МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Воронежский государственный технический университет»

Кафедра экономической безопасности

ЭКОНОМЕТРИКА

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к изучению дисциплины, выполнению лабораторных работ,
самостоятельной и курсовой работ для обучающихся
по специальности 38.05.01 "Экономическая безопасность"
(специализации "Экономико-правовое обеспечение экономической безопасности",
"Экономика и организация производства на режимных объектах")
всех форм обучения

Воронеж 2022

УДК 519.862.7:330.43(07)
ББК 65в6я7

Составитель
д-р экон. наук, проф. С. В. Амелин

Эконометрика: методические указания к изучению дисциплины, выполнению лабораторных работ, самостоятельной и курсовой работ для обучающихся по специальности 38.05.01 "Экономическая безопасность" (специализации "Экономико-правовое обеспечение экономической безопасности", "Экономика и организация производства на режимных объектах") всех форм обучения / ФГБОУ ВО «Воронежский государственный технический университет»; сост. С. В. Амелин. – Воронеж: Изд-во ВГТУ, 2022. – 41 с.

Основной целью указаний является изучение ключевых понятий дисциплины, приобретение умений и навыков обоснования принятия решений посредством применения эконометрического анализа и прогнозирования в целях обеспечения и повышения эффективности деятельности предприятий и организаций.

Предназначены для студентов 3 курса при изучении дисциплины «Эконометрика».

Методические указания подготовлены в электронном виде и содержатся в файле МУ_Эконометрика_ЭБ_2022. pdf.

Библиогр.: 5 назв.

УДК 519.862.7:330.43(07)
ББК 65в6я7

Рецензент - И. Ф. Елфимова, канд. экон. наук, доцент кафедры экономической безопасности ВГТУ

Издается по решению редакционно-издательского совета Воронежского государственного технического университета

ВВЕДЕНИЕ

Целью изучения дисциплины является формирование у обучающихся комплекса знаний, умений и практических навыков использования современных эконометрических методов и инструментария эконометрического моделирования при решении профессиональных задач в сфере обеспечения экономической безопасности предприятий и организаций.

Задачами дисциплины «Эконометрика» являются:

– применение статистико-математического инструментария построения эконометрических моделей для решения экономических задач, на основе выявления количественных характеристик тенденций и взаимосвязей;

– использование эконометрических моделей исследуемых процессов, явлений и объектов для решения задач, относящихся к области профессиональной деятельности, анализ, оценка, интерпретация полученных результатов и обоснование выводов;

– прогнозирование на основе экономико-математических, эконометрических моделей развития экономических процессов и явлений.

Результатом изучения дисциплины «Эконометрика» является освоение компетенций:

ОПК-1 - Способен использовать знания и методы экономической науки, применять статистико-математический инструментарий, строить экономико-математические модели, необходимые для решения профессиональных задач, анализировать и интерпретировать полученные результаты.

Методические указания включают содержание тем изучаемой дисциплины в соответствии с рабочей программой, перечень лабораторных работ, рекомендации по выполнению курсовой и самостоятельной работы, а также перечень рекомендуемой литературы и вопросы к экзамену и задания для самопроверки.

1. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ИЗУЧЕНИЮ ДИСЦИПЛИНЫ

<u>Тема 1. Парный регрессионный анализ</u>

Основные понятия эконометрического моделирования. Линейная парная регрессия. Метод наименьших квадратов (МНК). Коэффициент корреляции. Показатели качества уравнения регрессии. Оценка параметров парной регрессионной модели. Свойства коэффициентов регрессии. Теорема Гаусса-Маркова. Интервальная оценка функции регрессии и ее параметров. Оценка значимости уравнения регрессии. *Самостоятельное изучение* Коэффициент детерминации

<u>Тема 2. Статистические гипотезы и их проверка.</u>

Статистические гипотезы и основные принципы их проверки. Виды статистических критериев. Односторонние и двусторонние критерии. Мощность критерия. *Самостоятельное изучение* Распределения, используемые при проверке гипотез

Тема 3. Множественный регрессионный анализ.

Классическая нормальная линейная модель множественной регрессии. Оценка параметров классической регрессионной модели с помощью МНК. МНК в матричной форме. Ковариационная матрица и ее выборочная оценка. Оценка качества модели. Определение доверительных интервалов для коэффициентов и функции регрессии. Оценка значимости множественной регрессии. Коэффициенты детерминации R^2 и $R^{\wedge 2}$. Процедура шаговой регрессии. Проблема мультиколлинеарности факторов. Отбор наиболее существенных объясняющих переменных в регрессионной модели.

Тема 4. Временные ряды и их прогнозирование.

Стационарные временные ряды и их характеристики. Автокорреляционная функция. Аналитическое выравнивание (сглаживание) временного ряда (выделение неслучайной компоненты). Прогнозирование на основе моделей временных рядов. Экстраполяция временных рядов. Доверительные интервалы прогноза. Проверка адекватности выбранных моделей. Характеристика точности моделей.

Тема 5. Обобщенная линейная модель.

Гетероскедастичность и автокорреляция. Обобщенная линейная модель множественной регрессии. Обобщенный метод наименьших квадратов. Метод взвешенных наименьших квадратов. Стандартные ошибки и их корректировка. Гетероскедастичность пространственной выборки. Тесты на гетероскедастичность. Устранение гетероскедастичности. Автокорреляция остатков временного ряда. Положительная и отрицательная автокорреляция. Определение порядка ARMA моделей.

Тема 6. Системы эконометрических уравнений.

Системы эконометрических уравнений. Общий вид системы одновременных уравнений. Модель спроса и предложения. Структурная и приведенная формы одновременных уравнений. Косвенный и двухшаговый метод наименьших квадратов. Проблемы идентифицируемости. Метод инструментальных переменных. Методы оценивания параметров структурных моделей.

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ И ЛАБОРАТОРНЫХ РАБОТ

Методические указания по практическим занятиям и лабораторным работам содержатся в пособии 65(ББК) А615 Амелин С.В. Эконометрика: практикум: учеб. пособие / С.В. Амелин. Воронеж: ФГБОУ ВО «Воронежский государственный технический университет», 2016. 125 с. – Режим доступа https://education.cchgeu.ru/course/view R:\Литература\для СПЕЦИАЛИСТОВ\ Эконометрика

Перечень практических занятий для всех форм обучения.

Тема 1. Парный регрессионный анализ

Практическое занятие №1. Расчет ковариации, дисперсии, корреляции

Практическое занятие №2. Парная регрессия по методу наименьших квадратов. Построение нелинейных моделей

Тема 2. Статистические гипотезы и их проверка.

Практическое занятие №3. Оценка значимости уравнения регрессии. Проверка статистических гипотез

Тема 3. Множественный регрессионный анализ

Практическое занятие № 4. Расчет множественных регрессионных моделей. Регрессионные модели в матричной форме

Практическое занятие № 5. Построение доверительных интервалов

Практическое занятие № 6. Модели с переменной структурой

Практическое занятие № 7. Мультиколлинеарность. Частная корреляция

Тема 4. Временные ряды и их прогнозирование.

Практическое занятие № 8. Методы прогнозирования на основе временных рядов

Перечень лабораторных работ

Тема 1. Парный регрессионный анализ

Лабораторная работа №1. Расчет параметров линейной модели методом наименьших квадратов

Лабораторная работа №2. Расчет параметров линейной модели при аппроксимации опытных данных с помощью функций Excel

Лабораторная работа №3. Графическое моделирование линейных и нелинейных зависимостей

Тема 2. Статистические гипотезы и их проверка

Лабораторная работа №4. Применение инструмента электронной таблицы Excel *Анализ данных* при определении параметров линейных моделей. Проверка статистических гипотез

Тема 3. Множественный регрессионный анализ

Лабораторная работа №5. Определение параметров нелинейных зависимостей в форме, определенной пользователем

Лабораторная работа №6. Точечные и интервальные оценки линейной модели

Лабораторная работа №7. Регрессионная модель с гетероскедастичностью. Метод взвешенных наименьших квадратов

Тема 4. Временные ряды и их прогнозирование

Лабораторная работа №8. Регрессионный анализ временных рядов

Тема 6. Системы эконометрических уравнений

Лабораторная работа №9. Системы одновременных уравнений

3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Рабочей программой дисциплины «Эконометрика» предусмотрена самостоятельная работа студентов. Виды самостоятельной работы:

1) подготовка к лекциям и самостоятельная проработка материала; 2) подготовка к практическим занятиям и лабораторным работам; 3) подготовка к выполнению курсовой работы; 4) самоподготовка к итоговой проверке знаний; 5) выполнение домашних заданий.

Подготовка к лекциям и самостоятельная проработка материала является обязательным видом самостоятельной работы и предполагает предварительное ознакомление студента с вопросами предстоящей лекции с целью наиболее эффективного усвоения материала. Особое внимание следует уделить вопросам, выносимым на самостоятельное изучение.

Подготовка к практическим занятиям заключается в выполнении определенных заданий к каждому практическому занятию. Выполнение заданий в качестве подготовки к практическим занятиям является обязательным и оценивается преподавателем как элемент общей успеваемости студента.

Подготовка к лабораторным занятиям заключается в проработке теоретического материала к каждой лабораторной работе. Защита лабораторных работ является обязательной и оценивается преподавателем как элемент общей успеваемости студента.

Отчёты по лабораторным работам оформляются как индивидуальный отчет студента, на листах формата А4 в соответствии с требованиями нормокнтроля ВГТУ. Наличие титульного листа обязательно.

Самоподготовка к итоговой проверке знаний предполагает самостоятельную проработку материала, опираясь на содержание лекций и практических занятий, вопросы, выносимые на самостоятельное изучение.

Студент допускается к итоговой аттестации (экзамену) на основании посещения лекций и практических занятий, а также выполнения курсовой работы. На экзамен выносятся основные вопросы, изучаемые в течение семестра. Экзамен предполагает ответы на теоретические вопросы, выполнение стандартного и прикладного задания.

4. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ КУРСОВОЙ РАБОТЫ

В соответствии с учебным планом курсовая работа является обязательным структурным элементом изучения дисциплины «Эконометрика».

Задачи, решаемые при выполнении курсовой работы: систематизация и закрепление полученных теоретических значений и практических умений по дисциплине; углубление теоретических знаний в соответствии с выбранной темой; развитие навыков научно-исследовательской работы (развитие умения обобщать, критически оценивать теоретические положения, вырабатывать свою точку зрения); формирование профессиональных навыков, умение применять теоретические знания при решении поставленных задач; развитие творческой инициативы, самостоятельности.

Курсовая работа представляет собой пояснительную записку объемом 25-30 страниц, должна содержать следующие разделы: введение, теоретический раздел, расчётно-аналитический раздел, глоссарий терминов по теме исследования, заключение, список литературы, приложения. Указанные разделы являются обязательными. Введение и заключение не нумеруются как разделы.

Во *введении* курсовой работы обосновывается актуальность темы, цель и задачи работы, приводится краткая характеристика предмета исследования, излагается краткое содержание основных разделов работы.

В *теоретической части работы* раскрывается сущность и содержание исследуемой проблемы, рассматриваются теоретические и методические аспекты ее решения на основе обзора литературных источников по теме курсовой работы. В теоретическом разделе должны быть проанализированы различные точки зрения специалистов по проблеме, выявлены тенденции развития. Следует рассмотреть основные понятия по теме курсовой работы, выделить методические подходы, классификации, проанализировать факторы, содержание основных подходов и принципов построения теоретических концепций, определить проблемы, тенденции. Теоретический раздел выполняется в формате критического обзора точек зрения и взглядов разных специалистов (теоретиков и практиков) по выбранному направлению.

Выполнение *расчётно-аналитического раздела* курсовой работы, состоящего из двух заданий, осуществляется по вариантам в соответствии со списком группы (студент, находящийся по списку на 11 месте, выполняет 1 вариант и т.д.).

Содержание каждого параграфа данного раздела должно включать: постановку задачи в соответствии с решаемой проблемой, исходные данные, эконометрическую модель, расчётные формулы, решение задачи, графическое представление моделей, анализ полученных результатов, выводы. Скриншоты применения информационных технологий при решении задач размещаются в тексте параграфов. Все рисунки и таблицы должны иметь нумерацию и названия, формулы должны иметь пояснения к используемым переменным. Каждый параграф должен содержать библиографические ссылки на используемые источники литературы.

Глоссарий терминов состоит из тематического словаря используемых в работе терминов и понятий, каждое понятие или термин должен быть определен и дана ссылка на автора интерпретации этого понятия или определения.

