МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. Ломоносова
Экономический факультет

Ф. С. Картаев

СБОРНИК ЗАДАНИЙ ПО ЭКОНОМЕТРИКЕ

для подготовки к вступительным экзаменам в магистратуру экономического факультета МГУ имени М.В. Ломоносова

Задачник

Москва
2016

УДК 330.42
ББК 08.00.13
К27

Картаев Ф. С.

К27 Сборник заданий по эконометрике для подготовки к вступительным экзаменам в магистратуру экономического факультета МГУ имени М. В. Ломоносова: Задачник. — М.: Экономический факультет МГУ имени М. В. Ломоносова, 2016. – 36 с.

ISBN 978-5-906783-26-4

Сборник задач предназначен для студентов учебных курсов бакалавриата, подготовки к вступительным экзаменам в магистратуру.

Материалы сборника позволят приобрести навык решения задач базового уровня по курсу эконометрика, развить понимание следующих тем дисциплины эконометрика: классическая линейная модель парной и множественной регрессии, спецификация модели регрессии, обобщенная линейная модель множественной регрессии, гетероскедастичность, инструментальные переменные.

Сборник содержит не только условия задач и ответы к ним, но и подробные решения значительной их части, что позволит осуществлять эффективную самостоятельную подготовку.

УДК 330.42
ББК 08.00.13

9785906783264

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

В настоящее время желающим познакомиться с основами эконометрики доступно большое количество качественных учебников как на русском, так и на английском языках (с списком вводных книг по эконометрике, рекомендуемым автором этого сборника читатель может познакомиться в списке литературы в конце этой работы). В тоже время наблюдается некоторый дефицит тренировочных заданий базового уровня, которые могли бы помочь закрепить усвоенный материал, и убедиться в том, что он усвоен верно. Представленный сборник заданий призван восполнить этот пробел. Он содержит ряд тестов и задач по ключевым темам вводного курса эконометрики пространственных данных. В конце сборника приведены ответы ко всем заданиям. Для большей части задач приводится также полное решение.

Особенно полезен этот сборник для абитуриентов магистратуры экономического факультета МГУ, желающим подготовиться к эконометрическим заданиям общей части вступительного экзамена по экономической теории. Сборник содержит тесты и задачи из вступительных экзаменов прошлых лет. Все задания соответствуют программе вступительного экзамена, действующей на момент издания сборника. Тем не менее, эта программа может меняться год от года, поэтому автор рекомендует абитуриентам магистратуры ознакомиться с ее текущей версией на официальном сайте факультета в разделе с информацией для поступающих на магистерские программы[1].

Также сборник пригодится студентам, изучающим курс введения в эконометрику, а также преподавателям этого курса. (При этом следует подчеркнуть, что охват тем и уровень сложности заданий в этом сборнике недостаточен для подготовки к экзамену по курсу эконометрики в бакалавриате экономического факультета МГУ имени М. В. Ломоносова, так как в здесь отражены только задания базового уровня сложности по ключевым темам эконометрики пространственных данных).

Если в процессе решения заданий вы обнаружите в сборнике неточности и опечатки, то автор будет вам признателен за сообщение о них. Пишите на kartaev@gmail.com.

---

[1] http://www.econ.msu.ru/entrance/masters/

ПРОСТЫЕ ТЕСТЫ

Тест 1.1

Вариант 1. Исследователь получил следующие результаты оценивания параметров линейной модели множественной регрессии при помощи метода наименьших квадратов на основе данных о тысяче наблюдений:

$$\hat{y}_i = \underset{(0,12)}{0,21} + \underset{(1,95)}{6,72}\, x_i + \underset{(1,97)}{8,81}\, z_i .$$

В скобках под оценками коэффициентов указаны соответствующие стандартные ошибки. Используя пятипроцентный уровень значимости, тестируйте незначимость коэффициентов при переменных в этом уравнении. (Соответствующее критическое значение t-статистики при уровне значимости 5% составляет 1,96.)

а) Значим только коэффициент при переменной x_i.
б) Значим только коэффициент при переменной z_i.
в) Значимы и коэффициент при переменной x_i, и коэффициент при переменной z_i.
г) Ни один из коэффициентов не значим.

Вариант 2. Исследователь получил следующие результаты оценивания параметров линейной модели множественной регрессии при помощи метода наименьших квадратов на основе данных о двух тысячах наблюдений:

$$\hat{y}_i = \underset{(2,44)}{0,08} + \underset{(1,91)}{2,62}\, x_i + \underset{(1,99)}{1,69}\, z_i .$$

В скобках под оценками коэффициентов указаны соответствующие стандартные ошибки. Используя пятипроцентный уровень значимости, тестируйте незначимость коэффициентов при переменных в этом уравнении. (Соответствующее критическое значение t-статистики при уровне значимости 5% составляет 1,96.)

а) Значим только коэффициент при переменной x_i.
б) Значим только коэффициент при переменной z_i.
в) Значимы и коэффициент при переменной x_i, и коэффициент при переменной z_i.
г) Ни один из коэффициентов не значим.

Вариант 3. В рамках предпосылок классической линейной модели множественной регрессии исследователь при помощи метода наименьших квадратов на основе данных о десяти тысячах наблюдений получил следующие результаты оценивания параметров уравнения:

$$\widehat{\ln y_i} = 0{,}34 + 0{,}12 * \ln x_i + 0{,}71 * \ln z_i, \quad R^2 = 0{,}4.$$

Что можно сказать о значимости полученного уравнения?

(Соответствующее критическое значение F-статистики при уровне значимости 5% составляет 3,00, а при уровне значимости 1% равно 4,61.)

а) При использовании 5-процентного уровня значимости следует сделать вывод о значимости уравнения. При использовании 1-процентного уровня значимости следует сделать вывод о незначимости уравнения.

б) При использовании 1-процентного уровня значимости следует сделать вывод о значимости уравнения. При использовании 5-процентного уровня значимости следует сделать вывод о незначимости уравнения.

в) При использовании как 5-процентного, так и 1-процентного уровней значимости следует сделать вывод о значимости уравнения.

г) При использовании как 5-процентного, так и 1-процентного уровней значимости следует сделать вывод о незначимости уравнения

Вариант 4. В рамках предпосылок классической линейной модели множественной регрессии исследователь при помощи метода наименьших квадратов на основе данных о десяти тысячах наблюдений получил следующие результаты оценивания параметров уравнения:

$$\widehat{\ln y_i} = -0{,}01 + 0{,}06 * \ln x_i + 0{,}04 * \ln z_i, \quad R^2 = 0{,}2.$$

Что можно сказать о значимости полученного уравнения?

(Соответствующее критическое значение F-статистики при уровне значимости 5% составляет 3,00, а при уровне значимости 1% равно 4,61.)

а) При использовании 5-процентного уровня значимости следует сделать вывод о значимости уравнения. При использовании 1-процентного уровня значимости следует сделать вывод о незначимости уравнения.

б) При использовании 1-процентного уровня значимости следует сделать вывод о значимости уравнения. При использовании 5-процентного уровня значимости следует сделать вывод о незначимости уравнения.

в) При использовании как 5-процентного, так и 1-процентного уровней значимости следует сделать вывод о значимости уравнения.

г) При использовании как 5-процентного, так и 1-процентного уровней значимости следует сделать вывод о незначимости уравнения.

Тест 1.2

Вариант 1. Исследователь оценивает параметры модели $y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$ при помощи **двухшагового метода наименьших квадратов (2МНК)**, используя переменную ***w*** как инструмент для переменной ***x***. Каким требованиям должна удовлетворять инструментальная переменная, чтобы полученная оценка коэффициента β_2 была состоятельной?

а) $Cov(w_i, \varepsilon_i) = 0,\ Cov(w_i, x_i) \neq 0$

б) $Cov(w_i, \varepsilon_i) \neq 0,\ Cov(w_i, x_i) = 0$

в) $Cov(w_i, \varepsilon_i) \neq 0,\ Cov(w_i, x_i) \neq 0$

г) $Cov(w_i, \varepsilon_i) = 0,\ Cov(w_i, x_i) = 0$

Вариант 2. Исследователь оценивает параметры модели $y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$ при помощи **метода инструментальных переменных**, используя переменную ***q*** как инструмент для переменной ***x***. Каким требованиям должна удовлетворять инструментальная переменная, чтобы полученная оценка коэффициента β_2 была состоятельной?

