-основные тенденции моделируемого признака и факторных признаков.
3. Построение модели по разности между уровнями рядов: $\Delta y_{t+1} = a_0 + a_1\Delta x_{1,t+1} + a_2\Delta x_{2,t+1} + ... + a_p\Delta x_{p,t+1}$
4. Построение модели по отклонениям уровней от среднего уровня. Равнозначен методу коррелирования отклонений от тренда.
5. Введение времени в модель в качестве независимой переменной.

Модели рядов, содержащие сезонную компоненту

Периодичность тренда означает, что он в точности повторяет себя через определенный промежуток времени (т.е. *f(t)=f(t+λ)),* причем такое повторение абсолютно регулярно. Промежуток λ называют **периодом колебаний временного ряда.** Величина, обратная периоду, называется **частотой**. Она равна числу периодов (не обязательно целому), содержащемуся в единичном интервале. Наибольшее значение периодической функции называется **амплитудой** ***(ρ).***

Детерминированная составляющая модели временного ряда:

$$f(t) = \alpha\cos\left(\frac{2\pi}{\lambda}\right)t + \beta\sin\left(\frac{2\pi}{\lambda}\right)t = p\cos\left(\frac{2\pi}{\lambda} * t - \theta\right)$$

Здесь λ-период колебаний тригонометрического слагаемого, величина λ может не совпадать с периодом колебания временного ряда; α и β-неизвестные параметры; ρ-амплитуда.

Максимальное число тригонометрических составляющих:

$q_{\max} = \dfrac{n-1}{2}$ для временных рядов с нечетным периодом n;

$q_{\max} = \dfrac{n}{2} - 1$ - для временных рядов с четным периодом колебаний n.

Порядок тригонометрического слагаемого $\varphi_k(t)$ задается числом k (k=1,2,…,$q_{\max}$), тогда тригонометрическое составляющее порядка k (или **гармонику порядка k**) можно записать как:

$$\varphi_k(t) = a_k\cos\frac{2\pi k}{n}t + в_k\sin\frac{2\pi k}{n}t$$

Детерминированная составляющая периодических колебаний временного ряда (разложение в ряд Фурье):

$$f(t) = a_0\sum_{k=1}^{q}(a_k\cos\frac{2\pi k}{n}t + в_k\sin\frac{2\pi k}{n}t) \quad \text{(если n – нечетное)}$$

$$f(t) = a_0 + \sum_{k=1}^{q}(a_k\cos\frac{2\pi k}{n}t + в_k\sin\frac{2\pi k}{n}t) + a_{\frac{n}{2}}(-1)^t \quad \text{(если n - четное)}$$

Слагаемое $\alpha_{n/2}(-1)^t$ - периодическая функция с периодом 2.

Оценки параметров для данной функции определяются с помощью МНК как:

(k=1,..., q; t=1,...,T)

$$a_0 = \frac{\sum y_t}{T} \qquad a_k = \frac{2}{T}\sum y_t \cos\frac{2\pi k}{T}t; \quad b_k = \frac{2}{T}\sum y_t \sin\frac{2\pi k}{T}$$

Для временных рядов с четным n **оценка** $a_{\frac{n}{2}} = \frac{1}{T}\sum_{t=1}^{T} y_t(-1)^t$

Оценка дисперсии оценок параметров в модели сезонных колебаний вычисляется как $S^2_{a_0} = \frac{S^2_{ост}}{T}$ $\quad S^2_{a_k} = S^2_{b_k} = \frac{2S^2_{ост}}{T}$ $\quad S_{a^2_{n/2}} = \frac{S^2_{ост}}{T}$

Оценки амплитуды колебаний и фазы: $Rk = \sqrt{a_k^{\ 2} + b_k^2}$; $\quad \theta_k = arctg\left(\frac{b_k}{a_k}\right)$

Тема 6. СИСТЕМЫ ЛИНЕЙНЫХ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ
(4 часа)

Система взаимосвязанных регрессионных уравнений и тождеств, в которой одни и те же переменные в различных регрессионных уравнениях могут одновременно выступать и в роли результирующих показателей, и в роли объясняющих переменных, называется ***системой одновременных (эконометрических) уравнений.***

$$\beta_{i1}y_{1t}+\beta_{i2}y_{2t}+\ldots+\beta_{iG}y_{Gt}+\gamma_{i1}x_{1t}+\ldots+\gamma_{ik}x_{Kt}=u_{it} \qquad (t=1,n;\ i=1,G)$$

Здесь y_{it} - значения эндогенных переменных в момент *t;* x_{it} - значения экзогенных переменных в момент t и лаговых эндогенных переменных. Переменные x_{it} в момент времени t называются **предопределенными**. *Совокупность равенств данного вида есть* ***система одновременных уравнений в структурной форме***.