Заключение посвящается обобщению полученных результатов, оценке их эффективности.

Использование современных информационных технологий и программных *средств*. Каждая курсовая работа должна включать как подробные расчёты по соответствующим эконометрическим моделям, так и использование современных информационных технологий и программных средств, скриншоты с пояснениями проводимых расчётов включаются в соответствующие параграфы.

Курсовая работа должна завершаться списком использованных литературных источников. Используемые публикации должны быть современными, по дате опубликования не ранее, чем за пять предшествующих выполнению курсовой работы лет. *Список литературы* должен включать не менее 20 источников, в т.ч. периодические издания и ресурсы сети Интернет. Не допускается: общее описание практики обоснования принятия решений с помощью математического моделирования без анализа проблем и выявления слабых сторон; переписывание статей без ссылок и анализа. В список литературы включаются только использованные в тексте курсовой работы источники.

Примерное распределение объема курсовой работы по разделам: введение 1-2 с.; теоретическая часть 10-15 с.; расчётно-аналитическая часть 15-20 с.; глоссарий 1-2 с.; заключение 1-2 с.

Тематика курсовых работ

1. Эконометрические методы анализа временных рядов. Применение фиктивных переменных.
2. Нелинейные эконометрические модели регрессионного анализа в исследовании рынка
3. Системы одновременных эконометрических уравнений и область их применения.
4. Исследование проблемы мультиколлинеарности в эконометрике
5. Исследование проблемы автокорреляции случайных отклонений в эконометрике.
6. Исследование применения моделей с фиктивными переменными в эконометрике и тест Чоу.
7. Критерий Дарбина-Уотсона в эконометрике.
8. Тест ранговой корреляции Спирмена как способ обнаружения гетероскедастичности.
9. Эконометрические модели с распределённым лагом.
10. Исследование сезонных и циклических колебаний в эконометрике с использованием ряда Фурье.
11. Метод максимального правдоподобия в эконометрике.
12. Обобщённый метод наименьших квадратов.
13. Тобит-модели в эконометрике.
14. Динамические эконометрические модели.
15. Адаптивные модели временных рядов и сезонных явлений.
16. Двухшаговый метод наименьших квадратов.

17. Трёхшаговый метод наименьших квадратов.
18. Модели выбора с бинарной зависимой переменной. Logit и Probit модели.
19. Оценка стоимости объектов недвижимости: эконометрический подход.
20. Исследование производительности труда с помощью эконометрических моделей.

ИСХОДНЫЕ ДАННЫЕ ДЛЯ РАСЧЁТНО-АНАЛИТИЧЕСКОЙ ЧАСТИ

Индивидуальное задание № 1. Парный регрессионный анализ

$$y = b_0 + b_1x_1 + \varepsilon$$

Построить диаграммы рассеяния. Определить выборочную ковариацию, среднее квадратическое отклонение для величин X и Y, выборочную дисперсию переменной X, коэффициенты уравнения регрессии, коэффициент корреляции, выборочную остаточную дисперсию, 95% доверительный интервал для функции регрессии, 95% доверительный интервал для индивидуальных значений зависимой переменной (значение из середины таблицы – 6-я сторока), 95% доверительный интервал для параметров регрессионной модели (для коэффициента регрессии, для дисперсии и для среднего квадратического отклонения случайной составляющей), коэффициент детерминации. Оценить значимость уравнения регрессии, значимость коэффициентов регрессии, значимость коэффициента корреляции. Принять уровень значимости $\alpha = 0,05$. (В качестве исходных данных для первого задания взять значения переменной x_1 и y для варианта, соответствующего последней цифре в зачётной книжке плюс 1, например, для номера зачётной книжки 2509 выбирается вариант 10)

Индивидуальное задание № 2. Множественный регрессионный анализ

$$y = b_0 + b_1x_1 + b_2x_2 + \varepsilon$$

Задание 2. Провести оценку параметров уравнения связи для многофакторной модели, проверить значимость и адекватность полученного уравнения (по F–критерию Фишера) и каждого из его параметров (по t-критерию Стьюдента). Рассчитайте прогнозное значение результата, если прогнозные значения факторов составляют 80% от их максимальных значений. Принять уровень значимости $\alpha = 0,05$.

Найти 95% доверительные интервалы для параметров уравнения, для дисперсии, для функции регрессии и для индивидуальных значений. Провести анализ на мультиколлинеарность. Определить и проанализировать частные коэффициенты корреляции. Вычислить коэффициент множественной корреляции и коэффициент детерминации и проанализировать их. Определить и проанализировать стандартизованные коэффициенты регрессии и коэффициенты эластичности. (В качестве исходных данных для первого задания взять значения переменной x_1, x_2 и y для варианта, соответствующего последней цифре в зачётной книжке, например, для номера зачётной книжки 2509 выбирается вариант 9)

Вариант 1

<table>
<tr><td rowspan="13">Имеются следующие данные о выработке литья на одного работающего x_1 (т), браке литья x_2 (%) и себестоимости одной тонны литья y (руб.) по литейным цехам заводов:</td><td>№ п/п</td><td>x_1</td><td>x_2</td><td>y</td></tr>
<tr><td>1</td><td>14,6</td><td>4,2</td><td>239</td></tr>
<tr><td>2</td><td>13,5</td><td>6,7</td><td>254</td></tr>
<tr><td>3</td><td>21,5</td><td>5,5</td><td>262</td></tr>
<tr><td>4</td><td>17,4</td><td>7,7</td><td>251</td></tr>
<tr><td>5</td><td>44,8</td><td>1,2</td><td>158</td></tr>
<tr><td>6</td><td>111,9</td><td>2,2</td><td>101</td></tr>
<tr><td>7</td><td>20,1</td><td>8,4</td><td>259</td></tr>
<tr><td>8</td><td>28,1</td><td>1,4</td><td>186</td></tr>
<tr><td>9</td><td>22,3</td><td>4,2</td><td>204</td></tr>
<tr><td>10</td><td>25,3</td><td>0,9</td><td>198</td></tr>
<tr><td>11</td><td>56,0</td><td>1,3</td><td>170</td></tr>
<tr><td>12</td><td>40,2</td><td>1,8</td><td>173</td></tr>
</table>

Вариант 2

<table>
<tr><td rowspan="13">По предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).</td><td>№ п/п</td><td>Y</td><td>X_1</td><td>X_2</td></tr>
<tr><td>1</td><td>3,9</td><td>10,0</td><td>7,0</td></tr>
<tr><td>2</td><td>3,9</td><td>14,0</td><td>7,0</td></tr>
<tr><td>3</td><td>3,7</td><td>15,0</td><td>7,0</td></tr>
<tr><td>4</td><td>4,0</td><td>16,0</td><td>7,0</td></tr>
<tr><td>5</td><td>3,8</td><td>17,0</td><td>7,0</td></tr>
<tr><td>6</td><td>4,8</td><td>19,0</td><td>7,0</td></tr>
<tr><td>7</td><td>5,4</td><td>19,0</td><td>8,0</td></tr>
<tr><td>8</td><td>4,4</td><td>20,0</td><td>8,0</td></tr>
<tr><td>9</td><td>5,3</td><td>20.0</td><td>8,0</td></tr>
<tr><td>10</td><td>6,8</td><td>20,0</td><td>10,0</td></tr>
<tr><td>11</td><td>6,0</td><td>21,0</td><td>9,0</td></tr>
<tr><td>12</td><td>6,4</td><td>22,0</td><td>11,0</td></tr>
</table>

Вариант 3

<table>
<tr><td rowspan="13">Изучается влияние стоимости основных и оборотных средств на величину валового дохода торговых предприятий. Для этого по следующим предприятиям были получены данные</td><td>Номер предприятия</td><td>Валовой доход за год, млн руб.</td><td>Среднегодовая стоимость, млн руб. основных фондов</td><td>Среднегодовая стоимость, млн руб. оборотных средств</td></tr>
<tr><td>1</td><td>118</td><td>105</td><td>203</td></tr>
<tr><td>2</td><td>28</td><td>56</td><td>63</td></tr>
<tr><td>3</td><td>17</td><td>54</td><td>45</td></tr>
<tr><td>4</td><td>50</td><td>63</td><td>113</td></tr>
<tr><td>5</td><td>56</td><td>28</td><td>121</td></tr>
<tr><td>6</td><td>102</td><td>50</td><td>88</td></tr>
<tr><td>7</td><td>116</td><td>54</td><td>110</td></tr>
<tr><td>8</td><td>124</td><td>42</td><td>56</td></tr>
<tr><td>9</td><td>114</td><td>36</td><td>80</td></tr>
<tr><td>10</td><td>154</td><td>106</td><td>237</td></tr>
<tr><td>11</td><td>115</td><td>88</td><td>160</td></tr>
<tr><td>12</td><td>98</td><td>46</td><td>75</td></tr>
</table>

Вариант 4

Исследовать зависимость уровня рентабельности от производительности труда (x_1) и размера заработной платы (x_2) по данным следующих предприятий	№ п/п	Производительность труда, д.е (x_1)	Заработная плата, д.е. (x_2)	Уровень рентабельности, % (y)
	1	7134	149	5,20
	2	5415	142	4,41
	3	7633	151	5,23
	4	10259	165	6,72
	5	14620	175	7,14
	6	8736	155	4,40
	7	5590	144	3,78
	8	10212	165	6,83
	9	11586	171	6,07
	10	9156	161	6,10
	11	12501	173	7,10
	12	11274	168	6,21

Вариант 5

Исследовать зависимость производства валовой продукции, млн. р (y) от суммы оборотных производственных фондов (x_1) и среднегодовая численность рабочих (x_2).	X_1	X_2	y
	3	250	4,1
	5	280	5,6
	2	240	3,1
	6	290	6,4
	7	510	8,6
	5	270	5,5
	10	540	11
	8	300	7,9
	3	250	3,8
	7	510	8,5
	4	275	4,5
	8	450	9,9

Вариант 6

Исследовать зависимость уровня рентабельности (y, %) от удельного веса продукции собственного производства в товарообороте X_1 % и удельного веса покупной продукции в товарообороте X_2 -%	№ п/п	X_1	X_2	Y
	1	56,6	43,4	7,92
	2	51,6	48,4	8,17
	3	48,5	51,5	8,00
	4	63,2	36,8	7,04
	5	47,6	52,4	9,14
	6	60,8	39,2	9,00
	7	32,2	67,8	9,13
	8	53,2	46,8	7,81
	9	41,6	58,4	9,17
	10	76,1	23,9	9,01
	11	43,6	56,4	6,43
	12	52,8	47,2	5,64

Вариант 7

	№ п/п	y	x_1	x_2
Исследовать зависимость фондоотдачи (y) в процентах на единицу основных производственных фондов (ОПФ) от среднечасовой производительности оборудования (x_1) и удельного веса активной части ОПФ (x_2).	1	26	15	39
	2	33	36	40
	3	24	26	35
	4	29	24	48
	5	42	15	53
	6	24	33	42
	7	52	44	54
	8	26	34	54
	9	26	63	50
	10	45	44	53
	11	27	43	46
	12	54	31	50

Вариант 8

	№ п/п	y	x_1	x_2
По данным для машиностроительных предприятий провести регрессионный анализ зависимости индекса снижения себестоимости продукции (y) от трудоёмкости (x_1) и удельного веса покупных изделий (x_2).	1	204	0,23	0,40
	2	209	0,24	0,26
	3	222	0,19	0,40
	4	236	0,17	0,50
	5	62	0,23	0,40
	6	53	0,43	0,19
	7	172	0,31	0,25
	8	56	0,26	0,44
	9	52	0,49	0,17
	10	46	0,36	0,39
	11	53	0,37	0,33
	12	31	0,43	0,25

Вариант 9

	№ п/п	y	x_1	x_2
На основании данных по следующим предприятиям провести регрессионный анализ зависимости выработки на одного работающего y (шт/чел.) от средней часовой производительности оборудования первого участка x_1 (шт/ч) и средней часовой производительности оборудования второго участка x_2 (шт/ч).	1	996	37	46
	2	1362	23	44
	3	759	15	26
	4	1216	36	34
	5	1350	26	26
	6	1026	24	31
	7	1099	15	20
	8	1726	33	32
	9	1620	44	38
	10	3018	34	32
	11	1831	63	50
	12	1167	8	23

Вариант 10

	№ п/п	y	x_1	x_2
На основании данных по следующим предприятиям провести регрессионный анализ зависимости объёма выпуска продукции y (тыс.руб.) от количества занятых работников x_1 (чел) и стоимости основных фондов x_2 (тыс.руб.).	1	2320	40	430
	2	2500	40	410
	3	2560	42	530
	4	2700	45	560
	5	2893	47	480
	6	2941	49	430
	7	3020	52	440
	8	3648	55	510
	9	3400	59	550
	10	3110	44	580
	11	3150	41	630
	12	3246	51	570

ПРИМЕР ВЫПОЛНЕНИЯ ЗАДАНИЯ № 1

ПАРНЫЙ КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

Параметризация регрессионного уравнения.

