

**Е.Н.Лукаш, А.А.Туманов**

**ПРАКТИЧЕСКИЕ ЗАДАНИЯ ПО КУРСУ  
«ЭКОНОМЕТРИКА – 1»**

**ДЛЯ СТУДЕНТОВ 3 КУРСА**

**ПЛАН СЕМИНАРСКИХ ЗАНЯТИЙ  
В 5-М СЕМЕСТРЕ 2003/2004**

## ОГЛАВЛЕНИЕ

<b>ОТЧЕТНОСТЬ.....</b>	<b>3</b>
<b>ТРЕБОВАНИЯ К ИТОГОВОЙ КОМПЛЕКСНОЙ ЗАДАЧЕ «МОДЕЛИРОВАНИЕ КОНКРЕТНОГО СЕГМЕНТА РЫНКА» .....</b>	<b>3</b>
<b>ПРИМЕРНОЕ РАСПРЕДЕЛЕНИЕ БАЛЛОВ .....</b>	<b>4</b>
<b>КРИТЕРИИ ОЦЕНОК.....</b>	<b>4</b>
<b>План семинарских занятий.....</b>	<b>5</b>
<b>РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА .....</b>	<b>5</b>
<b>Семинар 1. Линейная модель парной регрессии.....</b>	<b>6</b>
<b>Семинар 2. Компьютерный семинар по модели парной регрессии. ....</b>	<b>7</b>
<b>Семинар 3. Линейная модель парной регрессии.....</b>	<b>9</b>
<b>Семинар 4. Компьютерный семинар по модели множественной регрессии. ....</b>	<b>10</b>
<b>Семинар 5. Модели нелинейной регрессии. Мультиколлинеарность. Фиктивные переменные.....</b>	<b>11</b>
<b>Семинар 6. Компьютерный семинар по темам семинара 5.....</b>	<b>13</b>
<b>Семинар 7. Гетероскедастичность .....</b>	<b>16</b>
<b>Семинар 8. Компьютерный семинар по гетероскедастичности.....</b>	<b>17</b>
<b>Семинар 9. Автокорреляция. ....</b>	<b>20</b>
<b>Семинар 10. Компьютерный семинар по автокорреляции. ....</b>	<b>21</b>
<b>Семинар 11. Модели бинарного выбора. Фиктивные переменные.....</b>	<b>22</b>
<b>Семинар 12. Компьютерный семинар по темам семинара 11.....</b>	<b>23</b>
<b>Семинар 13. Резерв.....</b>	<b>24</b>
<b>Семинар 14. Резерв.....</b>	<b>24</b>

## Отчетность:

1. Шесть домашних компьютерных заданий (сдаются каждые две недели)
2. Домашние задачи по теории (сдаются по выбору преподавателя)
3. Контрольная работа № 1 по моделям парной и множественной регрессии, МНК, ОМНК (октябрь)
4. Контрольная работа № 2 по оставшимся темам курса (декабрь)
5. Итоговая комплексная задача «Моделирование конкретного сегмента рынка»
6. Экзамен по теории

## Требования к итоговой комплексной задаче «Моделирование конкретного сегмента рынка»

1. Данная задача является зачетным курсовым проектом, охватывающим следующие темы:
  - ☐ Линейная модель множественной регрессии
  - ☐ Модели нелинейной регрессии.
  - ☐ Мультиколлинеарность.
  - ☐ Фиктивные переменные.
  - ☐ Гетероскедастичность
2. Проект может выполняться как индивидуально, так и творческим коллективом студентов из одной академической группы, не превышающей 3 человек. Состав коллектива должен быть утвержден у семинариста.
3. Тема проекта выбирается студентами самостоятельно, но в обязательном порядке должна быть согласована с семинаристом.
4. Расчеты должны проводиться по данным, содержащим не менее 200 наблюдений и являющихся «свежими». Информацию для расчетов студенты собирают самостоятельно. В тексте работы обязательно указание источников данных.
5. По желанию студентов, информация и тема работы могут быть предоставлены семинаристом. В этом случае оценка работы будет на 20% ниже, чем при использовании собственных данных.
6. **Сроки:** согласование темы и утверждение состава группы – не позднее семинара №8 (конец октября – начало ноября); сдача проекта – последнее семинарское занятие.
7. **Содержание работы:** в проекте должны быть продемонстрированы знания всех тем, указанных в п.1, приведена постановка задачи и анализ полученных результатов. Работа должна представлять законченное **оригинальное** исследование и должна заканчиваться выводами, а также указанием вклада каждого участника и предполагаемого распределения баллов между ними (в процентах) в случае, если проект выполнялся группой студентов.

Ссылки на литературу и исследования по выбранной теме приветствуются. Данные являются неотъемлемой частью проекта и должны быть сданы в электронном виде вместе с остальными материалами, представляемыми как в электронном, так и в печатном виде. В работе должно содержаться указание на источник информации. С образцом оформления работы можно ознакомиться у семинариста.

8. **Оценивание проекта:** проводится семинаристами в соответствии с указанными выше требованиями. Учитывается соответствие работы указанным в п.1 темам, новизна и оригинальность работы, корректность приводимых в работе расчетов и выводов. Баллы между участниками группы распределяются с учетом их вклада в проект.
9. **Максимальная стоимость проекта** определяется по следующей формуле:  $CT = \min(N, 3) * P$ , где CT – максимальная стоимость проекта, N – количество участников, P – максимальная стоимость на одного студента (утверждается лектором). Если количество участников превышает 3 человека, то при расчете максимальной стоимости проекта количество участников принимается равным 3.
10. **Ответственность за нарушения требований:**
  - a. п. 2, 3, 6 (в части сдачи проекта) и 7 (в части оригинальности исследования) – работа аннулируется и все авторы получают по 0 баллов. В случае выявления одинаковых работ – аннулируются обе работы.
  - b. п. 6 (в части согласования темы и состава группы) – штраф в размере 20% от оценки за проект и выполнение работы по теме, предложенной семинаристом.
  - c. п. 7 (в части предоставления данных) – штраф в размере 30% от оценки за проект.
  - d. п. 7 (в части требований к содержанию и оформлению работы) – штраф в размере до 95% от оценки за проект

### Примерное распределение баллов:

1. Экзамен	35%
2. Проект	35%
3. Домашние задания	20%
4. Контрольные	10%
5. Работа в течение семестра	5%

### Критерии оценок:

Оценки будут выставляться в соответствии с рейтингом студентов.

