Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Санкт-Петербургский государственный университет
промышленных технологий и дизайна»

Кафедра экономики и финансов

ЭКОНОМЕТРИКА

Задания по выполнению контрольных работ для студентов направлений подготовки 38.03.01 Экономика и 38.03.02 Менеджмент заочной формы обучения.

Составитель

А. И. Богданов

Санкт-Петербург
2019

Общие требования к выполнению контрольной работы

Контрольная работа по курсу «Эконометрика» выполняется для закрепления знаний и навыков применения математических методов в экономике. При самостоятельном изучении курса (для заочной формы обучения) следует руководствоваться указанными литературными источниками.

Задания к контрольной работе составлены в 10 вариантах. Каждый студент выполняет один вариант, номер которого соответствует последней цифре номера зачетной книжки. Если номер зачетной книжки заканчивается цифрой «0», то выполняется вариант № 10.

Расчеты должны быть представлены в развернутом виде со всеми формулами, пояснениями и выводами, соблюдая достаточную точность вычислений. Работа должна быть оформлена в соответствии с требованиями ГОСТ.

1. ПОНЯТИЕ О ПРОГНОЗИРОВАНИИ И МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ

1.1. ПОНЯТИЕ О ПРОГНОЗИРОВАНИИ И ПРОГНОСТИКЕ

В литературе имеется большое количество терминов и определений, связанных с проблемой прогнозирования, которые, к сожалению, трактуются далеко не однозначно. К ним относятся такие термины как предсказание (prediction), прогнозирование (forecasting),) и некоторое обобщающее понятие предвидение (prognousis).

Предсказание (prediction) представляет собой не вероятностное (практически достоверное) утверждение о будущем. Оно возможно на основании физических законов. Например, можно практически без ошибок рассчитать положение планет в обозримом будущем.

Прогнозирование (forecasting) представляет собой вероятностное утверждение о будущем.

При этом предсказание и прогнозирование являются частным случаем предвидения, которое в принципе представляет собой некоторое рассуждение о будущем.

Результатом предвидения является прогноз будущего. Прогнозы по своему содержанию могут быть качественными и количественными.

Качественные прогнозы могут быть получены как путем логических рассуждений, так и на основе количественных прогнозов процессов и явлений, оказывающих влияние на прогнозируемый процесс.

Количественный прогноз связан с какими-то численными параметрами прогнозируемого объекта, которые по сути являются случайными величинами (при прогнозировании). Поэтому с ними связаны такие характеристики случайных величин как математическое ожидание, дисперсия, наиболее вероятное значение и т.д.

При количественном прогнозировании различают точечные и интервальные прогнозы.

Под точечным прогнозом понимают оценку математического ожидания прогнозируемого параметра в заданный момент времени в будущем. Однако, как указывалось выше, мы никогда не сможем точно «угадать» будущую ситуацию. Поэтому значение точечного прогноза, как правило, не является достаточным и может рассматриваться как некоторый центр, около которого по некоторому закону будут группироваться будущие события. Поэтому дополнительно к точечному прогнозу рассматривается интервальный прогноз, характеризующий размер области, в которую с заданной вероятностью попадет будущее значение прогнозируемого параметра.

Интервал наблюдения – отрезок времени, на котором имеются статистические данные о значении прогнозируемой величины до настоящего момента времени.

Интервал упреждения – отрезок времени с момента осуществления прогноза до момента времени в будущем, для которого делается прогноз.

1.2. СОВРЕМЕННОЕ СОСТОЯНИЕ МЕТОДОЛОГИИ ПРОГНОЗИРОВАНИЯ

Современные методы прогнозирования можно классифицировать на три основные группы:

- методы экстраполяции;
- методы экспертных оценок;
- методы математического моделирования.

Суть метода экстраполяции состоит в распространении существующих закономерностей (тенденций и связей) на некоторый период в будущем. Динамическая (временная) экстраполяция основывается на предположении, что имеющийся временной ряд y_t представляет собой сумму двух составляющих – регулярной и случайной:

$$y_t = f(t) + \varepsilon_t, \tag{1.2.1}$$

где $f(t)$ - регулярная составляющая;

ε_t- случайная величина с нулевым математическим ожиданием.

В основе метода экстраполяции лежит интуитивное представление о какой-то очищенной от помех сущности анализируемого процесса. Применение этого метода фактически сводится к построению наилучшего в некотором смысле описания регулярной составляющей и экстраполяции его на прогнозный момент времени.

