

**Н.А. ДЕГТЯРЕВА, Д.С. ГОРДЕЕВА**

**МЕТОДЫ СТАТИСТИКИ**

**УЧЕБНОЕ ПОСОБИЕ**

**ЧЕЛЯБИНСК**

**2021 г**

УДК 330.115(021) : 311(021)

ББК 65в641я73 : 60.6я73

Д 26

ISBN 978-5-93162-520-1

Дегтярева, Н.А. Методы статистики [Текст]: учебное пособие для студентов вузов / Н.А. Дегтярева, Д.С. Гордеева – Челябинск: Изд-во ЗАО «Библиотека А. Миллера», 2021. – 63 с.

Учебное пособие содержит краткое изложение методического материала по использованию статистических методов, таких как корреляционный анализ и регрессионный анализ; методические указания по решению типовых задач и 35 вариантов задач для тренировочной, самостоятельной и контрольной работы студентов.

Учебное пособие предназначено для студентов высших учебных заведений, обучающихся по экономическим и неэкономическим специальностям и направлениям, как очной, так и заочной форм обучения.

Рецензенты:

А.С. Кутузов, к.физ.-мат.н., доцент.

Н.А. Берг, к.э.н., доцент.

ISBN 978-5-93162-520-1

Дегтярева Н.А, Гордеева Д.С. ,2021

## СОДЕРЖАНИЕ

	ВВЕДЕНИЕ.....	4
1	КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ В СТАТИСТИКЕ...	8
2	ПОСТРОЕНИЕ И АНАЛИЗ МОДЕЛИ ПАРНОЙ ЛИНЕЙНОЙ И НЕЛИНЕЙНОЙ РЕГРЕССИИ.....	21
3	ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ.....	49
4	ГЛОССАРИЙ.....	52
	БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	57

## ВВЕДЕНИЕ

Сегодня деятельность в любой области экономики (управлении, финансово – кредитной сфере, маркетинге, учете, аудите) требует от специалиста применения современных статистических методов работы, знания достижений мировой экономической мысли, понимания научного языка. Большинство новых методов основано на статистических моделях, концепциях и приемах [22,30].

В учебном пособии рассмотрим применение методов статистики в науке, которая дает количественное выражение взаимосвязей экономических явлений и процессов. Эта наука возникла в результате взаимодействия и объединения трех компонент: экономической теории, статистических и математических методов. Она занимается определением наблюдаемых в экономической жизни конкретных количественных закономерностей, применяя для этой цели статистические методы – методы парной и множественной регрессии, парной, частной и множественной корреляции, выделения тренда и других компонент временного ряда, на статистическом оценивании [31,7].

Специфической особенностью деятельности экономиста является работа в условиях постоянно меняющейся информации и неполноты исходных данных. Анализ такой информации требует специальных методов, которые составляют один из аспектов науки эконометрики [19,29,33].

Рыночная экономика требует улучшения использования статистической и экономической информации, проведения эконометрических расчетов будущих результатов хозяйственной

деятельности и обоснования путей их достижения. Для достижения этих целей необходимо провести построение эконометрической модели и определить возможности ее использования для описания, анализа и прогнозирования реальных экономических процессов [8,11,16].

В современных условиях управленческие решения должны приниматься лишь на основе тщательного анализа имеющейся информации, поэтому на практике все больший интерес возникает к статистическим методам, в частности, к корреляционно-регрессионному анализу, и компьютерным программам для их реализации (Econometric Views и Statistica, SPSS и др) [2, 3,36].

Компьютерные эконометрические пакеты сделали эконометрические методы более доступными и наглядными, т.к. наиболее трудоемкую (рутинную) работы по расчету различных статистик, параметров, характеристик, построению таблиц и графиков в основном стал выполнять компьютер [17,18].

Большинство процессов и явлений в экономике находятся в постоянной, взаимной, объективной связи. Исследование зависимостей и взаимосвязей между объективно существующими явлениями и процессами играет большую роль. Эффективная хозяйственная деятельность невозможна без оценки этих связей между различными факторами и результативными показателями, выявление их тенденций и разработки прогнозов [22].

По направлению и сложности связей между внутренними (эндогенными) переменными и внешними (экзогенными) переменными интерес вызывают регрессионные модели, т.е. экономико–математические модели факторного анализа, параметры

которой оцениваются средствами математической статистики. Эта модель выступает в качестве средства анализа и прогнозирования конкретных экономических процессов [13,15,20].

Методы статистики, методы статистического анализа, а именно корреляционный анализ позволяет установить связь между случайными переменными и оценить ее тесноту, а задачами регрессионного анализа являются установление формы зависимости между переменными, оценка функции регрессии и оценка неизвестных значений (прогноз) зависимой переменной [6,14].

Задачей практикума является дать, по сути, введение в регрессионный анализ и рассмотреть его как один из методов статистического анализа, не вдаваясь в тонкости.

Регрессионный анализ – совокупность методов, позволяющих исследовать вид односторонней зависимости объясняемой (зависимой) переменной  $Y$  от одной или нескольких объясняющих (независимых) переменных  $X = (X_1, \dots, X_p)$  и подбирать ее параметры.

В регрессионном анализе определяются точные количественные характеристики изменения  $Y$  и на этапе статистического анализа решаются следующие основные задачи [29,32]:

- 1) выбор общего вида функции регрессии  $\varphi(x^*, \theta)$ ;
- 2) отбор, если необходимо, наиболее информативных факторов;
- 3) оценивание параметров уравнение регрессии  $\theta = (\theta_1, \theta_2, \dots, \theta_i)$ ;

4) анализ точности полученного уравнения регрессии путем построения доверительных интервалов для коэффициентов регрессии, т.е. компонент вектора  $\theta$ , для условного среднего отклика  $\bar{y}(x)$  и для прогнозов наблюдений отклика  $y(x^*)$  при значениях факторов  $x^* = (x_1^*, \dots, x_p^*)$ .

Практическое значение знания регрессионной зависимости между случайными переменными  $X$  и  $Y$  заключается в возможности прогнозирования значения зависимой случайной переменной  $Y$ , когда независимая случайная переменная  $X$  принимает определенное значение. Прогноз не может быть безошибочным, однако можно определить границы вероятности ошибки прогноза.

В учебном пособии также проиллюстрируем основные идеи корреляционно-регрессионного анализа на примере исследования влияния на зависимую переменную  $Y$  одного фактора  $X$ , построение регрессионной модели (линейной и нелинейной) и на ее основе дается оценка зависимой переменной.

# 1 КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ В СТАТИСТИКЕ

Почти все наблюдаемые явления экономической жизни общества – следствие действия определенных факторов. Например, прибыль, получаемая предприятием, связана со множеством показателей: численностью работников, их образованием, стоимостью основных производственных фондов и др.

Общественная жизнь состоит из большого количества сложных явлений, которые формируются под влиянием многочисленных, разнообразных и взаимосвязанных факторов. Понять и изучить какое-либо явление можно, исследуя его во взаимосвязи с окружающими признаками. Для исследования интенсивности, вида и формы зависимостей широко применяется *корреляционно – регрессионный анализ*, который является методическим инструментарием при решении задач прогнозирования, планирования и анализа хозяйственной деятельности предприятий.

Корреляционно - регрессионный анализ является одним из наиболее распространенных методов обработки результатов наблюдений при изучении зависимостей в физике, биологии, экономике и других областях [9,12,25].

Основные требования к статистической информации, исследуемой методами корреляционного-регрессионного анализа:

- используемые для исследования наблюдения должны являться случайно выбранными из генеральной совокупности объектов;
- требование независимости наблюдений друг от друга;

- исходная совокупность данных должна быть однородной, без аномальных наблюдений;
- исходные данные должны подчиняться нормальному закону распределения;
- число наблюдений по которым устанавливается взаимосвязь признаков и строится модель регрессии, должно превышать количество факторных признаков хотя бы в 3-4 раза (лучше в 8-10 раз);
- факторные признаки не должны находится между собой в функциональной зависимости.

В теории вероятностей *регрессия* случайной величины  $Y$  по величине  $X$  (если известно совместное распределение случайной величины  $X$  и  $Y$ ) определяется как условное математическое ожидание случайной величины  $Y$ , вычисленное при условии  $X = x$  :  $M(Y/x) = y(x)$ . Уравнение  $y = y(x)$ , в котором  $x$  играет роль независимой переменной, называется уравнением регрессии  $Y$  по  $X$ . Точность, с которой уравнение регрессии  $Y$  по  $X$  отражает изменение  $Y$  в среднем при изменении  $x$ , измеряется условной дисперсией  $D(Y/x) = \sigma^2(x)$ .

Цель регрессионного анализа состоит в определении общего вида уравнения регрессии, построении статистических оценок неизвестных параметров, входящих в уравнение регрессии, построении статистических оценок неизвестных параметров, входящих в уравнение регрессии и проверке статистических гипотез о регрессии. При изучении связи между двумя случайными величинами по данным выработки  $(x_1, y_1), \dots, (x_n, y_n)$  в соответствии с теорией

регрессии предполагается, что одна из них  $Y$  имеет некоторое распределение вероятностей при фиксированном значении  $x$ ,

Рассмотрим самую простую линейную модель:

$$y = \alpha_0 + \alpha_1 x + \varepsilon$$

где  $\alpha_0, \alpha_1$  неизвестные параметры регрессии,  $\varepsilon$  – случайная ошибка наблюдений, причем  $M\varepsilon = 0$  и  $D\varepsilon = \sigma^2$ . Значение дисперсии ошибок наблюдений  $\sigma^2$  также не известно.

Линейная модель - экономико–математическая модель факторного анализа, параметры которой оцениваются средствами математической статистики. Эта модель выступает в качестве средства анализа и прогнозирования конкретных экономических процессов [24,34].

По направлению и сложности связей между внутренними (эндогенными) переменными и внешними (экзогенными) переменными интерес вызывают регрессионные модели.

Регрессионными называют модели, основанные на уравнении регрессии, или системе регрессионных уравнений, связывающих величины эндогенных и экзогенных переменных. Различают уравнения (модели) парной и множественной регрессии. Если для обозначения эндогенных переменных использовать букву  $y$ , а для экзогенных переменных букву  $x$ , то в случае линейной модели уравнение парной регрессии имеет вид  $y_t = a_0 + a_1 x_{t1} + \varepsilon_t$ , а уравнение множественной регрессии:  $y_t = a_0 + a_1 x_{t1} + a_2 x_{t2} + \dots + a_m x_{tm} + \varepsilon_t$ .

Основными этапами построения регрессионной модели являются:

- формирование перечня факторов и их логический анализ;
- сбор исходных данных и их первичная обработка;
- спецификация функции регрессии;
- оценка функции регрессии;
- отбор главных факторов;
- проверка адекватности модели;
- экономическая интерпретация;
- прогнозирование на основе модели регрессии.

Вопросы построения и использования эконометрических моделей множественной регрессии рассмотрим на примере линейной двухфакторной модели.

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессии используют метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака  $y$  от теоретических  $y_p$  минимальна, то есть

$$\sum_{t=1}^n ((y_t - (a_0 - a_1 x_t))^2 \rightarrow \min.$$

Для линейных уравнений, приводимых к линейным, решается система относительно  $a_0$  и  $a_1$ :

$$a_0 \cdot n + a_1 \cdot \sum_{t=1}^n x_t = \sum_{t=1}^n y_t$$

$$a_0 \cdot \sum_{t=1}^n x_t + a_1 \cdot \sum_{t=1}^n x_t^2 = \sum_{t=1}^n y_t \cdot x_t$$

Для упрощения расчетов используются готовые формулы, которые вытекают из этой системы:

$$a_0 = \bar{y} - a_1 \cdot \bar{x}$$

$$a_1 = \frac{\sum (y_t - \bar{y})(x_t - \bar{x})}{\sum (x_t - \bar{x})^2}$$

где  $\bar{y} = \frac{\sum_{t=1}^n y_t}{n}$ ,  $\bar{x} = \frac{\sum_{t=1}^n x_t}{n}$  - выборочные средние.

Основу математического аппарата для рассматриваемых моделей составляют такие разделы статистики, как корреляционный и регрессионный анализ. Эндогенные переменные в этих моделях будем называть результативными признаками (обозначать их буквой  $y$ ), а экзогенные переменные – факторными признаками (обозначать их буквой  $x$ ) [1,26,27].

Регрессионный анализ позволяет определить конкретный вид зависимости между переменной величиной  $y$  и, влияющими факторами, в виде некоторой параметрической функции.

Корреляционный анализ позволяет оценить степень тесноты статистической связи между двумя переменными величинами, которая может быть как линейной, так и нелинейной.

Методы корреляционно – регрессионного анализа позволяют решать три основные задачи: определение формы связи между результативным и факторными признаками, измерение тесноты связи между ними, анализ влияния отдельных факторных признаков.