Система указанных равенств в матричном виде: $By_t+Gx_t=u_t$, где B - матрица, состоящая из коэффициентов при текущих значениях эндогенных переменных; G - матрица, состоящая из коэффициентов при предопределенных переменных;

$y_t=(y_{1t},\ldots, y_{Gt})$; $x_t=(x_{1t},\ldots, x_{Kt})$; $u_t=(u_{1t},\ldots, u_{Gt})$ - вектора-столбцы.

Если матрица B невырождена, то систему можно разрешить относительно y_t: $y_t=Px_t+\eta_t$, где $P=-B^{-1}G$; $\eta_t=B^{-1}u_t$ - случайное возмущение. Такая форма записи называется **приведенной формой системы одновременных уравнений.**

* *Совместно-зависимые - это те переменные, которые в один и тот же момент времени выступают как объясняющие переменные в одних уравнениях и как зависимые - в других.*

В качестве одного из *критериев идентифицируемости*, удовлетворение требований которого обеспечивает однозначную идентифицируемость параметров системы уравнений, выступает **правило порядка.**

Правило порядка - число неизвестных, исключенных из уравнения, должно быть равно по меньшей мере числу уравнений минус единица, или *число исключенных из уравнения экзогенных переменных должно быть не меньше числа участвующих в нем эндогенных переменных, уменьшенного на единицу.*

Среди систем приведенных уравнений наиболее простые - **рекурсивные системы**, для оценивания коэффициентов которых можно применять метод наименьших квадратов. **Система одновременных уравнений *BY+GX=U* называется *рекурсивной*, если** матрица B является нижней треугольной матрицей (т.е. $\beta ij=0$ при j>i) и каждое ограничение на структурные коэффициенты относится к отдельному уравнению. Общий вид рекурсивной системы может быть представлен следующим образом:

$y_1=\beta_{11}x_1+\ldots+\beta_{1n}x_n+\varepsilon_1$

$y_2=-\alpha_{21}y_1+\beta_{21}x_1+\ldots+\beta_{2n}x_n+\varepsilon_2$

...

$y_m=-\alpha_{m1}y_1+\alpha_{mm-1}y_{m-1}+\beta_{m1}x_1+\ldots+\beta_{mn}x_n+\varepsilon_m$

Методы оценивания системы одновременных уравнений

Для оценивания параметров точно идентифицируемой системы применяется **косвенный метод наименьших квадратов**, состоящий в оценивании обычным МНК коэффициентов приведенной формы и подстановке оценок в выражения для коэффициентов структурной формы через коэффициенты приведенной формы, что приводит к смещенным, но состоятельным оценкам.

Для оценивания произвольных систем одновременных уравнений существует две группы методов:

1)методы, применяемые к каждому уравнению в отдельности (двухшаговый МНК, метод максимума правдоподобия с ограниченной информацией и др.);

2)методы, применяемые для оценивания всей системы в целом (трехшаговый МНК).

Тема 7. ОСНОВНЫЕ ПОНЯТИЯ И МОДЕЛИ ДИСПЕРСИОННОГО АНАЛИЗА (4 часа)

Дисперсионным анализом называется метод организации (планирования), статистического анализа и интерпретации результатов экспериментов, в которых изучается зависимость количественной переменной у от сочетания градаций качественных переменных X.