Выбор типа функции $f(t)$ осуществляется на основании визуального анализа динамики показателя. В качестве функций обычно используются:

- линейная $$f(t) = a_0 + a_1 t; \tag{1.2.2}$$
- полиномиальная $$f(t) = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n; \tag{1.2.3}$$
- экспоненциальная $$f(t) = ae^{bt}; \tag{1.2.4}$$
- логистическая кривая $$f(t) = \frac{a}{1 + be^{-ct}} \tag{1.2.5}$$

и другие простые математические зависимости.

Методы экстраполяции целесообразно использовать при прогнозировании показателей в тех случаях, когда имеется достаточно представительная статистика, позволяющая провести построение временных рядов, причем прогнозируемые события должны носить массовый характер.

К методам экстраполяции относится также группа методов прогнозирования, основанных на построении и использовании моделей связи прогнозируемого показателя y с различными факторами $X_1, X_2, ..., X_k$

$$y = f(X_1, X_2, ..., X_k). \tag{1.2.6}$$

Метод заключается в построении регрессии (математической зависимости) прогнозируемого показателя от ряда факторов (независимых переменных), прогнозировании будущих значений факторов и расчете прогнозных значений показателя по математической зависимости. При этом прогнозные значения независимых переменных определяются вне рамок регрессионной модели на основе дополнительной информации, как правило, с использованием экстраполяции трендов или экспертным путем.

Однако применение многофакторных моделей на практике сталкивается со сложностью отбора существенных факторов и трудностями в получении исходной информации.

Методы экспертного прогнозирования используют индивидуальные или групповые мнения о перспективах развития прогнозируемого объекта и могут использоваться при отсутствии или сложности получения статистической информации.

1.3. ПОНЯТИЕ О МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ И ЕГО РОЛИ В ПРОГНОЗИРОВАНИИ

Сложность и многообразие реальных процессов обуславливают необходимость их упрощения, схематизации и идеализации, т.е. абстрагирования от несущественных второстепенных деталей. Такое абстрагирование позволяет получить модель процесса. В последние десятилетия математическое моделирование широко проникло в теоретическую и прикладную экономику.

Термин «модель» широко используется в различных областях человеческой деятельности. Модель – материальный или мысленно представляемый объект, который в процессе исследования замещает объект – оригинал так, что его непосредственное изучение дает новые знания об объекте-оригинале.

Под моделированием понимается процесс построения, исследования и применения моделей. Главная особенность моделирования в том, что это метод опосредованного познания с помощью объектов-заменителей.

Необходимость использования метода моделирования определяется тем, что многие объекты непосредственно исследовать невозможно (например, ядро Земли), опасно (ядерный реактор, эпидемический процесс) или весьма трудоемко.

Математические модели описывают взаимосвязи между переменными, характеризующими объект, на языке математики (в виде формул и математических операций). Полезность математических моделей состоит в том, что с их помощью удается выразить зависимости между различными величинами, в частности в виде функций.

Математические модели - мощный инструмент познания реального мира, так как позволяют просчитывать на ЭВМ большое число различных вариантов. Они должны учитывать основные стороны и взаимосвязи объекта и пренебрегать несущественными. Разработка моделей – большое искусство.

Можно выделить два способа построения математических моделей:

- на основе статистической обработки результатов исследований;

- на основе известных законов физики или логических соображений.

Для эконометрики характерен первый способ построения моделей.

2. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ

2.1. ФУНКЦИОНАЛЬНАЯ И СТАТИСТИЧЕСКАЯ ЗАВИСИМОСТЬ

Во многих науках (физика, экономика и т. д.) используются модели, в которых некоторые переменные (не случайные) связаны функциональной зависимостью.

При статистической зависимости переменные (случайные величины) не связаны функционально. Однако закон распределения одной из них зависит от того, какое значение приняла другая случайная величина. Поэтому речь идет об условном распределении Y при заданном x.

В частности, можно рассматривать условное математическое ожидание M(Y/x) как некоторую функцию x. Такая зависимость называется регрессией.

При исследовании статистической зависимости между переменными пытаются ответить на следующие вопросы:

- существует ли статистическая связь между переменными;
- какова степень этой связи;
- какова форма связи.

Первые два вопроса решаются на основании корреляционного анализа.

2.2. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

В качестве меры тесноты связи между двумя случайными величинами обычно используется коэффициент корреляции - r

$$r = \frac{\mathrm{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{\mathrm{M}(XY) - m_x m_y}{\sigma_x \sigma_y}, \qquad (2.2.1)$$

где $cov(X,Y)$ - ковариация случайных величин X и Y;

σ_x - среднее квадратичное отклонение случайной величины X;

σ_y - среднее квадратичное отклонение случайной величины Y;

m_x – математическое ожидание случайной величины X;

m_y – математическое ожидание случайной величины Y;

$M(XY)$- математическое ожидание XY.