Для определения меры степени линейной связи между двумя переменными используется выборочный коэффициент корреляции ( $-1 \leq r_{xy} \leq 1$ ):

$$r_{yx} = \frac{\sum (y_t - \bar{y})(x_t - \bar{x})}{\sqrt{\sum (y_t - \bar{y})^2} \sqrt{\sum (x_t - \bar{x})^2}}$$

Если факторов несколько, то из них выбирается в модель один или два фактора, с помощью парных коэффициентов корреляции. Составляется матрица (таблица 1) парных коэффициентов корреляции, измеряющих тесноту связи каждого из факторов-признаков с результативным фактором и между собой:

Таблица 1– Матрица парных коэффициентов корреляции

Факторы	$y$	$x_1$	$x_2$	...	$x_m$
$y$	<b>1</b>	$r_{yx1}$	$r_{yx2}$	...	$r_{yxm}$
$x_1$	$r_{yx1}$	<b>1</b>		...	$r_{x1xm}$
$x_2$	$r_{yx2}$	$r_{x1x2}$	<b>1</b>	...	$r_{x2xm}$
...	...	...	...	...	...
$x_m$	$r_{yxm}$	$r_{x1xm}$	$r_{x2xm}$	...	<b>1</b>

Значение коэффициента парной корреляции лежит в интервале от -1 до +1. Его положительное значение свидетельствует о прямой связи ( $r_{yx} > 0$ ); отрицательное ( $r_{yx} < 0$ ) – об обратной. Чем ближе его значение к 1, тем теснее связь. Считается, что связь достаточно сильная, если коэффициент корреляции по абсолютной величине превышает 0,7, то есть  $0,7 < |r_{yx}| < 1$ ; слабая, если меньше 0,3, то есть  $|r_{yx}| < 0,3$ ; при равенстве его нулю, связь полностью отсутствует.

В модель включают те факторы, связь которых с зависимой переменной наиболее сильная.

Одним из условий регрессионной модели является предположение о линейной независимости объясняющих

переменных. Для экономических показателей это условие выполняется не всегда. Линейная или близкая к ней связь между факторами называется мультиколлинеарностью и приводит к линейной зависимости нормальных уравнений, что делает вычисление параметров либо невозможным, либо затрудняет содержательную интерпретацию параметров модели.

Для устранения мультиколлинеарности используют метод исключения переменных. Этот метод заключается в том, что высоко коррелированные объясняющие переменные (факторы) устраняются из регрессии, и она заново оценивается. Отбор переменных, подлежащих исключению, производится с помощью коэффициентов парной корреляции. Опыт показывает, что если ,

$|r_{хкх1}| \geq 0,7$  , то одну из переменных можно исключить.

Для оценки значимости коэффициента корреляции применяется t-критерий Стьюдента. При этом фактическое значение этого критерия определяется по формуле:  $t_{набл} = \sqrt{r^2_{xy}(n-2)/(1-r^2_{xy})}$

Вычисленное по этой формуле значение  $t_{набл}$  сравнивается с критическим значением t-критерия, которое берется из таблицы значений t-Стьюдента с учетом заданного уровня значимости ( $\alpha = 0,05$ ) и числа степеней свободы  $k = n - 2$

Если  $t_{набл} > t_{кр}$  , то полученное значение коэффициента корреляции признается значимым (то есть нулевая гипотеза, утверждающая равенство нулю коэффициента корреляции, отвергается). Делается вывод о том, что между исследуемыми переменными есть тесная статистическая взаимосвязь.

Качество модели оценивается по адекватности и точности на основе анализа остатков регрессии:  $E(t) = y_t - \hat{y}(t)$  - остаточный ряд.

Анализ остатков позволяет получить представление, насколько хорошо подобрана сама модель и насколько правильно выбран метод оценки коэффициентов.

Под адекватностью модели понимается соответствие ее реальному экономическому процессу.

Модель адекватна, если ее остаточный ряд  $E(t)$  является случайной компонентой. Это значит, что для остаточного ряда, как для случайной компоненты, должны выполняться четыре ее свойства [34]:

- 1) математическое ожидание значений остаточного ряда равно нулю;
- 2) значения остаточного ряда случайны;
- 3) уровни остаточного ряда независимы;
- 4) подчинены нормальному закону распределения.

Проверка равенства математического ожидания уровней ряда остатков нулю осуществляется с использованием  $t$  критерия Стьюдента:

$$t_p = \frac{|\bar{E}|}{\sigma_E} \cdot \sqrt{n}$$

где  $\sigma_E$  - среднеквадратическое отклонение уровней остаточного ряда:

$$\sigma_E = \sqrt{\frac{\sum_{t=1}^n (E(t) - \bar{E})^2}{n-1}}$$

где  $\bar{E}$  - среднее значение уровней остаточного ряда:  $\bar{E} = \frac{\sum_{t=1}^n E(t)}{n}$

Если  $t_p < t_\alpha$ , то  $M(E) \approx 0$ , где  $t_\alpha$  – критерий Стьюдента,  $\alpha$  - уровень значимости.

На практике, если  $|\bar{E}| < 0,5$ , то критерий Стьюдента выполняется и, значит, гипотеза о равенстве математического ожидания значений остаточного ряда нулю выполняется.

Проверка случайного характера отклонения может осуществляться различными методами: методом серий знаков; поворотных точек (пиков) и др.

Рассмотрим *метод поворотных точек*:

Значение  $E(t)$  - является поворотной точкой, если  $E(t) > E(t-1)$  и  $E(t) > E(t+1)$  или  $E(t) < E(t-1)$  и  $E(t) < E(t+1)$ :

Число поворотных точек  $P$  должно удовлетворять строгому

неравенству:

$$P > \left[ 2 \cdot \frac{(n-2)}{3} - 1,96 \sqrt{\frac{16n-29}{90}} \right]$$

(Квадратные скобки означают, что берётся целая часть числа, т. е. наибольшее целое число не превосходящее данное.)

Числом 1 отмечаются поворотные точки, числом 0 – неповоротные точки.

Уровни остаточного ряда независимы, если отсутствует автокорреляция. Наличие автокорреляции устанавливаются с помощью  $d$ -критерия Дарбина-Уотсона:

$$d = \frac{\sum_{t=2}^n (E(t) - E(t-1))^2}{\sum_{t=1}^n E(t)^2}.$$

Полученное значение  $d$  сравнивается с двумя критическими табличными значениями  $d_1$  и  $d_2$ . ( $n \leq 10$ ;  $d_1 = 1,08$ ,  $d_2 = 1,36$ .)

Если  $d > 2$ , то находят  $d' = 4 - d$  и к найденным  $d$  и  $d'$  применяют критерий Дарбина-Уотсона:

Наличие нормального закона распределения в остаточном ряду линейной функции устанавливается с помощью  $R/S$ - критерия:

$$RS = \frac{E_{\max} - E_{\min}}{S_E}$$

где  $E_{\max}$  - максимальный уровень ряда остатков;

$E_{\min}$  - минимальный уровень ряда остатков;

$$S_E = \sqrt{\frac{\sum E(t)^2}{n-1}} \quad - \quad \text{«исправленное» среднее квадратическое}$$

отклонение

Вычисленное значение сравнивают с табличными верхней и нижней границами;

Фрагмент таблицы для уровня значимости  $\alpha = 0,05$ :

$n$	Нижняя граница	Верхняя граница
10	2,67	3.68
20	3,18	4,49
30	3,47	4,89

Если значение этого критерия попадает между табулированными границами с заданным уровнем значимости, то гипотеза о нормальном законе распределения ряда остатков принимается

Если все четыре условия выполняются, то остаточный ряд является случайной компонентой, а модель адекватна [34].

В качестве критерия точности модели можно использовать:

среднее квадратическое отклонение (где  $\bar{S}_E$  – ошибка аппроксимации):

$$\overline{S_E} = \sqrt{\frac{\sum E(t)^2}{n-m-1}},$$

где  $m$  – число параметров модели; среднюю относительную ошибку:  $\overline{E_{отн}} = \frac{1}{n} \sum \frac{|E(t)|}{y_t} \cdot 100\%$ , если  $\overline{E_{отн}} < 5\%$ , то модель точная; модель допустимо точная, если  $\overline{E_{отн}} \leq 10\%$ , в противном случае модель не точная и по ней нельзя выполнять прогноз.

Оценку качества построенной модели также дает коэффициент детерминации:

$$R^2 = 1 - \frac{S_E^2}{S_y^2} = 1 - \frac{\sum e^2(t)}{\sum (y_t - \bar{y})^2}$$

Если  $R \approx 1$ , то модель качественная. Коэффициент  $R^2$  показывает долю вариации результативного признака под воздействием изучаемых факторов.

Используя коэффициент детерминации можно проверить значимость уравнения регрессии с помощью критерия Фишера.

Для этого выполняется сравнение фактического  $F_{факт}$  и критического (табличного)  $F_{табл}$  значений F - критерия Фишера.  $F_{факт}$  определяется из соотношения:

$$F_{факт} = \frac{R^2 / m}{(1 - R^2) / (n - m - 1)}$$

Если  $F_{табл} < F_{факт}$ , то гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность. Важную роль при оценке влияния факторов играют коэффициенты регрессионной модели. Однако непосредственно с их помощью нельзя сопоставить факторы по степени их влияния на результативный признак из-за различия

единиц измерения и разной степени колеблемости. Для устранения таких различий применяются средние частные коэффициенты эластичности и бета-коэффициент.

Для модели парной регрессии коэффициент эластичности рассчитывается по формуле:

$$\mathcal{E}_{yx_i}^{\wedge} = a_i \frac{\bar{x}_i}{\bar{y}}$$

Коэффициент эластичности показывает, на сколько процентов изменяется результирующий признак  $y$  при изменении факторного признака  $x_i$  на 1%;

Бета  $\beta$ -коэффициент вычисляют по формуле:

$$\beta_{yx_i}^{\wedge} = a_i \frac{S_{x_i}}{S_y} = a_i \sqrt{\frac{\sum (x_i - \bar{x}_i)^2}{\sum (y_t - \bar{y})^2}}$$

где  $S_{x_i}$  и  $S_y$  – средние квадратические ошибки выборки величин  $x_i$  и  $y$ .

Бета  $\beta$ -коэффициент показывает, на какую часть величины своего среднего квадратического отклонения изменится в среднем значение результирующего признака  $y$  при изменении факторного признака  $x_i$  на величину его среднеквадратического отклонения при фиксированном на постоянном уровне значений остальных независимых переменных.

Указанные коэффициенты позволяют проранжировать факторы по степени влияния факторов на зависимую переменную.

При использовании построенной модели для прогнозирования делается предположение о сохранении в период прогнозирования и на период упреждения существовавших ранее взаимосвязей переменных.

Прогноз состоит из двух частей: точечный и интервальный прогнозы.

*Точечный прогноз:* для прогнозирования зависимой переменной на  $k$  шагов вперед необходимо знать прогнозные значения всех входящих в модель факторов. Их оценки могут быть получены методами экспертных оценок или непосредственно заданы исследователем или могут быть получены на основе экстраполяционных методов, *например*, с использованием средних абсолютных приростов (САП) факторных признаков:

$$x(n+k) = x(n) + \text{САП} \cdot k, \quad \text{где } \text{САП} = (x_n - x_1)/(n-1)$$

Прогнозные значения факторов подставляют в модель и получают точечные прогнозные оценки изучаемого показателя:  $\hat{y}(n+k) = a_0 + a_1 \cdot x(n+k)$ .

*Интервальный прогноз:* для линейной модели регрессии область возможных значений результативного показателя при известных значениях факторов, то есть доверительный интервал прогноза рассчитывается следующим образом. Оценивается величина отклонения от линии регрессии (обозначим ее  $U$ ):

Величина  $U(k)$  для парной модели регрессии имеет вид:

$$U(k) = t_\alpha \cdot \overline{S_E} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x(n+k) - \bar{x})^2}{\sum (x_t - \bar{x})^2}}$$

где  $\overline{S_E}$  – ошибка аппроксимации, которая вычисляется по формуле:

$$\overline{S_E} = \sqrt{\frac{\sum E(t)^2}{n - m - 1}}, \quad n - \text{число наблюдений, } m - \text{число факторов в модели;}$$

$k$  – количество шагов;  $x_t$  – наблюдаемое значение факторного признака в момент  $t$ ;  $x(n+k)$  – прогнозное значение фактора на  $k$  шагов вперед;  $t_\alpha$  – табличное значение  $t$ -статистики Стьюдента при заданном уровне значимости  $\alpha$  и числа наблюдений  $n$ .

Если исследователь задает уровень вероятности попадания прогнозируемой величины внутрь доверительного интервала, равный 70%, то  $t_\alpha = 1,05$ ; если 95%, то  $t_\alpha = 1,96$ ; если 99%, то  $t_\alpha = 2,65$ .

Доверительный интервал прогноза для шага прогнозирования  $k$  имеет следующие границы:

верхняя граница прогноза:  $\hat{y}(n+k) + U(k)$ ,

нижняя граница прогноза:  $\hat{y}(n+k) - U(k)$ , причем  $U(1) < U(2)$ .

Если построенная регрессионная модель адекватна и прогнозные оценки факторов достаточно надежны, то с выбранной исследователем вероятностью можно утверждать, что при сохранении сложившихся закономерностей развития прогнозируемая величина попадет в интервал, образованный нижней и верхней границами.

## 2. ПОСТРОЕНИЕ И АНАЛИЗ МОДЕЛИ ПАРНОЙ ЛИНЕЙНОЙ И НЕЛИНЕЙНОЙ РЕГРЕССИИ

Проиллюстрируем основные идеи корреляционно-регрессионного анализа на примере исследования влияния на зависимую переменную  $Y$  одного фактора  $X$ .