а) $Cov(q_i, \varepsilon_i) \neq 0,\ Cov(q_i, x_i) \neq 0$

б) $Cov(q_i, \varepsilon_i) = 0,\ Cov(q_i, x_i) = 0$

в) $Cov(q_i, \varepsilon_i) = 0,\ Cov(q_i, x_i) \neq 0$

г) $Cov(q_i, \varepsilon_i) \neq 0,\ Cov(q_i, x_i) = 0$

Вариант 3. Исследователь оценивает параметры модели $y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$ при помощи двухшагового метода наименьших квадратов (2МНК), используя переменную ***w*** как инструмент для переменной ***x***. На основе данных о тысяче наблюдений исследователь вычислил значения выборочных коэффициентов ковариации между парами переменных: $\widehat{Cov}(x, y) = 45, \widehat{Cov}(x, w) = 30, \widehat{Cov}(y, w) = 90$, а также выборочную дисперсию $\widehat{Var}(x) = 9$. Используя те из доступных данных, которые вам необходимы, вычислите 2МНК-оценку коэффициента β_2.

а) 5
б) 4
в) 3
г) 2

Вариант 4. Исследователь оценивает параметры модели $y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$ при помощи двухшагового метода наименьших квадратов (2МНК), используя переменную ***q*** как инструмент для переменной ***x***. На основе данных о тысяче наблюдений исследователь вычислил значения выборочных коэффициентов ковариации между парами переменных: $\widehat{Cov}(x,y)=60$, $\widehat{Cov}(x,q)=20$, $\widehat{Cov}(y,q)=40$, а также выборочную дисперсию $\widehat{Var}(x)=10$. Используя те из доступных данных, которые вам необходимы, вычислите 2МНК-оценку коэффициента β_2.

а) 0,5
б) 2
в) 3
г) 6

Тест 1.3

Вариант 1. Имеются некоторые данные о переменных $\boldsymbol{x}$ и $\boldsymbol{y}$ (см. таблицу). Используя доступную информацию, найдите МНК-оценку коэффициента $\hat{\beta}_1$ в регрессии

$$\hat{y}_i=\hat{\beta}_1+\hat{\beta}_2 x_i.$$

y	10	5	9	11	12	13
x	0	0	0	1	1	1

а) 2
б) 4
в) 6
г) 8

Вариант 2. Имеются некоторые данные о переменных $\boldsymbol{x}$ и $\boldsymbol{y}$ (см. таблицу). Используя доступную информацию, найдите МНК-оценку коэффициента $\hat{\beta}_1$ в регрессии

$$\hat{y}_i=\hat{\beta}_1+\hat{\beta}_2 x_i.$$

y	12	7	11	13	15	14
x	0	0	0	1	1	1

а) 4
б) 8
в) 10
г) 15

Вариант 3. Имеются следующие данные о переменных $\boldsymbol{x}$ и $\boldsymbol{y}$:

$$\sum_{i=1}^{10} x_i=0,\ \sum_{i=1}^{10} x_i^2=10,\ \sum_{i=1}^{10} y_i=12,\ \sum_{i=1}^{10} y_i^2=20,\ \sum_{i=1}^{10} x_i * y_i=11.$$

Используя доступную информацию, найдите МНК-оценку коэффициента $\hat{\beta}_1$ в регрессии

$$\hat{y}_i=\hat{\beta}_1+\hat{\beta}_2 x_i.$$

а) 1,0
б) 1,2
в) 1,4
г) 1,6

Вариант 4. Имеются следующие данные о переменных ***x*** и ***y***:

$$\sum_{i=1}^{10} x_i = 0,\ \sum_{i=1}^{10} x_i^2 = 64,\ \sum_{i=1}^{10} y_i = 25,\ \sum_{i=1}^{10} y_i^2 = 50,\ \sum_{i=1}^{10} x_i * y_i = -6.$$

Используя доступную информацию, найдите МНК-оценку коэффициента $\hat{\beta}_1$ в регрессии

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i.$$

а) 1,0
б) 1,5
в) 2,0
г) 2,5

Тест 1.4

Вариант 1. Исследователь анализирует модель $y_i = \alpha + \beta x_i + \varepsilon_i$, для которой выполнены все предпосылки классической линейной модели множественной регрессии за одним исключением: дисперсия случайной ошибки ε_i имеет вид $V(\varepsilon_i) = \sigma_0^2 x_i^2$. Если исследователь воспользуется для оценивания коэффициентов модели обычным методом наименьших квадратов, то полученные им оценки будут:

а) смещенными и эффективными
б) смещенными и неэффективными
в) несмещенными и эффективными
г) несмещенными и неэффективными

Вариант 2. Исследователь анализирует модель $y_i = \theta x_i + \varepsilon_i$, для которой выполнены все предпосылки классической линейной модели множественной регрессии за одним исключением: дисперсия случайной ошибки ε_i имеет вид $V(\varepsilon_i) = \delta x_i^2, \delta > 0$. Если исследователь воспользуется для оценивания коэффициентов модели обычным методом наименьших квадратов, то полученные им оценки будут:

а) несмещенными и неэффективными
б) смещенными и эффективными
в) смещенными и неэффективными
г) несмещенными и эффективными

Вариант 3. Исследователь анализирует модель $y_i = \alpha + \beta x_i + \varepsilon_i$, для которой выполнены все предпосылки классической линейной модели множественной регрессии за одним исключением: дисперсия случайной ошибки ε_i имеет вид $V(\varepsilon_i) = \sigma_0^2 x_i^2$. В этом случае для получения эффективных оценок коэффициентов следует использовать:

а) метод наименьших квадратов
б) двухшаговый метод наименьших квадратов
в) взвешенный метод наименьших квадратов
г) метод инструментальных переменных

Вариант 2. Исследователь анализирует модель $y_i = \theta x_i + \varepsilon_i$, для которой выполнены все предпосылки классической линейной модели множественной регрессии за одним исключением: дисперсия случайной ошибки ε_i имеет вид $V(\varepsilon_i) = \delta x_i^2$, $\delta > 0$. В этом случае для получения несмещенных оценок коэффициентов можно использовать:

а) обычный метод наименьших квадратов
б) обычный метод наименьших квадратов в сочетании с расчетом стандартных ошибок оценок коэффициентов в форме Уайта (состоятельных в условиях гетероскедастичности стандартных ошибок)
в) взвешенный метод наименьших квадратов
г) любой из указанных выше вариантов

Тест 1.5

Вариант 1. Исследователь анализирует зависимость потребления некоторого товара от уровня дохода для однородной группы потребителей: $Y_i = \beta_1 + \beta_2 \ln INCOME_i + \varepsilon_i$, где Y_i — потребление некоторого товара (в килограммах), $INCOME_i$ — доход потребителя (в рублях). В ходе оценивания модели на основе данных о 400 потребителях получены следующие результаты: $\hat{Y}_i = 3{,}0 + 0{,}8 \ln INCOME_i$, $R^2 = 0{,}95$. Дайте интерпретацию коэффициента при переменной:

а) при увеличении дохода на 1% потребление товара увеличивается на 8%,
б) при увеличении дохода на один рубль потребление товара увеличивается на 8%,
в) при увеличении дохода на один рубль потребление товара увеличивается на 0,08%,
г) при увеличении дохода на 1% потребление товара увеличивается на 8 кг,
д) при увеличении дохода на 1% потребление товара увеличивается на 0,008 кг.

Вариант 2. Исследователь анализирует зависимость потребления некоторого товара от уровня дохода для однородной группы потребителей:

$\ln Y_i = \beta_1 + \beta_2 \ln INCOME_i + \varepsilon_i$, где Y_i — потребление некоторого товара (в килограммах), $INCOME_i$ — доход потребителя (в рублях). В ходе оценивания модели на основе данных о 400 потребителях получены следующие результаты: $\ln \hat{Y}_i = 0{,}5 + 2{,}0 \ln INCOME_i$, $R^2 = 0{,}94$. Дайте интерпретацию коэффициента при переменной:

а) при увеличении дохода на 1% потребление товара увеличивается на 2%,
б) при увеличении дохода на один рубль потребление товара увеличивается на 2%,
в) при увеличении дохода на один рубль потребление товара увеличивается на 0,02%,
г) при увеличении дохода на 1% потребление товара увеличивается на 2 кг,
д) при увеличении дохода на 1% потребление товара увеличивается на 0,02 кг.

Вариант 3. Руководство фирмы «АВС» решило исследовать эффективность курсов по повышению квалификации, которые иногда проводятся для ее сотрудников. В исследовании принимали участие 200 работников фирмы. Среди них случайным образом были отобраны 50 человек, для которых были проведены эти курсы, остальные сотрудники, участвующие в исследовании, курсов по повышению квалификации не проходили. После этого на основе полученных данных при помощи МНК было оценено следующее уравнение регрессии (в скобках указаны стандартные ошибки оценок коэффициентов):

$$\widehat{\ln y_i} = \underset{(0,4)}{2{,}0} + \underset{(0,1)}{0{,}5} x_i + \underset{(0,1)}{0{,}3} z_i, \quad R^2 = 0{,}8.$$

Здесь x_i — фиктивная переменная, которая равна 1 для сотрудников, прошедших курсы повышения квалификации, и равна 0 для остальных сотрудников, z_i — стаж работы i-го сотрудника, измеренный в годах, y_i — производительность труда i-го сотрудника. В соответствии с полученным уравнением на сколько процентов при прочих равных условиях увеличивается производительность труда работника в результате прохождения курсов повышения квалификации?