При этом $-1 \le r \le 1$. При $|r| = 1$связь становится функциональной.

Выборочный коэффициент корреляции $\hat{r}$ рассчитывается по формуле

$$\hat{r} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2 \sum_{i=1}^{n}(x_i - \overline{x})^2}}, \tag{2.2.2}$$

где x_i - значение случайной величины X для i-го наблюдения (объекта);

y_i - значение случайной величины Y для i-го наблюдения (объекта);

$\bar{x}, \bar{y}$ - выборочные средние значения случайных величин X и Y;

n – число наблюдений (объем выборки).

На практике используются следующие формулы для «ручных» вычислений

$$\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x}) = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n};$$

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^{\ 2} - \frac{(\sum_{i=1}^{n} x_i)^2}{n}; \tag{2.2.3}$$

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^{\ 2} - \frac{(\sum_{i=1}^{n} y_i)^2}{n}.$$

2.3. ПРОВЕРКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

После того, как вычислен выборочный коэффициент корреляции $\hat{r}$ следует проверить гипотезу об отсутствии корреляционной связи для генеральной совокупности H_0: $r = 0$.

Для этого вычисляется критерий

$$t = \frac{|\hat{r}|\sqrt{n-2}}{\sqrt{1-\hat{r}^2}} \tag{2.3.1}$$

и сравнивается с табличным значением $t_{n-2,\alpha}$ критерия Стьюдента с $\nu = n-2$ степенями свободы уровня значимости α.

Если $t > t_{n-2,\alpha}$, то с надежностью $1-\alpha$ можно отвергнуть гипотезу H_0 и считать, что корреляция имеется.

2.4. ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Следующий этап исследования корреляционной связи заключается в том, чтобы описать зависимость признака-результата от признака-фактора некоторым аналитическим выражением.

$$\tilde{y} = a_0 + a_1 x \quad ; \quad Y = \tilde{y} + \varepsilon ,$$

где $\tilde{y}$ – средний уровень показателя Y при данном значении x;

ε – случайная компонента.

Модель парной линейной регрессии может быть записана и в следующем виде

$$y_i = a_0 + a_1 x_i + \varepsilon_i (i = 1,...,n),$$

где x_i – значение переменной X в i-ом наблюдении;

y_i – значение переменной Y в i-ом наблюдении;

ε_i – значение случайной компоненты ε в i-ом наблюдении;

n – число наблюдений (объем выборки).

Основные предположения регрессионного анализа относятся к случайной компоненте ε и имеют решающее значение для правильного и обоснованного применения регрессионного анализа на практике.

В классической модели регрессионного анализа имеют место следующие предположения:

1) Величины ε_i (а также зависимые переменные y_i) являются случайными, а объясняющая переменная x_i – величина неслучайная.

2) Математическое ожидание случайной компоненты ε_i равно нулю ($\mathrm{M}(\varepsilon_i) = 0$).

3) Случайные величины ε_i имеют одинаковую дисперсию ($\mathrm{M}(\varepsilon_i{}^2) = \sigma^2$). Данное условие называется условием гомоскедастичности.

4) Случайные компоненты ε_i и ε_j являются некоррелированными между собой ($\mathrm{M}(\varepsilon_i \varepsilon_j) = 0$ при $i \neq j$).

Эти четыре предположения являются необходимыми для проведения регрессионного анализа в рамках классической модели.

Пятое предположение дает достаточные условия для обоснованного проведения проверки статистической значимости полученных регрессий и заключается в нормальности закона распределения ε_i.

В общем случае задача оценки параметров регрессии может решаться с помощью метода наименьших квадратов (МНК).

Рассмотрим использование метода наименьших квадратов для оценки параметров регрессии

$$\tilde{y} = a_0 + a_1 x . \qquad (2.4.1)$$

На практике имеется серия наблюдений $(x_i;y_i)$ $(i=1,..,n)$.