**Пример 1:** В таблице приведены статистические данные показателя  $y(t)$  и данные факторов  $x_1(t)$  и  $x_2(t)$ :

	30	28	33	38	40	42	44	49	47
$x_1(t)$	5	7	10	12	15	18	20	23	24
$x_2(t)$	12	15	16	19	17	20	24	25	28

**Требуется:**

*Задание 1.*

1. Построить матрицу коэффициентов парной корреляции  $y(t)$  с  $x_1(t)$  и  $x_2(t)$  и выбрать фактор наиболее тесно связанный с зависимой переменной  $y(t)$ .
2. Построить линейную однопараметрическую модель регрессии  $y(t) = a_0 + a_1x(t)$  для выбранного  $x(t)$ .
3. Оценить качество построенной модели.
4. Для модели регрессии рассчитать коэффициент эластичности и бетта коэффициент.
5. Построить точечный и интервальный прогнозы на два шага вперёд по модели регрессии (для вероятности  $P = 70\%$  используйте коэффициент  $t_\alpha = 1,05$ ) (прогнозные оценки фактора  $x(t)$  на два шага вперёд получить на основе среднего прироста от фактически достигнутого уровня).

*Задание 2. Найти параметры модели  $\hat{y} = a \cdot b^x$ , с выбранным ранее фактором.*

Проверить качество этой модели, вычислить коэффициент детерминации, проверить значимость модели с использованием F-критерия Фишера, вычислить  $\bar{E}_{\text{отн}}$ . Если по модели можно сделать прогноз, то вычислить  $\hat{y}_{\text{прогн.}} \pm U(k)$ , взяв прогноз для  $x$ , значение составленное: 110% от последнего уровня:  $x_{\text{прогн.}} = 110\% \cdot x_n$ .

Отобразить на графиках фактические данные, результаты расчетов и прогнозирования. Вычисления провести с одним знаком в дробной части. Основные промежуточные результаты вычислений представить в таблицах.

### **Решение**

*Задание 1:*

1. Построить матрицу коэффициентов парной корреляции  $y(t)$  с  $x_1(t)$  и  $x_2(t)$  и выбрать фактор наиболее тесно связанный с зависимой переменной  $y(t)$ .

Найдем средние значения результативного признака и наблюдаемых факторов:

$$\bar{y} = \frac{\sum_{t=1}^9 y_t}{n} = \frac{351}{9} = 39,0 \qquad \bar{x}_1 = \frac{\sum x_1}{n} = \frac{134}{9} = 14,9 \qquad \bar{x}_2 = \frac{\sum x_2}{n} = \frac{176}{9} = 19,6$$

Вычислим парные коэффициенты корреляции: (в том числе и между факторами), используя промежуточные вычисления, приведённые в таблице 2:

$$r_{yx_1} = \frac{\sum (y - \bar{y})(x_1 - \bar{x}_1)}{\sqrt{\sum (y - \bar{y})^2} \sqrt{\sum (x_1 - \bar{x}_1)^2}} = \frac{397}{\sqrt{438} \sqrt{376,8}} = \frac{397}{405,8} \approx 0,98$$

$$r_{yx_2} = \frac{\sum (y - \bar{y})(x_2 - \bar{x}_2)}{\sqrt{\sum (y - \bar{y})^2} \sqrt{\sum (x_2 - \bar{x}_2)^2}} = \frac{288,2}{\sqrt{438} \sqrt{218,24}} = \frac{288,2}{308,8} \approx 0,93$$

$$r_{x_1x_2} = \frac{\sum (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)}{\sqrt{\sum (x_1 - \bar{x}_1)^2} \sqrt{\sum (x_2 - \bar{x}_2)^2}} = \frac{274,6}{\sqrt{376,8} \sqrt{218,24}} = \frac{274,6}{287,3} \approx 0,96$$

Таблица 2 – Расчетная таблица

	y	x <sub>1</sub>	x <sub>2</sub>	y - $\bar{y}$	x <sub>1</sub> - $\bar{x}_1$	(y - $\bar{y}$ )(x <sub>1</sub> - $\bar{x}_1$ )	(y - $\bar{y}$ ) <sup>2</sup>	(x <sub>1</sub> - $\bar{x}_1$ ) <sup>2</sup>
1	30	5	12	-9	-9,9	89,1	81	98,0
2	28	7	15	-11	-7,9	86,9	121	62,4
3	33	10	16	-6	-4,9	29,4	36	24
4	38	12	19	-1	-2,9	2,9	1	8,4
5	40	15	17	1	0,1	0,1	1	0,01
6	42	18	20	3	3,1	9,3	9	9,61
7	44	20	24	5	5,1	25,5	25	26,0
8	49	23	25	10	8,1	81,0	100	65,6
9	47	24	28	8	9,1	72,8	64	82,8
Σ	351	134	176			397,0	438	376,8

(x <sub>2</sub> - $\bar{x}_2$ )	(x <sub>2</sub> - $\bar{x}_2$ ) <sup>2</sup>	(y - $\bar{y}$ )(x <sub>2</sub> - $\bar{x}_2$ )	(x <sub>1</sub> - $\bar{x}_1$ )(x <sub>2</sub> - $\bar{x}_2$ )
-7,6	57,76	68,4	75,24
-4,6	21,16	50,6	36,34
-3,6	12,96	21,6	17,64
-0,6	0,36	0,6	1,74
-2,6	6,76	2,6	-0,26
0,4	0,16	1,2	1,24
4,4	19,36	22,0	22,44
5,4	29,16	54,0	43,74
8,4	70,56	67,2	76,44
	218,24	288,2	274,6

Из полученных парных коэффициентов корреляции составим матрицу:

	<b>y</b>	<b>x<sub>1</sub></b>	<b>x<sub>2</sub></b>
<b>y</b>	<b>1</b>	0,98	0,93
<b>x<sub>1</sub></b>	0,98	<b>1</b>	0,96
<b>x<sub>2</sub></b>	0,93	0,96	<b>1</b>

Так как  $r_{yx_1} = 0,97 > 0$ , то связь прямая;  $|r_{yx_1}| > 0,7$  – высокая

Так как  $r_{yx_2} = 0,91 > 0$ , то связь прямая;  $|r_{yx_2}| > 0,7$  – высокая

Так как  $|r_{x_1x_2}| = 0,97 > 0,8$ , то имеет место мультиколлинеарность, оба фактора в модель включать нельзя. Так как  $|r_{yx_1}| > |r_{yx_2}|$ , то в модель выбираем фактор  $x_1$ , потому что он наиболее тесно связан с зависимой переменной  $y$ .

2. Построить линейную однопараметрическую модель регрессии:  $y(t) = a_0 + a_1x(t)$  для выбранного  $x(t)$ .

Уравнение вида:  $\hat{y} = a_0 + a_1x_1$  – линейное уравнение регрессии.

Коэффициенты найдём по формулам:

$$a_1 = \frac{\sum (y - \bar{y})(x_1 - \bar{x}_1)}{\sum (x_1 - \bar{x}_1)^2} = \frac{397}{376,8} \approx 1,1$$

$$a_0 = \bar{y} - a_1 \cdot \bar{x}_1 = 39,0 - 1,1 \cdot 14,9 = 22,6$$

Линейная однопараметрическая модель регрессии для выбранного  $x_1$  примет вид:  $\hat{y} = 22,6 + 1,1x_1$

3. Оценить качество построенной модели, исследовав её на адекватность и точность.

Для этого в таблице 3 приведём промежуточные вычисления и результаты использования линейной модели.

$$\hat{y} = 22,6 + 1,1x_1 \quad \hat{y}_p(1) = 22,6 + 1,1 \cdot 5 = 28,1 \quad \hat{y}_p(2) = 22,6 + 1,1 \cdot 7 = 30,3 \dots$$

$$E(t) = y_t - \hat{y}_p(t) \quad E(1) = 30 - 28,1 = 1,9 \quad E(2) = 28 - 30,3 = -2,3 \text{ и т.д.}$$

Таблица 3 – Расчетная таблица

$t$	$y_t$	$x_1$	$\hat{y}_p$	$E(t)$	Поворотная точка	$E(t)^2$	$E(t)-E(t-1)$	$(E(t)-E(t-1))^2$	$( E(t) /y_t)100\%$
1	30	5	28,1	1,9	–	3,61	–	–	6,3
2	28	7	30,3	-2,3	1	5,29	-4,2	17,6	8,2
3	33	10	33,6	-0,6	0	0,36	1,7	2,9	1,8
4	38	12	35,8	2,2	1	4,84	2,8	7,6	5,8
5	40	15	39,1	0,9	0	0,81	-1,3	1,7	2,3
6	42	18	42,4	-0,4	0	0,16	-1,3	1,7	1,0
7	44	20	44,6	-0,6	1	0,36	-0,2	0,04	1,4
8	49	23	47,9	1,1	1	1,21	1,7	2,9	2,3
9	47	24	49,0	-2,0	–	4,0	-3,1	9,6	4,3
$\Sigma$	351	134		0,2	4	20,6		44,24	33,4

*Проверим модель на адекватность*

1. Проверка равенства математического ожидания уровней ряда остатков нулю осуществим с использованием t- критерия Стьюдента:

$$M(E) = \bar{E} = \frac{\sum E(t)}{n} = \frac{0,2}{9} = 0,02$$

Если  $|\bar{E}| = 0,02 < 0,5$ , то  $t_p < t_{\alpha}$ -критерий Стьюдента выполняется, то есть  $M(E) \approx 0$ .

2. Проверку случайности уровней ряда остатков проведём на основе критерия поворотных точек:

$$P > \left[ 2 \cdot \frac{(n-2)}{3} - 1,96 \sqrt{\frac{16n-29}{90}} \right] = \left[ \frac{2 \cdot (9-2)}{3} - 1,96 \sqrt{\frac{16 \cdot 9 - 29}{90}} \right] = [2,4] = 2$$

Так как  $P = 4 > 2$ , то свойство случайности выполняется.

3. Независимость (отсутствие автокорреляции) проверим с помощью d-критерия Дарбина-Уотсона:

$$d = \frac{\sum_{t=2}^n (E(t) - E(t-1))^2}{\sum_{t=1}^n E(t)^2} = \frac{44,24}{20,6} \approx 2,15$$

Так как  $d = 2,15$  не попадает в интервал от  $d_2$  до 2, то есть в  $[1,36;2]$ , то найдем  $d' = 4 - 2,15 = 1,85 \in [1,36;2]$ , то есть уровни ряда являются независимыми (автокорреляции нет).

4.Соответствие ряда остатков нормальному закону распределения определим при помощи RS–критерия:

$$RS = \frac{E_{\max} - E_{\min}}{S_E}, \text{ где } S_E = \sqrt{\frac{\sum E(t)^2}{n-1}}, S_E - \text{СКО.}$$

$$\text{Вычислим } S_E = \sqrt{\frac{20,6}{8}} = \sqrt{2,58} \approx 1,6, \text{ тогда } RS = \frac{2,2 - (-2,3)}{1,6} = 2,8$$

Так как  $2,8 \in [2,7;3,7]$ , то есть значение RS–критерия попадает между табулированными границами с заданным уровнем вероятности (для  $n \leq 10$ ,  $p = 70\%$ ), то гипотеза о нормальном распределении ряда остатков принимается, что позволяет строить доверительный интервал прогноза.

Так как все четыре условия выполняются, то модель адекватна.

#### *Проверим модель на точность*

Для характеристики точности воспользуемся средней относительной ошибкой:  $\overline{E_{\text{оми}}} = \frac{1}{n} \sum \frac{|E(t)|}{y_t} \cdot 100\%$

Для нашего примера:  $\overline{E_{\text{оми}}} = \frac{33,4}{9} = 3,7\% < 5\%$ , что свидетельствует о "хорошей" точности модели.

Оценку качества построенной модели дает коэффициент детерминации:  $R^2 = 1 - \frac{S_E^2}{S_y^2} = 1 - \frac{\sum E(t)^2}{\sum (y - \bar{y})^2}$ ,  $R^2 = 1 - \frac{20,6}{438} \approx 0,95$

Коэффициент  $R^2$  показывает, что более 95% вариации зависимой переменной ( $y$ ) учтено в модели и происходит под влиянием фактора, включённого в модель, а остальные 5% приходится на другие факторы, не включённые в модель.

Так как  $R = \sqrt{0,95} = 0,98$ , то есть  $R \approx 1$ , то модель качественная.

4. Для модели регрессии рассчитать коэффициент эластичности и бета-коэффициент.

Вычислим коэффициент эластичности модели:  $\mathcal{E}_{yx_1}^{\wedge} = a_1 \frac{\bar{x}_1}{\bar{y}}$ ,

В нашем примере  $a_1 = 1,1$ ,  $\bar{x}_1 = 14,9$ ,  $\bar{y} = 39$ , поэтому  $\mathcal{E}_{yx_1}^{\wedge} = \frac{1,1 \cdot 14,9}{39} \approx 0,42$ , он показывает, что при увеличении  $x_1$  на 1% результативный признак  $y$  в среднем увеличивается на 0,42%.

Вычислим  $\beta$ -коэффициент по формуле:

$$\beta_{yx_1}^{\wedge} = a_1 \sqrt{\frac{\sum (x_1 - \bar{x}_1)^2}{\sum (y - \bar{y})^2}} = 1,1 \sqrt{\frac{376,8}{438}} = 1,1 \sqrt{0,96} = 1,1 \cdot 0,93 \approx 1,02$$

5. Построить точечный и интервальный прогнозы на два шага вперёд по модели регрессии (для вероятности  $P = 70\%$  используйте коэффициент  $t_\alpha = 1,05$ ) (прогнозные оценки фактора  $x(t)$  на два шага вперёд получить на основе среднего прироста от фактически достигнутого уровня).