а) на $\left(e^2 - 1\right) * 100\%$.
б) на $e^2 * 100\%$,
в) на $\left(\sqrt{e} - 1\right) * 100\%$,
г) на $\sqrt{e} * 100\%$.

Вариант 4. Руководство фирмы «АВС» решило исследовать эффективность курсов по повышению квалификации, которые иногда проводятся для ее сотрудников. В исследовании принимали участие 200 работников фирмы. Среди них случайным образом были отобраны 50 человек, для которых были проведены эти курсы, остальные сотрудники, участвующие в исследовании, курсов по повышению квалификации не проходили. После этого на основе полученных данных при помощи МНК было оценено следующее уравнение регрессии (в скобках указаны стандартные ошибки оценок коэффициентов):

$$\widehat{\ln y_i} = \underset{(0,2)}{0{,}5} + \underset{(0,3)}{2{,}0}\,x_i + \underset{(0,4)}{0{,}4}\,z_i, \quad R^2 = 0{,}9.$$

Здесь x_i — фиктивная переменная, которая равна 1 для сотрудников, прошедших курсы повышения квалификации, и равна 0 для остальных сотрудников, z_i — стаж работы i-го сотрудника, измеренный в годах, y_i — производительность труда i-го сотрудника. В соответствии с полученным уравнением, на сколько процентов при прочих равных условиях увеличивается производительность труда работника в результате прохождения курсов повышения квалификации?

а) на $(e^2 - 1) * 100\%$.

б) на $e^2 * 100\%$,

в) на $(\sqrt{e} - 1) * 100\%$,

г) на $\sqrt{e} * 100\%$.

Тест 1.6

Вариант 1. Исследователь получил следующие результаты оценивания параметров линейной модели множественной регрессии при помощи метода наименьших квадратов на основе данных о 200 наблюдениях:

$$\hat{y}_i = -\underset{(0,1)}{0{,}6} + \underset{(100,0)}{800{,}0}\,x_i + \underset{(2,4)}{9{,}9}\,z_i.$$

В скобках под оценками коэффициентов указаны соответствующие стандартные ошибки. Постройте 95-процентный доверительный интервал для коэффициента при переменной x_i. (Соответствующее критическое значение t-статистики при уровне значимости 5% составляет 1,96.)

а) (700, 900)

б) (798,04, 801,96)

в) (408, 1192)

г) (604, 996).

Вариант 2. Исследователь получил следующие результаты оценивания параметров линейной модели множественной регрессии при помощи метода наименьших квадратов на основе данных о 400 наблюдениях:

$$\hat{y}_i = \underset{(10,4)}{232{,}6} + \underset{(200,0)}{600{,}0}\,x_i - \underset{(2,8)}{7{,}5}\,z_i.$$

В скобках под оценками коэффициентов указаны соответствующие стандартные ошибки. Постройте 95-процентный доверительный интервал для коэффициента при переменной x_i. (Соответствующее критическое значение t-статистики при уровне значимости 5% составляет 1,96.)

а) (400, 800)
б) (598,04, 601,96)
в) (208, 992)
г) (404, 796).

Вариант 3. В рамках предпосылок классической линейной модели множественной регрессии исследователь анализирует модель:

$$y_i = \beta_1 + \beta_2 * x_i + \beta_3 * z_i + \varepsilon_i.$$

Собрав данные о 500 наблюдениях, он оценил параметры модели при помощи метода наименьших квадратов:

$$\hat{y}_i = \underset{(0,12)}{0,21} + \underset{(1,95)}{6,72} x_i + \underset{(0,94)}{0,81} z_i.$$

В скобках под оценками коэффициентов указаны соответствующие стандартные ошибки. Проверьте гипотезу $\beta_3 = 1$. (Соответствующее критическое значение t-статистики при уровне значимости 5% составляет 1,96, а при уровне значимости 1% равно 2,58.)

а) Тестируемая гипотеза не отклоняется как при уровне значимости 1%, так и при уровне значимости 5%.
б) Тестируемая гипотеза не принимается как при уровне значимости 1%, так и при уровне значимости 5%.
в) Тестируемая гипотеза не отклоняется при уровне значимости 1% и не принимается при уровне значимости 5%.
г) Тестируемая гипотеза не отклоняется при уровне значимости 5% и не принимается при уровне значимости 1%.

Вариант 4. В рамках предпосылок классической линейной модели множественной регрессии исследователь анализирует модель:

$$y_i = \beta_1 + \beta_2 * x_i + \beta_3 * z_i + \varepsilon_i.$$

Собрав данные о 500 наблюдениях, он оценил параметры модели при помощи метода наименьших квадратов:

$$\hat{y}_i = \underset{(2,44)}{0,69} + \underset{(2,91)}{3,44} x_i + \underset{(1,99)}{1,61} z_i.$$

В скобках под оценками коэффициентов указаны соответствующие стандартные ошибки. Проверьте гипотезу $\beta_3 = 2$. (Соответствующее критическое значение t-статистики при уровне значимости 5% составляет 1,96, а при уровне значимости 1% равно 2,58.)

а) Тестируемая гипотеза не отклоняется как при уровне значимости 1%, так и при уровне значимости 5%.
б) Тестируемая гипотеза не принимается как при уровне значимости 1%, так и при уровне значимости 5%.
в) Тестируемая гипотеза не отклоняется при уровне значимости 1% и не принимается при уровне значимости 5%.
г) Тестируемая гипотеза не отклоняется при уровне значимости 5% и не принимается при уровне значимости 1%.

ТЕСТЫ ПОСЛОЖНЕЕ

Тест 2.1

Вариант 1. Имеются следующие данные о 100 наблюдениях переменных $x^{(2)}$, $x^{(3)}$ и y:

$$\sum y_i = 200, \quad \sum x_i^{(2)} = 0, \quad \sum x_i^{(3)} = 0, \quad \sum x_i^{(2)} y_i = 300,$$

$$\sum \left(x_i^{(2)}\right)^2 = 100, \; \sum \left(x_i^{(3)}\right)^2 = 200, \; \sum x_i^{(2)} x_i^{(3)} = 100, \; \sum x_i^{(3)} y_i = 200.$$

Вычислите МНК-оценку коэффициента β_2 в регрессии $y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i$.

а) 2
б) 3
в) 4
г) 5

Вариант 2. Имеются следующие данные о 25 наблюдениях переменных $x^{(2)}$, $x^{(3)}$ и y:

$$\sum y_i = 600, \quad \sum x_i^{(2)} = 0, \quad \sum x_i^{(3)} = 0, \quad \sum x_i^{(2)} y_i = 500,$$

$$\sum \left(x_i^{(2)}\right)^2 = 50, \; \sum \left(x_i^{(3)}\right)^2 = 100, \; \sum x_i^{(2)} x_i^{(3)} = 50, \; \sum x_i^{(3)} y_i = 400.$$

Вычислите МНК-оценку коэффициента β_2 в регрессии $y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i$.

а) 10
б) 12
в) 14
г) 16

Вариант 3. Имеются следующие данные о 100 наблюдениях переменных $x^{(2)}$, $x^{(3)}$ и y:

$$\sum y_i = 200, \quad \sum x_i^{(2)} = 0, \quad \sum x_i^{(3)} = 0, \quad \sum x_i^{(2)} y_i = 300,$$

$$\sum \left(x_i^{(2)}\right)^2 = 100, \; \sum \left(x_i^{(3)}\right)^2 = 200, \; \sum x_i^{(2)} x_i^{(3)} = 100, \; \sum x_i^{(3)} y_i = 200.$$

При помощи обычного МНК оцениваются параметры уравнения регрессии $y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i$. Известно, что сумма квадратов остатков в оцененной регрессии равна 19400. Предполагая выполнение предпосылок классической линейной модели множественной регрессии и нормальность случайных ошибок, вычислите расчетное значение t-статистики для проверки гипотезы $\beta_2 = 0$.

а) 2
б) 3
в) 4
г) 5

Вариант 4. Имеются следующие данные о 25 наблюдениях переменных $x^{(2)}$, $x^{(3)}$ и y:

$$\sum y_i = 600, \quad \sum x_i^{(2)} = 0, \quad \sum x_i^{(3)} = 0, \quad \sum x_i^{(2)} y_i = 500,$$
$$\sum \left(x_i^{(2)}\right)^2 = 50, \quad \sum \left(x_i^{(3)}\right)^2 = 100, \quad \sum x_i^{(2)} x_i^{(3)} = 50, \quad \sum x_i^{(3)} y_i = 400.$$

При помощи обычного МНК оцениваются параметры уравнения регрессии $y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i$. Известно, что сумма квадратов остатков в оцененной регрессии равна 2200. Предполагая выполнение предпосылок классической линейной модели множественной регрессии и нормальность случайных ошибок, вычислите расчетное значение t-статистики для проверки гипотезы $\beta_2 = 20$.