Математическая модель **однофакторного дисперсионного анализа (**когда оценивается влияние одного качественного признака на количественную переменную)**:**

$y_{ij} = \bar{y} + \alpha_j + \varepsilon_{ij}$ где y_{ij} - значение результирующего показателя для i-го *(i=1,n$_j$)* наблюдения при уровне градации *j (j=1,P)* качественного признака; n_j - количество наблюдений, когда фактор находится на уровне *j* $(\Sigma n_j=N,\ j=1,P)$; $\bar{y}$ - среднее значение результирующего показателя по всем наблюдениям всех градаций качественного признака; α_j - эффект влияния фактора на j-ом уровне;

ε_{ij} - случайная компонента, отражающая влияние всех прочих факторов

Табличная форма представления исходных данных для проведения однофакторного ДА

Градации качественного фактора	**Значения результирующего показателя (y_{ij})**	**Количество наблюдений в группе (n_j)**	**Сумма значений наблюдений в группе ($\sum_{i=1}^{n_j} y_{ij}$)**	**Среднее значение наблюдений в группе** $\bar{y}_i = \frac{\sum y_{ij}}{n_j}$
1	$y_{11}, y_{12},\dots, y_{n11}$	n_1	$\sum_{i=1}^{n_1} y_{i1}$	$\bar{y}_i = \frac{\sum_{i=1}^{n_1} y_{ij}}{n_1}$,
…	…	…	**…**	**…**
J	$y_{1j}, y_{2j},\dots, y_{njj}$	n_j	$\sum_{i=1}^{n_j} y_{ij}$	$\bar{y} = \frac{\sum_{i=1}^{n_j} y_{ij}}{n_j}$
…	…	…	…	…
P	$y_{1P}, y_{2P},\dots, y_{nPP}$	n_P	$\sum_{i=1}^{n_p} y_{ip}$	$\bar{y} = \frac{\sum_{i=1}^{n_p} y_{ij}}{n_p}$
		$N = \sum_{j=1}^{p} n_j$	$\sum_{j=1}^{p}\sum_{i=1}^{n_j} y_{ij}$	$\bar{\bar{y}} = \frac{\sum_{j=1}^{p}\sum_{i=1}^{n_j} y_{ij}}{N}$

$D_{общ} = D_{внутригр} + D_{межгр}$ или

$$\sum_{j=1}^{P}\sum_{i=1}^{nj}(y_{ij}-\bar{y}) = \sum_{j=1}^{P}\sum_{i=1}^{nj}(y_{ij}-\bar{y}_j+\bar{y}_j-y)^2 = \sum_{j=1}^{P}\sum_{i-1}^{nj}(y_{ij}-\bar{y}_j)^2+\sum_{j=1}^{P}\sum_{i=1}^{nj}(\bar{y}_j-\bar{y})^2$$

Оценки дисперсий: $\frac{D_{общ}}{N-1} = S^2_{общ}$; $\frac{D_{внутригр}}{N-P} = S^2_{внутригр}$; $\frac{D_{межгр}}{P-1} = S^2_{межгр}$;

Проверка гипотезы об отсутствии влияния неколичественного фактора на результирующий показатель Y.

H_0: σ^2межгр=σ^2внутригр.; H_1: σ^2межгр>σ^2внутригр.

Для проверки строится статстика, имеющая распределение Фишера $F = \frac{S^2_{межгр}}{S^2_{внутригр}}$ Если

Fрасч>Fтабл,α,ν$_1$,ν$_2$ (где $\nu_1=P-1$, $\nu_2=N-P$), то нулевая гипотеза отвергается с уровнем значимости α и с вероятностью, равной *p=1-α,* делается вывод о существенности влияния данного качественного признака на результирующий показатель

Результаты процедуры однофакторного дисперсионного анализа

Источник вариации	Сумма квадратов	Число степеней свободы	Дисперсия (или средние квадраты)	Критерий F
Между градациями (влияние качественного признака)	$D_{межгр} = \sum_{j=1}^{P}(\bar{y}_j - \bar{y})^2$	P-1	$S^2_{межгр} = \frac{D_{межгр}}{P-1}$	
Ошибки (влияние неучтенных факторов)	$D_{внутригр} = \sum_{j=1}^{P}\sum_{i=1}^{nj}(y_{ij} - \bar{y}_j)^2$	N-P	$S^2_{внутригр} = \frac{D_{внутригр}}{N-P}$	
«Полная» сумма квадратов	$D_{общ} = \sum_{j=1}^{P}\sum_{i=1}^{nj}(y_{ij} - \bar{y})^2$	N-1		$F = \frac{S^2_{межгр}}{S^2_{внутр}}$