Классический подход к оцениванию параметров линейных зависимостей (параметризации регрессионных уравнений) рассматривается на примере линейной парной регрессии

$$\boldsymbol{y = b_0 + b_1x + \varepsilon}$$

$$\hat{y} = \boldsymbol{y_x = b_0 + b_1x} \ ,$$

где y – фактическое значение результативного признака;

$\hat{y}$ или $\boldsymbol{y_x}$ – теоретические значение результативного признака, найденные из уравнения регрессии, путём подстановки в него фактических значений фактора $\boldsymbol{x}$;

$\boldsymbol{b_0}$, $\boldsymbol{b_1}$ – параметры (коэффициенты) уравнения регрессии;

$\boldsymbol{\varepsilon}$ - случайная составляющая (возмущение, ошибка), характеризующая отклонение фактического значения результативного признака от теоретического, найденного по уравнению регрессии.

Имеются два ряда эмпирических (полученных из опыта) данных $\boldsymbol{x}$ ($\boldsymbol{x}_1$, $\boldsymbol{x}_2$, …, $\boldsymbol{x}_n$) и $\boldsymbol{y}$ ($\boldsymbol{y}_1$, $\boldsymbol{y}_2$, …, $\boldsymbol{y}_n$), отображение соответствующих им точек с координатами ($\boldsymbol{x_i}$, $\boldsymbol{y_i}$), где $\boldsymbol{i}$ = 1, 2, …, n, на координатной плоскости называется *полем корреляции*.

По расположению эмпирических точек можно предположить вид корреляционной зависимости. Например, наличие линейной корреляционной зависимости между переменными $\boldsymbol{x}$ и $\boldsymbol{y}$.

Построение линейной регрессии предполагает оценку её параметров $\boldsymbol{b_0}$ и $\boldsymbol{b_1}$ с помощью метода наименьших квадратов (МНК).

Согласно МНК неизвестные параметры $\boldsymbol{b_0}$ и $\boldsymbol{b_1}$ получают таким образом, чтобы сумма квадратов отклонений фактических значений $\boldsymbol{y_i}$ от значений $\hat{y}_i$, найденных по уравнению регрессии была бы минимальной

$$S = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \to min.$$

Таким образом, из множества возможностей, положение линии регрессии на графике выбирается таким образом, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была минимальной

$$\varepsilon_i = y_i - \hat{y}_i \ ,$$

$$S = \sum_{i=1}^{n}\varepsilon_i^{\,2} \to min.$$

Для поиска минимума функции, необходимо вычислить частные производные по каждому из параметров $\boldsymbol{b_0}$ и $\boldsymbol{b_1}$ и приравнять их к нулю

$$S=\sum_{i=1}^{n}(y_i-\tilde{y}_i)^2=\sum_{i=1}^{n}(y_i-b_0-b_1x_i)^2;$$

$$\begin{cases}\dfrac{\partial S}{\partial b_0}=-2\sum\limits_{i=1}^{n}(y_i-b_0-b_1x_i)=\sum\limits_{i=1}^{n}y_i-nb_0-b_1\sum\limits_{i=1}^{n}x_i=0\\ \dfrac{\partial S}{\partial b_1}=-2\sum\limits_{i=1}^{n}(y_i-b_0-b_1x_i)x_i=\sum\limits_{i=1}^{n}x_iy_i-b_0\sum\limits_{i=1}^{n}x_i-b_1\sum\limits_{i=1}^{n}x_i^2=0.\end{cases}$$

В результате преобразований получается следующая система нормальных уравнений для оценки параметров b_0 и b_1

$$\begin{cases}nb_0+b_1\sum\limits_{i=1}^{n}x_i=\sum\limits_{i=1}^{n}y_i\\ b_0\sum\limits_{i=1}^{n}x_i+b_1\sum\limits_{i=1}^{n}x_i^2=\sum\limits_{i=1}^{n}x_iy_i.\end{cases}$$

Искомые оценки параметров b_0 и b_1 находят решая систему нормальных уравнений методом подстановки, последовательного исключения переменных либо методом определителей. Так,

$$b_1=\frac{\sum\limits_{i=1}^{n}x_iy_i-n\cdot\bar{x}\cdot\bar{y}}{\sum\limits_{i=1}^{n}x_i^2-n\cdot\bar{x}^2}.$$

Разделив обе части уравнений системы на n, получим

$$\begin{cases}b_0+b_1\bar{x}=\bar{y}\\ b_0\bar{x}+b_1\overline{x^2}=\overline{xy}.\end{cases}$$

Из первого уравнения системы получим

$$b_0=\bar{y}-b_1\bar{x},$$

После подстановки во второе уравнение получим

$$b_1=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\overline{x^2}-\bar{x}^2}=\frac{C\hat{o}v(x,y)}{S_x^2},$$

где $C\hat{o}v(x,y)$ – выборочная ковариация признаков (корреляционный момент)

$$C\hat{o}v(x,y)=\overline{xy}-\bar{x}\cdot\bar{y}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y});$$

S_x^2 – дисперсия признака x

$$S_x^2=\overline{x^2}-\bar{x}^2=\frac{\sum\limits_{i=1}^{n}x_i^2}{n}-\left(\frac{\sum\limits_{i=1}^{n}x_i}{n}\right)^2=\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{n}.$$

Решение системы нормальных уравнений может быть осуществлено методом определителей

$$b_0=\frac{\Delta b_0}{\Delta},\ b_1=\frac{\Delta b_1}{\Delta},$$

где Δ – определитель системы;

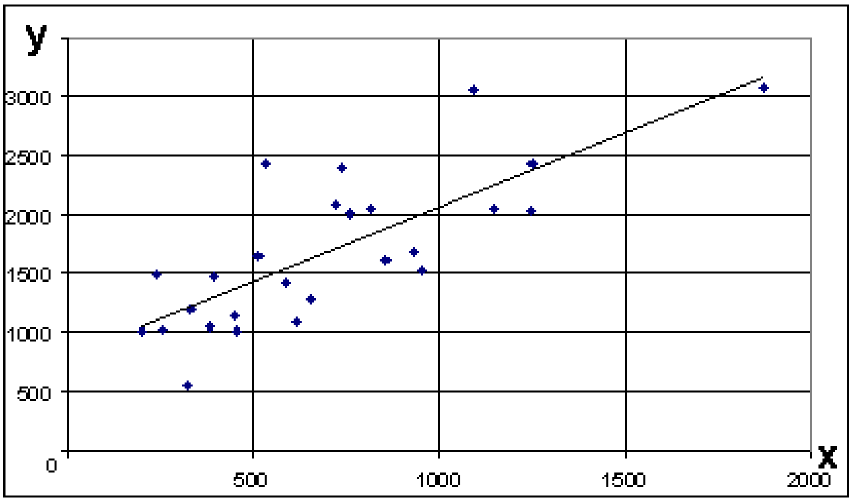
Δb₀, ***Δb₁*** – частные определители, получаемые путём замены соответствующего столбца матрицы определителя системы данными правой части исходной системы нормальных уравнений;

$$\Delta = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix}, \ \Delta b_0 = \begin{vmatrix} \sum y & \sum x \\ \sum xy & \sum x^2 \end{vmatrix}, \ \Delta b_1 = \begin{vmatrix} n & \sum y \\ \sum x & \sum xy \end{vmatrix}.$$

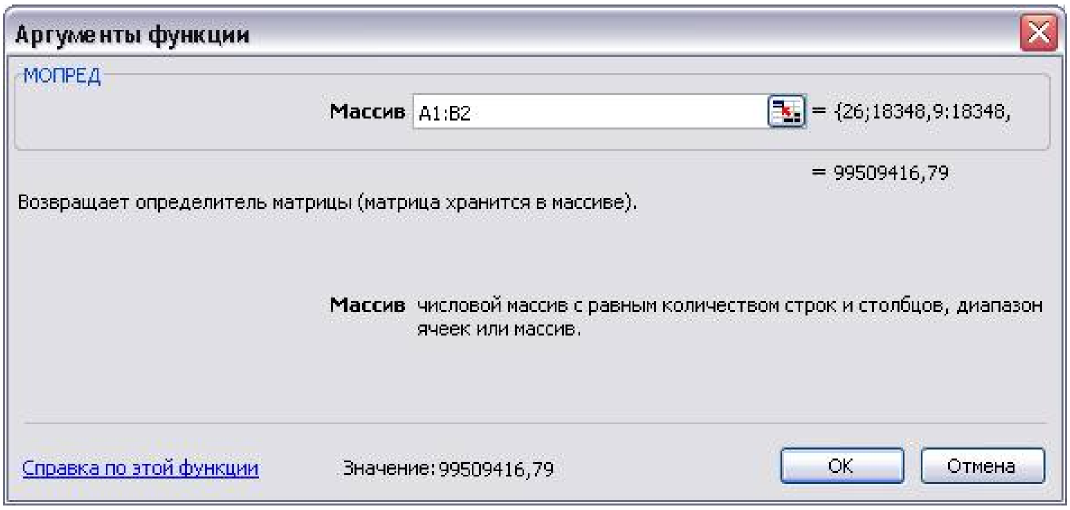
Данные о стоимости основных фондов x и продукции предприятий (фирм) y, млн руб.

фирма	x	y	Σxy	x^2	y^2
1	201,6	1011,3	203878,1	40642,56	1022728
2	242,6	1490,4	361571	58854,76	2221292
3	255,4	1024,5	261657,3	65229,16	1049600
4	323,7	559,9	181239,6	104781,7	313488
5	331,9	1195,1	396653,7	110157,6	1428264
6	384,6	1050,1	403868,5	147917,2	1102710
7	397,7	1482,8	589709,6	158165,3	2198696
8	450,7	1151,7	519071,2	203130,5	1326413
9	457,6	1020,6	467026,6	209397,8	1041624
10	515,3	1648	849214,4	265534,1	2715904
11	533,8	2441,9	1303486	284942,4	5962876
12	587,8	1424,6	837379,9	345508,8	2029485
13	614,9	1095,4	673561,5	378102	1199901
14	655,1	1278,5	837545,4	429156	1634562
15	720,1	2091,4	1506017	518544	4373954
16	741,5	2403,5	1782195	549822,3	5776812
17	760,9	2010	1529409	578968,8	4040100
18	814,1	2042,3	1662636	662758,8	4170989
19	859,2	1607,9	1381508	738224,6	2585342
20	931	1683,2	1567059	866761	2833162
21	953,8	1529	1458360	909734,4	2337841
22	1092,6	3063,9	3347617	1193775	9387483
23	1148,9	2048,4	2353407	1319971	4195943
24	1247,5	2034,4	2537914	1556256	4138783
25	1253,1	2435,9	3052426	1570260	5933609
26	1873,5	3082,1	5774314	3510002	9499340
Сумма	18348,9	43906,8	35838726	16776598	84520903

Построение корреляционного поля и проведение линии регрессии

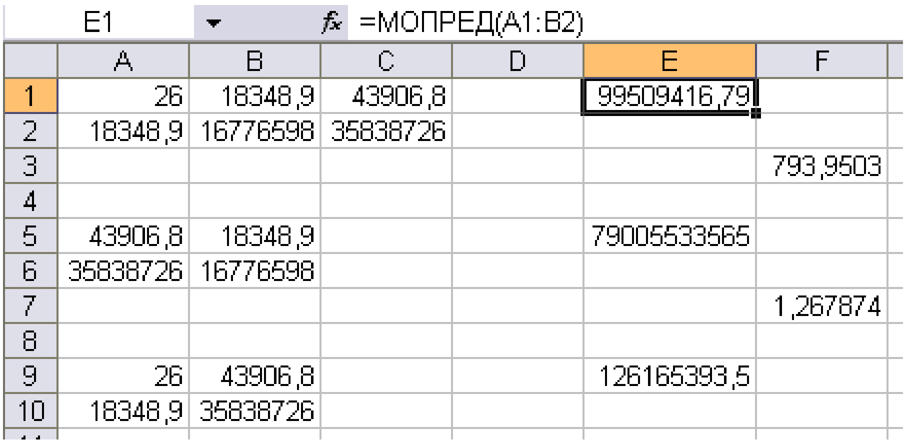


Для нахождения параметров уравнения регрессии используем функцию Excel МОПРЕД, позволяющую рассчитать определитель матрицы



Так, определитель системы в целом равен ***Δ* = =99509416,79**, частный определитель ***Δb*₀ = 79005533565**, частный определитель ***Δb*₁ = 126165393,5**.

$$b_0 = \frac{\Delta b_0}{\Delta} = 793{,}9503, \; b_1 = \frac{\Delta b_1}{\Delta} = 1{,}267874.$$



E1 f_x =МОПРЕД(A1:B2)

	A	B	C	D	E	F
1	26	18348,9	43906,8		99509416,79	
2	18348,9	16776598	35838726			
3						793,9503
4						
5	43906,8	18348,9			79005533565	
6	35838726	16776598				
7						1,267874
8						
9	26	43906,8			126165393,5	
10	18348,9	35838726				

Уравнение регрессии имеет вид

$$y = 793{,}9503 + 1{,}26 \cdot x + e.$$

Анализ результатов показывает, что величина коэффициента b_0 = 793,9503 в данной модели экономически не интерпретируется, поскольку означает объём выпуска продукции при нулевом значении основных производственных фондов, что реально невозможно. Однако, эта константа характеризует смещение линии регрессии по оси y (точку пересечения линии регрессии с осью y). Коэффициент уравнения регрессии b_1= 1,26 означает рост объёма выпуска продукции на 1,26 млн.руб. при увеличении основных производственных фондов в стоимостном выражении на 1 млн.руб.