Вычислим САП и получим прогнозные оценки фактора на основе величины его среднего прироста:  $x(n+k) = x(n) + \text{САП} \cdot k$ , где  $\text{САП} = (x_n - x_1)/(n-1)$ :

$$\text{САП} = (24 - 5)/(9 - 1) = 19/8 \approx 2,4$$

$$k=1 \quad x(9+1) = x(10) = x(9) + 2,4 \cdot 1 = 24 + 2,4 = 26,4$$

$$k=2 \quad x(9+2) = x(11) = x(9)+2,4 \cdot 2 = 24 + 4,8 = 28,8$$

Для получения прогнозных (точечных) оценок зависимой переменной по модели:  $\hat{y}(n+k) = a_0 + a_1 x(n+k)$ , подставим в неё найденные прогнозные значения фактора:

$$\hat{y}(10) = 22,6 + 1,1 \cdot 26,4 = 51,6; \quad \hat{y}(11) = 22,6 + 1,1 \cdot 28,8 = 54,3$$

Доверительный интервал прогноза имеет следующие границы:

при  $k = 1$

$$\text{Верхняя граница прогноза: } y_p(9+1) + U(1) = 51,6 + 3,7 = 55,3$$

$$\text{Нижняя граница прогноза: } y_p(9+1) - U(1) = 51,6 - 3,7 = 47,9$$

Величина  $U(k)$  для найденной модели регрессии имеет вид:

$$k=1 \quad U(1) = 1,05 \cdot 2,9 \sqrt{1 + \frac{1}{9} + \frac{(26,4 - 14,9)^2}{376,8}} = 3,04 \cdot 1,2 \approx 3,7$$

Доверительный интервал прогноза имеет следующие границы:

при  $k = 2$

$$\text{Верхняя граница прогноза: } y_p(9+4) + U(2) = 54,3 + 3,9 = 58,2$$

$$\text{Нижняя граница прогноза: } y_p(9+2) - U(2) = 54,3 - 3,9 = 50,4$$

$$k=2 \quad U(2) = 1,05 \cdot 2,9 \sqrt{1 + \frac{1}{9} + \frac{(28,8 - 14,9)^2}{376,8}} = 3,04 \cdot 1,3 \approx 3,9$$

$$\text{где } \bar{s}_E = \sqrt{\frac{20,6}{9-1-1}} = \sqrt{7,34} \approx 2,9$$

Сведем полученные данные в таблицу прогноза:

$k$	$x(n+k)$	$\hat{y}(n+k)$	Нижняя граница	Верхняя граница
1	26,4	51,6	47,9	55,3
2	28,8	54,3	50,4	58,2

Отообразим на графиках фактические данные, модель и результаты прогнозирования.

Для построения модели достаточно двух точек:

при  $t = 1$  ( $x_1 = 5$ ),  $\hat{y}(1) = 22,6 + 1,1 \cdot 5 = 28,1$

при  $t = 9$  ( $x_1 = 24$ ),  $\hat{y}(9) = 22,6 + 1,1 \cdot 24 = 49,0$

С вероятностью 0,7 можно гарантировать, что при сохранении сложившихся закономерностей развития прогнозируемая величина  $y$  в следующем году примет значение из интервала (47,9;55,3), а ещё через год-(50,4;58,2).

*Задание 2. Найти параметры модели  $\hat{y} = a \cdot b^x$ , с выбранным ранее фактором.*

Нелинейную модель регрессии  $\hat{y} = a \cdot b^x$  приведём к линейной. Для этого прологарифмируем обе части равенства по основанию 10:  
 $lg y = lg a + x lg b$ . Обозначим:  $lg y = Y$ ,  $lg a = A$ ,  $lg b = B$

Запишем предыдущее равенство в виде:  $Y = A + B \cdot x$ , таким образом получили линейное уравнение регрессии, где  $B = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}$ ,

$$A = \bar{y} - B \cdot \bar{x}$$

Для вычисления коэффициентов  $A$  и  $B$ , составим таблицу, в которой приведем промежуточные вычисления:

Таблица – Расчетные значения

$t$	$y_t$	$x$	$y=lg y$	$x \cdot y$	$x^2$	$y - \bar{y}$	$(y - \bar{y})^2$
1	30	5	1,4771	7,3855	25	-0,1054	0,0111
2	28	7	1,4472	10,1304	49	-0,1353	0,0183
3	33	10	1,5185	15,185	100	-0,064	0,0041
4	38	12	1,5682	18,8184	144	-0,143	0,0002
5	40	15	1,6021	24,0315	225	0,0196	0,0004
6	42	18	1,6232	29,218	324	0,0407	0,0017
7	44	20	1,6435	32,87	400	0,061	0,0037
8	49	23	1,6902	38,875	529	0,1077	0,0116
9	47	24	1,6721	40,1304	576	0,0896	0,0080
Сумма	351	134	14,2421	216,64	2372		0,0591

$$\overline{x \cdot y} = \frac{\sum xy}{n} = \frac{216,64}{9} = 24,07 \quad \overline{x^2} = 2372:9 = 263,6 \quad \overline{x} = 134:9 = 14,9$$

$$\overline{y} = 14,2421:9 = 1,5825$$

$$\text{Вычислим } B = \frac{24,07 - 14,9 \cdot 1,5825}{263,66 - 14,9^2} = \frac{0,491}{41,65} = 0,012$$

$$A = 1,5825 - 0,012 \cdot 14,9 = 1,401$$

Так как  $A = \lg a = m$ , то  $a = 10^m$ ;  $B = \lg b = p$ , то  $b = 10^p$

Отсюда,  $\hat{y}(t) = 10^{1,401} \cdot (10^{0,012})^x$  - модель.

Проверим качество этой модели, с помощью коэффициента

детерминации:  $R^2 = 1 - \frac{\sum E(t)^2}{\sum (y - \overline{y})^2}$

Таблица – Промежуточные вычисления

$t$	$y_t$	$x$	$\hat{y}_p$	$E(t)$	$E(t)^2$	$ E(t) /y_t \cdot 100\%$
1	30	5	28,9	1,1	1,21	3,7
2	28	7	30,5	-2,5	6,25	8,9
3	33	10	33,2	-0,2	0,04	0,6
4	38	12	35,1	2,9	8,41	7,6
5	40	15	38,1	1,9	3,61	4,8
6	42	18	41,4	0,6	0,36	1,4
7	44	20	43,7	0,3	0,09	0,7
8	49	23	47,5	1,5	2,25	3,1
9	47	24	51,4	-4,4	19,36	9,4
$\Sigma$	351	134			41,58	40,2

$$\hat{y}(t) = 10^{1,401} \cdot (10^{0,012})^x; \quad E(1) = 30 - 28,9 = 1,1$$

$$\hat{y}_p(1) = 10^{1,401} \cdot (10^{0,012})^5 = 10^{1,401} \cdot 10^{0,06} = 10^{1,461} = 28,9; \quad E(2) = 28 - 30,5 = -2,5$$

$$\hat{y}_p(2) = 10^{1,401} \cdot (10^{0,012})^7 = 10^{1,485} = 30,5 \quad E(3) = 33 - 33,2 = -0,2$$

$$\hat{y}_p(3) = 10^{1,401} \cdot (10^{0,012})^{10} = 10^{1,521} = 33,2 \quad E(4) = 38 - 35,1 = 2,9$$

$$\hat{y}_p(4) = 10^{1,401} \cdot (10^{0,012})^{12} = 10^{1,545} = 35,1$$

$$E(5) = 40 - 38,1 = 1,9$$

$$\hat{y}_p(5) = 10^{1,401} \cdot (10^{0,012})^{15} = 10^{1,581} = 38,1$$

$$E(6) = 42 - 41,4 = 0,6$$

$$\hat{y}_p(6) = 10^{1,401} \cdot (10^{0,012})^{18} = 10^{1,617} = 41,4$$

$$E(7) = 44 - 43,7 = 0,3$$

$$\hat{y}_p(7) = 10^{1,401} \cdot (10^{0,012})^{20} = 10^{1,641} = 43,7$$

$$E(8) = 49 - 47,5 = 1,5$$

$$\hat{y}_p(8) = 10^{1,401} \cdot (10^{0,012})^{23} = 10^{1,677} = 47,5$$

$$E(9) = 47 - 51,4 = -4,4$$

$$\hat{y}_p(9) = 10^{1,401} \cdot (10^{0,012})^{24} = 10^{1,689} = 48,9$$

$$R^2 = 1 - \frac{36,78}{438} = 0,92$$

Так как  $R = \sqrt{0,92} = 0,96 \approx 1$ , то модель качественная.

Вычислим  $\bar{E}_{отн}$ :  $\overline{E_{отн}} = \frac{1}{n} \sum \frac{|E(t)|}{y_t} \cdot 100\%$ ,  $\overline{E_{отн}} = \frac{1}{9} \cdot 40,2 = 4,47\% < 5\%$ ,

что свидетельствует о "хорошей" точности модели; по ней можно строить прогноз.

Проверим значимость модели регрессии с использованием F-критерия Фишера:

$$F_p = \frac{R^2 / m}{(1 - R^2) / (n - m - 1)} = \frac{0,92 / 1}{(1 - 0,92) / (9 - 1 - 1)} = \frac{0,92}{0,08 : 7} = 80,5$$

Если  $F_p > F_{табл}$ , то уравнение значимое ( $F_{табл} = 5,59$ ). Так как  $80,5 > 5,59$ , то модель признаётся значимой.

Для построения точечного прогноза, сначала получим прогнозные оценки фактора на основе величины:  $x_{прогн} = 110\% \cdot x_n$  с помощью формулы:  $x(n+k) = 1,1 \cdot x(n)$

$$k = 1; \quad x(9+1) = x(10) = 1,1 \cdot x(9) = 1,1 \cdot 24 = 26,4$$

$$k = 2; \quad x(9+2) = x(11) = 1,1 \cdot x(10) = 1,1 \cdot 26,4 = 29,1$$

Для получения прогнозных оценок зависимой переменной по модели, подставим в неё найденные прогнозные значения фактора.

$$\hat{y}(n+k) = 10^{1,401} \cdot (10^{0,012})^{x(n+k)}$$

$$\hat{y}(10) = 10^{1,401} \cdot (10^{0,012})^{26,4} = 10^{1,401} \cdot 10^{0,3168} = 10^{1,7178} = 52,2$$

$$\hat{y}(11) = 10^{1,401} \cdot (10^{0,012})^{29,1} = 10^{1,401} \cdot 10^{0,3492} = 10^{1,7502} = 56,3$$

Доверительный интервал прогноза имеет следующие границы:

$$\text{Верхняя граница: } \hat{y}_p(n+k) + U(k)$$

$$\text{Нижняя граница: } \hat{y}_p(n+k) - U(k)$$

$$\text{Найдём } U(1) = 1,05 \cdot 2,43 \sqrt{1 + \frac{1}{9} + \frac{(26,4 - 14,9)^2}{376,8}} = 2,55 \cdot 1,21 = 3,1$$

$$(\overline{S_E}) = \sqrt{\frac{\sum E(t)^2}{n-m-1}} = \sqrt{\frac{41,58}{9-1-1}} \approx 2,43$$

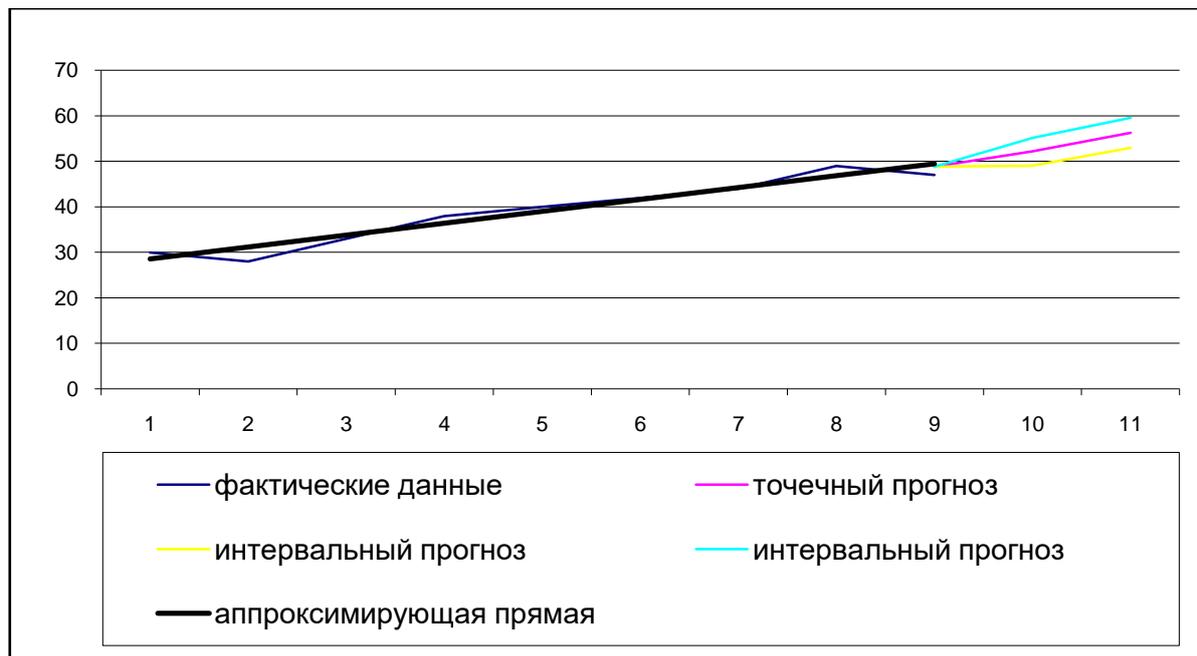
$$U(2) = 1,05 \cdot 2,43 \sqrt{1 + \frac{1}{9} + \frac{(29,1 - 14,9)^2}{376,8}} = 2,55 \cdot 1,28 \approx 3,3$$

Таблица прогноза

$k$	$x(n+k)$	$\hat{y}(n+k)$	Нижняя граница	Верхняя граница
1	26,4	52,2	59,1	55,2
2	29,1	56,3	53	59,6

С вероятностью 0,7 можно гарантировать, что в следующем году у примет значение из интервала (50,9;57,1), а ещё через год- (53,7;60,3).