## План семинарских занятий

Семинар №	Темы семинарских занятий	Кол-во часов	Компьютерный семинар
1	Линейная модель парной регрессии	2	
2	Линейная модель парной регрессии	2	+
3	Линейная модель парной регрессии	2	
4	Линейная модель множественной регрессии	2	+
5	Модели нелинейной регрессии. Мультиколлинеарность. Фиктивные переменные.	2	
6	Модели нелинейной регрессии. Мультиколлинеарность. Фиктивные переменные.	2	+
7	Гетероскедастичность	2	
8	Гетероскедастичность	2	+
9	Автокорреляция	2	
10	Автокорреляция	2	+
11	Модели бинарного выбора. Фиктивные переменные.	2	
12	Модели бинарного выбора. Фиктивные переменные	2	+
13		2	
14		2	+

## Рекомендуемая литература:

1. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. – М.: Дело, 2000.
2. Доугерти К. Введение в эконометрику. - М.: ИНФРА-М, 1997.
3. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998.
4. Green, William. 2003. *Econometric analysis*. 5<sup>th</sup> Ed. - Prentice Hall.
5. Verbeek, Marno. 2000. *A Guide to Modern Econometrics*. – Wiley.
6. Молчанов И.Н., Герасимова И.А. Компьютерный практикум по начальному курсу эконометрики (реализация на Eviews). Практикум – Ростов-на-Дону, 2001.  
[HTTP://MOLCHANOV.NAROD.RU/ECONOMETRICS.HTML](http://molchanov.narod.ru/econometrics.html)
7. Johnson, R. R. *A Guide to Using EViews with Using Econometrics: A Practical Guide*.
8. Lee, Byung-Joo. 2000. *Hitchhiker's Guide to EViews and Econometrics*.
9. Schmidheiny, Kurt. 2001. *A Quick Guide to Eviews 3.1*.

## Семинар 1. Линейная модель парной регрессии.

- Ковариация. Свойства ковариации. Доказательства свойств. (Доугерти. Глава 1)
- МНК: алгебраический подход. Вывод формул для оценок  $\alpha$  и  $\beta$ .
- *Геометрическая картинка для МНК:  $n$ -мерное пространство, проекция вектора  $(y_1, \dots, y_n)^T$  на плоскость векторов  $(x_1, \dots, x_n)^T$  и  $(1, \dots, 1)^T$ .*
- Интерпретация коэффициентов.
- Свойства оценок. Несмещенность, состоятельность.

### Задачи для разбора на семинаре:

- показать эквивалентность оценок в матричном  $[(X^T X)^{-1} X^T Y, \sigma^2 (X^T X)^{-1}]$  и явном  $[\text{Cov}(x, y)/\text{Var}(x), \sigma^2/(n \text{Var}(x))]$  через  $\sum x_i$  и  $y_i$  видах.
- показать, что  $\sum e_i = 0, \sum e_i x_i = 0$ .
- показать несмещенность и состоятельность оценок: вычислить математическое ожидание оценок коэффициентов и их дисперсию.
- проходит ли регрессия через точку средних значений  $\bar{x}$  и  $\bar{y}$ . А в случае отсутствия в модели константы?

### Домашнее задание:

1. Как изменится качество оценки модели парной регрессии, если вместо переменной  $X$  будет использоваться переменная  $Z = c * X$ , где  $c$  – константа?
2. Пусть в модели  $y_i = \alpha + \beta x_i + \varepsilon_i, i = 1, \dots, n$  ошибки  $\varepsilon$  удовлетворяют классическим условиям. Рассматривается следующая оценка коэффициента  $\beta$ :

$$\tilde{\beta} = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \bar{y}}{x_i - \bar{x}}$$

- а) Является ли оценка  $\tilde{\beta}$  несмещенной? Является ли она линейной (по  $y$ )? *Подсказка: показать, что ее можно представить в виде  $\tilde{\beta} = \sum c_i y_i$ .*

- б) Вычислите дисперсию оценки  $\tilde{\beta}$ . *Ответ:  $V(\tilde{\beta}) = \sigma^2 \sum_{i=1}^n \frac{1}{n^2} \left( \frac{1}{x_i} - \frac{1}{n} \sum_{s=1}^n \frac{1}{x_s} \right)^2$*

- в) Проверьте теорему Гаусса-Маркова, сравнив полученную дисперсию оценки  $\tilde{\beta}$  с дисперсией МНК-оценки  $\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$ . *Подсказка: воспользуйтесь неравенством Коши-*

*Буняковского* -  $\sum a_i^2 \cdot \sum b_i^2 \geq \left( \sum (a_i \cdot b_i) \right)^2$

## Семинар 2. Компьютерный семинар по модели парной регрессии.

**Пример 1. Генерация двумерных выборок. Оценка коэффициентов модели парной линейной регрессии.** Оценить модель парной линейной регрессии зависимости  $y$  от  $x$  с использованием EXCEL.

Модель:  $y_i = 20 - 3x_i + \varepsilon_i, i=1, \dots, n, n=200$

Объясняющая переменная  $x$  - детерминированная (неслучайная), принимает значения 1, 2, 3 ... 10 по 20 раз каждое ( $n = 200$ ).

Ошибки  $\varepsilon_i$  – независимые, нормально распределенные случайные величины с параметрами  $a=0$  и  $\sigma^2=1$

### I. Генерация двумерных выборок.

1) Ввод значений переменной  $x$  в столбец А (ячейки А2–А201): в ячейку А2 ввести 1 и скопировать ее в А3–А21, затем ячейку А22 с формулой  $=A2 + 1$  копировать вниз (в ячейки А23 по А201)

2) Ввод значений «истинной» регрессии:  $y_i = 20 - 3x_i$  в ячейки В2–В201: в ячейку В2 ввести формулу  $=20+3*A2$  и копировать ее в ячейки В3–В201

3) Генерация ошибок наблюдений:

а) Предварительно решить две вспомогательные задачи:

1. Если случайная величина  $\xi$  имеет непрерывную функцию распределения  $y=F(x), x \in R^1$ , то случайная величина  $\eta=F(\xi)$  имеет равномерное распределение на отрезке  $[0,1]$

2. Пусть  $\eta$  – равномерная на  $[0,1]$  случайная величина, а  $\Phi(x) = \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$  – функция

распределения стандартной нормальной величины  $N(0,1)$ . Тогда случайная величина  $\xi=\Phi^{-1}(\eta)$  имеет стандартное нормальное распределение.