а) -1
б) -2
в) -3
г) -4

Тест 2.2

Вариант 1. Пусть истинная (но не известная исследователю) зависимость между переменными y, $x^{(1)}$ и $x^{(2)}$ описывается уравнением: $y_i = \beta_1 + \beta_2\, x_i^{(1)} + \beta_3\, x_i^{(2)} + \varepsilon_i$, где $\beta_1 > 0$, $\beta_2 > 0$, $\beta_3 > 0$ и выборочная ковариация между переменными $x^{(1)}$ и $x^{(2)}$ **больше нуля**. Исследователь ошибочно не включил в уравнение переменную $x^{(2)}$ и при помощи обычного МНК оценил парную регрессию: $\widehat{y_i} = \hat{\beta}_1 + \hat{\beta}_2 x_i^{(1)}$. Выберите верное утверждение об оценке коэффициента при переменной $x_i^{(1)}$ в этой парной регрессии.

а) Оценка является несмещенной
б) Оценка является смещенной и завышенной
в) Оценка является смещенной и заниженной
г) Оценка является смещенной, однако направление ее смещения невозможно определить, так как оно зависит от соотношения выборочных дисперсий переменных $x^{(1)}$ и $x^{(2)}$

Вариант 2. Пусть истинная (но не известная исследователю) зависимость между переменными y, $x^{(1)}$ и $x^{(2)}$ описывается уравнением:

$y_i = \beta_1 + \beta_2\ x_i^{(1)} + \beta_3\ x_i^{(2)} + \varepsilon_i$, где $\beta_1 < 0,\ \beta_2 > 0,\ \beta_3 < 0$ и выборочная ковариация между переменными $x^{(1)}$ и $x^{(2)}$ **меньше нуля**. Исследователь ошибочно не включил в уравнение переменную $x^{(2)}$ и при помощи обычного МНК оценил парную регрессию: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i^{(1)}$. Выберите верное утверждение об оценке коэффициента при переменной $x_i^{(1)}$ в этой парной регрессии.

а) Оценка является смещенной и завышенной
б) Оценка является смещенной и заниженной
в) Оценка является смещенной, однако направление ее смещения невозможно определить, так как оно зависит от соотношения выборочных дисперсий переменных $x^{(1)}$ и $x^{(2)}$
г) Оценка является несмещенной

Вариант 3. Пусть истинная (но не известная исследователю) зависимость между переменными y, $x^{(1)}$ и $x^{(2)}$ описывается уравнением: $y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$, где $\beta_1 > 0,\ \beta_2 > 0$. Исследователь ошибочно включил в уравнение еще одну переменную и при помощи обычного МНК оценил регрессию: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i^{(1)} + \hat{\beta}_3 x_i^{(2)}$. Известно, что выборочная ковариация между переменными $x^{(1)}$ и $x^{(2)}$ **больше нуля**. Выберите верное утверждение об оценке коэффициента при переменной $x_i^{(1)}$ в этой множественной регрессии.

а) Оценка является несмещенной
б) Оценка является смещенной и завышенной
в) Оценка является смещенной и заниженной
г) Оценка является смещенной, однако направление ее смещения невозможно определить, так как оно зависит от соотношения выборочных дисперсий переменных $x^{(1)}$ и $x^{(2)}$

Вариант 4. Пусть истинная (но не известная исследователю) зависимость между переменными y, $x^{(1)}$ и $x^{(2)}$ описывается уравнением: $y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$, где $\beta_1 > 0,\ \beta_2 < 0$. Исследователь ошибочно включил в уравнение еще одну переменную и при помощи обычного МНК оценил регрессию: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i^{(1)} + \hat{\beta}_3 x_i^{(2)}$. Известно, что выборочная ковариация между переменными $x^{(1)}$ и $x^{(2)}$ **меньше нуля**. Выберите верное утверждение об оценке коэффициента при переменной $x_i^{(1)}$ в этой множественной регрессии.

а) Оценка является смещенной и завышенной
б) Оценка является смещенной и заниженной
в) Оценка является смещенной, однако направление ее смещения невозможно определить, так как оно зависит от соотношения выборочных дисперсий переменных $x^{(1)}$ и $x^{(2)}$
г) Оценка является несмещенной

Тест 2.3

Вариант 1. Рассмотрим простую модель закрытой экономики, которая описывается следующими уравнениями:

$$C_t = \beta_1 + \beta_2 GDP_t + \varepsilon_t \quad (*)$$

$$GDP_t = C_t + I_t + G_t \quad (**)$$

GDP_t — ВВП в году t, C_t — совокупное потребление, I_t — совокупные инвестиции, G_t — государственные закупки, ε_t — случайные шоки потребления. Государственные закупки и инвестиции не коррелированы с ε_t. Исследователь рассматривает два способа оценки предельной склонности к потреблению в рассматриваемой экономике: **(I)** оценить уравнение (*) при помощи двухшагового МНК, используя переменную I_t в качестве инструмента для переменной GDP_t; **(II)** оценить уравнение (*) при помощи двухшагового МНК, используя переменную $Z_t = I_t + G_t$ в качестве инструмента для переменной GDP_t. Какой из указанных способов позволит исследователю получить состоятельную оценку?

а) Ни один из указанных способов
б) Только способ (I)
в) Только способ (II)
г) Любой из двух указанных способов

Вариант 2. Рассмотрим простую модель закрытой экономики, которая описывается следующими уравнениями:

$$C_t = \beta_1 + \beta_2 GDP_t + \varepsilon_t \quad (*)$$

$$GDP_t = C_t + I_t + G_t$$

GDP_t — ВВП в году t, C_t — совокупное потребление, I_t — совокупные инвестиции, G_t — государственные закупки, ε_t — случайные шоки потребления. Государственные закупки и инвестиции не коррелированы с ε_t. Исследователь рассматривает два способа оценки предельной склонности к потреблению в рассматриваемой экономике: **(I)** оценить уравнение (*) при помощи двухшагового МНК, используя переменную G_t в качестве инструмента для переменной GDP_t; **(II)** оценить уравнение (*) при помощи обычного МНК. Какой из указанных способов позволит исследователю получить состоятельную оценку?

а) Ни один из указанных способов
б) Только способ (I)
в) Только способ (II)
г) Любой из двух указанных способов

Вариант 3. В некоторой стране функция спроса на мороженое в i-м регионе имеет вид: $Q_i = \beta_1 + \beta_2 P_i + \beta_3 Inc_i + \beta_4 T_i + \varepsilon_i$. Функция предложения мороженого описывается соотношением:

$$Q_i = \gamma_1 + \gamma_2 P_i + u_i \quad (*)$$

Q_i — количество мороженого, продаваемое в i-м регионе, P_i — цена мороженого в i-м регионе, Inc_i — доход на душу населения в i-м регионе, T_i — среднегодовая температура воздуха в i-м регионе. Предполагается, что, $\beta_2 < 0, \beta_3 > 0, \beta_4 > 0, \gamma_2 > 0$, ε_i — независимые и одинаково распределенные случайные величины с дисперсией σ_ε^2, u_i — независимые и одинаково распределенные случайные величины дисперсией σ_u^2, u_i не коррелированы с доходом и температурой.

Исследователь рассматривает два способа оценки функции предложения в анализируемой экономике: **(I)** оценить уравнение (*) при помощи обычного МНК; **(II)** оценить уравнение (*) при помощи двухшагового МНК, используя переменную T_i в качестве инструмента для переменной P_i. Какой из указанных способов позволит исследователю получить состоятельные оценки параметров функции предложения?

а) Ни один из указанных способов
б) Только способ (I)
в) Только способ (II)
г) Любой из двух указанных способов

Вариант 4. В некоторой стране функция спроса на мороженое в i-м регионе имеет вид: $Q_i = \beta_1 + \beta_2 P_i + \beta_3 Inc_i + \beta_4 T_i + \varepsilon_i$. Функция предложения мороженого описывается соотношением:

$$Q_i = \gamma_1 + \gamma_2 P_i + u_i \quad (*)$$

Q_i — количество мороженого, продаваемое в i-м регионе, P_i — цена мороженого в i-м регионе, Inc_i — доход на душу населения в i-м регионе, T_i — среднегодовая температура воздуха в i-м регионе. Предполагается, что, $\beta_2 < 0, \beta_3 > 0, \beta_4 > 0, \gamma_2 > 0$, ε_i — независимые и одинаково распределенные случайные величины с дисперсией σ_ε^2, u_i — независимые и одинаково распределенные случайные величины дисперсией σ_u^2, u_i не коррелированы с доходом и температурой.