Отообразим на графиках фактические данные, модель и результаты прогнозирования.



**Пример 2:** В таблице приведены статистические данные показателя  $y(t)$  и данные фактора  $x_1(t)$  :

$x(t)$	45,1	59,0	57,2	61,8	58,8	47,2	55,2
$y(t)$	68,8	61,2	59,9	56,7	55,0	54,3	49,3

**Требуется:**

*Задание №1.* Для характеристики зависимости  $Y$  от  $X$  рассчитать параметры.

*Задание №2.* Оценить тесноту связей с помощью коэффициента корреляции и коэффициента детерминации.

*Задание №3.* Дать с помощью среднего коэффициента эластичности сравнительную оценку силы связи фактора с результатом.

*Задание №4.* Оценить с помощью средней ошибки аппроксимации качество уравнения.

Задание №5. Оценить с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования.

### Решение

Таблица – Расчетная таблица

$t$	$X_t$	$Y_t$	$x_t = 1/X_t$	$Y_t^2$	$x_t^2$	$x_t \cdot Y_t$	$\hat{Y}_t$	$e_t = Y_t - \hat{Y}_t$	$e_t^2$	$Y_t - \bar{Y}$	$(Y_t - \bar{Y})^2$	$ Y_t - \hat{Y}_t /Y_t \cdot 100\%$
1	45,1	68,8	0,0222	4733,44	0,0005	1,5255	61,8205	6,9795	48,7137	10,91	119,12	10,1447
2	59,0	61,2	0,0169	3745,44	0,0003	1,0373	56,3111	4,8889	23,9014	3,31	10,9845	7,9884
3	57,2	59,9	0,0175	3588,01	0,0003	1,0472	56,8736	3,0264	9,1590	2,01	4,0573	5,0524
4	61,8	56,7	0,0162	3214,89	0,0003	0,9175	55,5012	1,1988	1,4371	-1,19	1,4059	2,1143
5	58,8	55,0	0,0170	3025,00	0,0003	0,9354	56,3719	1,3719	1,8821	-2,89	8,3273	2,4944
6	47,2	54,3	0,0212	2948,49	0,0004	1,1504	60,7800	6,4800	41,9909	-3,59	12,8573	11,9338
7	55,2	49,3	0,0181	2430,49	0,0003	0,8931	57,5417	8,2417	67,9252	-8,59	73,7145	16,7174
$\Sigma$	384,3	405,2	0,1291	23685,76	0,0024	7,5064	61,8205	0,0000	195,0094		230,4686	56,4453
Ср.зн.	54,90	57,89	0,0184	3383,68	0,0003	1,0723						8,0636

Задание №1. Для характеристики зависимости  $Y$  от  $X$  рассчитать параметры гиперболической модели парной регрессии.

$$Y_t = \alpha + \frac{\beta}{X_t} + \varepsilon_t \text{ - теоретическая модель;}$$

$$\hat{Y}_t = a_0 + \frac{a_1}{X_t} \text{ - выборочная модель.}$$

Введем замену  $x_t = \frac{1}{X_t}$  и получим следующую линейную модель  $\hat{Y}_t = a_0 + a_1 \cdot x_t$ .

$$a_0 = \bar{Y} - a_1 \cdot \bar{x}; \quad a_1 = \frac{\overline{xY} - \bar{x} \cdot \bar{Y}}{\overline{x^2} - (\bar{x})^2};$$

$$a_1 = 1054,6699 \quad a_0 = 38,4353$$

$$\hat{Y}_t = 38,4353 + \frac{1054,6699}{X_t}.$$

*Задание №2.* Оценить тесноту связей с помощью коэффициента корреляции и коэффициента детерминации.

$$\rho_{xy} = \sqrt{1 - \frac{\sum (Y_t - \hat{Y}_t)^2}{\sum (Y_t - \bar{Y})^2}}$$

Так как 0,3922, то связь между X и Y тесная.

*Задание №3.* Дать с помощью среднего коэффициента эластичности сравнительную оценку силы связи фактора с результатом.

$$\bar{\varepsilon} = f'(\bar{X}) \cdot \frac{\bar{X}}{\bar{Y}}; \quad f'(X) = \frac{1054,6699}{X^2};$$

$$\rho_{xy} = 0,3922 \quad f'(\bar{X}) = 0,3499 \quad \bar{\varepsilon} = 0,3319 \%$$

На 0,3319% в среднем увеличился результат Y от своей средней величины при изменении фактора X на 1% от своего среднего значения.

*Задание №4.* Оценить с помощью средней ошибки аппроксимации качество уравнения.

$$\bar{E} = \frac{1}{n} \sum \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100\%;$$

$$\bar{E} = 8,0636$$

Так как  $\bar{E} = 8,0636 > 5\%$ , то модель допустимо точная.

*Задание №5.* Оценить с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования.

$$H_0 : F = 0;$$

$$H_1 : F \neq 0.$$

$$F_{\text{набл}} = \frac{R^2(n-2)}{1-R^2} \square F_{v_1, v_2}, \text{ где } v_1 = 1, v_2 = n - 2 \text{ (для парной регрессии),}$$

$$R^2 = 1 - \frac{\sum (Y_t - \hat{Y}_t)^2}{\sum (Y_t - \bar{Y})^2} \quad R^2 = 0,1539$$

$$F_{\text{набл}} = 0,9092$$

Так как  $0,9092 < 6.61$ . то гипотеза принимается, уравнение регрессии незначимо.

**Пример 3:** Дана динамика развития показателя:  $Y_t$ : 33, 35, 40, 41, 45, 47, 45, 51, 53

**Требуется:**

*Задание №1.* Сгладить  $Y(t)$  простой скользящей средней.

*Задание №2.* Определить величину тренда  $Y(t)$ .

*Задание №3.* Построить линейную модель  $Y(t) = a_0 + a_1t$ , параметры которой оценить МНК.

*Задание №4.* Оценить адекватность построенной модели на основе исследования:

- случайности остаточной компоненты по критерию пиков;
- независимости уровней ряда остатков по  $d$ -критерию (в качестве критических используйте уровни  $d_1=1,08$  и  $d_2=1,36$ ) или по первому коэффициенту корреляции, критический уровень которого  $r(1)=0,36$ ;
- нормальности распределения остаточной компоненты по  $R/S$ -критерию с критическими уровнями  $2,7 - 3,7$ .
- для оценки точности модели используйте среднеквадратическое отклонение и среднюю по модулю ошибку.

*Задание №5.* Построить точечный и интервальный прогнозы на два шага вперед (по вероятности  $P=70\%$  используйте коэффициент  $t_\alpha = 1.05$ ) по модели  $y = 31.3 + 2.4t$ .

**Решение:**

*Задание №1.* Сгладить  $Y(t)$  простой скользящей средней.

Метод простой скользящей средней заключается в следующем: каждый уровень ряда начиная со второго при  $m=3$  заменяется новым средним значением:

$$\bar{y}_2 = \frac{y_1 + y_2 + y_3}{3}, \text{ и т.д.}$$

$$\bar{y}_3 = \frac{y_2 + y_3 + y_4}{3}$$

Сглаженный ряд: —, 36, 38,7, 42, 44,3, 45,7, 47,7, 49,7, —,

*Задание №2.* Определить величину тренда  $Y(t)$ .

Определим тренд методом Фостера-Стьюарта:

I этап: Ставим две последовательности:

$$u(t) = \begin{cases} 1, & \text{если } Y(t) \text{ больше всех предыдущих уровней} \\ 0, & \text{если } Y(t) \text{ меньше всех предыдущих уровней} \end{cases}$$

$$v(t) = \begin{cases} 1, & \text{если } Y(t) \text{ меньше всех предыдущих уровней} \\ 0, & \text{если } Y(t) \text{ больше всех предыдущих уровней} \end{cases}$$

$$u(t) \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 1$$

$$v(t) \quad 0 \quad 0$$

II этап:

$$s = \sum_{t=1}^n (u(t) + v(t)); \quad d = \sum_{t=1}^n (u(t) - v(t)).$$

$$s = (1+0) + (1+0) + \dots + (1+0) = 8;$$

$$d = (1-0) + (1-0) + \dots + (1-0) = 8.$$

III этап:

Вычислим два значения:

$$t_s = \frac{s - \mu}{\sigma_s}, \quad t_d = \frac{d}{\sigma_d},$$

где  $\mu$  - математическое ожидание,

$\sigma_s, \sigma_d$  - среднее квадратичное отклонение.

Значения  $\mu, \sigma_s, \sigma_d$  затабулированы для различной длины временных рядов.

При  $n=9$

$$t_s = \frac{8 - 3.858}{1.288} = 3.216, \quad t_d = \frac{8}{1.964} = 4.073.$$

IV этап:

Найденные значения  $t_s$  и  $t_d$  сравнивают с табличными значениями  $t_\alpha$  - Стьюдента с  $k = n - 2$  - степенями свободы,  $\alpha$  - уровень значимости.

$$k = n - 2 = 9 - 2 = 7.$$

$$|t_s| = 3.216 > t_{\alpha=0.05} = 1.90$$

$$|t_d| = 4.073 > t_{\alpha=0.05} = 1.90$$

Так как  $|t_s| > t_\alpha$  и  $|t_d| > t_\alpha$ , то тренд есть.

*Задание №3.* Построить линейную модель  $Y(t) = a_0 + a_1 t$ , параметры которой оценить МНК.

Для полинома первой степени  $\hat{y} = a_0 + a_1 t$  параметры вычислим по формулам:

$$a_1 = \frac{\sum (y_t - \bar{y})(t - \bar{t})}{\sum (t - \bar{t})^2}; \quad a_0 = \bar{y} - a_1 \bar{t},$$

$$\text{где } \bar{y} = \frac{\sum_{t=1}^n y_t}{n}, \quad \bar{t} = \frac{\sum_{t=1}^n t}{n} - \text{выборочные средние.}$$

Таблица – Промежуточные расчеты

$t$	$Y(t)$	$t-\bar{t}$	$(t-\bar{t})^2$	$y_t - \bar{y}$	$(y_t - \bar{y})(t-\bar{t})$	Расчет $Y_p(t)$	Отклонение $E(t)$
1	33	-4	16	-10,3	41,2	33,7	-0,7
2	35	-3	9	-8,3	24,9	36,1	-1,1
3	40	-2	4	-3,3	6,6	38,5	1,5
4	41	-1	1	-2,3	2,3	40,9	0,1
5	45	0	0	1,7	0	43,3	1,7
6	47	1	1	3,7	3,7	45,7	1,3
7	45	2	4	1,7	3,4	48,1	-3,1
8	51	3	9	7,7	23,1	50,5	0,5
9	53	4	16	9,7	38,8	52,9	0,1
$\Sigma 45$	390	0	60	0,3	144	389,7	0,3

$$a_1 = \frac{144}{60} = 2.4; \quad a_0 = 43.3 - 2.4 \cdot 5 = 31.3.$$

Таким образом, линейная модель примет вид:  $\hat{y} = 31.3 + 2.4t$ .

Последовательно подставляя в модель вместо  $t$  его значения от 1 до  $n$ , получаем расчетные значения уровней  $Y_p(t)$ :

$$t = 1 \quad Y_p(1) = 31.3 + 2.4 \cdot 1 = 33.7$$

$$t = 2 \quad Y_p(2) = 31.3 + 2.4 \cdot 2 = 36.1 \text{ и т.д.}$$

Вычислим отклонения расчетных значений от фактических наблюдений

$$E(t) = Y(t) - Y_p(t), t = \overline{1, n}.$$

*Задание №4.* Оценить адекватность построенной модели на основе исследования:

- случайности остаточной компоненты по критерию пиков;
- независимости уровней ряда остатков по  $d$ -критерию (в качестве критических используйте уровни  $d_1=1,08$  и  $d_2=1,36$ ) или по первому коэффициенту корреляции, критический уровень которого  $r(1)=0,36$ ;

- нормальности распределения остаточной компоненты по  $R/S$ -критерию с критическими уровнями 2,7 – 3,7.
- для оценки точности модели используйте среднеквадратическое отклонение и среднюю по модулю ошибку.

Результаты исследования адекватности модели отразим в таблице.