б) в ячейках С2–С201 получить независимые реализации случайной величины  $u$ , имеющей равномерное распределение на отрезке  $[0,1]$ : формула  $=случис()$  из ячейки С2 копируется вниз

в) в ячейках D2–D201 получить реализации ошибок наблюдений  $\varepsilon_i$ , имеющих  $(0,1)$ –нормальное распределение: формулу  $=нормобр(С2;0;1)$  копировать из ячейки D2 вниз

4) Генерация «наблюденных» (фактических)  $y$  [ $= 20 - 3x + e$ ]

- Заморозить  $y$  (формулы перевести в числа): копирование и специальная вставка-значения

- Из таблицы убрать лишнее (все кроме столбцов  $x$  и  $y$ ).

## II. Оценка коэффициентов парной линейной регрессии.

Исходя из выборки  $\{(x_i, y_i), i=1, \dots, n\}$  восстановить (оценить) коэффициенты  $\alpha=20$  и  $\beta=-3$ , значения которых нам неизвестны (мы их забыли). Используем формулы для МНК-оценок.

*Что должно быть посчитано:*

- коэффициент парной корреляции (посчитанный «вручную»)
- оценка коэффициентов через формулы ковариации
- оценка дисперсии ошибок и дисперсий оценок коэффициентов ( $\sigma^2, \hat{V}(\alpha), \hat{V}(\beta)$ ).
- коэффициент детерминации  $R^2$ .
- Нарисовать график облака и линии регрессии, используя прогнозные значения модели (*график – точечная – с отрезками – двойной щелчок на линии фактических данных – линия отсутствует*).
- проверить, что результаты расчетов коэффициентов, сделанные «вручную», будут такими же, если использовать встроенные функции Excel.
- можно экспортировать ряды  $x, y$  в Eviews и построить регрессию.

**Домашнее задание 1.** Должно быть сдано через **2 недели на Семинаре 4.**

**Задача 1.** В Excel подготовить таблицу, в которой параметры уравнения  $(\alpha, \beta, \sigma^2)$  вынесены в отдельные ячейки для возможности их изменения. Имитацией убедиться в том, что при разных истинных значениях параметров оценки получаются близкими к ним. Ответить на вопрос «Как влияет изменение  $\sigma^2$  на точность оценивания, т.е. на оценки дисперсии оценок?». Дать интерпретацию коэффициентам регрессии. Решение оформить в текстовом файле: постановка задачи, результаты, графики, выводы.

**Вариант 1.**  $X$  – детерминированная (неслучайная) величина, принимающая значения 5, 6, 7 ... 14 по 20 раз каждое ( $n = 200$ )

**Вариант 2.**  $X$  – стохастический регрессор (случайный). Равномерная величина на отрезке  $[6; 14]$ . Проверьте, коррелированы ли случайный регрессор  $x$  и случайная ошибка  $\varepsilon$ .

**Вариант 3.**  $X$  – стохастический регрессор (случайный). Нормальная величина  $N(10, 15)$ . Проверьте, коррелированы ли случайный регрессор  $x$  и случайная ошибка  $\varepsilon$ .

**Задача 2.** К данным, сформированным в задаче 1, применить (заведомо неправильную) модель вида  $y = \beta x + \varepsilon$ . Рассчитать коэффициент детерминации двумя способами ( $\frac{RSS}{TSS}$  и  $1 - \frac{ESS}{TSS}$ ). Объяснить полученные результаты.



### Семинар 3. Линейная модель парной регрессии.

- Мера качества подгонки (коэффициент детерминации  $R^2$ ). (Мы используем обозначения из книги Магнуса и др. (в Доугерти обозначения противоположные)).
- Свойства коэффициента детерминации. Крайние случаи:  $R^2=0$ ,  $R^2=1$ . Связь между
  - а) коэффициентом детерминации  $R^2$  и парным коэффициентом корреляции  $\hat{r}_{xy}$
  - б) коэффициентом детерминации  $R^2$  и парным коэффициентом корреляции  $\hat{r}_{yy}$
- Выражение углового коэффициента регрессии через коэффициент корреляции.
- Теорема Гаусса-Маркова.
- Проверка гипотез.
- F-тест на качество оценивания

#### Задачи для разбора на семинаре:

Существуют ли другие методы оценивания коэффициентов кроме МНК?

Рассмотрим следующие методы (по очереди):

- **Метод крайних точек.** Упорядочим имеющиеся наблюдения по  $x$ . Возьмем минимальный ( $x_{\min}$ ) и максимальный ( $x_{\max}$ )  $x$  и соответствующие им  $y$  ( $y_{\min}$  и  $y_{\max}$ ). Проведем через эти точки прямую. Оценка углового коэффициента будет иметь вид:  $\beta = \frac{y_{\max} - y_{\min}}{x_{\max} - x_{\min}}$ . Чем она плоха? А чем хороша? Проверить, выполняются ли свойства несмещенности, состоятельности. Вычислить дисперсию оценки и сравнить ее с дисперсией оценки МНК.
- **Метод крайних групп (Бартлетт).** Пусть число наблюдений  $n$  кратно 3. Упорядочим имеющиеся наблюдения по  $x$ . Возьмем крайние группы наблюдений, а в них – средние ( $\bar{x}_L$  и  $\bar{x}_H$ ). Через эти две точки проводим прямую и берем тангенс наклона как оценку  $\beta$ . Оценка будет иметь вид:  $\beta = \frac{\bar{y}_H - \bar{y}_L}{\bar{x}_H - \bar{x}_L}$ . Чем она плоха? А чем хороша? Проверить, выполняются ли свойства несмещенности, состоятельности. Вычислить дисперсию оценки и сравнить ее с дисперсией оценки МНК.
- Предположим, что модель  $y_i = \alpha + \beta x_i + \varepsilon_i$ ,  $i = 1, \dots, n$  удовлетворяет условиям классической регрессии. Рассматривается следующая оценка коэффициента  $\beta$ :

$$\beta = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \bar{y}}{x_i - \bar{x}}$$

- а) Является ли оценка  $\beta$  несмещенной? Является ли она линейной?

б) Вычислите дисперсию оценки  $\beta$ .

в) Проверьте теорему Гаусса-Маркова, сравнив полученную дисперсию оценки  $\beta$

с дисперсией МНК-оценки  $\frac{\sigma^2}{\sum_{t=1}^n (x_t - \bar{x})^2}$

## Семинар 4. Компьютерный семинар по модели множественной регрессии.

**Сдать сделанное задание 1.**

**Статистический пакет:** Eviews 3.1

**Пример 2: cheburek.xls**

После финансового кризиса спрос на чебуреки упал, и менеджер был вынужден тратить часть средств на рекламу. Для изучения зависимости объема продаж от цены и расходов на рекламу менеджер использует следующую модель:

$$q_t = \beta_1 + \beta_2 p_t + \beta_3 a_t + \varepsilon_t$$

В файле приведены данные наблюдений за 20 недель ( $t$  – номер недели,  $q_t$  – количество проданных чебуреков,  $p_t$  – цена одного чебурека,  $a_t$  – количество рекламных объявлений (например, количество минут на телевидении в неделю) (цена одного объявления – 100 рублей)).