При этом

$$y_i = a_0 + a_1 x_i + \varepsilon_i .$$

Тогда

$$Q = \sum_{i=1}^{n} \varepsilon_i^{\,2} = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \to \min . \qquad (2.4.2)$$

Возьмем частные производные Q по параметрам a_0 и a_1 и приравняем их нулю

$$\begin{cases} \dfrac{\partial Q}{\partial a_0} = 2\sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)(-1) = 0 , \\ \dfrac{\partial Q}{\partial a_1} = 2\sum_{i=1}^{n} - x_i (y_i - a_0 - a_1 x_i) = 0 . \end{cases} \qquad (2.4.3)$$

Первое уравнение системы (2.4.3) можно преобразовать к виду

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} a_1 x_i = 0$$

или

$$n a_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i .$$

Второе уравнение можно преобразовать к виду

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} x_i y_i .$$

Таким образом, мы имеем систему уравнений

$$\begin{cases} n a_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i ; \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} x_i y_i , \end{cases} \qquad (2.4.4)$$

Разделив обе части уравнений (2.4.4) на n, получим систему уравнений в виде

$$\begin{cases} a_0 + a_1 \overline{x} = \overline{y} \\ a_0 \overline{x} + a_1 \overline{x^2} = \overline{xy} \end{cases} \qquad (2.4.5)$$

где соответствующие средние определяются по формулам

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} ; \qquad \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} ;$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \quad ; \qquad \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^{\,2}}{n} . \qquad (2.4.6)$$

Подставляя значение $a_0 = \bar{y} - a_1 \bar{x}$
из первого уравнения системы (2.4.5) во второе, получим

$$\hat{a}_1 = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\hat{\mathrm{cov}}(X,Y)}{s_x^{\,2}} , \qquad (2.4.7)$$

где $s_x^{\,2}$ - выборочная дисперсия переменной *X:*

$$s_x^{\,2} = \overline{x^2} - \bar{x}^2 = \frac{\sum_{i=1}^{n} x_i^{\,2}}{n} - (\bar{x})^2 , \qquad (2.4.8)$$

$\hat{\mathrm{cov}}(X,Y)$ - выборочная ковариация:

$$\hat{\mathrm{cov}}(X,Y) = \overline{xy} - \bar{x}\,\bar{y} = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}\,\bar{y} . \qquad (2.4.9)$$

Отметим, что линия регрессии проходит через точку $(\bar{x}, \bar{y})$, то есть $\bar{y} = \hat{a}_0 + \hat{a}_1 \bar{x}$.

В заключение приведем удобные для расчета оценок параметров формулы:

$$\begin{cases} \hat{a}_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} , & (2.4.10) \\ \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} . & (2.4.11) \end{cases}$$

Если рассчитан выборочный коэффициент корреляции $\hat{r}$, то коэффициенты a_0 и a_1 могут быть определены следующим образом

$$\hat{a}_1 = \hat{r} \frac{s_y}{s_x} , \qquad \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} , \qquad (2.4.12)$$

где $s_y^{\,2}$ - выборочная дисперсия переменной *Y:*

$$s_y{}^2 = \overline{y^2} - \overline{y}^2 = \frac{\sum_{i=1}^{n} y_i{}^2}{n} - (\overline{y})^2$$

В качестве оценки дисперсии случайной компоненты σ^2 используется

$$s^2 = \frac{\sum_{i=1}^{n} \varepsilon_i{}^2}{n-2}. \qquad (2.4.13)$$

Проверка статистической значимости оценок коэффициентов регрессии

В действительности может оказаться, что фактор *X* не влияет на результирующий признак *Y*, что эквивалентно условию $a_1 = 0$. Однако при этом $\hat{a}_1 \neq 0$. Для проверки существенности отклонения a_1 от 0 служит статистический тест. Рассматривается гипотеза H_0: $a_1=0$ при альтернативной гипотезе H_1: $a_1 \neq 0$.

Проверка значимости оценок с помощью критерия Стьюдента проводится путем сопоставления вычисленных значений оценок с величиной их среднего квадратичного отклонения.

Статистика теста имеет вид

$$t(\hat{a}_1) = \frac{\hat{a}_1}{\hat{\sigma}(\hat{a}_1)} = \frac{\hat{a}_1}{\sqrt{\hat{D}\{\hat{a}_1\}}}. \qquad (2.4.14)$$

Известно, что дисперсия оценки $\hat{a}_1$ определяется следующим образом

$$\sigma^2(\hat{a}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad (2.4.15)$$

Поэтому для вычисления оценки $\hat{\sigma}(\hat{a}_1)$ можно использовать следующую формулу

$$\hat{\sigma}(\hat{a}_1) = \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}. \qquad (2.4.15а)$$

Если $|t(\hat{a}_1)| \le t_{n-2,\alpha}$, то принимается гипотеза H_0, в противном случае принимается гипотеза H_1. В случае принятия гипотезы H_0 фактор X исключается из модели и принимается, что $\tilde{y} = a_0$.