Таблица – Результаты исследования адекватности модели

$t$	$Y(t)$	$E(t)$	Точка поворота	$E(t)^2$	$E(t) - E(t-1)$	$(E(t) - E(t-1))^2$	$\frac{ E(t) }{y(t)} \cdot 100\%$
1	33	-0,7	-	0,49	-	-	2,1
2	35	-1,1	1	1,21	-0,4	0,16	3,1
3	40	1,5	1	2,25	2,6	6,76	3,8
4	41	0,1	1	0,01	-1,4	1,96	0,2
5	45	1,7	1	2,89	1,6	2,56	3,8
6	47	1,3	0	1,69	-0,4	0,16	2,8
7	45	-3,1	1	9,61	-4,4	19,36	6,9
8	51	0,5	1	0,25	3,6	12,96	1
9	53	0,1	-	0,01	-0,4	0,16	0,2
$\sum 45$	390	0,3	6	18,41	-	44,08	23,9

Проверка равенства математического ожидания уровней ряда остатков нулю:

вычислим  $\bar{E}$  - среднее значение уровней остаточного ряда:

$$\bar{E} = \frac{\sum_{t=1}^n E(t)}{n} = \frac{0,3}{9} = 0,03.$$

Так как  $|\bar{E}| = |0,03| < 0,5$ , следовательно, критерий Стьюдента выполняется автоматически,  $\Rightarrow M(E) \approx 0, P = 70\%$ .

Таким образом, гипотеза о равенстве математического ожидания значений остаточного ряда нулю выполняется.

Проверку случайности уровней ряда остатков проведем на основе критерия поворотных точек.

«1» обозначим поворотные точки, а «0» - неповоротные точки.

Число поворотных точек  $p$  должно удовлетворять строгому неравенству:

при  $n = 9$  имеем:

$$p > \left[ \frac{2(9-2)}{3} - 1.96 \sqrt{\frac{16 \cdot 9 - 29}{90}} \right] = [4.67 - 2.23] = [2.44] = 2.$$

Так как неравенство  $p = 6 > 2$  выполняется, следовательно, свойство случайности выполняется.

При проверке независимости (отсутствия автокорреляции) определяется отсутствие в ряду остатков систематической составляющей.

Вычислим коэффициент  $d$ -Дарбина-Уотсона:

$$d = \frac{\sum_{t=2}^n (E(t) - E(t-1))^2}{\sum_{t=1}^n E(t)^2}.$$

Полученное значение  $d$  сравним с двумя критическими табличными уровнями (нижним  $d_1$ , и верхним  $d_2$ ).

Для линейной модели при  $n \leq 10$  возьмем в качестве критических табличных уровней величины  $d_1 = 1,08$  и  $d_2 = 1,36$ .

$$d = 44.08 \div 18.41 = 2.4.$$

Расчетное значение  $d$  превышает 2, то это свидетельствует об отрицательности корреляции и поэтому его величину надо преобразовать:  $d = 4 - d = 4 - 2.39 = 1.61$ . Наше значение попадает в интервал от  $d_2$  до 2 (  $1.36 < 1.61 < 2$  ), то уровни ряда являются независимыми (автокорреляции нет).

Наличие нормального закона распределения в остаточном ряду линейной функции установим с помощью  $R/S$ -критерия:

$$PS = \frac{E_{\max} - E_{\min}}{S_E}.$$

$$E_{\max} = 1.7 \text{ и } E_{\min} = -3.1$$

$$S_E = \sqrt{\frac{\sum_{t=1}^n E(t)^2}{n-1}} = \sqrt{\frac{18.41}{8}} \approx 1.5.$$

$$RS = (2.3 - (-3.1)) \div 1.5 = 3.6$$

Так как  $RS = 3.6 \in [2.7; 3.7]$ , т.е. значение  $RS$ -критерия попадает между табулированными границами с заданным уровнем вероятности, то гипотеза о нормальном распределении ряда остатков принимается.

Таким образом все четыре условия выполняются, то остаточный ряд является случайной компонентой, а модель адекватна.

Для оценки точности нашей модели воспользуемся средней относительной ошибкой аппроксимации:

$$\bar{E}_{\text{отн}} = \frac{1}{n} \sum_{t=1}^n \frac{|E(t)|}{y(t)} \cdot 100\%,$$

$$\bar{E}_{\text{отн}} = \frac{1}{9} \cdot 23.9 = 2.6\%.$$

Если  $\bar{E}_{\text{отн}} = 2.6\% < 5\%$ , то модель точная.

*Задание №5.* Построить точечный и интервальный прогнозы на два шага вперед (по вероятности  $P = 70\%$  используйте коэффициент  $t_\alpha = 1.05$ ) по модели  $y = 31.3 + 2.4t$ .

Точечные прогнозы получим, подставляя, в уравнение модели

$$\begin{aligned} \hat{y}_{10} &= 31.3 + 2.4 \cdot 10 = 55.3 \\ \hat{y}_{11} &= 31.3 + 2.4 \cdot 11 = 57.7 \end{aligned}$$

значения  $t = 10$  и  $t = 11$ :

Доверительный интервал прогноза будет иметь следующие границы:

$$\text{верхняя граница прогноза: } \hat{y}(n+k) + U(k),$$

$$\text{нижняя граница прогноза: } \hat{y}(n+k) - U(k),$$

$$\text{где } U(k) = \bar{S}_E \cdot t_\alpha \sqrt{1 + \frac{1}{n} + \frac{(n+k-t_{cp})^2}{\sum_{t=1}^n (t-t_{cp})^2}},$$

$\bar{S}_E$  - среднее квадратическое отклонение от линии тренда:

$$\bar{S}_E = \sqrt{\frac{\sum_{t=1}^n E(t)^2}{n-2}} = \sqrt{\frac{18.41}{7}} = 1.6$$

Следовательно:

$$k = 1 \quad (t = 10)$$

$$U(1) = 1.6 \cdot 1.05 \cdot \sqrt{1 + \frac{1}{9} + \frac{(9+1-5)^2}{60}} = 1.68 \cdot \sqrt{1.1 + 0.42} = 1.68 \cdot \sqrt{1.52} = 1.68 \cdot 1.23 = 2.1.$$

$$k = 2 \quad (t = 11)$$

$$U(2) = 1.6 \cdot 1.05 \cdot \sqrt{1 + \frac{1}{9} + \frac{(9+2-5)^2}{60}} = 1.68 \cdot \sqrt{1.1 + 0.6} = 1.68 \cdot \sqrt{1.7} = 1.68 \cdot 1.30 = 2.2.$$

В таблицу 5 сведем результаты расчетов прогнозных оценок по линейной модели по формуле:

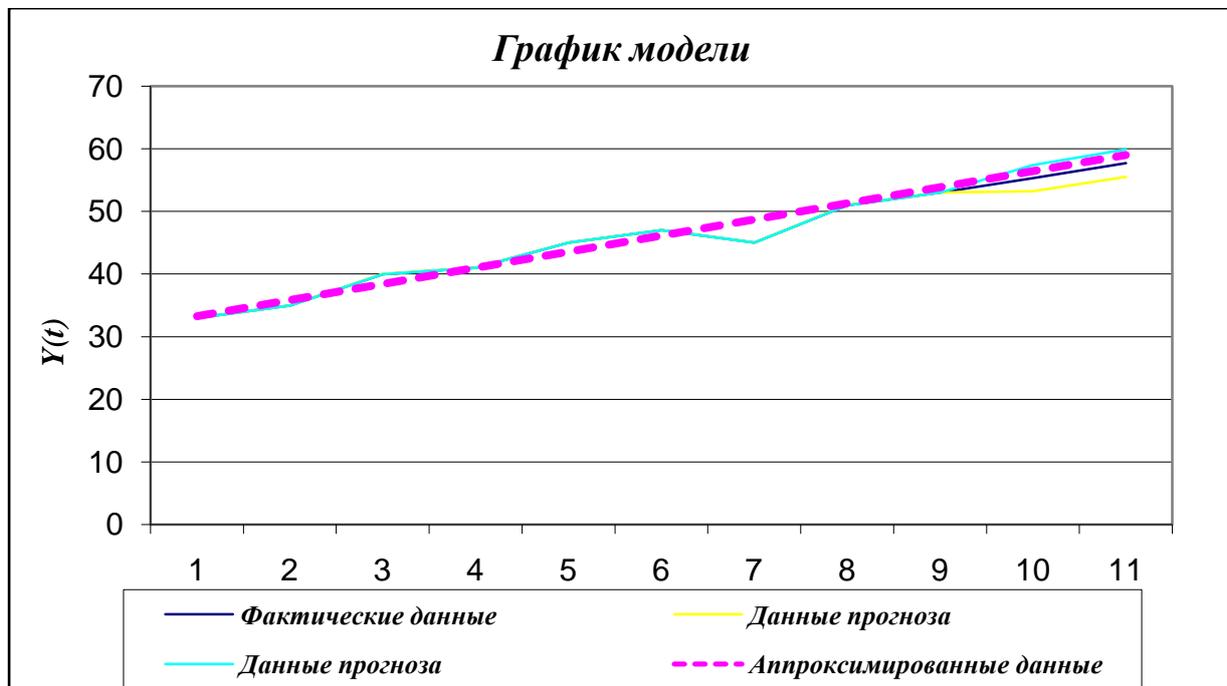
$$Y_p(9+1) \pm U(1) = Y_p(10) \pm U(1) = 55.3 \pm 2.1;$$

$$Y_p(9+2) \pm U(2) = Y_p(11) \pm U(2) = 57.7 \pm 2.2.$$

Таблица – Прогнозные оценки

Время $t$	Шаг $k$	Прогноз $Y_p(t)$	Нижняя граница	Верхняя граница
10	1	55,3	53,2	57,4
11	2	57,7	55,5	59,9

Отообразим на графике фактические данные, результаты аппроксимации и прогнозирования по модели  $\hat{y} = 31.3 + 2.4t$ .



**Пример 4:** Имеются выборочные данные по 10 однородным предприятиям, где  $x_t$  - электровооруженность труда на 1 рабочего (кВт.ч.),  $y_t$  – выпуск готовой продукции на 1 рабочего (т.) Построить однофакторную регрессионную модель по данным наблюдения:

*Задание 1.* Определить вид зависимости  $y(x)$ :

№	1	2	3	4	5	6	7	8	9	10
$x_t$	2	5	3	7	2	6	4	9	8	4
$Y_t$	3	6	4	6	4	8	6	9	9	5

(2;3) (5;6) (3;4) (7;6) (2;4) (6;8) (4;6) (9;9) (8;9) (4;5) – точки поля корреляции, построим их в системе координат X0Y.

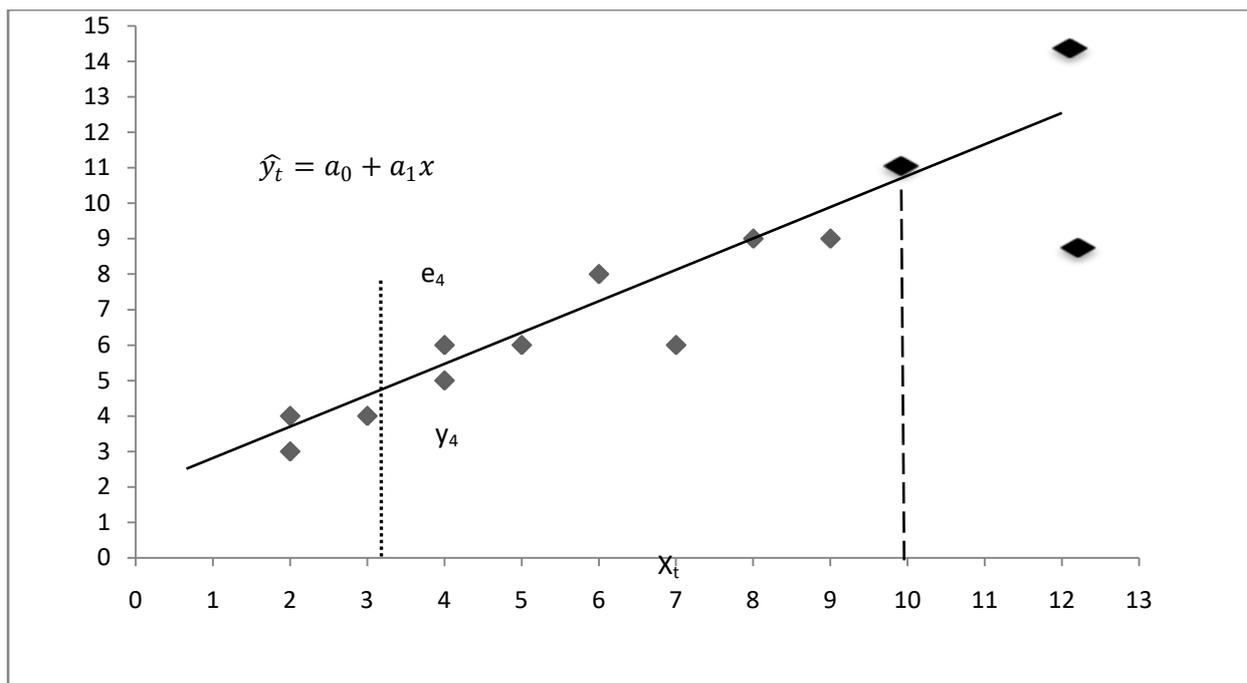


Рисунок – Поле корреляции

$$y_4 - \hat{y}_t = e_t$$

Вид зависимости линейный -  $\hat{y}_t = a_0 + a_1 x$

*Задание 2* Определить параметры  $a_0$  и  $a_1$ , используя метод наименьших квадратов

$$a_1 = \frac{n \sum x_t y_t - \sum x_t \sum y_t}{n \sum x_t^2 - (\sum x_t)^2} = \frac{10 \cdot 343 - 50 \cdot 60}{10 \cdot 304 - 2500} = \frac{3430 - 3000}{3040 - 2500} = \frac{430}{540} = 0,796$$

$$a_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{34,3 - 5 \cdot 6}{30,4 - 25} = \frac{4,3}{5,4} = 0,796$$

$$a_0 = \frac{1}{n} \sum y_t - \frac{1}{n} \sum x_t \cdot a_1 = \frac{1}{10} \cdot 60 - \frac{1}{10} \cdot 50 \cdot 0,796 = 6 - 5 \cdot 0,796 = 6 - 3,98 = 2,02$$

$$a_0 = \bar{y} - a_1 \bar{x} = 6 - 0,796 \cdot 5 = 6 - 3,98 = 2,02$$

$\hat{y}_t = 2.02 + 0.796x_t$  уравнение регрессии (парная модель/однофакторная регрессионная модель/выборочное уравнение парной регрессии)