### Задания:

1. Ввод данных в Eviews. Файл должен быть сохранен в формате Excel 95.

- Создать новый рабочий файл в памяти *File – New – Workfile – Undated – (End observation = 20) – OK*
- Импорт данных в рабочий файл *File – Import – (Read Text-Lotus-Excel) – (выберите файл cheburek.xls) – Upper left data cell = A2, By observations, Number of series = 4, OK*

2. Какие знаки коэффициентов вы ожидаете получить? Интерпретация коэффициентов.

3. Найти оценки коэффициентов регрессии. Метод наименьших квадратов.

- *ls q c p adv*

Dependent Variable: Q  
Method: Least Squares  
Sample: 1 20  
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1247.343	191.8093	6.503033	0.0000
P	-65.00053	32.01886	-2.030070	0.0583
ADV	-76.39330	16.74429	-4.562350	0.0003
R-squared	0.654514	Mean dependent var	555.2000	
Adjusted R-squared	0.613869	S.D. dependent var	143.4274	

4. Как устроена таблица с результатами расчетов?

5. Чем уравнение хорошо? Чем плохо? Соответствует ли оно вашим ожиданиям?
6. Проверка коэффициентов на значимость.
7. Устраивает ли Вас с содержательной точки зрения предположение о линейности связи между расходами на рекламу и объемом продаж.
8. Рассмотрим модифицированную модель:  $q_t = \beta_1 + \beta_2 p_t + \beta_3 a_t + \beta_4 a^2 + \varepsilon_t$ . Какой знак ожидается при переменной  $a^2$ ?

- создание новой переменной *adv2*: *genr adv2=adv^2*
- *ls q c p adv adv2*

Dependent Variable: Q  
Method: Least Squares  
Sample: 1 20  
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	957.3348	129.2183	7.408663	0.0000
P	-110.1528	21.32749	-5.164828	0.0001
ADV	255.5294	62.22484	4.106550	0.0008
ADV2	-43.37607	8.020249	-5.408319	0.0001
R-squared	0.877839	Mean dependent var		555.2000
Adjusted R-squared	0.854934	S.D. dependent var		143.4274

9. Как изменились оценки коэффициентов? Что вы можете сказать о модели?
10. Найти оптимальную цену чебуреков при объемах рекламы в 2.8 объявлений при условии, что себестоимость производства одного чебурека равна 2 руб.

**Домашнее задание 2.** Должно быть сдано через 2 недели на Семинаре 6.

## Семинар 5. Модели нелинейной регрессии. Мультиколлинеарность. Фиктивные переменные.

- Линеаризация. Преобразования переменных, приводящие к линейности модели.
- Логарифмическая, полул로그арифмическая и транслогарифмическая модели.
- Интерпретация коэффициентов модели.
- Границы применимости коэффициента детерминации. Цель исследования: прогнозирование или выявление связи ( $R^2$  не обязательно должен быть высоким). Невозможность сравнений моделей по коэффициенту детерминации при разных зависимых переменных. ( $y$  и  $\log y$ ).
- Скорректированный коэффициент детерминации. Аналоги данного коэффициента: информационные критерии AIC и BIC.
- Мультиколлинеарность.

- Фиктивные переменные. Интерпретация коэффициентов. F-тест на ограничения модели. Тест Чоу.

### Задачи для разбора на семинаре:

- $y = AK^\alpha L^\beta$  (Формула Кобба – Дугласа). Коэффициенты – эластичность.
- $y = Ae^{\beta x}$  (Интерпретация  $\beta$  – формула). Коэффициент – полуэластичность.
- **Задача 5.3.** Предположим, что вы оцениваете линейную функцию потребления  $c_t = \alpha + \beta y_t + \varepsilon_t$ . Как учесть возможный сдвиг этой функции при переходе от городского к сельскому потреблению, если вы считаете, что предельная склонность к потреблению постоянна, в то время как средняя склонность к потреблению может меняться? Как проверить гипотезу о том, что предельные склонности к потреблению индивидуумов с доходом выше и ниже уровня  $y^*$  различаются?
- **Задача 5.4.** На основе квартальных данных с 1971 по 1976 г. с помощью МНК получено следующее уравнение (в скобках указаны стандартные отклонения):

$$y_t = 1.12 - 0.0098x_t^1 - 5.62x_t^2 + 0.044x_t^3$$

(2.14) (0.0034) (3.42) (0.009)

RSS = 110,32; ESS = 21,43

- Проверьте значимость одного из коэффициентов.
- Были раздельно проведены две регрессии на основе данных: 1-й квартал 1971 г. – 1-й квартал 1975 г. и 2-й квартал 1975 г. – 4-й квартал 1976 г. Соответственно получены следующие значения сумм квадратов остатков:  $ESS_1 = 12.25$ ,  $ESS_2 = 2.32$ . Проверьте вашу гипотезу о том, что между 1-м и 2-м кварталами 1975 г. произошло структурное изменение.

### Домашнее задание

- На основании условий задачи 5.4:
  - Проверить значимость каждого из коэффициентов
  - Найти коэффициент детерминации
  - Протестировать значимость регрессии в целом
  - Когда в уравнение были добавлены три фиктивные переменные, соответствующие трем первым кварталам года, величина RSS выросла до 118.20. Проверить гипотезу о наличии сезонности, сформулировав необходимые предположения о виде этой сезонности.

## Семинар 6. Компьютерный семинар по темам семинара 5.

Сдать сделанное задание 2.

**Пример 3.** Стоимость колготок в Московских оптовых торговых фирмах, осень 1997 г.

Файл **tights.wf1** (n = 74)

*Переменные:*

<b>N</b> - номер по порядку.	<b>cotton</b> - % хлопка
<b>Price</b> - цена колготок в рублях 1997 г.	<b>wool</b> - % шерсти
<b>DEN</b> - плотность в DEN.	<b>firm</b> - фирма-производитель:
<b>polyamid</b> - % полиамида.	0 - Levante, 1 - Golden Lady
<b>lykra</b> - % лайкры.	

Вопрос: Какая из фирм устанавливает более высокие цены и на сколько?