Использование регрессионной модели для прогноза. Дисперсия ошибки прогноза.

Основное назначение регрессионной модели – использование ее для прогноза экономического показателя Y. Прогноз осуществляется подстановкой значения фактора x_l в оценку детерминированной составляющей:

$$\hat{y}(x_l) = \hat{a}_0 + \hat{a}_1 x_l . \qquad (2.4.16)$$

Найдем дисперсию прогноза $\hat{y}(x_l)$, представляющего собой выборочную оценку $M(Y/x_l)$. С этой целью уравнение (2.4.16) представим в виде:

$$\hat{y}(x_l) = \bar{y} + \hat{a}_1 (x_l - \bar{x}) . \qquad (2.4.17)$$

Дисперсия $\hat{y}(x_l)$ равна сумме дисперсий двух независимых (доказательство этого факта опускается) слагаемых выражения (2.4.17).

$$\sigma^2\big(\hat{y}(x_l)\big) = \sigma^2(\bar{y}) + \sigma^2(\hat{a}_1)(x_l - \bar{x})^2 \qquad (2.4.18)$$

Здесь учтено, что $(x_l - \bar{x})$ – неслучайная величина, при вынесении которой за знак дисперсии ее необходимо возвести в квадрат.

Дисперсия выборочной средней $\bar{y}$

$$\sigma^2(\bar{y}) = \sigma^2\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{\sum_{i=1}^{n} \sigma^2(y_i)}{n^2} = \frac{\sum_{i=1}^{n} \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

(2.4.19)

Для нахождения дисперсии $\sigma^2(\hat{a}_1)$ воспользуемся формулой (2.4.15), т.е.

$$\sigma^2(\hat{a}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Найдем оценку дисперсии $\hat{y}(x_l)$, учитывая (2.4.18), (2.4.19), (2.4.15) и заменяя σ^2 ее оценкой s^2.

$$s^2\big(\hat{y}(x_l)\big) = s^2\Big(\frac{1}{n} + \frac{(x_l - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\Big) \qquad (2.4.20)$$

Из формулы (2.4.20) видно, что дисперсия $\hat{y}(x_l)$ зависит от значения объясняющей переменной x_l: при $x_l = \bar{x}$ она минимальна, а по мере удаления x_l от $\bar{x}$ она увеличивается.

Ошибку прогноза можно представить следующим образом

$$y(x_l) - \hat{y}(x_l) = \big(y(x_l) - M(Y/x_l)\big) + \big(M(Y/x_l) - \hat{y}(x_l)\big) \qquad (2.4.22)$$

Первая скобка представляет собой ошибку, вызванную наличием случайной составляющей ε. Ее дисперсия равна σ^2.

Вторая скобка представляет собой ошибку оценки среднего уровня. Ее дисперсия определяется по формуле (2.4.20).

Таким образом, для оценки дисперсии ошибки прогноза можно пользоваться следующим выражением

$$\widehat{D}\{y(x_l) - \hat{y}(x_l)\} = s^2\Big(1 + \frac{1}{n} + \frac{(x_l - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\Big). \qquad (2.4.23)$$

Эта формула учитывает как погрешность оценки $\hat{y}(x_l)$, так и отклонение $y(x_l)$ от своего математического ожидания, обусловленное наличием случайной составляющей.

Из (2.4.23) следует, что с ростом $|x_l - \bar{x}|$ дисперсия ошибки прогноза увеличивается.

Проверка значимости уравнения регрессии

Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Проверка значимости уравнения регрессии производится на основе дисперсионного анализа.

Обозначим $\hat{y}(x_i) = \hat{y}_i$. Согласно основной идее дисперсионного анализа

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

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

(2.4.24)

или

$$Q = Q_R + Q_e \qquad (2.4.25)$$

где Q – общая сумма квадратов отклонений зависимой переменной от средней, а Q_R и Q_e - соответственно сумма квадратов, обусловленная регрессией, и остаточная сумма квадратов, характеризующая влияние неучтенных факторов.

Можно показать, что пропущенное в (2.4.25) третье слагаемое $Q_3 = 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i)$ равно 0.

Схема дисперсионного анализа имеет вид, представленный в *табл.* 1.