Таблица - Расчетная таблица

t	$x_t$	$y_t$	$x_t y_t$	$x_t^2$	$y_t^2$	$\hat{y}_t$	$y_t - \hat{y}_t = e_t$	$e_t^2$	$y_t - \bar{y}$	$(y_t - \bar{y})^2$	$\left  \frac{e_t}{y_t} \right  \cdot 100\%$	$(x_t - \bar{x})^2$
1	2	3	6	4	9	3,61	-0,61	0,3721	-3	9	20,34	9
2	5	6	30	25	36	6,01	-0,01	0,0001	0	0	0,17	0
3	3	4	12	9	16	4,41	-0,41	0,1681	-2	4	10,25	4
4	7	6	42	49	36	7,59	-1,53	2,5281	0	0	26,50	4
5	2	4	8	4	16	3,61	0,39	0,1521	-2	4	9,75	9
6	6	8	48	36	64	6,80	1,20	1,4400	2	4	15,00	1
7	4	6	24	16	36	5,20	0,80	0,6400	0	0	13,33	1
8	9	9	81	81	81	9,18	-0,18	0,0324	3	9	2,00	16
9	8	9	72	64	81	8,39	0,61	0,3721	3	9	6,78	9
10	4	5	20	16	25	5,20	-0,20	0,0400	1	1	4,00	1
$\Sigma$	50	60	343	304	400			5,7540		40	108,12	54
Сред	5	6	34,3	30,4	40							

Задание 3. Определить средний коэффициент эластичности

$$\bar{\varepsilon} = a_1 \cdot \frac{\bar{x}}{\bar{y}} = 0,796 \cdot \frac{5}{6} = 0,66\%$$

Вывод: выпуск готовой продукции увеличится на 0,66% при изменении электровооруженности труда на 1%

Задание 4. Оценить тесноту (силу корреляционной связи) между выпуском продукции и электровооруженностью труда с помощью: линейного коэффициента корреляции; теоретического корреляционного отношения; коэффициента детерминации.

Линейный коэффициент корреляции:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{y_t^2 - \bar{y}^2} \cdot \sqrt{x_t^2 - \bar{x}^2}} = \frac{34,3 - 5 \cdot 6}{\sqrt{40 - 36} \cdot \sqrt{30,4 - 25}} = \frac{4,3}{4,647} = 0,926$$

т.к.  $r_{xy} \approx 1$ , то связь тесная (сильная)

т.к.  $r_{xy} = 0,926 > 0$ , то связь прямая

Теоретическое корреляционное отношение:

$$\eta = \sqrt{1 - \frac{\sum(y_t - \hat{y}_t)^2}{\sum(y_t - \bar{y})^2}} = \sqrt{1 - \frac{5,754}{40}} = \sqrt{0,856} = 0,926$$

Найдем  $\hat{y}_t = 2,02 + 0,796 \cdot x_t$

$$t = 1 \quad \hat{y}_t = 2,02 + 0,756 \cdot 2 = 3,61 \qquad t = 2 \quad \hat{y}_t = 2,02 + 0,756 \cdot 5 = 6,01$$

Найдем  $e_t = y_t - \hat{y}_t$

$$t = 1 \quad e_1 = y_1 - \hat{y}_1 = 3 - 3,61 = -0,61 \qquad t = 2 \quad e_2 = 6 - 6,01 = -0,01$$

Вывод: так как  $\eta = 0,926 \approx 1$ , то связь тесная.

Коэффициент детерминации:

$$R^2 = 1 - \frac{\frac{1}{n} \sum (y_t - \hat{y}_t)^2}{\frac{1}{n} \sum (y_t - \bar{y})^2} = 1 - \frac{5,754}{40} = 0,856, \text{ т.к. } R^2 = 0,856 \approx 1, \text{ то качество}$$

подгонки хорошее; модель точная.

*Задание 5.* Оценить качество построенной модели с помощью средней ошибки аппроксимации:

$$\bar{E} = \frac{1}{n} \cdot \sum \left| \frac{e_t}{y_t} \right| \cdot 100\% = \frac{1}{n} \cdot \sum \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100\% = \frac{1}{10} \cdot 107,91 = 10,8, \text{ т.к. } \bar{E} =$$

$10,8 > 10$ , то модель допустимо точная; в среднем расчетные значения для линейной модели отличаются от фактических значений на 10,8%

### 3 ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Номер вашего варианта соответствует двум последним цифрам зачетной книжки. В соответствии с ним из таблицы выберите показатель  $y$ , а данные факторов  $x_1(t)$  и  $x_2(t)$  возьмите из следующих по порядку строк. Если число, которое образуют последние две цифры номера Вашей зачетной книжки, превышает 50, то предварительно вычтите из него эту величину, а затем выберите соответствующие данные указанным способом.

#### Задача 1:

**Требуется:** (с использованием пакета EViews)

1. Рассчитать выборочный коэффициент корреляции  $Y(t)$  с  $X(t)$  и оценить их связь.

2. Построить выборочную модель линейной парной регрессии:

$$\hat{Y}(t) = a + bX(t).$$

3. Рассчитать коэффициент эластичности, бета коэффициент и пояснить их экономический смысл.

4. Найти стандартную ошибку оценки коэффициентов регрессии.

5. Оценить значимость коэффициентов регрессии через t- критерий Стьюдента.

6. Оценить качество построенной модели с помощью средней ошибки аппроксимации и коэффициента детерминации.

7. Проверить значимость уравнения регрессии на 5% - ном уровне по F – критерию Фишера.

8. Рассчитайте прогнозное значение зависимой переменной, если прогнозное значение фактора увеличится на 10% от его среднего уровня.

9. Найти 95% - ный доверительный интервал для индивидуальных значений зависимой переменной.

10. Найти с надежностью 0,95 интервальные оценки коэффициента регрессии  $a_1$  и дисперсии  $\sigma^2$ .

Отобразить на графике фактические данные, модель и результаты прогнозирования.

### Задача 2:

По предприятиям легкой промышленности региона получена информация, характеризующая зависимость объема выпуска продукции (Y, млн. руб.) от объема капиталовложений (X, млн. руб.).

**Требуется:** (с использованием пакета EViews)

1. Для характеристики зависимости Y от X рассчитать параметры уравнений: а) гиперболической, б) степенной, в) показательной парной регрессии.

2. Оцените тесноту связи с помощью показателей корреляции и детерминации.

3. Дайте с помощью среднего коэффициента эластичности сравнительную оценку силы связи фактора с результатом.

4. Оцените с помощью средней ошибки аппроксимации качество уравнений.

5. Оцените с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования.

6. По значениям характеристик, рассчитанных в пп. 3,4,5, выберите лучшее уравнение регрессии и дайте его обоснование (для этого составить сводную таблицу, включая линейную модель).

7. Рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 10% от его среднего уровня.

8. Определите доверительный интервал прогноза для уровня значимости  $\alpha = 0,05$ .

Таблица

Номер показателя	Номер наблюдения								
	1	2	3	4	5	6	7	8	9
1	10	14	21	24	33	41	44	47	49
2	43	47	50	48	54	57	61	59	65
3	3	7	10	11	15	17	21	25	23
4	30	28	33	37	40	42	44	49	47
5	5	7	10	12	15	18	20	23	26
6	12	15	16	19	17	20	24	25	28
7	20	27	30	41	45	51	51	55	61
8	8	13	15	19	25	27	33	35	40
9	45	43	40	36	38	34	31	28	25
10	33	35	40	41	45	47	45	51	53
11	10	15	21	23	25	34	32	37	41
12	16	20	22	20	25	23	25	28	30
13	12	17	20	21	25	27	24	28	31
14	20	22	24	26	25	29	35	38	43
15	25	30	36	41	38	43	47	45	50
16	80	75	78	72	69	70	64	61	59
17	24	22	26	29	33	31	28	33	36
18	30	34	40	38	42	48	50	52	53
19	88	85	84	86	81	80	83	78	76
20	50	52	54	59	57	60	63	68	70
21	25	27	30	31	35	41	42	45	47
22	75	77	73	70	66	63	67	63	61
23	28	34	32	36	39	42	45	41	46
24	15	20	24	30	33	37	36	40	42
25	70	74	76	75	78	78	83	85	87
26	82	79	78	72	69	70	64	61	59
27	25	27	26	29	32	32	30	33	35
28	32	34	38	40	42	46	50	52	53
29	90	87	85	86	82	80	81	78	76
30	55	57	54	59	57	60	63	66	64
31	12	15	18	22	25	31	32	37	41
32	26	30	32	30	35	33	35	38	40
33	62	67	80	81	85	87	84	88	91
34	18	21	24	26	25	29	34	38	41
35	28	32	36	40	38	43	45	48	50

## 4 ГЛОССАРИЙ

*Корреляционно-регрессионный анализ* - раздел статистической науки, занимающийся исследованием причинных связей между социально-экономическими явлениями и процессами, имеющими количественное выражение.

*Задачи корреляционного анализа:* выявление статистической связи между случайными переменными; установление ее направления; количественная оценка ее тесноты (силы); отбор факторов, оказывающих наиболее существенное влияние на результативный признак, на основании измерения тесноты связи между явлениями.

*Задачи регрессионного анализа:* установление формы зависимости между переменными (линейная или нелинейная, положительная или отрицательная и т.д.); оценка функции регрессии (в виде математического уравнения того или иного типа); оценка качества построенной модели; оценка неизвестных значений (прогноз значений) зависимой переменной.

Регрессионные модели с одним уравнением – модели, в которых зависимая (объясняемая) переменная  $y$  представляется в виде функции  $f(x, \beta) = f(x_1, \dots, x_k, \beta_1, \dots, \beta_p)$ , где  $x_1, \dots, x_k$  – независимые (объясняющие) переменные, а  $\beta_1, \dots, \beta_p$  – параметры.

*Модель множественной регрессии* – уравнение вида:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i, \text{ или } Y_i = \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i, \text{ где } i=1, \dots, n$$

*Метод наименьших квадратов* состоит в минимизации суммы квадратов остатков, т.е.  $\sum \varepsilon_i^2 \rightarrow \min$ .

Оценка  $\hat{\beta}$  параметра  $\beta$  называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру, т.е.  $M(\hat{\beta}) = \beta$ ; в противном случае оценка называется *смещенной*. Требование несмещенности гарантирует отсутствие систематических ошибок при оценивании.

Оценка параметра называется *состоятельной*, если ее предел по вероятности при увеличении числа наблюдений стремится к истинному значению параметра, т.е.  $\lim \hat{\beta} = \beta$ , при  $n \rightarrow \infty$ .

$\sum(Y_i - \bar{Y})^2 - TSS$  – вся дисперсия;

$\sum(Y_i - \hat{Y}_i)^2 - ESS$  – необъясненная часть всей дисперсии;

$\sum(\hat{Y}_i - \bar{Y})^2 - RSS$  – объясненная часть всей дисперсии.

Оценка дисперсии ошибок  $\sigma^2$ . Распределение  $S^2$ .

$$\sigma^2 = S^2 = \frac{e'e}{n-k} = \frac{\sum e_i^2}{n-k}$$

Коэффициент детерминации  $R^2$  (доля объясненной дисперсии):

$$R^2 = 1 - (ESS \div TSS) = RSS \div TSS$$

Статистика Стьюдента

$$t_c = \frac{\hat{\beta}_{ols,i} - \beta_1}{S\hat{\beta}_1} \approx t(n-k)$$

Статистика  $F$ :

$$F = \frac{R^2}{1-R^2} \times \frac{n-k}{k-1} = \frac{RSS}{ESS} \times \frac{n-k}{k-1} = \frac{Y_*' Y_* / (k-1)}{e'e / (n-k)} \approx F(k-1, n-k)$$

*Мультиколлинеарность* – линейная взаимосвязь двух или нескольких объясняющих переменных.

Коэффициент корреляции между двумя переменными, очищенный от влияния других переменных, называется *частным коэффициентом корреляции*.

Выборочное линейное уравнение парной регрессии:  $\hat{y} = a + bx$ .

По МНК определяются неизвестные коэффициенты  $a$  и  $b$ :

$$a = \bar{y} - b\bar{x} \quad \text{или} \quad a = \frac{1}{n} \left( \sum y_t - b \cdot \sum x_t \right)$$

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} \quad \text{или} \quad b = \frac{n \sum x_t \cdot y_t - \sum x_t \cdot \sum y_t}{n \sum x_t^2 - (\sum x_t)^2}$$

Коэффициент парной корреляции:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}}$$

Если  $r_{xy} = \pm 1$ , то корреляционная связь представляет функциональную зависимость.

Если  $r_{xy} = 0$ , то линейная корреляционная связь отсутствует

Коэффициент эластичности (средний):  $\bar{\varepsilon} = f'(x) \cdot \frac{\bar{x}}{\bar{y}}$ .