1. На основании анализа средних значений по группам скажите, различаются ли цены на колготки у разных производителей? Какие недостатки такого подхода?
2. Постройте уравнение зависимости цены колготок от их плотности, состава и производителя.

*LS PRICE C COTTON DEN FIRM LYKRA POLYAMID WOOL*

Dependent Variable: PRICE

Method: Least Squares

Sample: 1 74

Included observations: 74

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	15525.48	63702.90	0.243717	0.8082
COTTON	-51.46878	630.5319	-0.081628	0.9352
DEN	180.0664	39.88580	4.514549	0.0000
FIRM	1300.192	1468.435	0.885427	0.3791
LYKRA	208.1137	583.4898	0.356671	0.7225
POLYAMID	-109.1990	645.0302	-0.169293	0.8661
WOOL	58.29509	1245.402	0.046808	0.9628
R-squared	0.496571	Mean dependent var	15841.89	
Adjusted R-squared	0.451488	S.D. dependent var	7629.898	

3. Можно ли считать, что выборки колготок разных производителей различаются? Проведите тест Чоу.

- a. Отсортируем наблюдения по переменной «фирма»:

*SORT FIRM*

- b. Оценим регрессию без переменной «фирма»:

*LS PRICE C COTTON DEN FIRM LYKRA WOOL*

- c. Выполним тест Чоу:

*View – Stability Tests – Chow Breakpoint Test*

d. Укажем первое наблюдение, относящееся ко второй подвыборке:

41 - OK

e. Результаты теста отвергает гипотезу о наличие разрыва при 10% уровне значимости

Chow Breakpoint Test: 41

F-statistic	0.789353	Probability	0.561228
Log likelihood ratio	4.428270	Probability	0.489538

4. Как можно объяснить большое количество незначимых переменных в уравнении из п.2? Может быть, существует мультиколлинеарность? Как правило, состав колготок равняется 100%? Для проверки этого создадим новую переменную, представляющую собой остаток:

$GENR\ REST = 100 - COTTON - LYKRA - POLYAMID - WOOL$

Посмотрим состав данных: двойной щелчок на переменной *rest* – *view* – *descriptive statistics* – *histogram and stats*

«Неправильные» точки: можно включить в уравнение новую переменную («золотая нить») или выкинуть эти наблюдения (ошибка при вводе). Оценивать по двум точкам бессмысленно, поэтому выкинем их:

$sample - IF\ rest = 0$

5. Снова построим уравнение регрессии, исключив из нее полиамид как наиболее «массовую» и дешевую составляющую:

$LS\ PRICE\ C\ COTTON\ DEN\ FIRM\ LYKRA\ WOOL$

Dependent Variable: PRICE  
Method: Least Squares  
Sample: 1 74 IF REST=0  
Included observations: 72

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3102.770	2323.322	1.335489	0.1863
COTTON	66.17462	71.77907	0.921921	0.3599
DEN	180.3160	39.36984	4.580055	0.0000
FIRM	1840.025	1485.296	1.238827	0.2198
LYKRA	424.3863	110.8591	3.828160	0.0003
WOOL	-659.9838	783.2637	-0.842607	0.4025
R-squared	0.458383	Mean dependent var	15561.11	
Adjusted R-squared	0.417351	S.D. dependent var	7307.194	

6. Можно ли что-то еще сделать для улучшения качества уравнения? Посмотрим на данные поподробнее. По шерсти только два наблюдения не равны 0. Скорее всего – это другой вид колготок, который описывается другой моделью, поэтому исключим их. ( $wool = 0$ ).

$LS\ PRICE\ C\ COTTON\ DEN\ FIRM\ LYKRA\ WOOL$

Результат – чистая мультиколлинеарность факторов, поэтому шерсть надо исключить.

*LS PRICE C COTTON DEN FIRM LYKRA*

Dependent Variable: PRICE  
Method: Least Squares  
Sample: 1 74 IF REST=0 AND WOOL=0  
Included observations: 70

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3102.770	2333.105	1.329889	0.1882
COTTON	66.17462	72.08131	0.918055	0.3620
DEN	180.3160	39.53562	4.560850	0.0000
FIRM	1840.025	1491.550	1.233633	0.2218
LYKRA	424.3863	111.3259	3.812109	0.0003
R-squared	0.409964	Mean dependent var	15204.29	
Adjusted R-squared	0.373654	S.D. dependent var	7077.366	

7. Коэффициент при переменной *FIRM* по-прежнему незначим, а его значимость – наша цель. Посмотрим на переменную *COTTON*. Только 3 наблюдения принимают гораздо более высокие значения. Скорее всего, это совершенно другой вид колготок. Исключаем их:  $COTTON < 40$ .

*LS PRICE C COTTON DEN FIRM LYKRA*

Dependent Variable: PRICE  
Method: Least Squares  
Sample: 1 74 IF REST=0 AND WOOL=0 AND COTTON<40  
Included observations: 67

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2728.327	2112.034	1.291801	0.2012
COTTON	2203.028	521.8824	4.221311	0.0001
DEN	166.2228	35.76868	4.647162	0.0000
FIRM	2566.686	1373.875	1.868209	0.0665
LYKRA	329.3805	102.7272	3.206360	0.0021
R-squared	0.500748	Mean dependent var	14856.72	
Adjusted R-squared	0.468538	S.D. dependent var	6908.103	

8. Коэффициент при фиктивной переменной оказывается значимым, но на уровне 6.65%, а хотелось бы не больше 5%. Тем не менее можно сделать предварительные выводы.
9. Для улучшения модели изменим функциональную форму модели на полулогарифмическую ( $\log(PRICE)$ ). Объяснение причин изменение формы зависимости.

*LS LOG(PRICE) C COTTON DEN FIRM LYKRA*

Dependent Variable: LOG(PRICE)  
Method: Least Squares  
Sample: 1 74 IF REST=0 AND WOOL=0 AND COTTON<40  
Included observations: 67

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.445910	0.131615	64.17123	0.0000
COTTON	0.148974	0.032522	4.580716	0.0000
DEN	0.011582	0.002229	5.196298	0.0000
FIRM	0.286642	0.085616	3.348012	0.0014
LYKRA	0.035277	0.006402	5.510641	0.0000
R-squared	0.608580	Mean dependent var	9.496674	
Adjusted R-squared	0.583327	S.D. dependent var	0.486186	

10. Интерпретация коэффициентов. С помощью построенной модели ответьте на вопрос, верно ли, что цены колготок двух фирм-производителей различаются статистически достоверно. Какая из фирм устанавливает более высокие цены?