Идентификация и верификация модели парной регрессии

На этапе идентификации полученных моделей парной регрессии рассчитываются показатели тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает линейный коэффициент корреляции (r_{xy}):

$$r_{xy} = b_1 \cdot \frac{S_x}{S_y} = \frac{Cov(x,y)}{S_x \cdot S_y} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{S_x \cdot S_y} =$$

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

$$= \frac{26 \cdot 35838726 - 43906{,}8 \cdot 18348{,}9}{\sqrt{\left(26 \cdot 16776598 - 18348{,}9^2\right) \cdot \left(26 \cdot 84520903 - 43906{,}8^2\right)}} = 0{,}77$$

Корреляционная связь между переменными называется прямой, если $r_{xy} > 0$, и обратной, если $r_{xy} < 0$.

Коэффициент корреляции принимает значения от -1 до 1, т. е. $-1 \le r_{xy} \le 1$. Чем ближе коэффициент корреляции r_{xy} к единице, тем связь теснее. Для качественной характеристики силы связи используют шкалу Чеддока:

Показатель тесноты связи	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99
Характеристика силы связи	слабая	умеренная	заметная	высокая	весьма высокая

Связь между переменными x и y модели тесная (высокая), прямо пропорциональная.

x	y	y_x	$(y - \bar{y})^2$	$(y_x - \bar{y})^2$	$(y - y_x)^2$	$(x - \bar{x})^2$	x^2
201,6	1011,3	1049,6	458902	408537,6	1463,342	254144	40642,56
242,6	1490,4	1101,5	39332,04	344788,1	151214,8	214486,5	58854,76
255,4	1024,5	1117,8	441192,3	325992,8	8698,412	202794,3	65229,16
323,7	559,9	1204,4	1274242	234606,6	415330,1	145944,6	104781,7
331,9	1195,1	1214,8	243663,7	224643,2	386,4227	139746,6	110157,6
384,6	1050,1	1281,6	407839,4	165769,9	53580,49	103122,5	147917,2
397,7	1482,8	1298,2	42404,31	152521	34083,16	94880,59	158165,3
450,7	1151,7	1365,4	288393,8	104550,1	45659,6	65038,73	203130,5
457,6	1020,6	1374,1	446388,4	98969,18	124983	61566,97	209397,8
515,3	1648	1447,3	1658,369	58291,99	40286,22	36262,41	265534,1
533,8	2441,9	1470,7	567275,5	47516,01	943149	29558,87	284942,4
587,8	1424,6	1539,2	69761	22355,18	13134,67	13906,76	345508,8
614,9	1095,4	1573,6	352032,3	13261,16	228642,7	8249,53	378102
655,1	1278,5	1624,5	168283	4120,171	119739,9	2563,085	429156
720,1	2091,4	1706,9	162148,7	332,0869	147804,6	206,5853	518544
741,5	2403,5	1734,1	510906	2057,144	448124,7	1279,713	549822,3
760,9	2010	1758,7	103218,9	4893,354	63163,96	3044,068	578968,8
814,1	2042,3	1826,1	125016,6	18879,7	46730,99	11744,72	662758,8
859,2	1607,9	1883,3	6532,37	37863,14	75849,35	23553,99	738224,6
931	1683,2	1974,3	30,50438	81577,57	84763,06	50747,96	866761
953,8	1529	2003,3	25511,46	98926,23	224911,6	61540,25	909734,4
1092,6	3063,9	2179,2	1891112	240596,4	782642,1	149670,8	1193775
1148,9	2048,4	2250,6	129367,5	315717,7	40889,17	196402,4	1319971
1247,5	2034,4	2375,6	119492,5	471831,6	116433,2	293518,1	1556256
1253,1	2435,9	2382,7	558273,4	481636,1	2827,775	299617,3	1570260
1873,5	3082,1	3169,3	1941499	2192144	7605,97	1363694	3510002
Сумма			10374477	6152378	4222098	3827285	16776598

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции r^2_{xy}, который называется коэффициентом детерминации. Он характеризует долю результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака:

$$d = r^2_{xy} = \frac{S^2_{y_x}}{S^2_y} = 1 - \frac{S^2_e}{S^2_y}.$$

Величина $1 - r^2$ характеризует долю дисперсия y, вызванную влиянием остальных не учтённых в модели факторов.

Для получения несмещённой оценки общей дисперсии на одну степень свободы, общую сумму квадратов отклонений делят не на число единиц совокупности, а на число степеней свободы

$$S^2_y = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = 10374477 / (26 - 1) = 414979{,}1,$$

факторная дисперсия

$$S_{y_x}^2 = \frac{\sum_{i=1}^{n}(y_{x_i} - \bar{y})^2}{p} = 6152378 / 1 = 6152378,$$

остаточная дисперсия

$$S_e^2 = \frac{\sum_{i=1}^{n}(y_i - y_{x_i})^2}{n-(p+1)} = \frac{\sum_{i=1}^{n} e_i^2}{n-(p+1)} = 4222098 / (26-2) = 175920{,}8,$$

дисперсия факторной переменной

$$S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = 3827285 / (26-1) = 153091{,}4 \;,$$

где n – число наблюдений, p – число оцениваемых параметров уравнения при независимых переменных, ($p+1$) – число оцениваемых параметров уравнения регрессии, включая константу b_0, y_i – наблюдаемые значения зависимой переменной (i = 1, … , n), y_{x_i} или $\hat{y}_i$ - расчётные значения зависимой переменной, x_i – наблюдаемые значения независимой переменной.

Среднеквадратические отклонения:
общее S_y = **644,1887** , факторное Sy_x= **2480,399** , остаточное S_e = =**419,4291** , независимой переменной S_x= **391,269** .

Вычисляя отношение факторной и остаточной дисперсии в расчете на одну степень свободы, получают величину F-критерия Фишера. Если нулевая гипотеза справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Для H_0 необходимо опровержение, чтобы факторная дисперсия превышала остаточную в несколько раз, т.е. выполнялась бы гипотеза $H_1 : S_{y_x}^2 > S_e^2$.

Значение F-критерия Фишера для парной линейной регрессии:

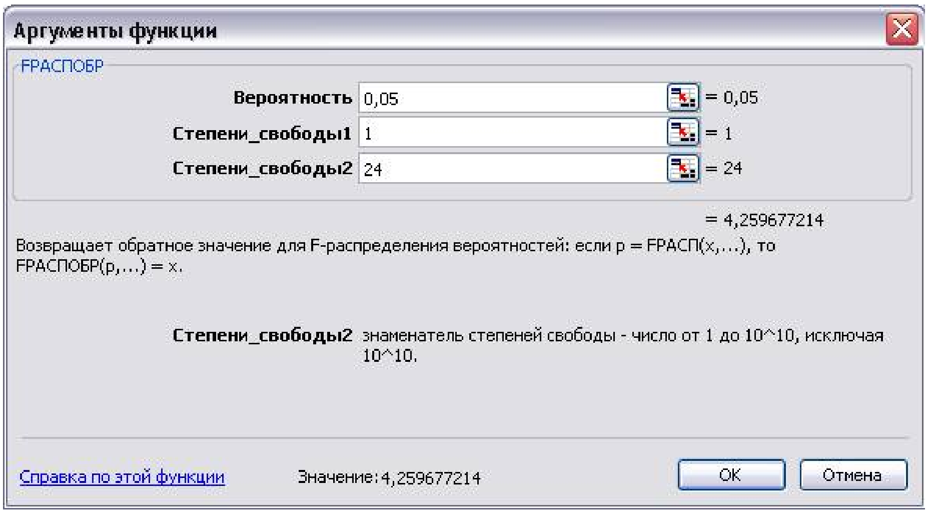
$$F = \frac{r^2}{1-r^2} \cdot (n-2) = \frac{0{,}77^2}{1-0{,}77^2} \cdot (26-2) = 34{,}95.$$

Вычисленное значение F-критерия признается достоверным (отличающимся от единицы), если оно больше табличного. Табличное значение F-критерия это максимальная величина отношения дисперсий, которая имеет место при случайном их расхождении для данного уровня вероятности наличия нулевой гипотезы. В этом случае нулевая гипотеза (H_0) отвергается и делается вывод о существенности изучаемой связи: $F_{факт} > F_{табл}$.

Если величина $F_{факт}$ оказывается меньше $F_{табл}$, то вероятность нулевой гипотезы выше заданного уровня (α = 0,05 или 0,01) и она не может быть отклонена без риска неверного вывода о наличии связи. В этом случае уравнение регрессии признается статистически незначимым.

Полученное значение $F_{факт} = 34{,}95$ сравниваем с табличным критическим значением, которое при уровне значимости α = 0,05 и числе степеней свободы $k_1 = p$ = 1 (регрессионном) и $k_2 = n - p$ – 1 = 26 – 1 – 1 = 24 (остаточном) составит $F_{табл}$= 4,259677. Так как $F_{факт} > F_{табл}$, то делаем вывод о существенности изучаемой связи по линейной модели, т.е. о статистической значимости модели в целом.

Критическое (табличное) значение ***F***-критерия Фишера можно определить с помощью функции Excel FРАСПОБР из категории *Статистические*.



В последующих реализациях Excel используется функция =F.ОБР.ПХ(α;*df1*;*df2*), где α - уровень значимости, *df1*- регрессионное число степеней свободы, *df2* – остаточное число степеней свободы.

В линейной регрессии обычно оценивается значимость не только уравнения в целом, но и отдельных его параметров.

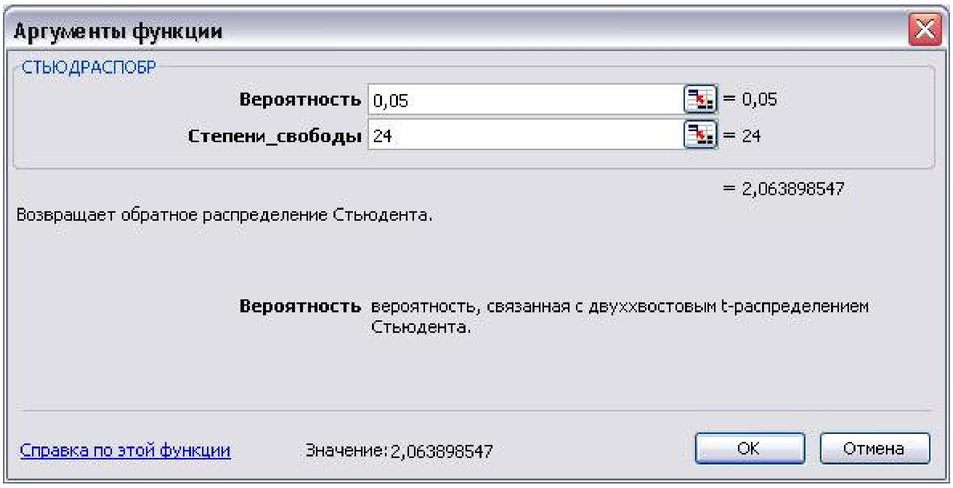
Процедура оценивания существенности параметров b_0 и b_1 базируется на расчете фактических значений ***t***-критерия Стьюдента:

$$t_{b_0} = \frac{b_0}{S_{b_0}} = b_0/S_e\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \frac{793{,}95}{419{,}429\cdot\sqrt{\frac{16776598}{26\cdot 3827285}}} = \mathbf{4{,}435949}$$

$$t_{b_1} = \frac{b_1}{S_{b_1}} = \frac{b_1}{S_e}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} = b_1\frac{\sqrt{n-2}\cdot S_x}{S_e} = 1{,}268\cdot\frac{\sqrt{26-2}\cdot 391{,}269}{418{,}429} = \mathbf{5{,}838067},$$

которые затем сравниваются с табличным (критическим) значением для заданного уровня значимости α = 0,05 и числа степеней свободы (n–p–1) = (n – 2). Если фактическое значение ***t***-критерия Стьюдента (по модулю) превышает табличное ($t_{крит}$), то гипотезу о несущественности коэффициента регрессии можно отклонить.