Т а б л и ц а 1. Схема дисперсионного анализа

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Средние квадраты
Регрессия	$Q_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	*m-1*	$s_R^{\ 2} = \frac{Q_R}{m-1}$
Остаточная	$Q_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	*n-m*	$s^{\ 2} = \frac{Q_e}{n-m}$
Общая	$Q = \sum_{i=1}^{n}(y_i - \bar{y})^2$	*n-1*	

Средние квадраты $s_R^{\ 2}$ и $s^{\ 2}$ представляют собой несмещенные оценки дисперсий зависимой переменной, обусловленных соответственно регрессией или объясняющей переменной X и воздействием неучтенных случайных факторов и ошибок; m – число оцениваемых параметров уравнения регрессии; n – число наблюдений.

При отсутствии линейной зависимости между зависимой и объясняющими переменными случайные величины Q_R и Q_e имеют χ^2-распределение соответственно с *m-1* и *n-m* степенями свободы, а отношение $s_R^{\ 2}$ к $s^{\ 2}$ – распределение Фишера с теми же степенями свободы. Поэтому уравнение регрессии значимо на уровне α, если фактически наблюдаемое значение статистики

$$F = \frac{Q_R(n-m)}{Q_e(m-1)} = \frac{{s_R}^2}{s^2} > F_{\alpha,m-1,n-m}\ , \qquad (2.4.26)$$

где $F_{\alpha,m-1,n-m}$ - табличное значение критерия Фишера, определенное на уровне значимости α при m-1 и n-m степенях свободы.

В случае линейной парной регрессии $m=2$, и уравнение регрессии значимо на уровне α, если

$$F = \frac{Q_R(n-2)}{Q_e} = \frac{{s_R}^2}{s^2} > F_{\alpha,1,n-2} \qquad (2.4.27)$$

Мерой качества регрессионной модели, характеристикой прогностической силы регрессионной модели является коэффициент детерминации, определяемый по формуле

$$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_e}{Q} \qquad (2.4.28)$$

Величина R^2 показывает, какая часть (доля) вариации зависимой переменной обусловлена вариацией объясняющей переменной.

Так как $0 \le Q_R \le Q$, то $0 \le R^2 \le 1$.

Чем ближе R^2 к единице, тем лучше регрессия аппроксимирует эмпирические данные, тем теснее наблюдения примыкают к линейной регрессии. Если R^2=1, то эмпирические точки (x_i,y_i) лежат на линии регрессии и между переменными Y и X существует линейная функциональная зависимость. Если R^2=0, то вариация зависимой переменной полностью обусловлена воздействием неучтенных в модели переменных, и линия регрессии параллельна оси абсцисс.

Пример

Исследуем зависимость розничного товарооборота магазинов (млрд р.) от среднесписочного числа работников. Обозначим:

х – число работников;

у – товарооборот.

Исходные данные и результаты расчетов приведены в *табл*. 2.

Т а б л и ц а 2. Исходные данные и результаты расчетов

Номер магазина	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	$\hat{y}_i$	ε_i	ε_i^2
1	79	0,5	39,5	6 241	0,25	0,48	0,02	0,0004
2	85	0,7	59,5	7 225	0,49	0,61	0,09	0,0081
3	102	0,9	91,8	10 404	0,81	0,96	-0,06	0,0036
4	115	1,1	126,5	13 225	1,21	1,23	-0,13	0,0169
5	122	1,4	170,8	14 884	1,96	1,37	0,03	0,0009
6	126	1,4	176,4	15 876	1,96	1,45	-0,05	0,0025
7	134	1,7	227,8	17 956	2,89	1,62	0,08	0,0064
8	147	1,9	279,3	21 609	3,61	1,88	0,02	0,0004
Итого	910	9,6	1171,6	107 420	13,18	-	-	0,0392

Вычислим выборочный коэффициент корреляции, используя удобные для вычисления формулы (2.2.3):

$$\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} = 1171{,}6 - \frac{910 \cdot 9{,}6}{8} = 79{,}6\,;$$

$$\sum_{i=1}^{n} x_i^{\,2} - \frac{(\sum_{i=1}^{n} x_i)^2}{n} = 107420 - \frac{910^2}{8} = 3907{,}5\,;$$

$$\sum_{i=1}^{n} y_i^{\,2} - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 13{,}18 - \frac{9{,}6^2}{8} = 1{,}66\,.$$

Тогда

$$\hat{r} = \frac{79{,}6}{\sqrt{3907{,}5 \cdot 1{,}66}} = 0{,}988.$$

Проверим значимость выборочного коэффициента корреляции. Для этого вычислим статистику t:

$$t = \frac{|\hat{r}|\sqrt{n-2}}{\sqrt{1-\hat{r}^2}} = \frac{|0{,}988| \cdot \sqrt{8-2}}{\sqrt{1-0{,}988^2}} = 15{,}65.$$

Табличное значение критерия Стьюдента для n-2 = 6 степеней свободы и уровня значимости $\alpha = 0{,}05$ составляет

$$t_{n-2,\alpha} = 2{,}45.$$

Так как $t > t_{n-2,\alpha}$, то нулевая гипотеза об отсутствии корреляции отвергается, и полученный коэффициент корреляции статистически значим.