Исследователь рассматривает два способа оценки функции предложения в анализируемой экономике: **(I)** оценить уравнение (*) при помощи двухшагового МНК, используя переменную Q_i в качестве инструмента для переменной P_i; **(II)** оценить уравнение (*) при помощи двухшагового МНК, используя переменную Inc_i в качестве инструмента для переменной P_i. Какой из указанных способов позволит исследователю получить состоятельные оценки параметров функции предложения?

а) Ни один из указанных способов
б) Только способ (I)
в) Только способ (II)
г) Любой из двух указанных способов

Тест 2.4

Вариант 1. Исследователь оценил коэффициенты уравнения регрессии:

$$y_i = \beta_1 + \beta_2\, x_i^{(1)} + \beta_3\, x_i^{(2)} + \varepsilon_i.$$

Число наблюдений равно 250. После этого он решил провести тест Уайта на гетероскедастичность. Он оценил необходимое для этого вспомогательное уравнение, и коэффициент в этом вспомогательном уравнении оказался равен 0,16. Выберите верное утверждение.

Примечание: соответствующее критическое значение χ^2 при уровне значимости 5% составляет 11,07.

а) Расчетное значение тестовой статистики для теста Уайта составляет 40 и на основании этого теста следует сделать вывод о наличии в модели гетероскедастичности,

б) Расчетное значение тестовой статистики для теста Уайта составляет 40 и на основании этого теста следует сделать вывод об отсутствии в модели гетероскедастичности,

в) Расчетное значение тестовой статистики для теста Уайта составляет 4 и на основании этого теста следует сделать вывод о наличии в модели гетероскедастичности,

г) Расчетное значение тестовой статистики для теста Уайта составляет 4 и на основании этого теста следует сделать вывод об отсутствии в модели гетероскедастичности.

Вариант 2. Исследователь оценил коэффициенты уравнения регрессии:

$$y_i = \beta_1 + \beta_2\, x_i^{(1)} + \beta_3\, x_i^{(2)} + \beta_4\, x_i^{(3)} + \varepsilon_i.$$

Число наблюдений равно 600. После этого он решил провести тест Уайта на гетероскедастичность. Он оценил необходимое для этого уравнение, и коэффициент в этом уравнении оказался равен 0,02. Выберите верное утверждение.

Примечание: соответствующее критическое значение χ^2 при уровне значимости 5% составляет 21,7.

а) Расчетное значение тестовой статистики для теста Уайта составляет 24 и на основании этого теста следует сделать вывод о наличии в модели гетероскедастичности,

б) Расчетное значение тестовой статистики для теста Уайта составляет 24 и на основании этого теста следует сделать вывод об отсутствии в модели гетероскедастичности,

в) Расчетное значение тестовой статистики для теста Уайта составляет 12 и на основании этого теста следует сделать вывод о наличии в модели гетероскедастичности, г) Расчетное значение тестовой статистики для теста Уайта составляет 12 и на основании этого теста следует сделать вывод об отсутствии в модели гетероскедастичности.

Вариант 3. На основе данных, представленных в таблице, при помощи метода наименьших квадратов оцениваются параметры модели линейной регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$. Вычислите оценку коэффициента β_1.

y	2	3	4	5	16
x	–2	–1	0	1	2
z	3	3	–2	–2	–2

а) 0
б) 2
в) 4
г) 6

Вариант 4. На основе данных, представленных в таблице, при помощи метода наименьших квадратов оцениваются параметры модели линейной регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$. Вычислите оценку коэффициента β_1.

y	0	1	4	5	5
x	–3	–2	0	2	3
z	4	5	–3	–3	–3

а) 0
б) 1
в) 2
г) 3

ЗАДАЧИ

Задача 1. В вашем распоряжении имеются следующие данные о заработной плате сотрудников компании «ABC» в июне 2013 года.

Сотрудник	**Заработная плата (тысяч рублей)**
Александр Сергеевич	18
Лев Николаевич	28
Михаил Юрьевич	28
Петр Сергеевич	33
Марк Ильич	33
Елена Владимировна	25
Людмила Игоревна	30
Светлана Васильевна	35
Анна Петровна	20
Юлия Сергеевна	20

Рассматривается классическая линейная модель парной регрессии: $y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$, где y_i — заработная плата i-го работника, x_i — фиктивная переменная равная единице, если i-тый работник мужчина, и нулю, если i-ый работник женщина.

(а) Найдите МНК-оценки для коэффициентов модели.

(б) Вычислите стандартную ошибку для оценки коэффициента β_2. Проверьте гипотезу $\beta_2 = 0$. Можно ли утверждать, что пол работника статистически значимо влияет на уровень его заработной платы в фирме «ABC»?

Примечание: соответствующее критическое значение t-статистики при уровне значимости 5% составляет 2,3.

(в) Естественно ожидать, что на заработную плату работника влияет также его образование. Предположим, что истинная модель для заработной платы имеет вид: $y_i = \beta_1 + \beta_2 * x_i + \beta_3 * z_i + \varepsilon_i$, где z_i — переменная, характеризующая количество лет обучения, для i-го работника. Предположим также, что более образованные работники в среднем получают более высокую заработную плату, и что в фирме «ABC» женщины в среднем более образованны, чем мужчины.

Что в этой ситуации можно сказать об оценке коэффициента β_2, полученной в пункте (б)? Будет ли она смещена? Если ответ «да», то будет ли она завышена или занижена? Формально обоснуйте свой ответ.

Задача 2. В вашем распоряжении имеются следующие данные о заработной плате сотрудников компании «ABC» в июне 2013 года.

Сотрудник	Заработная плата (тысяч рублей)
Иван Петрович	30
Сергей Васильевич	20
Василий Иванович	26
Петр Сергеевич	23
Марк Ильич	26
Елена Владимировна	25
Людмила Игоревна	19
Светлана Васильевна	21
Анна Петровна	24
Юлия Сергеевна	21

Рассматривается классическая линейная модель парной регрессии: $y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$, где y_i — заработная плата i-го работника, x_i — фиктивная переменная равная единице, если i-тый работник женщина, и нулю, если i-ый работник мужчина.

(а) Найдите МНК-оценки для коэффициентов модели.

(б) Вычислите стандартную ошибку для оценки коэффициента β_2. Проверьте гипотезу $\beta_2 = 0$. Можно ли утверждать, что пол работника статистически значимо влияет на уровень его заработной платы в фирме «ABC»?

Примечание: соответствующее критическое значение t-статистики при уровне значимости 5% составляет 2,3.

(в) Естественно ожидать, что на заработную плату работника влияет также его образование. Предположим, что истинная модель для заработной платы имеет вид: $y_i = \beta_1 + \beta_2 * x_i + \beta_3 * z_i + \varepsilon_i$, где z_i — переменная, характеризующая количество лет обучения, для i-го работника. Предположим также, что в фирме «ABC» мужчины и женщины в среднем имеют одинаковый уровень образования и при этом более образованные работники в среднем получают более высокую заработную плату, чем менее образованные.

Что в этой ситуации можно сказать об оценке коэффициента β_2, полученной в пункте (б)? Будет ли она смещена? Если ответ «да», то будет ли она завышена или занижена? Формально обоснуйте свой ответ.

Задача 3

Имеются следующие данные о 100 наблюдениях переменных $x^{(2)}$, $x^{(3)}$ и y:

$$\sum y_i = 200, \quad \sum x_i^{(2)} = 0, \quad \sum x_i^{(3)} = 0, \quad \sum x_i^{(2)} y_i = 300,$$

$$\sum \left(x_i^{(2)}\right)^2 = 100, \quad \sum \left(x_i^{(3)}\right)^2 = 200, \quad \sum x_i^{(2)} x_i^{(3)} = 100, \quad \sum x_i^{(3)} y_i = 200.$$

(а) **(5)** Вычислите МНК-оценки коэффициентов в регрессии

$$y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i.$$

(б) **(5)** Пусть также известно, что сумма квадратов остатков в оцененной регрессии равна 19400. Выпишите оценку ковариационной матрицы вектора $\hat{\beta}$.

(в) **(10)** Используя результаты предыдущих пунктов, проверьте гипотезу $\beta_3 = 0$. (Соответствующее критическое значение при уровне значимости 5% равно 1,99.)

Задача 4

Имеются следующие данные о 25 наблюдениях переменных $x^{(2)}$, $x^{(3)}$и y:

$$\sum y_i = 600, \quad \sum x_i^{(2)} = 0, \quad \sum x_i^{(3)} = 0, \quad \sum x_i^{(2)} y_i = 500,$$

$$\sum \left(x_i^{(2)}\right)^2 = 50, \quad \sum \left(x_i^{(3)}\right)^2 = 100, \quad \sum x_i^{(2)} x_i^{(3)} = 50, \quad \sum x_i^{(3)} y_i = 400.$$

(а) **(5)** Вычислите МНК-оценки коэффициентов в регрессии

$$y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i.$$

(б) **(5)** Пусть также известно, что сумма квадратов остатков в оцененной регрессии равна 2200. Выпишите оценку ковариационной матрицы вектора $\hat{\beta}$.