Критическое (табличное) значение ***t***-критерия Стьюдента можно определить с помощью функции Excel СТЬЮДРАСПОБР из категории *Статистические*.



В последующих реализациях Excel используется функция =СТЬЮДЕНТ.ОБР.2Х(α;*df*), где α - уровень значимости, *df* – остаточное число степеней свободы. Поскольку $t_{\alpha;\, n\text{-}m\text{-}1} = t_{0{,}05;\, 24} = t_{кр} = t_{табл} =$ 2,063899 , то в обоих случаях | $t_{факт}$ | > $t_{табл}$, делаем вывод, что значения параметров b_0 и b_1 не случайно отклоняются от нуля, т.е. статистически значимы и отражают реальную природу взаимосвязей между рассматриваемыми переменными в рамках разработанной модели парной линейной регрессии.

Значимость коэффициента корреляции проверяется также на основе расчета фактического значения ***t***-критерия Стьюдента:

$$t_r = r \cdot \sqrt{\frac{n-2}{1-r^2}} = 0{,}77 \cdot \sqrt{\frac{26-2}{1-0{,}77^2}} = 5{,}9.$$

В парной линейной регрессии $t_r^2 = F$, следовательно, оба способа проверки значимости модели (с помощью ***t*** и ***F***-критерия) для линейной парной регрессии равносильны. Кроме того, $t_{b_1}^2 = t_r^2$.

Для проверки нулевой гипотезы H_0 необходимо сравнить фактическое значение t_r (по модулю) с табличным значением (при заданном уровне значимости α), если | t_r | > $t_{крит}$, то коэффициент корреляции значимо отличается от нуля.

Оценить качество синтезированной модели в целом можно основываясь на минимальности отклонения фактических значений результативного признака от теоретических, рассчитанных по уравнению регрессии. Величина отклонений эмпирических и расчетных значений ($y - y_x$) по каждому наблюдению представляет собой абсолютную ошибку аппроксимации. Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации, как среднюю арифметическую простую из относительных ошибок аппроксимации:

$$\overline{A} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y_i - y_{x_i}}{y_i} \right| \cdot 100 = 21{,}9\ \%.$$

Полученное значение ошибки может быть использовано для сравнения моделей различных форм зависимостей.

При подстановке в уравнение регрессии $\boldsymbol{y} = \boldsymbol{b_0} + \boldsymbol{b_1 x}$ соответствующего значения $\boldsymbol{x}$ можно определить предсказываемое (прогнозируемое) значение $\boldsymbol{y_p}$, как вариант точечного прогноза. Примем $\boldsymbol{x_p}$ равным медиане (среднему значению из двух чисел, стоящих в центре ранжированного ряда величины $\boldsymbol{x}$) $\boldsymbol{x_p = 635}$. Тогда

$$\boldsymbol{y_p = 793{,}95 + 1{,}27 * 635 = 1599{,}05.}$$

Поскольку точечный прогноз является усреднённой оценкой, то вероятные значения прогнозируемой величины будут находиться в некотором интервале. Поэтому точечный прогноз необходимо дополнить расчетом стандартной ошибки $\boldsymbol{S_{yx}}$ и соответственно оценкой доверительного интервала теоретических значений результативного признака $\boldsymbol{y}$:

$$y_x - t_k S_{y_x} \le y_p \le y_x + t_k S_{y_x} .$$

Средняя стандартная ошибка расчетного значения результативного признака по уравнению регрессии:

$$S^2_{y_x} = \frac{\hat{S}^2}{n} + \frac{\hat{S}^2}{\sum (x_i - \bar{x})^2} \cdot (x_{прогн} - \bar{x})^2 = \hat{S}^2 \cdot \left[\frac{1}{n} + \frac{(x_{прогн} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right];$$

$$S_{y_x} = \hat{S} \cdot \sqrt{\frac{1}{n} + \frac{(x_{прогн} - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

где $\hat{S}^2 = \frac{\sum_{i=1}^{n} (y_i - y_{x_i})^2}{n - (p+1)}$ - остаточная дисперсия результативного признака в расчете на одну степень свободы, $\hat{S}$ - остаточное среднеквадратическое отклонение результативного признака.

$$\hat{S}^2 = \frac{\sum_{i=1}^{n} (y_i - y_{x_i})^2}{n - (p+1)} = \frac{4222098}{26 - 1 - 1} = 175920{,}8; \quad \hat{S} = 419{,}429 \text{ млн.руб.}$$

$$S_{y_x} = \hat{S} \cdot \sqrt{\frac{1}{n} + \frac{(x_{прогн} - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 419{,}429 \cdot \sqrt{\frac{1}{26} + \frac{(635 - 705{,}7)^2}{3827285}} = \boldsymbol{83{,}64}$$

С надёжностью $\boldsymbol{\alpha} = 0{,}05$ (табличное значение $\boldsymbol{t_k} = 2{,}0639$) доверительный интервал для $\boldsymbol{y_p}$ при заданном $\boldsymbol{x_p} = 635$ составит

$$y_p = y_x \pm \hat{S} \cdot \sqrt{\frac{1}{n} + \frac{(x_{прогн} - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 1599{,}05 \pm 2{,}0639 \cdot 83{,}64 .$$

Получим, что с вероятностью 95% доверительный интервал для результативного признака – объёма выпускаемой продукции при заданном значении фактора $\boldsymbol{x_p} = 635$ млн. руб. – величины основных производственных фондов, будут находиться в пределах от $\boldsymbol{y_p(min) = 1426{,}42}$ млн. руб. для нижней границы и до $\boldsymbol{y_p(max) = 1771{,}68}$ млн. руб. для верхней границы.

Средняя ошибка предсказанного индивидуального значения $\boldsymbol{y}$ при $\boldsymbol{x_p = x_k}$ составит:

$$S_{y_i(x_k)} = S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x - \bar{x})^2}} .$$

Средняя ошибка прогнозируемого индивидуального значения $\boldsymbol{y}$

$$S_{y_x} = \hat{S} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{прогн} - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 419{,}429 \cdot \sqrt{1 + \frac{1}{26} + \frac{(635 - 705{,}7)^2}{3827285}} = 427{,}7$$

Тогда при заданном уровне значимости доверительный интервал для y^*_p при $\boldsymbol{x_p = x_k}$ составит:

$$y_p = y_k \pm t_k S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x - \bar{x})^2}} ,$$

где $\boldsymbol{t_k}$ - критическое (табличное) значение $\boldsymbol{t}$-критерия Стьюдента для соответствующего уровня значимости и числа степеней свободы ($\boldsymbol{n}$ - 2), $\boldsymbol{t_k S}$ – предельная ошибка прогнозируемой величины.

Доверительный интервал для индивидуальных значений результативного признака

$$y_p = y_x \pm \hat{S} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{прогн} - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 1599{,}05 \pm 2{,}0639 \cdot 427{,}7 .$$

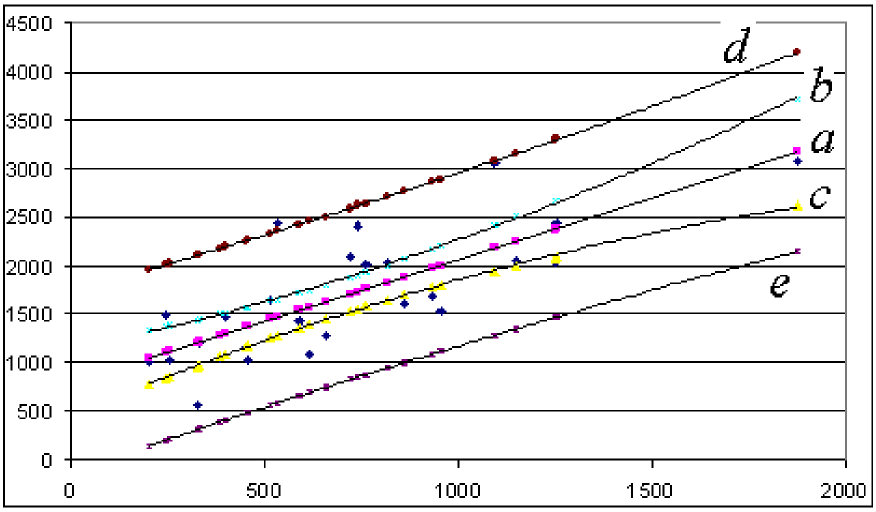
Получим, что с вероятностью 95% доверительный интервал для индивидуальных значений результативного признака – объёма выпускаемой продукции при заданном значении фактора $\boldsymbol{x_p}$ = 635 млн. руб. – величины основных производственных фондов, будут находиться в пределах от $\boldsymbol{y^*_p(min) = 716{,}346}$ млн. руб. для нижней границы и до $\boldsymbol{y^*_p(max) = 2481{,}755}$ млн. руб. для верхней границы.

Для построения графического представления доверительных интервалов для линии регрессии и индивидуальных значений результативного признака y, проведём расчёт теоретических значений величины y_x , в соответствии расчётными значениями коэффициентов уравнения регрессии и границ доверительных интервалов для всех значений факторной переменной x. Результаты представим в следующей таблице.

Расчётные значения y_x и границ доверительных интервалов для линии регрессии и индивидуальных значений результативного признака.

x	y	y_x	$y_p(min)$	$y_p(max)$	$y^*_p(min)$	$y^*_p(max)$
201,6	1011,3	1049,554	769,2288	1329,879	139,6373	1959,47
242,6	1490,4	1101,536	835,4213	1367,652	195,897	2007,176
255,4	1024,5	1117,765	855,9867	1379,544	213,3905	2022,14
323,7	559,9	1204,361	964,7842	1443,938	306,1613	2102,561
331,9	1195,1	1214,758	977,727	1451,788	317,2337	2112,282
384,6	1050,1	1281,575	1060,186	1502,963	388,0544	2175,095
397,7	1482,8	1298,184	1080,471	1515,897	405,567	2190,8
450,7	1151,7	1365,381	1161,528	1569,234	476,0433	2254,719
457,6	1020,6	1374,129	1171,951	1576,308	485,1738	2263,085
515,3	1648	1447,286	1257,755	1636,816	561,1213	2333,45
533,8	2441,9	1470,741	1284,706	1656,777	585,3178	2356,165
587,8	1424,6	1539,207	1361,598	1716,815	655,5153	2422,898
614,9	1095,4	1573,566	1399,104	1748,028	690,5016	2456,63
655,1	1278,5	1624,535	1453,293	1795,776	742,1008	2506,968
720,1	2091,4	1706,946	1537,058	1876,835	824,7741	2589,119
741,5	2403,5	1734,079	1563,573	1904,585	851,7875	2616,37
760,9	2010	1758,676	1587,159	1930,192	876,1885	2641,163
814,1	2042,3	1826,126	1649,714	2002,539	942,6747	2709,578
859,2	1607,9	1883,308	1700,459	2066,156	998,5482	2768,067
931	1683,2	1974,341	1777,47	2171,211	1086,578	2862,104
953,8	1529	2003,248	1801,083	2205,414	1114,296	2892,201
1092,6	3063,9	2179,229	1938,135	2420,324	1280,624	3077,835
1148,9	2048,4	2250,611	1991,234	2509,988	1346,928	3154,293
1247,5	2034,4	2375,623	2081,869	2669,377	1461,48	3289,766
1253,1	2435,9	2382,723	2086,944	2678,503	1467,927	3297,519
1873,5	3082,1	3169,312	2625,412	3713,212	2146,966	4191,659

Графически доверительные границы для y представляют собой дугообразные линии, расположенные по обе стороны от линии регрессии.