Модель двухфакторного дисперсионного анализа:

$y_{ijk} = \bar{y} + \alpha_i + \beta_j + \theta_{ij} + \varepsilon_{ijk}$, *i=1,P; j=1,Q; k=1,n.*

Результаты наблюдений для полного двухфакторного ДА:

Градации фактора A	Градации фактора B				Средние
	B_1	B_2	B_j	B_q	
	y_{111}, $y_{112},\ldots,y_{11n}$	y_{121}, $y_{122},\ldots,y_{12n}$	y_{1j1}, $y_{1j2},\ldots,$ y_{1jn}	y_{1q1}, $y_{1q2},\ldots,y_{1qn}$	$\bar{y}_1 = \dfrac{\sum_{j=1}^{Q}\sum_{k=1}^{n} y_{1jk}}{Q \cdot n}$
A_2	y_{211}, $y_{212},\ldots,y_{21n}$	y_{221}, $y_{222},\ldots,y_{22n}$	y_{2j1}, $y_{2j2},\ldots,$ y_{2jn}	y_{2q1}, $y_{2q2},\ldots,y_{2qn}$	$\bar{y}_2 = \dfrac{\sum_{j=1}^{Q}\sum_{k=1}^{n} y_{2jk}}{Q \cdot n}$
…	…	…	…	…	…
A_i	y_{i11}, $y_{i12},\ldots,$ yi1n	y_{i21} $y_{i22},\ldots,y_{i2n}$	y_{ij1}, $y_{ij2},\ldots,$ y_{ijn}	y_{iq1}, $y_{iq2},\ldots,y_{iqn}$	$\bar{y}_i = \dfrac{\sum_{j=1}^{Q}\sum_{k=1}^{n} y_{ijk}}{Q \cdot n}$
…	…	…	…	…	…
A_P	y_{p11}, $y_{p12},\ldots,y_{p1n}$	y_{p21}, $y_{P22},\ldots,y_{P2n}$	y_{Pj1}, $y_{Pj2},\ldots,$ y_{Pjn}	y_{pq1}, $y_{pq2},\ldots,y_{PQn}$	$\bar{y}_p = \dfrac{\sum_{j=1}^{Q}\sum_{k=1}^{n} y_{pjk}}{Q \cdot n}$
Средние	$\bar{y}_1 = \dfrac{\sum_{i=1}^{P}\sum_{k=1}^{n} y_{i1k}}{P \cdot n}$	$\bar{y}_2 = \dfrac{\sum_{i=1}^{P}\sum_{k=1}^{n} y_{i2k}}{P \cdot n}$	$\bar{y}j = \dfrac{\sum_{i=1}^{P}\sum_{k=1}^{n} y_{ijk}}{P \cdot n}$	$\bar{y}q = \dfrac{\sum_{i=1}^{p}\sum_{k=1}^{n} y}{P \cdot n}$	***Общее среднее*** $\bar{y} = \dfrac{\sum_{i=1}^{p}\sum_{j=1}^{Q}\sum_{k=1}^{n} y_{ijk}}{P \cdot Q \cdot n}$

Среднее значение для сочетания факторов (i,j) определяется как: $\bar{y}_{ij} = \dfrac{\sum_{k=1}^{n} y_{ijk}}{n}$

Общая сумма квадратов отклонений наблюдений зависимой переменной (Sy) раскладывается на:

- сумму квадратов, обусловленную влиянием фактора A (S_A);
- сумму квадратов, обусловленную влиянием фактора B (S_B);
- сумму квадратов, обусловленную влиянием взаимодействия факторов A и B (S_{AB});
- остаточную сумму квадратов (Sост).