(в) **(10)** Используя результаты предыдущих пунктов, проверьте гипотезу $\beta_2 = 20$. (Соответствующее критическое значение при уровне значимости 5% равно 2,07.)

Задача 5

На основе 20 наблюдений была оценена следующая модель регрессии (в скобках указаны стандартные ошибки оценок коэффициентов):

$$\hat{y}_i = \underset{(0,6)}{2{,}4} + \underset{(0,3)}{6{,}9}\, x_i + \underset{(9,8)}{5{,}1}\, w_i$$

Кроме того, известно, что TSS=2000, а сумма квадратов остатков равна 200.

(а) Вычислите значение коэффициента R^2, значение скорректированного коэффициента R^2_{adj} и стандартную ошибку регрессии.

(б) Проверьте значимость уравнения в целом: сформулируйте и проверьте гипотезу о том, что все коэффициенты при переменных уравнения одновременно равны нулю.

(в) Значим ли коэффициент при переменной $\boldsymbol{x}$? Сформулируйте и проверьте соответствующую гипотезу.

(г) Проверьте гипотезу о том, что коэффициент при переменной $\boldsymbol{x}$ равен 7.

(д) Постройте 99-процентный доверительный интервал для коэффициента при переменной $\boldsymbol{x}$.

(е) После того, как исследователь добавил в модель еще две переменных (***p*** и ***s***), R^2 в этой модели увеличился до 0,95. Осуществив соответствующий тест, определите, стоило ли добавлять в модель эти переменные?

Примечание: все гипотезы в этой задаче проверяйте при уровне значимости 1%. Приводите расчетные значения используемых t- и F-статистик, а также используемые критические значения из соответствующих таблиц с указанием числа степеней свободы.

Задача 6

Оценивание модели $y_i = \beta_1 + \beta_2 * x_i^{(2)} + \beta_3 * x_i^{(3)} + \varepsilon_i$ методом наименьших квадратов по 30 наблюдениям дало следующие результаты (в скобках указаны стандартные отклонения оценок коэффициентов):

$$\textbf{Модель 1: } \widehat{y_i} = \underset{(3,1)}{9{,}1} - \underset{(2,1)}{10{,}2} * x_i^{(2)} + \underset{(3,0)}{4{,}5} * x_i^{(3)},\ R^2 = 0{,}21$$

Оценивание модели $y_i = \beta_1 + \beta_2 * x_i^{(2)} + \beta_3 * x_i^{(3)} + \beta_4 * x_i^{(4)} + \beta_5 * x_i^{(5)} + \varepsilon_i$ методом наименьших квадратов по тем же самым 30 наблюдениям дало следующие результаты:

$$\textbf{Модель 2: } \widehat{y_i} = 8{,}2 - 8{,}5 * x_i^{(2)} + 9{,}0 * x_i^{(3)} + 5{,}0 * x_i^{(4)} + 6{,}1 * x_i^{(5)},$$
$$R^2 = 0{,}23$$

(а) Для модели №1 проверьте значимость уравнения в целом при уровне значимости 5%.

(б) Для модели №1 проверьте значимость переменной $\boldsymbol{x}^{(3)}$ при уровне значимости 5%.

(в) Какая из моделей лучше соответствует данным? Сравните модели №1 и №2, используя в качестве критерия скорректированный коэффициент детерминации R^2_{adj}. Почему не стоит использовать для сравнения этих моделей обычный R^2?

(г) Какая из моделей лучше соответствует данным? Сравните модели №1 и №2, используя в качестве критерия соответствующий статистический тест. Не забудьте сформулировать тестируемую гипотезу.

Задача 7

Исследуется зависимость среднедушевого потребления алкоголя по странам мира от различных факторов:

Модель №1:

$$ALCO_i = \beta_1 + \beta_2 * GDP_i + \beta_3 * MUSL_i + \beta_4 * BUDD_i + \beta_5 * HINDU_i + \varepsilon_i,$$

где $ALCO_i$ — среднедушевое потребление чистого спирта на человека (л), GDP_i — ВВП на душу населения (долларов США), $MUSL_i$, $BUDD_i$,

$HINDU_i$ — доли населения исповедующего, соответственно, мусульманство, буддизм и индуизм (в % от общей численности населения). В ходе МНК-оценивания модели на основе данных о 180 странах получены следующие результаты: сумма квадратов остатков[1] ESS=200, объясненная сумма квадратов RSS=300.

Также для проверки гипотезы о том, что религия не оказывает существенного влияния на потребление алкоголя, были оценены параметры второй модели:

$$\textbf{Модель №2: } ALCO_i = \beta_1 + \beta_2 * GDP_i + \varepsilon_i,$$

Во второй модели, по сравнению с первой, значение объясненной суммы квадратов RSS изменилось на 100.

(а) Вычислите R^2 в модели №1.

(б) Во второй модели, по сравнению с первой, значение RSS увеличилось или уменьшилось? Почему? Вычислите R^2 в модели №2.

(в) Влияет ли религия на потребление алкоголя? Сделайте вывод на основе соответствующего статистического теста.

Задача 8

На основе данных о 200 квартирах города Готэм было оценено следующее уравнение регрессии (все переменные оказались значимыми, $R^2 = 0{,}94$):

$$\ln \widehat{P}_i = 1{,}00 + 0{,}90 * \ln S_i + 0{,}20 * Center_i * \ln S_i + 0{,}03 * Center_i + \\ + 0{,}04 * Metro_i + 0{,}05 * Metro_i * Center_i$$

P_i — цена i-ой квартиры, тысяч долларов, S_i — площадь i-ой квартиры, квадратных метров, $Center_i$ — фиктивная переменная равная единице, если i-ая квартира расположена в центре города, и равная нулю в противном случае, $Metro_i$ — фиктивная переменная равная единице, если i-ая квартира расположена в пешей доступности от метро, и равная нулю если от квартиры до метро надо добираться на общественном транспорте.

(а) На сколько процентов при прочих равных условиях увеличивается цена квартиры в центре города, при увеличении ее площади на 1%?

(б) Для квартир, расположенных в центре города, на сколько процентов дороже квартира рядом с метро по сравнению с такой же квартирой, расположенной не рядом с метро?

---

[1] Обратите внимание, что в разных учебниках сумма квадратов остатков обозначается по-разному: где-то ESS (error sum of squares), где-то наоборот RSS (residuals sum of squares), а где-то другими способами. Мы, в рамках нашего пособия, используем вслед за учебником, авторами которого являются *Магнус, Катышев, Пересецкий,* обозначение ESS, но призываем читателя соблюдать бдительность при чтении разных источников.

(в) Для квартир, расположенных не в центре города, на сколько процентов дороже квартира рядом с метро по сравнению с такой же квартирой, расположенной не рядом с метро?

Задача 9

На рынке телевизоров некоторого города продаются телевизоры только трех фирм: «Альфа», «Бета» и «Гамма». Исследователь анализирует зависимость цены телевизора от диагонали экрана и марки производителя. В его распоряжении имеется информация о 100 моделях телевизоров. Для каждого наблюдения ему известна цена телевизора в долларах (обозначим ее P), длина диагонали экрана в дюймах (обозначим ее $Diag$) и марка производителя. Исследователь ввел следующие фиктивные переменные:

$Alfa_i$ — фиктивная переменная равная единице, если i-ая модель телевизора произведена фирмой «Альфа» и равная нулю во всех остальных случаях.

$Beta_i$ — фиктивная переменная равная единице, если i-ая модель телевизора произведена фирмой «Бета» и равная нулю во всех остальных случаях.

На основе доступных данных было оценено следующее уравнение регрессии (все переменные оказались значимыми):

$$\ln \widehat{P_i} = 2{,}00 + 0{,}05 * Diag_i + 0{,}07 * Alfa_i + 0{,}06 * Beta_i + 0{,}03 * Diag_i * Beta_i,$$

$$R^2 = 0{,}95$$

(а) На сколько процентов увеличивается цена телевизора фирмы «Бета» при увеличении диагонали его экрана на один дюйм?

(б) На сколько процентов увеличивается цена телевизора фирмы «Гамма» при увеличении диагонали его экрана на один дюйм?

(в) Как можно интерпретировать коэффициент при переменной $Alfa_i$? Заполните пропуски в формулировке: *При прочих равных условиях телевизоры фирмы «Альфа» на* ____________ *процентов дороже, чем* ____.

Задача 10

Рассматривается модель $y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i$, для которой выполнены все предположения классической линейной модели множественной регрессии, за одним исключением: в модели присутствует гетероскедастичность. Известна ее функциональная форма: $\sigma_i^2 = c^2 * x_i^2$.