**Домашнее задание 3.** Должно быть сдано через 2 недели на Семинаре 8.

## Семинар 7. Гетероскедастичность

- Смешанное влияние переменных. Интерпретация коэффициента при переменной «балкон» в логлинейной модели. Почему нельзя в этой модели пользоваться произведением?
- Тест ранговой корреляции Спирмена (см. стр. 206, Доугерти).
- Тест Голдфелда-Квандта.
- Тест Уайта.
- Устранение гетероскедастичности. Корректировка Уайта. Преобразование переменных.

### Задачи для разбора на семинаре:

- Показать, что МНК на преобразованных данных эквивалентно ВЗМНК на исходных данных.
- По имеющимся данным о 30 семьях, разбитых на 8 доходных групп ( $X_j$ ) разной численности ( $n_j$ ), и среднего потребления товаров в каждой группе ( $\bar{Y}_j$ ), оценить предельную склонность к потреблению и построить доверительный интервал.  
Указания: рассмотрите парную модель линейной регрессии  $y_j = \alpha + \beta x_j + \varepsilon_j$ .  
Какими свойствами обладают остатки? (Рассмотрите «нереализуемую» модель вида  $y_i = \alpha + \beta x_i + u_i$ ,  $i = 1, 2, \dots, 30$ .)

j	$\bar{Y}_j$	$X_j$	$n_j$
1	610	300	8
2	1360	500	7
3	2450	700	5
4	4110	900	4
5	5780	1100	2
6	6200	1400	2
7	7700	1800	1
8	9030	2000	1

**Домашнее задание 4а.** Должно быть сдано через 3 недели на Семинаре 10.

- Разобраться с тестом Глейзера. (стр. 208, Доугерти).



- Реализовать тест Спирмена и Голдфелда-Квандта на примере из Доугерти (стр. 207).
- Файл **credits.xls** содержит данные об уровнях запасов  $I$ , объемах продаж  $S$  (млн. долл.) и процентной ставки по кредитам  $R$  в 35 фирмах некоторой отрасли. Экономическая интуиция подсказывает, что  $I$  должно быть положительно связано с  $S$  и отрицательно с  $R$ .
  - a. Проведите регрессию  $I$  на  $S$  и  $R$  и тесты на гетероскедастичность.
  - b. Если в п. а выявлена гетероскедастичность, осуществите коррекцию на гетероскедастичность, предполагая, что дисперсия ошибки пропорциональна  $S^2$ .

## Семинар 8. Компьютерный семинар по гетероскедастичности.

### Сдать сделанное задание 3.

#### Пример 4. Гетероскедастичность. Файл **flat98\_1.wf1**.

Переменные:

*В файле содержатся данные по однокомнатным квартирам в Москве в 1998г.*

<b>N</b> - номер по порядку.	<b>floor</b> – этаж (0,1). 0 – первый/последний, 1 – нет.
<b>Price</b> - цена квартиры, тыс. USD.	<b>brick</b> – категория дома. 1 – кирпичный, 0 – нет.
<b>totsq</b> – общая площадь квартиры, кв. м.	<b>tel</b> – телефон. 0 – нет, 1 – есть.
<b>livsq</b> - жилая площадь квартиры, кв. м.	<b>bal</b> – балкон. 0 – нет, 1 – есть.
<b>kitsq</b> – площадь кухни, кв. м.	<b>metrdist</b> – расстояние пешком до метро, мин.
<b>dist</b> – расстояние до центра, км.	<b>walk</b> – 1- пешком до метро, 0 – иначе.

1. Будем рассматривать только подвыборку g8 (g8=1).
2. Введем остаток площади квартиры – dopsp.
3. Анализ дескриптивных статистик и корреляционных матриц.
4. Рассмотрим простую линейную модель от всех переменных

*LS PRICE C LIVSP DOPSP KITSP DIST WALK METRDIST TEL BRICK BAL FLOOR*

Dependent Variable: PRICE

Method: Least Squares

Sample(adjusted): 1202 1406 IF G8=1

Included observations: 205 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.173893	5.775609	-1.242102	0.2157
LIVSP	1.159641	0.255088	4.546037	0.0000
DOPSP	1.682448	0.174538	9.639409	0.0000
KITSP	1.529205	0.282332	5.416338	0.0000
DIST	-0.993102	0.124227	-7.994262	0.0000
WALK	5.193251	1.049761	4.947079	0.0000
METRDIST	-0.286520	0.111640	-2.566454	0.0110
TEL	-0.208753	1.268045	-0.164626	0.8694
BRICK	2.859552	1.199194	2.384563	0.0181
BAL	3.003189	1.024034	2.932705	0.0038

FLOOR	3.107897	1.162288	2.673946	0.0081
R-squared	0.685033	Mean dependent var	38.63244	
Adjusted R-squared	0.668798	S.D. dependent var	11.05564	

5. Все ли коэффициенты имеют разумную величину? Например, коэффициент при *WALK*?  
Как можно улучшить модель?
6. Рассмотрим модель с включением произведения переменных. Смысл коэффициентов.

*LS PRICE C LIVSP DOPSP KITSP DIST\*TOTSP WALK\*TOTSP METRDIST\*TOTSP  
TEL\*TOTSP BAL\*TOTSP BRICK\*TOTSP FLOOR\*TOTSP*

Dependent Variable: PRICE

Method: Least Squares

Sample(adjusted): 1202 1406 IF G8=1

Included observations: 205 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.084724	5.079738	-1.591563	0.1131
LIVSP	1.199646	0.248687	4.823925	0.0000
DOPSP	1.648613	0.179567	9.181067	0.0000
KITSP	1.577957	0.282328	5.589096	0.0000
DIST*TOTSP	-0.027833	0.003359	-8.285747	0.0000
WALK*TOTSP	0.135932	0.028433	4.780845	0.0000
METRDIST*TOTSP	-0.008595	0.003022	-2.844628	0.0049
TEL*TOTSP	-0.003784	0.034749	-0.108893	0.9134
BAL*TOTSP	0.087132	0.028177	3.092377	0.0023
BRICK*TOTSP	0.089409	0.033018	2.707937	0.0074
FLOOR*TOTSP	0.093357	0.032507	2.871936	0.0045
R-squared	0.704356	Mean dependent var	38.63244	
Adjusted R-squared	0.689116	S.D. dependent var	11.05564	

7. Исключение незначимых коэффициентов.