Доверительный интервал линии регрессии:

a - линия регрессия $y_x = b_0 + b_1 x$;

b, c - верхняя и нижняя границы доверительного интервала для y_p ;

d, e - доверительный интервал для индивидуальных значений y

ПРИМЕР ВЫПОЛНЕНИЯ ЗАДАНИЯ № 2.

Провести оценку параметров уравнения связи для многофакторной модели, проверить значимость и адекватность полученного уравнения и каждого из его параметров. Рассчитайте прогнозное значение результата, если прогнозные значения факторов составляют 70% от их максимальных значений. Принять уровень значимости $\alpha = 0{,}05$.

Найти 95% доверительные интервалы для параметров уравнения. Провести анализ на мультиколлинеарность. Определить и проанализировать частные коэффициенты корреляции. Вычислить коэффициент множественной корреляции и коэффициент детерминации и проанализировать их. Определить и проанализировать частные коэффициенты эластичности.

Имеются данные по десяти предприятиям, добывающим полезные ископаемые. Изучается влияние мощности (толщины) пласта (м) полезных ископаемых и уровня механизации работ (%) на сменную добычу на одного работника (т).

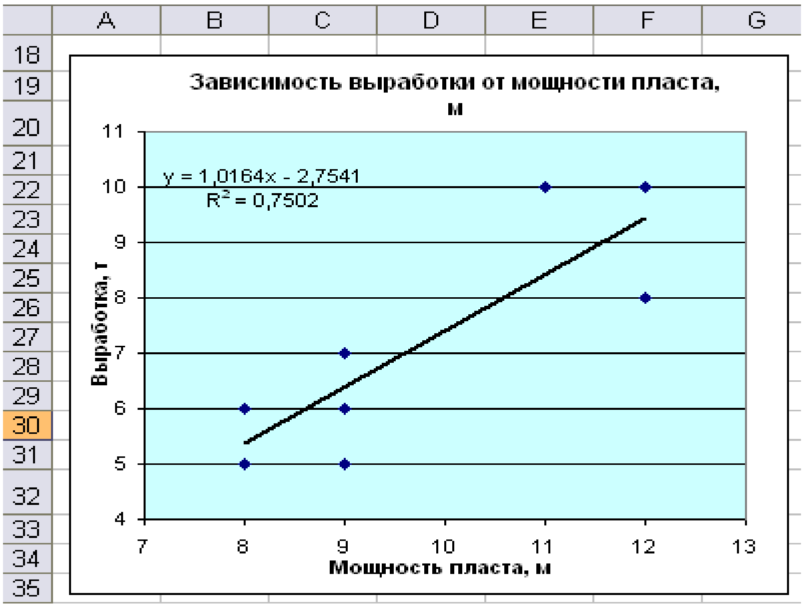
Исходные данные для расчёта

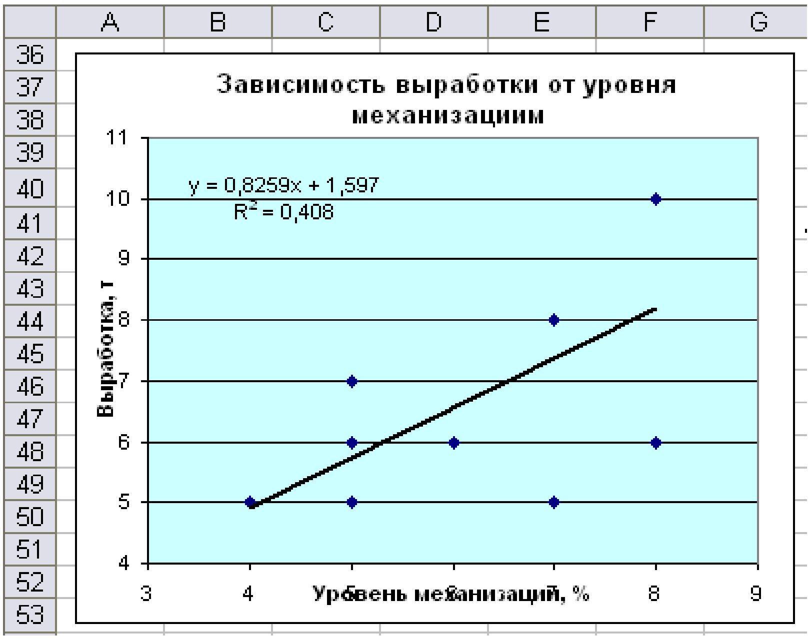
Предприятие	Мощность пласта, м X_1	Уровень механизации, % X_2	Сменная добыча на одного работника, т У
1	8	5	6
2	9	6	6
3	8	7	5
4	12	8	10
5	8	5	5
6	12	7	8
7	9	4	5
8	8	8	6
9	9	5	7
10	11	8	10

1. Построение линейной зависимости на основе поля корреляции

Используя возможности Excel построим точечные диаграммы зависимости у от X_1 и от X_2. Добавим к построенным диаграммам, отображающим поля корреляции – линии тренда для линейных регрессионных моделей. Поместим на диаграммы соответствующие парные уравнения линейной регрессии и коэффициенты детерминации.

Проведём анализ полученных зависимостей, таким же образом, как и для первой модели парной линейной регрессии.





2. Определение параметров уравнения регрессии в матричной форме **$B=(X'X)^{-1}X'Y$**

Сформировать матрицу X объясняющих переменных размером 10x3, добавив столбец с единичными элементами перед столбцами данных по факторным переменным. Этот столбец получается как единичное значение переменной x_0, умножаемой на коэффициент b_0. Столбец зависимой переменной составляет вектор Y.

	A	B	C	D	E	F	G
2		1	8	5	6		
3		1	9	6	6		
4		1	8	7	5		
5		1	12	8	10		
6		1	8	5	5		
7		1	12	7	8		
8		1	9	4	5		
9		1	8	8	6		
10		1	9	5	7		
11		1	11	8	10		
12							
13			X		Y		
14							
15			X1ср	X2ср	Yср		
16			9,4	6,3	6,8		
17	=СРЗНАЧ(C2:C11)		=СРЗНАЧ(D2:D11)		=СРЗНАЧ(E2:E11)		

Определим с помощью функции **=ТРАНСП** (из категории *Ссылки и массивы*) транспонированную матрицу X′. Для этого выделим массив ячеек 3х10 и введём в него функцию транспонирования, указав в качестве аргументов исходную матрицу X, включающую и первый столбец из единиц. Для получения массива результатов по этой функции следует в завершении нажать комбинацию клавиш *Ctrl+Shift+Enter* или же повторить эту комбинацию при активизации строки формул (щёлкнуть левой кнопкой мыши в строке формул).

Перемножим транспонированную матрицу **X′** с исходной матрицей **X**, используя функцию **=МУМНОЖ** из категории математические. Для вывода результатов предварительно должен быть выделен массив ячеек 3х3. Полученная матрица должна быть симметричной.

Найдём обратную матрицу $(\mathbf{X'X})^{-1}$, используя математическую функцию **=МОБР**, аргументом которой является матрица **X′X**. Поскольку результат также представляет собой симметричную матрицу 3-го порядка, то предварительно необходимо выделить массив ячеек 3х3.

Умножим транспонированную матрицу X′ на вектор Y, выделив для этого столбец из трёх ячеек.

Перемножение результатов этих действий (обратной матрицы $(\mathbf{X'X})^{-1}$ на вектор **X′Y**) даёт вектор коэффициентов уравнения регрессии ***B***. Для получения массива результатов по всем этим функциям следует нажать комбинацию клавиш *Ctrl+Shift+Enter*.

	G	H	I	J	K	L	M	N	O	P
2				X'	=ТРАНСП(B2:D11)					
3										
4	1	1	1	1	1	1	1	1	1	1
5	8	9	8	12	8	12	9	8	9	11
6	5	6	7	8	5	7	4	8	5	8
7										
8		$X'X$	=МУМНОЖ(G4:P6;B2:D11)				$(X'X)^{-1}$	=МОБР(H9:J11)		
9		10	94	63			4,02006	-0,3234	-0,1396	
10		94	908	603			-0,3234	0,05377	-0,0289	
11		63	603	417			-0,1396	-0,0289	0,06528	
12										
13	$X'Y$	=МУМНОЖ(G4:P6;E2:E11)				$B=(X'X)^{-1}X'Y$		=МУМНОЖ(M9:O11;H14:H16)		
14		68						-3,5393		
15		664						0,85393		
16		445						0,36704		

Уравнение регрессии имеет вид:

$$\hat{y} = \mathbf{-3{,}5393 + 0{,}85393}x_1 \mathbf{+ 0.6704}x_2$$

3. Анализ на наличие мультиколлинеарности. Используем функцию **=КОРРЕЛ** для определения парных коэффициентов корреляции. Поскольку коэффициент корреляции между x_1 и x_2 равен 0,48768, что меньше 0,8 , проблема коллинеарности факторов отсутствует. В тоже время, коэффициенты парной корреляции между факторами и результирующей переменной У имеют высокие значения (0,86614 и 0,63876), что свидетельствует об их тесной зависимости.

	O	P	Q	R	S	T
34		X1	X2	У		
35	X1	1	0,48768	0,86614		
36	X2		1	0,63876		
37	У			1		
38			=КОРРЕЛ(C2:C11;D2:D11)			
39				=КОРРЕЛ(C2:C11;E2:E11)		
40				=КОРРЕЛ(D2:D11;E2:E11)		

4. Для определения влияния параметров уравнения регрессии на зависимую величину, найдём средние арифметические значения всех переменных с помощью функции =СРЗНАЧ. Определим также средние значения переменных, возведённых в квадрат и используем для расчёта дисперсий факторных переменных и результирующей переменной

$$S_{x_j}^2 = \overline{x^2} - \overline{x}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2; \; S_y^2 = \overline{y^2} - \overline{y}^2 = \frac{\sum_{i=1}^{n} y_i^2}{n} - \left(\frac{\sum_{i=1}^{n} y_i}{n}\right)^2;$$

S^2x_1= 2,44 , S^2x_2= 2,01 S^2y= 3,36

Определим средние квадратические отклонения для факторов и результативного признака

Sx_1= 1,56205 , Sx_2= 1,41774 Sy= 1,83303

Определение стандартизованных коэффициентов регрессии

$$b'_j = b_j \frac{S_{x_j}}{S_y};$$

b'_1= 0,72769 b'_2=0,28389

Таким образом, увеличение мощности пласта и уровня механизации работ только на одно среднее квадратическое отклонение Sx_1 и Sx_2 увеличит сменную добычу на 0,72769 Sy и 0.28389 Sy соответственно.

Определение коэффициентов эластичности

$$E_j = b_j \frac{\bar{x}_j}{\bar{y}};$$

E_1= 1,18044 E_2= 0,34005

Увеличение этих переменных на 1% от своих средних значений приводит в среднем к росту добычи соответственно на 1,18% и 0,34%. На сменную добычу большее влияние оказывает фактор мощности пласта.