Y	1,00	1,00	1,50	1,50	1,00
X	1,00	1,00	0,50	0,50	0,25

(а) Покажите, что если от исходной модели перейти к взвешенной модели:

$$\frac{y_i}{x_i} = \beta_1 * \frac{1}{x_i} + \beta_2 + u_i,$$

то в этой новой модели гетероскедастичность будет отсутствовать.

(б) Используя данные таблицы, вычислите оценки обобщенного метода наименьших квадратов параметров β_1 и β_2.

ОТВЕТЫ И РЕШЕНИЯ

Ответы к тестам

Тест	Вариант 1	Вариант 2	Вариант 3	Вариант 4
1.1	В	Г	В	В
1.2	А	В	В	Б
1.3	Г	В	Б	Г
1.4	Г	А	В	Г
1.5	Д	А	В	А
1.6	Г	В	А	А
2.1	В	Б	А	Г
2.2	Б	А	А	Г
2.3	Г	Б	В	В
2.4	А	Г	Г	Г

Ответы к задаче 1

(а) $\hat{y}_i = 26 + 2 * x_i$,

(б) $se\left(\hat{\beta}_2\right) = 4$, Расчетное значение тестовой статистики равно 0,5, тестируемая гипотеза не отклоняется.

(в) Пусть истинная модель имеет вид:

$$y_i = \beta_1 + \beta_2 x_i^{(1)} + \beta_3 x_i^{(2)} + \varepsilon_i$$

Исследователь ошибочно не включил в уравнение $x^{(2)}$ (в нашем случае это уровень образования) и оценил регрессию:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x^{(1)}$$

Каковы последствия?

$$E(\hat{\beta}_2) = E\left(\frac{Cov(x^{(1)}, y)}{Var(x^{(1)})}\right) = E\left(\frac{Cov(x^{(1)}, \beta_1 + \beta_2 x^{(1)} + \beta_3 x^{(2)} + \varepsilon)}{Var(x^{(1)})}\right) =$$

$$= \beta_2 + \frac{\beta_3 Cov(x^{(1)}, x^{(2)})}{Var(x^{(1)})}$$

По условию $\beta_3 > 0$, так как образование положительно влияет на зарплату, и ковариация отрицательная (мужчины менее образованны, чем женщины). Следовательно, величина смещения отрицательная, то есть математическое ожидание оценки ниже, чем истинное значение коэффициента. Следовательно, оценка **занижена**.

Ответы к задаче 2

(а) $\hat{y}_i = 25 - 3 * x_i$,

(б) $se\left(\hat{\beta}_2\right) = 2$, Расчетное значение тестовой статистики равно $-1,5$, тестируемая гипотеза не отклоняется.

(в) Смещение равно нулю, то есть оценка не смещена.

Ответы к задаче 3

(а) $\widehat{\beta_1} = 2$, $\widehat{\beta_2} = 4$, $\widehat{\beta_3} = -1$.

(б) Оценка ковариационной матрицы вектора оценок коэффициентов имеет вид:

2	0	0
0	4	–2
0	–2	2

(в) При уровне значимости 5% гипотеза не отклоняется.

Ответы к задаче 4

(а) $\widehat{\beta_1} = 24$, $\widehat{\beta_2} = 12$, $\widehat{\beta_3} = -2$.

(б) Оценка ковариационной матрицы вектора оценок коэффициентов имеет вид:

4	0	0
0	4	–2
0	–2	2

(в) При уровне значимости 5% гипотеза отклоняется.

Решение задачи 5

(а) $R^2 = 1 - \dfrac{200}{2000} = 0,9$

$$R^2_{adj} = R^2 - \frac{k-1}{n-k}\left(1 - R^2\right) = 0,9 - \frac{3-1}{20-3} * \left(1 - 0,9\right) = 0,89$$

Стандартная ошибка регрессии: $\sqrt{\dfrac{200}{20-3}} = 3,43$

(б) Если обозначить коэффициенты в рассматриваемой модели стандартным образом: $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \varepsilon_i$, то тестируемая гипотеза может быть записана так:

$$H_0 : \beta_2 = \beta_3 = 0.$$

Расчетное значение: $F=\frac{R^2}{1-R^2}*\frac{n-k}{k-1}=\frac{0,9}{0,1}*\frac{17}{2}=76,5.$

Критическое значение при уровне значимости 1% $F(2,17)=6,11$. 76,5>6,11 => вывод: гипотеза не принимается, уравнение в целом значимо.

(в) Проверяемая гипотеза: $H_0:\beta_2=0$. Расчетное значение 6,9/0,3=23. Критическое значение при уровне значимости 1% $t(20-3)=2,898$. $23>2,898$ => вывод: гипотеза не принимается, переменная значима.

(г) Проверяемая гипотеза: $H_0:\beta_2=7$. Расчетное значение $\frac{6,9-7}{0,3}=-0,33$. Критическое значение при уровне значимости 1% $t(20-3)=2,898$. $0,3<2,898$ => вывод: мы не можем отклонить гипотезу о том, что коэффициент β_2 равен 7.

(д) $\beta_2\in(6,9-0,3*2,898,\quad 6,9+0,3*2,898)$

$\beta_2\in(6,03,\quad 7,77)$.

(е) Если обозначить коэффициенты в новой модели стандартным образом: $y_i=\beta_1+\beta_2x_i+\beta_3w_i+\beta_4p_i+\beta_5s_i+\varepsilon_i$, то тестируемая гипотеза может быть записана так:

$$H_0:\beta_4=\beta_5=0.$$

Расчетное значение: $F=\frac{R_{UR}^2-R_R^2}{1-R_{UR}^2}*\frac{n-k}{q}=\frac{0,95-0,9}{1-0,95}*\frac{20-5}{2}=7,5.$

Критическое значение при уровне значимости 1% $F(2,15)=6,36$. 7,5>6,36 => Вывод: гипотеза не принимается. Длинная регрессия значимо лучше, чем короткая. То есть переменные добавлять стоило. (Хотя, конечно, в реальных исследованиях лучше не оценивать уравнение с четырьмя переменными всего по 20 точкам.)

Решение задачи 6

(а) Если выписать модель №1 стандартным образом:

$$y_i=\beta_1+\beta_2*x_i^{(2)}+\beta_3*x_i^{(3)}+\varepsilon_i,$$

то тестируемая гипотеза может быть записана так:

$$H_0:\beta_2=\beta_3=0.$$

Расчетное значение: $F=\frac{R^2}{1-R^2}*\frac{n-k}{k-1}=\frac{0,21}{0,79}*\frac{27}{2}=3,589.$

Критическое значение при уровне значимости 5% $F(2,30-3)=3,35$. 3,589>3,35 => вывод: нулевая гипотеза не принимается, уравнение в целом значимо.

(б) Тестируемая гипотеза для проверки значимости переменной $x^{(3)}$ может быть записана так:

$$H_0 : \beta_3 = 0.$$

Расчетное значение 4,5/3=1,5. Критическое значение при уровне значимости 5% $t(30-3)=2{,}052$. $1{,}5<2{,}052$ => вывод: нулевая гипотеза не отклоняется, переменная $x^{(3)}$ не оказывает значимого влияния на y.

(в) Для сравнения моделей №1 и №2 не стоит использовать обычный R^2, так как в моделях различно количество регрессоров. Чем больше в модели регрессоров, тем выше значение обычного R^2. Следовательно, такой способ сравнения моделей с разным числом регрессоров на предмет лучшего соответствия данным некорректен. Скорректированный R^2_{adj} же позволяет сравнивать модели с одинаковой зависимой переменной, но с разным числом регрессоров, учитывая штраф за число переменных. Будем использовать скорректированный R^2_{adj}.

$$R^2_{adj(1)} = R_1^2 - \frac{k-1}{n-k} * \left(1-R_1^2\right) = 0{,}21 - \frac{2}{27} * 0{,}79 = 0{,}15$$

$$R^2_{adj(2)} = R_2^2 - \frac{k-1}{n-k} * \left(1-R_2^2\right) = 0{,}23 - \frac{4}{25} * 0{,}77 = 0{,}11$$

Видим, что $R^2_{adj(1)} > R^2_{adj(2)}$, следовательно, модель №1 лучше соответствует данным.

(г) Теперь сравним модели №1 и №2 на предмет лучшего соответствия данным, используя тест на «короткую» и «длинную» регрессии. Соответствующая гипотеза может быть записана следующим образом:

$$H_0 : \beta_4 = \beta_5 = 0.$$

Расчетное значение: $F = \dfrac{R^2_{UR} - R^2_R}{1-R^2_{UR}} * \dfrac{n-k}{q} = \dfrac{0{,}23-0{,}21}{1-0{,}23} * \dfrac{30-5}{2} = 0{,}32.$

Критическое значение при уровне значимости 5% $F(2,25)=3{,}385$. 0,32<3,385 => вывод: мы не можем отклонить нулевую гипотезу. Переменные $x^{(4)}$ и $x^{(5)}$ не оказывают значимого влияния на y. Следовательно, нет смысла их добавлять в регрессию, можно спокойно пользоваться «короткой» регрессией.