*LS PRICE C LIVSP DOPSP KITSP DIST\*TOTSP WALK\*TOTSP METRDIST\*TOTSP  
BAL\*TOTSP BRICK\*TOTSP FLOOR\*TOTSP*

Dependent Variable: PRICE

Method: Least Squares

Sample(adjusted): 1202 1406 IF G8=1

Included observations: 205 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.009595	5.019899	-1.595569	0.1122
LIVSP	1.193789	0.242185	4.929252	0.0000
DOPSP	1.647284	0.178697	9.218309	0.0000
KITSP	1.572733	0.277516	5.667177	0.0000
DIST*TOTSP	-0.027880	0.003323	-8.388859	0.0000
WALK*TOTSP	0.136105	0.028316	4.806611	0.0000
METRDIST*TOTSP	-0.008600	0.003014	-2.853822	0.0048
BAL*TOTSP	0.087142	0.028105	3.100584	0.0022
BRICK*TOTSP	0.088714	0.032313	2.745495	0.0066
FLOOR*TOTSP	0.093273	0.032415	2.877464	0.0045
R-squared	0.704338	Mean dependent var	38.63244	
Adjusted R-squared	0.690692	S.D. dependent var	11.05564	

8. Интерпретация коэффициентов.

9. Тест Вальда на равенство коэффициентов. Проверим гипотезу о том, что квадратный метр кухни и коридора стоят одинаково:

*View-Coefficient tests-Wald coefficient restrictions – C(3) = C(4)*

Wald Test:

Null Hypothesis: C(3)=C(4)

F-statistic	0.042924	Probability	0.836084
Chi-square	0.042924	Probability	0.835868

Т.к. вероятность высокая, то гипотеза не отвергается. Значит, можно объединить их в одну переменную.

Можно ли рассмотреть еще какие-то пары коэффициентов? Рассматривать гипотезу о равенстве коэффициентов перед METRDIST и TEL нельзя, т.к. показатели имеют разную размерность и при их возможном объединении теряют всякий смысл.

Можно тестировать сразу несколько гипотез:

10. Пересчет уравнения с объединенными переменными

Dependent Variable: PRICE

Method: Least Squares

Sample(adjusted): 1202 1406 IF G8=1

Included observations: 205 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.064600	5.000619	-1.612720	0.1084
LIVSP	1.186805	0.239241	4.960713	0.0000
DOPSP+KITSP	1.622987	0.134502	12.06664	0.0000
DIST*TOTSP	-0.027730	0.003236	-8.568418	0.0000
WALK*TOTSP	0.135408	0.028047	4.827907	0.0000
METRDIST*TOTSP	-0.008649	0.002997	-2.885773	0.0043
BAL*TOTSP	0.085803	0.027285	3.144677	0.0019
BRICK*TOTSP	0.090734	0.030733	2.952367	0.0035
FLOOR*TOTSP	0.092468	0.032102	2.880395	0.0044
R-squared	0.704273	Mean dependent var	38.63244	
Adjusted R-squared	0.692202	S.D. dependent var	11.05564	

11. Тест Уайта на гетероскедастичность.

*View--Residual Test -- White heteroscedasticity(no cross terms)*

White Heteroskedasticity Test:

F-statistic	5.297923	Probability	0.000000
Obs*R-squared	63.70713	Probability	0.000000

Гомоскедастичность отвергается, т.к. вероятность меньше 5%.

## 12. Коррекция ошибок в форме Уайта.

*Estimate – Options – ☒ Heteroscedasticity Consistent covariances(White) – OK – OK.*

Dependent Variable: PRICE

Method: Least Squares

Sample(adjusted): 1202 1406 IF G8=1

Included observations: 205 after adjusting endpoints

White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.009595	5.362836	-1.493537	0.1369
LIVSP	1.193789	0.244662	4.879336	0.0000
DOPSP	1.647284	0.212577	7.749119	0.0000
KITSP	1.572733	0.330470	4.759081	0.0000
DIST*TOTSP	-0.027880	0.003828	-7.283467	0.0000
WALK*TOTSP	0.136105	0.026574	5.121755	0.0000
METRDIS*TOTSP	-0.008600	0.002842	-3.026353	0.0028
BAL*TOTSP	0.087142	0.029133	2.991174	0.0031
BRICK*TOTSP	0.088714	0.035435	2.503564	0.0131
FLOOR*TOTSP	0.093273	0.033035	2.823466	0.0052
R-squared	0.704338	Mean dependent var	38.63244	
Adjusted R-squared	0.690692	S.D. dependent var	11.05564	

## 13. Проверка гипотез о равенстве коэффициентов.

Dependent Variable: PRICE

Method: Least Squares

Sample(adjusted): 1202 1406 IF G8=1

Included observations: 205 after adjusting endpoints

White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.064600	5.336703	-1.511158	0.1324
LIVSP	1.186805	0.243703	4.869888	0.0000
DOPSP+KITSP	1.622987	0.174646	9.293014	0.0000
DIST*TOTSP	-0.027730	0.003674	-7.547576	0.0000
WALK*TOTSP	0.135408	0.025761	5.256305	0.0000
METRDIS*TOTSP	-0.008649	0.002798	-3.091502	0.0023
BAL*TOTSP	0.085803	0.028660	2.993845	0.0031
BRICK*TOTSP	0.090734	0.033098	2.741352	0.0067
FLOOR*TOTSP	0.092468	0.031127	2.970662	0.0033
R-squared	0.704273	Mean dependent var	38.63244	
Adjusted R-squared	0.692202	S.D. dependent var	11.05564	

## 14. Сравните t- и F- статистики двух моделей.

**Домашнее задание 4b.** Должно быть сдано через 2 недели на Семинаре 10.

**Семинар 9. Автокорреляция.**

- Обнаружение автокорреляции. Статистика Дарбина-Уотсона (для AR(1)).
- Корректировка Ньюи-Уэста для стандартных ошибок МНК-оценок.
- Модель AR(1) для остатков. Коррекционная процедура Кохрейна – Оркатта.
- Автокорреляция с лаговой зависимой переменной. h-статистика Дарбина.

- Автокорреляция, связанная с неправильной спецификацией модели.

#### Задачи для разбора на семинаре:

- задача 7.9 (Доугерти, стр. 226)
- задача 7.10 (Доугерти, стр. 228)
- пример (Доугерти, стр. 231)
- задача 7.12 (Доугерти, стр. 233)
- задача 7.13 (Доугерти, стр. 233)

### Семинар 10. Компьютерный семинар по автокорреляции.

#### Сдать сделанное задание 4.

#### Пример 5. Файл `unemploy.xls`.

Рассмотрим модель, связывающую количество вакансий  $w_t$  и уровень безработицы  $u_t$ :

$$\ln w_t = \alpha + \beta \ln u_t + \varepsilon_t.$$

Ошибки  $\varepsilon_t$  независимы и нормально распределены  $N(0, \sigma_\varepsilon^2)$ .