Найдем коэффициенты парной линейной регрессии, предварительно вычислив средние значения и выборочные дисперсии переменных *x* и *y*

$$\bar{x} = \frac{910}{8} = 113{,}8 \; ; \; \bar{y} = \frac{9{,}6}{8} = 1{,}2.$$

$$s_x^{\,2} = \frac{\sum_{i=1}^{n} x_i^{\,2}}{n} - (\bar{x})^2 = \frac{107420}{8} - 113{,}8^2 = 477{,}06 \; ; \; s_x = 21{,}84 \; ;$$

$$s_y^{\,2} = \frac{\sum_{i=1}^{n} y_i^{\,2}}{n} - (\bar{y})^2 = \frac{13{,}18}{8} - 1{,}2^2 = 0{,}2075 \; ; \; s_y = 0{,}456.$$

Найдем оценки параметров регрессии:

$$\hat{a}_1 = \hat{r}\frac{s_y}{s_x} = 0{,}988 \cdot \frac{0{,}456}{21{,}84} = 0{,}0206 \; ;$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1\bar{x} = 1{,}2 - 0{,}0206 \cdot 113{,}8 = -1{,}144.$$

Таким образом, регрессия имеет вид

$$\hat{y}(x) = -1{,}144 + 0{,}0206x \; .$$

Рассчитаем значения $\hat{y}_i = \hat{y}(x_i) = -1{,}144 + 0{,}0206\, x_i$ для всех магазинов и занесем результаты вычислений в таблицу, найдя при этом $\varepsilon_i = y_i - \hat{y}_i$ и ε_i^2.

Найдем оценку дисперсии случайной компоненты **ε**

$$s^2 = \frac{\sum_{i=1}^{n} \varepsilon_i^{\,2}}{n-2} = \frac{0{,}0392}{8-2} = 0{,}0065 \; ; s = \sqrt{0{,}0065 =} \; 0{,}081.$$

Проверим статистическую значимость параметра $\hat{a}_1$

$$\hat{\sigma}(\hat{a}_1) = \frac{s}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} = \frac{0{,}081}{\sqrt{3907{,}5}} = 0{,}0013.$$

$$t(\hat{a}_1) = \frac{\hat{a}_1}{\hat{\sigma}(\hat{a}_1)} = \frac{0{,}0206}{0{,}0013} = 15{,}8.$$

Так как $|t(\hat{a}_1)| > 2{,}45$, то параметр $\hat{a}_1$ статистически значим.

Проверим статистическую значимость регрессии в целом по критерию Фишера.

Известно, что

$$Q = \sum_{i=1}^{n}(y_i - \overline{y})^2 = 1,66;$$

$$Q_e = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = 0,0392.$$

Тогда

$$Q_R = Q - Q_e = 1,66 - 0,0392 = 1,621;$$

$$F = \frac{Q_R(n-2)}{Q_e} = \frac{1,621 * 6}{0,0392} = 248,1.$$

Табличное значение критерия Фишера уровня значимости α=0,05 при 1 и 6 степенях свободы составит

$$F_{\alpha,1,n-2} = 5,99.$$

Так как $F > F_{\alpha,1,n-2}$, то уравнение регрессии значимо по критерию Фишера.

Прогнозное значение розничного товарооборота при $x_l = 200$ составит

$$\widehat{y}(x_l) = -1,144 + 0,0206 \cdot 200 = 2,976.$$

При этом оценка дисперсии ошибки прогноза

$$\widehat{D}\{y(x_l) - \widehat{y}(x_l)\} = s^2(1 + \frac{1}{n} + \frac{(x_l - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}) = 0,0065 * (1 + \frac{1}{8} + \frac{(200 - 113,8)^2}{3907,5}) =$$

=0,0197.