	H	I	J	K	L	M	N	O	P	Q
18	$\hat{y}$	$= -3,5393 + 0,85393x_1 + 0.6704x_2$								
19										
20		x_1^2	x_2^2	y^2		$S^2_{x_i} = \overline{x^2} - \bar{x}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2;$				
21			=D2^2							
22	=C2^2	64	25	36	=E2^2					
23	=C3^2	81	36	36	=E3^2		2,44	2,01	3,36	
24	=C4^2	64	49	25	=E4^2		=I33-C16^2	=J33-D16^2	=K33-E16^2	
25	=C5^2	144	64	100	=E5^2		Sx1	Sx2	Sy	
26	=C6^2	64	25	25	=E6^2		1,56205	1,41774	1,83303	
27	=C7^2	144	49	64	=E7^2		=N23^0,5	=O23^0,5	=P23^0,5	
28	=C8^2	81	16	25	=E8^2					
29	=C9^2	64	64	36	=E9^2		$b'_j = b_j \frac{S_{x_i}}{S_y};$			
30	=C10^2	81	25	49	=E10^2					
31	=C11^2	121	64	100	=E11^2		0,72769	0,28389		
32	средн	x_1^2	x_2^2	y^2		=N15*N26/P26		=N16*O26/P26		
33		90,8	41,7	49,6						
34	=СРЗНАЧ(I22:I31)		=СРЗНАЧ(J22:J31)		=СРЗНАЧ(K22:K31)					
35										
36		E1	E2							
37		1,18044	0,34005							
38	=N15*C16/E16		=N16*D16/E16							

5. Определение суммы квадратов отклонений $\boldsymbol{Q}$, дисперсии на степень свободы $\boldsymbol{S}^2$ и средние квадратические отклонения $\boldsymbol{S}$ (общие, объяснённые регрессией - $\boldsymbol{r}$ и остаточные - $\boldsymbol{e}$)

$$Q_{общ} = Q_r + Q_e$$

$Q_{общ} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ =33,6; $S^2_{общ} = \frac{Q_{общ}}{n-1}$ =3,73333; $S_{общ}$=1,96218

$Q_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ =6,32959; $S^2_e = S^2 = \frac{Q_e}{n-p-1}$ =0,90423; S=0,95091

$Q_r = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ =33,6 – 6,32989 = 27,2704; $S^2_r = \frac{Q_r}{p}$ = 13,6352; S_r = 3,69259

	A	B	C	D	E	F	G	H	I
77									
78		$Q_{общ} = \sum_{i=1}^{n}(y_i - \bar{y})^2$		33,6		$S^2_{общ} = \frac{Q_{общ}}{n-1}$		3,73333	
79								Sобщ=	1,93218
80									
81		$Q_r = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$		27,2704		$S_r^2 = \frac{Q_r}{p}$		13,6352	
82								Sr=	3,69259
83									
84		$Q_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$		6,32959	$S_e^2 = S^2 = \frac{Q_e}{n-p-1}$			0,90423	
85								S=	0,95091

Определение дисперсии и средние квадратические отклонения для параметров уравнения регрессии, используя диагональные элементы обратной матрицы $(\mathbf{X'X})^{-1}$ (см. стр. 28). Расчёт величин остатков $\boldsymbol{e}$, квадратов остатков $\boldsymbol{e^2}$ и их суммы (см. стр. 33).

$$S^2_{b_j} = S^2[(X'X)^{-1}]_{jj} \qquad S^2 = \frac{e'e}{n-(p+1)} = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1};$$

Sb_0=1,90658; Sb_1= 0.2205; Sb_2= 0.24295

Определим значения t-статистики Стьюдента

$t_j = \frac{|b_j|}{S_{b_j}}$; $\boldsymbol{t_0}$ = 1,85637; $\boldsymbol{t_1}$ = 3,87263; $\boldsymbol{t_2}$ = 1,51078

Критическое (табличное) значение критерия Стьюдента определяется с помощью функции **=СТЬЮДРАСПОБР**, в качестве аргументов которой вводится вероятность – уровень значимости α = 0,05 и остаточное число степеней свободы $\boldsymbol{df} = \boldsymbol{n} - \boldsymbol{p} - 1$. В последующих реализациях Excel используется функция =СТЬЮДЕНТ.ОБР.2Х (α; *df*).

Получим критическое значение $\boldsymbol{t_{крит}}$= 2,36462, следовательно, значимым оказался только коэффициент b_1, а b_0 и b_2 – статистически незначимы. Таким образом в модели следует отказаться от использования фактора x_2 и константы b_0.

Определим ***P***-значение, вероятность ошибки с помощью функции **=СТЬЮДРАСП**, аргументами которой являются: расчётное значение t-статистики (по модулю), число степеней свободы $\boldsymbol{df} = \boldsymbol{n} - \boldsymbol{p} - 1$, хвосты = 2. В последующих реализациях Excel используется функция =СТЬЮДЕНТ.РАСП.2Х(ABS(t_j); *df*).

$P(b_0)$ = 0,10577; $P(b_1)$ = 0,00611; $P(b_2)$ = 0,1746

Вероятность ошибки не должна превышать 0,05. Вероятность ошибки по коэффициенту b_0 – более 10%, b_1 – составляет 0,611% , что менее 5% и b_2 – более 17%. Это подтверждает значимость только коэффициента b_1.

	H	I	J	K	L	M	N	O	P	Q	R
39											
40					S^2	S	Sbo	Sb1	Sb2		
41	$S^2 = \frac{e'e}{n-(p+1)} = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1};$				0,90423	0,95091	1,90658	0,2205	0,24295		
42				=J55/(10-2-1)		=L41^0,5	=M41*(M9^0,5)				
43	y расч	e	e^2					=M41*(N10^0,5)			
44	5,12734	0,87266	0,76153			to	t 1	t 2	=M41*(O11^0,5)		
45	6,34831	-0,3483	0,12132			1,85637	3,87263	1,51078			
46	5,86142	-0,8614	0,74205		=ABS(N14)/N41	=ABS(N15)/O41		=ABS(N16)/P41			
47	9,64419	0,35581	0,1266								
48	5,12734	-0,1273	0,01622				2,36462	=СТЬЮДРАСПОБР(0,05;7)			
49	9,27715	-1,2772	1,63112			to< t кр	t 1> t кр	t 2< t кр			
50	5,61423	-0,6142	0,37728		Р-значен	0,10577	0,00611	0,1746	=СТЬЮДРАСП(ti; df; 2)		
51	6,22846	-0,2285	0,0522								
52	5,98127	1,01873	1,0378		$S_{\hat{y}} = S\sqrt{X_0'(X'X)^{-1}X_0}$						
53	8,79026	1,20974	1,46347								
54									1		y расч
55		Σe^2	6,32959		1	8	6		8		5,49438
56									6		
57		$Q_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$			0,5947	-0,0666	0,02087				
58				=МУМНОЖ(L55:N55;M9:O11)				0,187	=МУМНОЖ(L57:N57;P54:P56)		
59									0,4112	Sy	
60									=M41*(O58^0,5)		
61											
62											
63				$S_{\hat{y}_0} = S\sqrt{1 + X_0'(X'X)^{-1}X_0}$					1,03601	Syo	
64									=M41*(1+O58)^0,5		

6. Определим 95%-ные доверительные интервалы для коэффициентов регрессии, теоретической линии регрессии, индивидуальных значений и дисперсии. Допустим 70%-ные значения от максимальных величин факторов равны $x_1 = 8$, $x_2 = 6$. Определим расчётное значение функции для этих значений.

$$\hat{y} = \mathbf{-3{,}5393 + 0{,}85393}\boldsymbol{x_1} \mathbf{+ 0.6704}\boldsymbol{x_2} \mathbf{= -3{,}5393 + 0{,}85393 \cdot 8 + +0.6704 \cdot 6 = 5{,}49438.}$$

Определим стандартное отклонение для линии регрессии.

$$S_{\hat{y}} = S\sqrt{X_0'(X'X)^{-1}X_0}$$

Для этого умножим вектор $\boldsymbol{X'}_0 = (1 \;\; 8 \;\; 6)$ на обратную матрицу $(\mathbf{X'X})^{-1}$, выделив строку из трех ячеек. Результат (0,5947 -0,0666 0,02087) умножим на вектор-столбец $\boldsymbol{X}_0$. В результате получим значение 0,187. Стандартное отклонение равно 0,4112.

Стандартное отклонение для индивидуальных значений

$$S_{\hat{y}} = S\sqrt{1 + X_0'(X'X)^{-1}X_0} = 1{,}03601.$$

	A	B	C	D	E	F	G	H	I	J
57	$b_j - t_{1-\alpha;n-p-1} \cdot S_{b_j} \le \beta_j \le b_j + t_{1-\alpha;n-p-1} \cdot S_{b_j}$									
58										
59		=N14-N48*N41		-8,0477	≤ βo ≤	0,96902		=N14+N48*N41		
60		=N15-N48*O41		0,33252	≤ β1 ≤	1,37534		=N15+N48*O41		
61		=N16-N48*P41		-0,2074	≤ β2 ≤	0,94152		=N16+N48*P41		
62										
63	$\hat{y} - t_{1-\alpha;n-p-1} \cdot S_{\hat{y}} \le M_x(Y) \le \hat{y} + t_{1-\alpha;n-p-1} \cdot S_{\hat{y}}$									
64										
65		=R55-N48*P59		4,52204	≤Mx(Y)≤	6,46673		=R55+N48*P59		
66	$\hat{y}_0 - t_{1-\alpha;n-p-1} \cdot S_{\hat{y}_0} \le y_0^* \le \hat{y}_0 + t_{1-\alpha;n-p-1} \cdot S_{\hat{y}_0}$									
67										
68										
69				3,04461	≤ yo* ≤	7,94416				
70			=R55-N48*P63			=R55+N48*P63				
71	=ХИ2ОБР(0,025;7)		$\frac{nS^2}{\chi^2_{\alpha/2;n-p-1}} \le \sigma^2 \le \frac{nS^2}{\chi^2_{1-\alpha/2;n-p-1}}$				=ХИ2ОБР(0,975;7)			
72		16,0128						1,68987		
73										
74				0,56469	≤ σ2 ≤	5,35087				
75			=10*L41/B72			=10*L41/H72				

Определим доверительные интервалы для коэффициентов уравнения регрессии

$$b_j - t_{1-\alpha;n-p-1} \cdot S_{b_j} \le \beta_j \le b_j + t_{1-\alpha;n-p-1} \cdot S_{b_j}$$

Для коэффициентов β_0 и β_2 доверительные интервалы накрывают диапазон от отрицательных значений до положительных, включая и 0 ($-8{,}0477 \le \beta_0 \le 0{,}96902$ и $-0{,}2074 \le \beta_2 \le 0{,}94152$). Поэтому достоверно нельзя судить о том влияет данный фактор отрицательно, положительно или вообще не влияет на результирующую переменную, - невозможно. Это подтверждает статистическую незначимость данных коэффициентов. Истинное значение коэффициента β_1 с вероятностью 95% лежит в пределах от 0,33252 до 1,37534.

Определим доверительный интервал для линии регрессии

$$\hat{y} - t_{1-\alpha;n-p-1} \cdot S_{\hat{y}} \le M_x(Y) \le \hat{y} + t_{1-\alpha;n-p-1} \cdot S_{\hat{y}}$$

С вероятностью 95% выработка на одного работающего для предприятий с мощностью пласта 8 м и уровнем механизации 6% лежит в пределах **$4{,}52204 \le M_x(Y) \le 6{,}46673$** тонн.

Определим доверительный интервал для индивидуальных значений

$$\hat{y}_0 - t_{1-\alpha;n-p-1} \cdot S_{\hat{y}_0} \le y_0^* \le \hat{y}_0 + t_{1-\alpha;n-p-1} \cdot S_{\hat{y}_0}$$

С вероятностью 95% индивидуальные значения выработки на одного работающего для предприятий с мощностью пласта 8 м и уровнем механизации 6% лежит в пределах **$3{,}04461 \le y_0^* \le 7{,}94416$** тонн.

Определим доверительный интервал для дисперсии

$$\frac{nS^2}{\chi^2_{\alpha/2;n-p-1}} \le \sigma^2 \le \frac{nS^2}{\chi^2_{1-\alpha/2;n-p-1}}$$

Значения величин ХИ-квадрат распределения определим с помощью функции =ХИ2ОБР. Истинное значение дисперсии с вероятностью 95% лежит в пределах $0{,}56469 \le \sigma^2 \le 5{,}35087$.

7. Определим значимость модели в целом

	I	J	K	L	M	N	O	P	Q	R	S
66											
67			$Q_{общ} = \sum_{i=1}^{n}(y_i - \bar{y})^2$				$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_e}{Q}$			$R = \sqrt{R^2}$	
68		=E2-E16	-0,8	0,64	=K68^2						
69		=E3-E16	-0,8	0,64	=K69^2		0,81162	=1-J55/L79		0,9009	
70		=E4-E16	-1,8	3,24	=K70^2						
71		=E5-E16	3,2	10,24	=K71^2		$\hat{R}^2 = 1 - \frac{n-1}{n-p-1}(1-R^2)$				
72		=E6-E16	-1,8	3,24	=K72^2						
73		=E7-E16	1,2	1,44	=K73^2		0,7578	=1-(1-O69)*9/7=1-(1-O69)*9/7			
74		=E8-E16	-1,8	3,24	=K74^2						
75		=E9-E16	-0,8	0,64	=K75^2		$F = \frac{R^2 \cdot (n-p-1)}{(1-R^2) \cdot p} > F_{\alpha;p;n-p-1}$				
76		=E10-E16	0,2	0,04	=K76^2						
77		=E11-E16	3,2	10,24	=K77^2						
78							15,0794	>	4,73741		
79			Σ	33,6		=O69*7/((1-O69)*2)			=FРАСПОБР(0,05;2;7)		
80							Р-значение	0,0029	=FРАСП(O78;2;7)		

Выборочный коэффициент множественной детерминации рассчитывается по формуле

$$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_e}{Q} = 0{,}81162 \quad ,$$

где $Q = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad Q_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \quad Q_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Полученное значение показывает, что на 81,16% с помощью линейного уравнения множественной регрессии удалось объяснить вариацию зависимой переменной влиянием включённых факторов. Оставшиеся 18,84% - влиянием неучтённых в модели и случайных факторов.