Тогда $Sy = \sum_{i-1}^{P}\sum_{j=1}^{Q}\sum_{k=1}^{n}\left(y_{ijk} - \bar{y}\right)^2$ или

$$\sum_{i=1}^{P}\sum_{j=1}^{Q}\sum_{k=1}^{n}\left(\bar{y}_i - \bar{y}\right)^2 + \sum_{i=1}^{P}\sum_{j=1}^{Q}\sum_{k=1}^{n}\left(\bar{y}_j - \bar{y}\right)^2 + \sum_{i=1}^{P}\sum_{j=1}^{Q}\sum_{k=1}^{n}\left(\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}\right)^2 \text{ или}$$

$$Q \cdot n\sum_{i=1}^{P}\left(\bar{y}_i - \bar{y}\right)^2 + P \cdot n\sum_{j=1}^{Q}\left(\bar{y}_j - \bar{y}\right)^2 + n\sum_{i=1}^{P}\sum_{j=1}^{Q}\left(\bar{y}_{ij} - \bar{y}\right)^2 + \sum_{i=1}^{P}\sum_{j=1}^{Q}\sum_{k=1}^{n}\left(\bar{y}_{ijk} - \bar{y}_{ij}\right)^2$$

Результаты двухфакторного ДА:

Источник изменчивости	Сумма квадратов	Число степеней свободы	«Средние» квадраты (дисперсия зависимой переменной)	F
Фактор A	$S_A = Q \cdot n\sum_{i=1}^{P}\left(\bar{y}_i - \bar{y}\right)$	*P-1*	$D_{A^2} = \dfrac{S_A}{P-1}$	$F_A = \dfrac{S_{A^2}}{S_{ocm^2}}$
Фактор B	$S_B = P \cdot n\sum_{j=1}^{Q}\left(\bar{y}_j - \bar{y}\right)^2$ J	*Q-1*	$D_{B^2} = \dfrac{S_B}{Q-1}$	$F_B = \dfrac{S_{B^2}}{S_{ocm^2}}$
Взаимодействия AB	$S_{AB} = n\sum_{i=1}^{P}\sum_{j=1}^{Q}\left(\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}\right)^2$	$(P-1)\cdot(Q-1)$	$D_{AB^2} = \dfrac{S_{AB}}{(P-1)\cdot(Q-1)}$	$F_{AB} = \dfrac{S_{AB^2}}{S_{ocm^2}}$
Остаточная вариация	$S_{ocm} = \sum_{i=1}^{P}\sum_{j=1}^{Q}\sum_{k=1}^{n}\left(y_{ijk} - \bar{y}_{ij}\right)^2$	*N-PQ*	$D_{ocm^2} = \dfrac{S_{ocm}}{N-PQ}$	-
«Полная» сумма квадратов	$S_y = \sum_{i=1}^{P}\sum_{j=1}^{Q}\sum_{k=1}^{n}\left(y_{ijk} - \bar{y}\right)^2$	*N-1*	-	-

Для степеней свободы выполняется балансовое соотношение:

N-1=(P-1)+(Q-1)+(P-1)(Q-1)+N-PQ

Оценка значимости влияния каждого фактора, а также их взаимодействия на зависимый показатель:

H0: все $\alpha i=0$ (тогда $\sigma^2_A=\sigma^2 ocm$);
H0: все $\beta j=0$ (тогда $\sigma^2_B=\sigma^2 ocm$);
H0: все $\theta ij=0$ (тогда $\sigma^2_{AB}=\sigma^2 ocm$).
Если F_Aрасч > F_Aтабл(α, ν_1=P-1, ν_2=N-PQ);
F_Bрасч > F_Bтабл(α, ν_1=Q-1, ν_2=N-PQ);
F_{AB}расч > F_{AB}табл(α, $\nu_1=(P-1)\cdot(Q-1)$, ν_2=N-PQ);

то нулевые гипотезы отвергаются и делается вывод о существенности влияния факторов (либо их взаимодействия) на зависимый показатель.