Решение задачи 7

(а) $R_1^2 = \dfrac{RSS}{TSS} = \dfrac{RSS}{RSS+ESS} = \dfrac{300}{300+200} = 0{,}6.$

(б) Значение TSS в модели №2 при уменьшении числа регрессоров не изменилось, поскольку оно зависит лишь от $ALCO_i$ и $\overline{ALCO}$, а не от набора используемых регрессоров (TSS $= \sum_{i=1}^{180}\left(ALCO_i - \overline{ALCO}\right)^2$).

Значение ESS увеличилось (при уменьшении количества регрессоров увеличиваются отклонения предсказанных значений от истинных и, следовательно, сумма их квадратов). RSS, соответственно, уменьшилось на 100, так как TSS=RSS+ESS.

Поэтому имеем: $R_2^2 = \frac{200}{500} = 0{,}4.$

(в) Для того чтобы выяснить, влияет ли религия на потребление алкоголя, сравним модель №1 и №2 с помощью теста на «короткую» и «длинную» регрессии.

Имея модель №1 в виде:

$$ALCO_i = \beta_1 + \beta_2 * GDP_i + \beta_3 * MUSL_i + \beta_4 * BUDD_i + \beta_5 * HINDU_i + \varepsilon_i,$$

запишем тестируемую гипотезу:

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0.$$

Расчетное значение: $F = \frac{R_{UR}^2 - R_R^2}{1 - R_{UR}^2} * \frac{n-k}{q} = \frac{0{,}6 - 0{,}4}{1 - 0{,}6} * \frac{180-5}{3} = 29{,}17.$

Критическое значение при уровне значимости 5% $F(3{,}175) = 2{,}605$. 29,17>2,605 => Вывод: гипотеза не принимается. Следовательно, религия влияет на потребление алкоголя, и лучше пользоваться моделью №1.

Решение задачи 8

(а) Если рассматриваются только квартиры в центре города, то переменная $Center_i = 1$, а исходное оцененное уравнение регрессии принимает вид:

$$\ln \widehat{P}_i = 1{,}03 + 1{,}1 * \ln S_i + 0{,}09 * Metro_i.$$

$$\frac{\Delta P}{P} \approx 1{,}1 * \frac{\Delta S}{S}$$

Следовательно, при прочих равных условиях при увеличении площади квартиры, расположенной в центре города, на 1%, ее цена возрастает **на 1,1%**.

(б) Для преобразованной в пункте (а) модели

$$\frac{\Delta P}{P} \approx 0{,}09 * \Delta Metro.$$

Следовательно, при прочих равных условиях, квартира, расположенная в центре рядом с метро, стоит **на 9% дороже** аналогичной квартиры, расположенной не рядом с метро.

(в) Если рассматриваются только квартиры, расположенные не в центре города, то переменная $Center_i = 0$, а исходное оцененное уравнение регрессии принимает вид:

$$\ln \widehat{P}_i = 1{,}00 + 0{,}9 * \ln S_i + 0{,}04 * Metro_i.$$

$$\frac{\Delta P}{P} \approx 0{,}04 * \Delta Metro.$$

Следовательно, при прочих равных условиях, квартира, расположенная не в центре рядом с метро, стоит **на 4% дороже** аналогичной квартиры, расположенной не рядом с метро.

Решение задачи 9

(а) Поскольку в данном пункте рассматриваются только телевизоры фирмы «Бета», то переменная $Beta_i = 1$, а $Alfa_i = 0$. Тогда исходное уравнение принимает вид:

$$\ln \widehat{P}_i = 2{,}06 + 0{,}08 * Diag_i.$$

$$\frac{\Delta P}{P} \approx 0{,}08 * \Delta Diag$$

Следовательно, при увеличении диагонали экрана телевизора фирмы «Бета» на один дюйм, его цена увеличивается **на 8%.**

(б) Поскольку в данном пункте рассматриваются только телевизоры фирмы «Гамма», то переменная $Beta_i = 0$, а $Alfa_i = 0$. Тогда исходное уравнение принимает вид:

$$\ln \widehat{P}_i = 2{,}00 + 0{,}05 * Diag_i.$$

$$\frac{\Delta P}{P} \approx 0{,}05 * \Delta Diag$$

Следовательно, при увеличении диагонали экрана телевизора фирмы «Гамма» на один дюйм, его цена увеличивается **на 5%.**

(в) Нам нужно сравнить телевизоры марки «Альфа» с телевизорами какой-нибудь другой марки. Очевидно, что удобнее всего сравнивать с телевизорами марки «Гамма» (телевизоры марки «Гамма» — так называемая *эталонная* категория). Поэтому:

При прочих равных условиях телевизоры фирмы «Альфа» на 7 ***процентов*** *дороже, чем* ***телевизоры марки «Гамма»****.*

Обратите внимание, что в этой задаче мы имеем дело с логарифмически-линейной моделью. В этой модели использование приближенных формул корректно только если коэффициент при переменной не слишком велик (меньше одной десятой). В нашей задаче это так. В противном случае пришлось бы пользоваться точной формулой.

Решение задачи 10

(а) Рассмотрим две модели:

$$y_i = \beta_1 + \beta_2 * x_i + \varepsilon_i,$$

$$\frac{y_i}{x_i} = \beta_1 * \frac{1}{x_i} + \beta_2 + u_i.$$

Вторая модель получается с помощью деления обеих частей уравнения регрессии на x_i. Тогда случайные ошибки этой модели имеют вид $u_i = \frac{\varepsilon_i}{x_i}$, а $Var(u_i) = Var\left(\frac{\varepsilon_i}{x_i}\right) = E\left(\frac{\varepsilon_i}{x_i}\right)^2 = \frac{1}{x_i^2} * E(\varepsilon_i)^2 = \frac{c^2 * x_i^2}{x_i^2} = c^2 = const.$ В модифицированной модели дисперсия случайной ошибки постоянна, и, следовательно, в данной модели гетероскедастичность отсутствует.

(б) Используя данные из условия, вычислим оценки обобщенного метода наименьших квадратов параметров β_1 и β_2.

Пусть $y_i^* = \frac{y_i}{x_i}$, $z_i = \frac{1}{x_i}$, $u_i = \frac{\varepsilon_i}{x_i}$. Тогда в новых обозначениях модель принимает вид: $y_i^* = \beta_2 + \beta_1 * z_i + u_i$, а МНК-оценки коэффициентов равны:

$$\widehat{\beta_1} = \frac{\overline{y^* * z} - \overline{y^*} * \overline{z}}{\overline{z^2} - (\overline{z})^2}, \quad \widehat{\beta_2} = \overline{y^*} - \widehat{\beta_1} * \overline{z}.$$

	y_i	x_i	$y_i^* = \frac{y_i}{x_i}$	$z_i = \frac{1}{x_i}$	$y_i^* * z_i$	z_i^2
1	1	1	1	1	1	1
2	1	1	1	1	1	1
3	1,5	0,5	3	2	6	4
4	1,5	0,5	3	2	6	4
5	1	0,25	4	4	16	16
Сумма			**12**	**10**	**30**	**26**
Среднее			**2,4**	**2**	**6**	**5,2**

Тогда МНК-оценки в модифицированной модели равны:

$$\widehat{\beta_1} = \frac{6 - 2{,}4 * 2}{5{,}2 - 2^2} = 1, \quad \widehat{\beta_2} = 2{,}4 - 1 * 2 = 0{,}4.$$

Оцененное модифицированное уравнение регрессии имеет вид: $\widehat{y_i^*} = 0{,}4 + z_i$, а уравнение в исходных обозначениях: $\widehat{y_i} = 1 + 0{,}4 * x_i$.

Ответ на пункт (б): $\widehat{y_i} = 1 + 0{,}4 * x_i$.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. *Вербик М.* Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Научн. ред. и предисл. С. А. Айвазяна. — М.: Научная книга, 2008.
2. *Доугерти К.* Введение в эконометрику: Учебник. 3-е изд. / Пер. с англ. — М.: ИНФРА-М, 2009.
3. *Магнус Я. Р.*, *Катышев П. К.*, *Пересецкий А. А.* Эконометрика. Начальный курс: Учеб. — 6-е изд., перераб. и доп. — М.: Дело, 2004.
4. *Stock J.*, *Watson M.* Introduction to econometrics. Third Edition. — Pearson, Addison Wesley. 2010.

Картаев Филипп

Сборник заданий по эконометрике
для подготовки к вступительным экзаменам
в магистратуру экономического факультета
МГУ имени М.В. Ломоносова
(задачник)

Научное электронное издание