Выборочный множественный коэффициент корреляции $R = \sqrt{R^2}$ **= 0,9009** свидетельствует о высокой степени зависимости между результирующей переменной и влияющими на неё факторами.

Скорректированный (адаптированный) коэффициент детерминации равен

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1}(1 - R^2) = \mathbf{0{,}7578.}$$

Полученное значение характеризует значимость уравнения регрессии в целом.

Оценим статистическую значимости полученного уравнения регрессии в целом с помощью F-критерия Фишера. Уравнение множественной регрессии значимо или нулевая гипотеза Но о равенстве одновременно нулю всех параметров регрессионной модели, т.е. ***Н*о**: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_p = 0$ отвергается, если расчётное значение критерия Фишера будет больше критического (табличного) значения. Для проверки значимости модели в целом используем выражение для оценки значимости коэффициента детерминации

$$F = \frac{R^2 \cdot (n-p-1)}{(1-R^2) \cdot p} > F_{\alpha;p;n-p-1},$$

где R^2 – коэффициент множественной детерминации, n – объём выборки, p – количество факторных переменных в модели.

Расчётное значение ***F***-статистики 15,0794, критическое значение для доверительного уровня γ = 0,95 (уровня значимости α = 0,05) определяется с помощью статистической функции **=FРАСПОБР(*α*; *p*; *n*–*p*–1)**. $F_{кр}$= 4,73741. Поскольку расчётное значение критерия Фишера больше критического (табличного) значения, то с вероятностью 95% уравнение статистически значимо.

Определим значимость ***F***-статистики (вероятность ошибки) с помощью функции **=FРАСП(*F*; *p*; *n*–*p*–1)** = 0,0029. Поскольку полученное значение меньше величины α = 0,05, вывод о значимости уравнения регрессии подтверждается.

Но поскольку данное уравнение содержит и незначимые параметры, использование его как многофакторной модели может привести к неверным выводам при анализе ситуации и прогнозировании её развития.

5. ЗАДАНИЯ ДЛЯ САМОПРОВЕРКИ

5.1 Тесты

1. Пространственные данные – это …
1) данные по какому-либо экономическому показателю, полученные от разных однотипных объектов, но относящиеся к одному и тому же моменту времени
2) данные, характеризующие один и тот же объект в различные моменты времени
3) данные об эффективности использования производственных или торговых площадей

2. Временные ряды – это …
1) данные по какому-либо экономическому показателю, полученные от разных однотипных объектов, но относящиеся к одному и тому же моменту времени
2) данные, характеризующие один и тот же объект в различные моменты времени
3) данные о времени наработки на отказ производственного оборудования

3. К классам эконометрических моделей относятся … [ ]
1) модели временных рядов
2) регрессионные модели с одним уравнением
3) системы одновременных уравнений
4) экономические модели с функциональной зависимостью

4. Оценка является несмещённой оценкой параметра если...
1) её математическое ожидание равно оцениваемому параметру
2) она стремится к истинному значению параметра с

увеличением объема выборки
3) её дисперсия с увеличением выборки не изменяется
4) её дисперсия меньше дисперсии других оценок

5. Оценка является состоятельной оценкой параметра если...
1) ее математическое ожидание равно оцениваемому параметру
2) она стремится к истинному значению параметра с увеличением объёма выборки
3) её дисперсия с увеличением выборки не изменяется
4) её дисперсия меньше дисперсии других оценок

6. Оценка является эффективной оценкой параметра если...
1) ее математическое ожидание равно оцениваемому параметру
2) она стремится к истинному значению параметра с увеличением объема выборки
3) её дисперсия с увеличением выборки не изменяется
4) её дисперсия меньше дисперсии других оценок

7. В эконометрических моделях с *m* независимыми переменными наблюдаемые значения зависимой переменной y_i, i=1, 2,…, n, отличаются от модельных $\hat{y}_i$ на величину e_i, ($y_i = \hat{y}_i + e_i$). В данных обозначениях формула для расчета оценки общей дисперсии зависимой переменной $D_{общ}$ имеет вид:

1) $D = \frac{\sum (y_i - \bar{y})^2}{n-1}$ 2) $D = \frac{\sum (\hat{y}_i - \bar{y})^2}{m}$

3) $D = \frac{\sum (y_i - \hat{y}_i)^2}{n-m-1}$ 4) $D = \frac{\sum e_i^2}{n-m-1}$

8. В эконометрических моделях с *m* независимыми переменными наблюдаемые значения зависимой переменной y_i, i=1, 2,…, n, отличаются от модельных $\hat{y}_i$ на величину e_i, ($y_i = \hat{y}_i + e_i$). В данных обозначениях формула для расчета оценки факторной (обусловленной регрессией) дисперсии зависимой переменной $D_{факт}$ имеет вид:

1) $D = \frac{\sum (y_i - \bar{y})^2}{n-1}$ **2)** $D = \frac{\sum (\hat{y}_i - \bar{y})^2}{m}$

3) $D = \frac{\sum (y_i - \hat{y}_i)^2}{n-m-1}$ 4) $D = \frac{\sum e_i^2}{n-m-1}$

9. В эконометрических моделях с m независимыми переменными наблюдаемые значения зависимой переменной y_i, i=1, 2,…, n, отличаются от модельных $\hat{y}_i$ на величину e_i, ($y_i = \hat{y}_i + e_i$). В данных обозначениях оценки остаточной дисперсии $D_{ост}$ имеют вид: [ ]

1) $D = \dfrac{\sum (y_i - \bar{y})^2}{n-1}$ 2) $D = \dfrac{\sum (\hat{y}_i - \bar{y})^2}{m}$

3) $D = \dfrac{\sum (y_i - \hat{y}_i)^2}{n-m-1}$ 4) $D = \dfrac{\sum e_i^2}{n-m-1}$

10. В эконометрических моделях наблюдаемые значения зависимой переменной y_i,. i=1, 2,..., n, отличаются от модельных $\hat{y}_i$ на величину e_i ($y_i = \hat{y}_i + e_i$). В данных обозначениях формула для расчета общей суммы квадратов отклонений имеет вид:

1) $\sum (y_i - \hat{y}_i)^2$ 2) $\sum e_i^2$

3) $\sum (y_i - \bar{y})^2$ 4) $\sum (\hat{y}_i - \bar{y})^2$

5.2. Вопросы для подготовки к экзамену

1.Использование математического инструментария эконометрического моделирования для решения экономических задач

2. Точечные и интервальные оценки параметров эконометрических моделей

3. Доверительный интервал для функции регрессии и для индивидуальных значений зависимой переменной

4. Доверительный интервал для параметров регрессионной модели

5. Оценка качества уравнения регрессии

6. Коэффициент парной корреляции и оценка его значимости

7. Коэффициент детерминации и оценка его значимости

8. Оценка параметров классической регрессионной модели методом наименьших квадратов

9. Предпосылки множественного регрессионного анализа

10. Оценка значимости коэффициентов регрессии

11. Мультиколлинеарность. Расчет коэффициентов корреляции

21. Модели с переменной структурой. Фиктивные переменные

13. Тест Чоу

14. Нелийейные модели регрессии

15. Корреляция для нелинейной регрессии. Коэффициенты эластичности

16. Частная корреляция

17. Временные ряды. Автокорреляция уровней временного ряда

18. Аналитическое сглаживание временного ряда

19. Прогнозирование на основе временных рядов. Интервальные оценки прогноза

20. Авторегрессионные модели и модели скользящей средней

21. Обобщенный метод наименьших квадратов

22.Гетероскедастичность случайной составляющей. Обнаружение и устранение гетероскедастичности. Метод Голдфельда-Квандта

23. Метод взвешенных наименьших квадратов

24. Автокорреляция случайных составляющих. Тест Дарбина-Уотсона

25. Система одновременных уравнений. Проблемы идентификации

26. Косвенный и двухшаговый метод наименьших квадратов

27. Проведение анализа результатов эконометрического моделирования при решении профессиональных задач

28. Осуществление интерпретации результатов эконометрического моделирования при оценке параметров взаимосвязей и прогнозировании тенденций

ЗАКЛЮЧЕНИЕ

Изучение дисциплины «Эконометрика» направлено на направлено на получение знаний и общих представлений об использования современных эконометрических методов и моделей для прогнозирования при обосновании управленческих решений; освоение различных способов выражения связей и закономерностей развития экономических процессов и явлений через эконометрические модели и методы проверки их адекватности, основанные на данных статистических наблюдений; освоение современного инструментария эконометрического моделирования.

Кроме того, дисциплина «Эконометрика» ориентирована на развитие умений и навыков обучающихся в части поиска информации, сбора, анализа и обработки данных, необходимых для решения профессиональных задач, работы с литературой, выбора инструментальных средства для обработки экономических данных в соответствии с поставленной задачей, анализа результатов расчетов и обоснования полученных выводов в процессе принятия решений.

Изучение дисциплины «Эконометрика» является одним из основных этапов в формировании высококвалифицированных кадров, которые необходимы предприятиям и организациям в современных условиях.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Амелин С.В. Эконометрика [Текст]: учебное пособие / ФГБОУ ВО «Воронеж. гос. техн. ун-т», 2016. 142 с.. – Режим доступа http://bibl.cchgeu.ru/MarcWeb2/Found.asp Сервер каф. ЭБ: R:\Литература\для СПЕЦИАЛИСТОВ \ Эконометрика

2. Агаларов З.С. Эконометрика : учебник / Агаларов З.С., Орлов А.И.. — Москва : Дашков и К, 2021. — 380 c. — ISBN 978-5-394-04075-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: https://www.iprbookshop.ru/107834.html (дата обращения: 03.11.2021). — Режим доступа: для авторизир. пользователей до 18.05.2024

3. Амелин С.В. Эконометрика: практикум [Текст]: учебное пособие / ФГБОУ ВО «Воронежский государственный технический университет», 2016. 125 с. – Режим доступа http://bibl.cchgeu.ru/MarcWeb2/Found.asp Сервер каф. ЭБ: R:\Литература\для СПЕЦИАЛИСТОВ \ Эконометрика

4. Амелин С.В. Методические указания по выполнению самостоятельной работы и индивидуальных заданий по дисциплине "Эконометрика" для студентов, обучающихся по направлению 38.03.01 «Экономика» очной формы обучения / Сост. С. В. Амелин. - ФГБОУ ВПО "Воронежский государственный технический университет", 2011. - 43 с. – Режим доступа http://bibl.cchgeu.ru/MarcWeb2/Found.asp Сервер каф. ЭБ: R:\Литература\для СПЕЦИАЛИСТОВ \ Эконометрика

5. Шаравова О.И. Эконометрика : учебно-методическое пособие по выполнению лабораторных работ / Шаравова О.И.. — Москва : Московский технический университет связи и информатики, 2018. — 14 c. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: https://www.iprbookshop.ru/92489.html (дата обращения: 03.11.2021). — Режим доступа: для авторизир. пользователей до 12.02.2025

ПРИЛОЖЕНИЕ

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Воронежский государственный технический университет»
(ФГБОУ ВО «ВГТУ», ВГТУ)

Факультет экономики, менеджмента и информационных технологий
Кафедра экономической безопасности

КУРСОВАЯ РАБОТА

по дисциплине: Эконометрика

на тему: ____________________

Выполнил(а)

студент ____________________ (____________________)
(курс, группа, подпись) (инициалы, фамилия)

Руководитель
_д.э.н, проф____________ (_С.В. Амелин___)
(должность, подпись) (инициалы, фамилия)

Дата сдачи «____ » ___________20__ г. Оценка ____________________

Воронеж 2022

ОГЛАВЛЕНИЕ

ЭКОНОМЕТРИКА

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к изучению дисциплины, выполнению лабораторных работ,
самостоятельной и курсовой работ для обучающихся
по специальности 38.05.01 "Экономическая безопасность"
(специализации "Экономико-правовое обеспечение экономической безопасности",
"Экономика и организация производства на режимных объектах")
всех форм обучения

Составитель
Амелин Станислав Витальевич

В авторской редакции

Подписано к изданию 20.06.2022.
Уч.–изд. л. 2,2.

ФГБОУ ВО «Воронежский государственный технический
университет»
394006 Воронеж, ул. 20-летия Октября, 84