- а) Найдите МНК-оценки параметров  $\alpha, \beta$ , а также 95%-доверительный интервал для  $\beta$ .

LS LOG(W) C LOG(U)

Dependent Variable: LOG(W)

Method: Least Squares

Sample: 1 24

Included observations: 24

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.299704	0.185973	12.36580	0.0000
LOG(U)	-0.779056	0.113324	-6.874566	0.0000
R-squared	0.682355	Mean dependent var	1.071638	
Adjusted R-squared	0.667916	S.D. dependent var	0.439606	
S.E. of regression	0.253331	Akaike info criterion	0.171412	
Sum squared resid	1.411880	Schwarz criterion	0.269583	
Log likelihood	-0.056948	F-statistic	47.25966	
Durbin-Watson stat	1.102614	Prob(F-statistic)	0.000001	

- б) Рассмотрите статистику Дарбина-Уотсона. Что ее значение говорит об исходном предположении об ошибках  $\varepsilon_t$ ? Что можно сказать о доверительном интервале, найденном ранее?

в) Проведите корректировку Ньюи-Уэста.

Dependent Variable: LOG(W)

Method: Least Squares

Sample: 1 24

Included observations: 24

Newey-West HAC Standard Errors & Covariance (lag truncation=2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.299704	0.308304	7.459212	0.0000
LOG(U)	-0.779056	0.153912	-5.061710	0.0000
R-squared	0.682355	Mean dependent var		1.071638
Adjusted R-squared	0.667916	S.D. dependent var		0.439606
S.E. of regression	0.253331	Akaike info criterion		0.171412
Sum squared resid	1.411880	Schwarz criterion		0.269583
Log likelihood	-0.056948	F-statistic		47.25966
Durbin-Watson stat	1.102614	Prob(F-statistic)		0.000001

г) Оцените модель, используя модель автокорреляции первого порядка для ошибок регрессии. Найдите 95%-доверительный интервал для  $\beta$ . Сравните результат с интервалом, полученным ранее.

LS LOG(W) C LOG(U) AR(1)

Dependent Variable: LOG(W)

Method: Least Squares

Sample(adjusted): 2 24

Included observations: 23 after adjusting endpoints

Convergence achieved after 6 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.394881	0.171736	13.94512	0.0000
LOG(U)	-0.831240	0.095384	-8.714635	0.0000
AR(1)	0.467556	0.203784	2.294370	0.0327
R-squared	0.729098	Mean dependent var		1.094400
Adjusted R-squared	0.702007	S.D. dependent var		0.434786
S.E. of regression	0.237344	Akaike info criterion		0.082494
Sum squared resid	1.126642	Schwarz criterion		0.230601
Log likelihood	2.051324	F-statistic		26.91368
Durbin-Watson stat	2.161687	Prob(F-statistic)		0.000002

**Домашнее задание 5.** Должно быть сдано через 2 недели на Семинаре 12.

## Семинар 11. Модели бинарного выбора. Фиктивные переменные.

- Линейная модель вероятности. Недостатки и достоинства (гетероскедастичность остатков).
- Логит- и пробит- модели. Оценивание и интерпретация коэффициентов.
- Свойства оценок, полученных методом максимального правдоподобия.
- Учет качественных переменных в модели регрессии (с использованием фиктивных переменных). Случай, когда невозможно восстановление значений категоризованных переменных (кластерный анализ).

- Проблема мультиколлинеарности.
- Интерпретация коэффициентов при фиктивных переменных.
- Учет сезонности.

#### Задачи для разбора на семинаре:

- учет сезонности в аддитивной и в мультипликативной форме (см. Айвазяна).
- учет значимого события (например, кризиса) во времени.
- задачки из Доугерти

## Семинар 12. Компьютерный семинар по темам семинара 11.

### Собрать сделанное задание 5.

#### Пример 6. Выбор карьеры в Великобритании в 16 лет. Файл **choice.wfl**.

В Великобритании выбор карьеры подростками делается в 16 лет, когда они сдают национальные экзамены. Несколькими месяцами позже они должны решить: продолжить ли обучение в школе или бросить дневное обучение. В последнем случае у них есть выбор между полноценной работой и совмещением работы с учебой. Цель данного примера – выявление факторов, определяющих выбор карьеры.

<b>At16</b>	Принятое в 16 лет решение о: <b>1</b> – продолжении обучения в школе, <b>2</b> – совместительстве, <b>3</b> – работе.
<b>Able7</b>	Результат теста на общие способности (General ability test), измеренный в 7-летнем возрасте.
<b>Loginc</b>	Логарифм семейного дохода (измеренного в возрасте 16).
<b>Ctratio</b>	Показатель качества школ: количество учеников на одного учителя.
<b>Oldsib</b>	Количество старших сестер или братьев (измеренный в 16 лет).
<b>Yngsib</b>	Количество младших сестер или братьев (измеренный в 16 лет).
<b>Etot</b>	Количество «5» полученных на национальном экзамене в 16 лет (предшествующего принятому решению).
<b>Female</b>	1 – девушка, 0 – молодой человек.

1. Рассмотрите статистику по девушкам и молодым людям отдельно. Дайте интерпретацию результатам и используйте их для проверки отсутствия в переменных ошибок.
2. Создайте переменную ( $School = 1$  if  $At16 = 1$ ). Постройте линейную модель выбора. Проинтерпретируйте результаты.
3. Постройте логит-модель и дайте интерпретацию результатам.
4. Повторите процедуру, используя вместо логит-модели пробит.
5. Сделайте выбор между логит- и пробит-моделями. Проверьте гипотезу о том, что младшие и старшие братья или сестры одинаково влияют на выбор.
6. Разбейте выборку на девушек и молодых людей и оцените логит-модель отдельно для каждой подвыборки. Проинтерпретируйте различия в результатах.

7. Проверьте гипотезу о значимости такого разделения выборки.
8. Используя логит-модель предскажите вероятность принятия решения о продолжении образования для подростков со средними характеристиками. Вычислите ожидаемое влияние одного дополнительного младшего брата (сестры) на эту вероятность.

**Домашнее задание 6.** *Должно быть сдано через 2 недели на Семинаре 14.*

- Продолжить пример 6. Построить модели для двух остальных случаев. Ответить в каждом случае на те же вопросы.
- Разобраться с задачей 13.8 из Магнуса.

### **Семинар 13. Резерв.**

Обзор ошибок в домашних работах. Ответы на вопросы.

### **Семинар 14. Резерв.**

**Сдать сделанное задание 6 и проект.**