Оценки главных эффектов и взаимодействия факторов в модели двухфакторного ДА равны:

$a_i = \bar{y}_i - \bar{y}$, где i=1, P;

$b_j = \bar{y}_j - \bar{y}$, где j=1,Q;

$c_{ij} = \bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}$, где i=1,P; j=1,Q

ПРИЛОЖЕНИЯ (фрагменты основных статистических таблиц)

1. Распределение t-Стьюдента

Границы t-распределения с ν степенями свободы $t_{\alpha,\nu}$

α ν	0.1	0.05	0.01
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.043
6	1.953	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
21	1.721	2.080	2.831
22	1.717	2.074	2.819
23	1.714	2.069	2.807
24	1.711	2.064	2.797
25	1.708	2.060	2.787
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.697	2.042	2.750
40	1.684	2.021	2.704
60	1.671	2.000	2.660
120	1.658	1.980	2.617
∞	1.645	1.960	2.576

2. Критические значения (dн и dв) статистики Дарбина-Уотсона для уровня значимости α=0.05

	m=2		m=3		m=4		m=5		m=6	
N	dн	Dв	Dн	dв	dн	dв	Dн	dв	dн	dв
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

3. Распределение Фишера – Снедекора (F-распределение)

Значения Fтабл, удовлетворяющие условию P(F> Fтабл) для α=0.05, ν1 – число степеней свободы числителя, ν2 – знаменателя.

ν1 ν2	1	2	3	4	5	6	8	12	24	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	5,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,99
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,74	3,34	3,26	3,11	2,85	2,70	2,53	2,31	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,82
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,03	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,39
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,03
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

5. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ КУРСА

Список литературы

Основной

1. Доугерти Кр. Введение в эконометрику. / Пер. с англ. М., ИНФРА-М, 1997.
2. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. М., Дело, 1997, 2000.
3. Лапо В.Ф. Теория вероятностей, математическая статистика, эконометрика. Учебное пособие / Кн.2, КрасГУ, 1999.

Дополнительной

4. Замков О.О., Толстопятенко А.В., Черемных Ю.Н. Математические методы в экономике. М., ДиС, 1 изд.-1997, 2-изд.-1999.
5. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М., ЮНИТИ, 1998.
6. Высшая математика для экономистов /Под ред. Н.Ш. Кремера. – М.: ЮНИТИ, 1998.
7. Головина Л.И. Линейная алгебра и некоторые ее приложения. – М.: Наука, 1985.
8. Джонстон Дж. Эконометрические методы : Пер. с англ. – М.: Статистика, 1980.
9. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистика, 1998.
10. Канторович Г.Г. Эконометрика // Методические материалы по экономическим дисциплинам для преподавателей средних школ и вузов. Экономическая статистика. Эконометрика. Программы, тесты, задачи, решения /Под ред. Л.С. Гребнева . – М.: ГУ-ВШЭ, 2000.
11. Карасев А.И., Кремер Н.Ш., Савельева Т.И. Математические методы и модели в планировании. – М.: Экономика, 1987.
12. Крамер Г. Математические методы статистики: Пер. с англ. – М.: Мир, 1975.
13. Кремер Н.Ш. Теория вероятностей и математическая статистика. –М.: ЮНИТИ, 2000.
14. Кремер Н.Ш. Математическая статистика. – М.: Экономическое образование, 1992.
15. Кремер Н.Ш., Путко Б.А. Эконометрика. – М.: ЮНИТИ-ДАНА, 2002.
16. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере / Под ред. В.Э. Фигурнова. – М.: Инфра-М, 1998.
17. Уотшем Т. Дж., Паррамоу К. Количественные методы в финансах: Пер. с англ.-М.:. ЮНИТИ, 1999.
18. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа: Пер.с нем. – М.: Финансы и статистика, 1982.
19. Четыркин Е.М., Калихман И.Л. Вероятность и статистика –М.: Финансы и статистика, 1982.
20. Эконометрика /Под ред. Н.И. Елисеевой.-М.: Финансы и статистика, 2001.
21. Экономика-математические модели и прикладные модели. / Под ред. В.В. Федосеева. – М.: ЮНИТИ, 1999.

Эконометрика

Евгения Викторовна Зандер

Редактор О.Ф. Александрова

Корректура автора

Подписано в печать 06.02.2003

Тиражируется на электронных носителях

Заказ 221

Дата выхода 13.02.2003

Адрес в Internet: www.lan.krasu.ru/studies/editions.asp

Отдел информационных ресурсов управления информатизации КрасГУ

660041 г. Красноярск, пр. Свободный, 79, ауд. 22-05, e-mail: info@lan.krasu.ru

Издательский центр Красноярского государственного университета

660041 г. Красноярск, пр. Свободный, 79, e-mail: rio@lan.krasu.ru