Коэффициент  $\beta$  вычисляется по формуле:  $\beta = b \cdot \sqrt{\frac{\sum (x_t - \bar{x})^2}{\sum (y_t - \bar{y})^2}}$ .

$\beta$  - коэффициент показывает: на какую часть величины своего среднего квадратичного отклонения изменится в среднем значение результативного признака  $y$  при изменении факторного признака  $x$  на величину его среднего квадратичного отклонения при фиксированном на постоянном уровне, значения остальных независимых переменных.

Стандартная ошибка оценки коэффициентов регрессии:

$$S_b = \sqrt{\frac{\sum e_t^2}{n(n-2) \cdot (\overline{x^2} - (\bar{x})^2)}} \quad S_a = \sqrt{\overline{x^2} \cdot S_b^2}$$

*Оценка качества построенной модели с помощью средней ошибки аппроксимации и коэффициента детерминации:*

$$\bar{E} = \frac{1}{n} \sum \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100\% . \quad \text{Если } \bar{E} < 5\% - \text{ модель точная, } \bar{E} < 10\% - \text{ модель}$$

допустимо точная,  $\bar{E} > 10\%$  - модель не точная.

$$\text{Коэффициент детерминации: } R^2 = 1 - \frac{\sum (y_t - \hat{y}_t)^2}{\sum (y_t - \bar{y})^2} \quad \text{Если } r_{xy}^2 = R^2 -$$

высокое качество построенного уравнения.

*Фиктивная (искусственная) переменная (D)* – переменная, которая отражает 2 противоположных состояния качественного фактора.

*ANOVA* – модель (модели дисперсионного анализа) - регрессионная модель, содержащая лишь качественные объясняющие переменные.

*ANCOVA* – модель (модели ковариационного анализа) - модель, в которой объясняющие переменные носят как количественный, так и качественный характер.

*Статистика Дарбина-Уотсона (DW):*

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

*Прогнозирование* – построение оценки зависимой переменной.

*Точечное прогнозирование:* оценка – конкретное число.

*Интервальное прогнозирование:* оценка – интервал, в котором истинное значение переменной находится с заданным уровнем доверия.

*Безусловное* – прогнозирование, при котором объясняющие переменные известны точно.

*Условное* – прогнозирование, при котором объясняющие переменные известны приближенно.

*Функция Кобба – Дугласа*:  $Y = AK^\alpha L^\beta$ , где  $Y$  – объем выпуска;  $K, L$  – затраты капитала, труда;  $A, \alpha, \beta$  – параметры модели.

*Функция Энгеля*:  $Y = AX^\beta$ , где  $A, \beta$  – параметры модели.

*Двойная логарифмическая модель (модель постоянной эластичности)*:  $\ln Y = \beta_0 + \beta \ln X + \varepsilon$ .

*Полулогарифмические модели* – это модели вида:

$$\ln Y = \beta_0 + \beta X + \varepsilon, \quad Y = \beta_0 + \beta \ln X + \varepsilon.$$

*Лог – линейная модель*:  $\ln Y_t = \beta_0 + \beta t + \varepsilon_t$ .

*Линейно – логарифмическая модель*:  $Y = \beta_0 + \beta \ln X + \varepsilon$ .

*Обратная модель*:  $Y = \beta_0 + \beta_1(1/X) + \varepsilon$ .

*Степенная модель*:  $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_n X^n + \varepsilon$ .

*Показательная модель*:  $Y = \beta_0 e^{\beta x}$ .

*Эндогенные переменные* – переменные, значения которых определяются внутри модели.

*Экзогенные* – внешние по отношению к модели переменные. Их значения определяются вне модели и поэтому их считают фиксированными.

*Временной ряд (динамический ряд, ряд динамики)* – последовательность наблюдений некоторого признака (случайной величины)  $Y$  в последовательные моменты времени. Отдельные наблюдения называются *уровнями* ряда, которые обозначают  $y_t$ .

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Алферова Л.А. Экономическая теория. Часть 1. Микроэкономика: учебное пособие/ Алферова Л.А. – Электрон.текстовые данные. – Томск: Эль Контент, Томский государственный университет систем управления и радиоэлектроники, 2017. – 250 с.
2. Воробьева И.П. Экономика: учебник / И.П. Воробьева, А.С. Громова, М.В. Рыжкова. – Электрон. текстовые данные. – Томск: Томский политехнический университет, 2017. – 198 с.
3. Вуколов, Э.А. Основы статистического анализа : практикум по статистическим методам и исследованию операций с использованием пакетов Statistica и Excel / Э.А. Вуколов. - М. : ИНФРА - М, 2004. - 464 с.
4. Гальперин В.М. Микроэкономика: учебник для студентов ВУЗов. В 3 – х т. / В.М. Гальперин, С.М. Игнатьев, В.И. Моргунов. – СПб.: «Экон.шк. ГУ – ВШЭ: Экономикус»: Омега – Л, 2008.
5. Дегтярева Н.А. Конкурентоспособность регионов и ее оценка / Н.А. Дегтярева, К.О. Латышева // Конкурентоспособность и развитие социально – экономических систем, Третьей Всероссийской научной конференции памяти А.И. Татаркина. Под общей редакцией В.И. Бархатова, Д.С.Бенц, Челябинск, 2019, С.44 – 45.
6. Дегтярева Н.А. Анализ социально – экономического развития регионов РФ на основе исследования уровня жизни населения / Н.А. Дегтярева, Д.С. Гордеева, Д.Н. Корнеев, И.И. Плужникова // Азимут научных исследований. Серия: Экономика и управление. – 2018. Т.7, №2 (23). – С.142 – 146.

7. Дегтярева Н.А. Информатизация общеобразовательных школ // Функциональная и прикладная наука: сборник научных статей по итогам научно – исследовательской работы за 2014 г. / под науч. ред. М.В. Потаповой. – Челябинск: Изд – во Челяб.гос.пед.ун – т. – 2015. – С. 91 – 94.

8. Дегтярева Н.А. Исследование зависимости количества безработных от социально – экономических факторов на основе модели множественной регрессии // Фундаментальная и прикладная наука. – Челябинск: Из – во Челяб.гос.пед.ун – та. – 2016. – № 2. – С. 13 – 17.

9. Дегтярева Н.А. Исследование индекса развития человеческого потенциала социально – экономической системы / Н.А. Дегтярева, К.О. Латышева // Ключевые элементы развития человеческого потенциала, экономики и обеспечения экономической безопасности: сборник статей участников Международной научно – практической конференции V Уральского вернисажа науки и бизнеса / под общ. ред. Е.П. Велихова. – Челябинск: Изд – во Челяб. гос. ун – та, 16 марта 2018. – С. 104 – 108.

10. Дегтярева Н.А. Исследование экономических процессов с применением сетевых моделей: монография / Н.А. Дегтярева. – Челябинск: Изд – во ЗАО «Библиотека А.Миллера», 2019. – 160 с.

11. Дегтярева Н.А. Модели анализа и прогнозирования на основе временных рядов: монография / Н.А. Дегтярева. – Челябинск: Изд – во ЗАО «Библиотека А.Миллера», 2018. – 160 с.

12. Дегтярева Н.А. Повышение экономической эффективности функционирования крестьянских (фермерских) хозяйств в условиях рынка: диссертация на соискание ученой степени кандидата

экономических наук / Дегтярева Н.А.; Челябинская государственная агроинженерная академия. – Челябинск, 2000. – 218 с.

**13.** Дегтярева Н.А. Практикум по экономико – математическим методам и моделям: учебное пособие для студентов / Н.А. Дегтярева. – Челябинск: Изд – во ЗАО «Библиотека А.Миллера», 2019. – 97 с.

**14.** Дегтярева Н.А. Применение статистических методов исследования в сельском хозяйстве / Н.А. Дегтярева, Н.А.Берг // Известия высших учебных заведений. Уральский регион. – 2017. – № 1. – С.42 – 47.

**15.** Дегтярева Н.А. Применение экономико – математических моделей для принятия управленческих решений / Н.А. Дегтярева // В сборнике: Конкурентоспособность и развитие социально – экономических систем Материалы Всероссийской научной конференции памяти академика А. И. Татаркина. Под ред. В. И. Бархатова, Д. А. Плетнёва. 2017. С. 62 – 64.

**16.** Дегтярева Н.А. Принятие эффективных управленческих решений на основе эконометрического прогнозирования / Н.А. Дегтярева, Н.А. Берг // Вестник Челябинского государственного университета. Серия: «Экономические науки». – № 4 (414) 2018. вып. 61. – С. 176 – 183.

**17.** Дегтярева Н.А. Сборник задач по статистике: учебное пособие для студентов / Н.А. Дегтярева. – Челябинск: Цицеро, 2017. – 90 с.

**18.** Дегтярева Н.А. Сборник задач по экономико – математическим методам и моделям: учебное пособие для студентов / Н.А. Дегтярева. – Челябинск: Цицеро, 2017. – 77 с.

**19.** Дегтярева Н.А. Эконометрическое моделирование производственной эффективности / Н.А. Дегтярева, Д.С. Гордеева, Матвеева П.А., Матвеев В.В., Борисенко Я.М. Федосеев // Азимут научных исследований. Серия: «Экономика и управление». – 2019. – Т. 8, № 2 (27). – С.133 – 135.

**20.** Дегтярева Н.А. Экономико – математические методы и модели. Конспекты лекций: учебное пособие для студентов / Н.А. Дегтярева. – Челябинск: Изд – во ЗАО «Библиотека А.Миллера», 2019. – 82 с.

**21.** Дегтярева Н.А. Экономические аспекты развития системы образования Российской Федерации: учебное пособие / Н.А. Дегтярева, Д.С. Гордеева – Челябинск: Цицеро, 2018. – 145 с.

**22.** Дегтярева Н.А. Эконометрические модели анализа и прогнозирования: монография / Н.А. Дегтярева. – Челябинск: Цицеро, 2017. – 170 с.

**23.** Дегтярева, Н.А. Исследование зависимости количества безработных от социально-экономических факторов на основе модели множественной регрессии // Фундаментальная и прикладная наука. – Челябинск: Из-во Челяб.гос.пед.ун-та. - 2016. - № 2. - С. 13-17.

**24.** Дегтярева, Н.А. Модели анализа и прогнозирования на основе временных рядов: монография / Н.А. Дегтярева. - Челябинск: Изд-во ЗАО «Библиотека А.Миллера», 2018.- 160 с. ISBN 978-5-93162-080-0.

**25.** Дегтярева, Н.А. Модели принятия управленческих решений в сфере информатизации муниципальных образовательных учреждений / Н.А. Дегтярева, Е.М. Колмакова, И.Д. Колмакова, //

Вестник Челябинского государственного университета. Серия: «Экономические науки». - № 5 (401) 2017. - вып. 57. - С. 15 - 23.

**26.** Дубина, И.Н. Математико-статистические методы в эмпирических социально-экономических исследованиях / И.Н. Дубина. - М.: Финансы и статистика, Инфра, 2010. - 416 с.

Дубров, А.М. Последовательный анализ в статистической обработке информации / А.М. Дубров. - М.: Статистика, 2004. - 160 с.

**27.** Колмакова И.Д. К вопросу о применении эконометрических методов исследования в социальных процессах / И.Д. Колмакова, Н.А. Дегтярева, Е.М. Колмакова // Вестник Челябинского государственного университета. Серия: «Экономика». – 2016. вып. 55. – С. 51 – 57.

**28.** Орлов А.И. Эконометрика : учебник для вузов / А.И. Орлов. - 3-е изд., перераб. и доп.- М. : Издательство «Экзамен», 2004. - 576 с.

**29.** Палий И.А. Прикладная статистика: Учебное пособие. – М.: Издательско–торговая корпорация "Дашков и К", 2008. – 224 с.

**30.** Практикум по эконометрике: учебное пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др.; под ред. И.И. Елисеевой. - 2-е изд., перераб. и доп. - М. : Финансы и статистика, 2007. - 344 с.

**31.** Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. 2–у изд., испр. – Т. 2: Айвазян С.А. Основы эконометрики. – М.: ЮНИТИ–ДАНА, 2001. – 432 с.

**32.** Эконометрическое прогнозирование социально-экономического развития региона / Н.А. Дегтярева, К.О. Латышева // В книге: Конкурентоспособность и развитие социально-

экономических систем. Четвертой международной научной конференции памяти А.И. Татаркина. Под общей редакцией В.И. Бархатова, Д.С.Бенц, Челябинск, 2020, С.93 – 94.

**33.** Экономико-математические методы и прикладные модели : учебное пособие / В.В. Федосеев и др. - М. : ЮНИТИ - ДАНА, 2012. - 304 с.

**34.** Экономическая теория: учебник / Г.П. Журавлева [и др.].– Электронные текстовые данные.– Москва: Дашков и К, 2019.– 934 с.

*Учебное издание*

**НИНА АДАМОВНА ДЕГТЯРЕВА  
ДАРЬЯ СЕРГЕЕВНА ГОРДЕЕВА**

**МЕТОДЫ СТАТИСТИКИ**  
УЧЕБНОЕ ПОСОБИЕ

Изд – во ЗАО «Библиотека А.Миллера»,  
454080, г. Челябинск, Свердловский пр., 60

Подписано к печати 21.09.2021  
Формат 60x84 1/16 Объем 3,9 уч-изд.л.  
Заказ № 537 Тираж 100 экз  
Отпечатано на ризографе в типографии ФГБОУ ВО ЮУрГГПУ  
454080, г. Челябинск, пр. Ленина, 69