Задание № 1

В *табл. 3* приведены 5 показателей деятельности торговых предприятий. В соответствии с номером варианта выберите номера 2-х показателей

Номер варианта	Номер 1-го показателя	Номер 2-го показателя
1	1	2
2	1	3
3	1	4
4	1	5
5	2	3
6	2	4
7	2	5

8	3	4
9	3	5
10	4	5

Т а б л и ц а 3. Показатели деятельности торговых предприятий за год

Номер предприятия	Численность работников	Средняя зарплата, тыс. р.	Дебиторская задолженность на конец года, тыс. р.	Балансовая прибыль, тыс. р.	Собственные оборотные средства, тыс. р.
	(1)	(2)	(3)	(4)	(5)
1	20	17,3	7,0	80	320
2	50	20,2	5,1	105	611
3	80	19,1	1,2	100	840
4	35	17,0	7,1	94	482
5	115	20,3	2,2	112	1050
6	40	19,1	5,3	108	499
7	40	19,2	4,0	100	505
8	50	19,2	4,1	88	521
9	30	17,0	7,8	92	412
10	35	17,1	7,3	90	405
11	70	19,3	2,2	92	788
12	120	21,0	1,0	101	1280
13	100	20,0	2,3	98	990
14	70	19,7	7,4	95	810
15	65	19,2	5,6	90	750
16	80	19,1	2,0	95	924
17	150	21,3	1,5	109	1950
18	50	18,0	5,3	90	590
19	60	20,0	3,2	97	722
20	50	19,1	5,8	90	540

С помощью корреляционного и регрессионного анализа изучить связь между показателями, указанными в Вашем варианте.

1. Рассчитать значение выборочного коэффициента корреляции.
2. Проверить статистическую значимость полученного коэффициента корреляции.
3. Сделать вывод о наличии и тесноте статистической связи.
4. Найти коэффициенты парной линейной регрессии.
5. Проверить статистическую значимость параметра $\hat{a}_1$ по критерию Стьюдента.
6. Проверить статистическую значимость уравнения регрессии в целом по критерию Фишера.
7. Сделать прогноз признака-результата, если признак-фактор принимает свое среднее значение.
8. Найти оценку дисперсии ошибки прогноза.

3. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ

3.1. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Временной ряд представляет собой ряд числовых значений какого-либо показателя в последовательные моменты или периоды времени. Числовые значения, составляющие временной ряд, называются уровнями ряда.

По способу построения ряд может быть моментным, когда уровни ряда представлены на определенные моменты времени (конец квартала, начало года и т.д.) и интервальным, когда уровни ряда соответствуют определенным интервалам времени.

Изучение различных процессов на основе временных рядов включает следующие этапы:
- сбор исходной информации и построение временных рядов;
- визуальный анализ временного ряда и формирование набора возможных моделей прогнозирования;
- идентификация (подбор) модели;
- оценка параметров моделей;
- осуществление прогноза по математической модели.

В практике анализа временных рядов принято считать, что значения уровней временных рядов складываются из следующих компонент:
- тренд;
- сезонная составляющая;
- циклическая составляющая;
- случайная составляющая.

Под трендом (тенденцией) понимают изменения, определяющие общее направление развития изучаемого показателя. Это систематическая составляющая долговременного действия. Для описания тренда используют плавно меняющиеся, гладкие функции.

Наряду с долговременными тенденциями во временных рядах часто имеют место более или менее регулярные колебания – периодические составляющие рядов динамики. Если период колебаний не превышает одного года, то их называют сезонными. Причины сезонных колебаний могут быть связаны с природно-климатическими условиями, могут носить социальный характер (например, увеличение покупок в выходные дни, увеличение платежей в конце квартала и т. д.). Для описания сезонной компоненты используют периодические функции.

При большом периоде колебаний считают, что во временных рядах имеется циклическая составляющая. Примерами могут служить демографические, деловые, инвестиционные и другие циклы.

Если из временного ряда удалить тренд и периодические составляющие, то останется нерегулярная компонента. Часто причиной нерегулярных колебаний является действие большого числа различных факторов. Эта компонента рассматривается как случайная.

3.2. ПОКАЗАТЕЛИ ДИНАМИКИ ВРЕМЕННЫХ РЯДОВ

Показатели динамики – это величины, характеризующие изменения уровней временного ряда. К ним относятся абсолютный прирост, коэффициент роста и коэффициент прироста.

Различают базисные и цепные показатели динамики. Базисные показатели – это результат сравнения текущего уровня ряда с одним фиксированным уровнем, принятым за базу (обычно это начальный уровень ряда). Цепные показатели – это результат сравнения текущего уровня ряда с предшествующим уровнем.

Формулы для расчета показателей представлены в *табл*. 4.

Т а б л и ц а 4. Показатели динамики

Базисные	Цепные
Абсолютный прирост	
$A_i=y_i-y_1$	$a_i=y_i-y_{i-1}$
Коэффициент роста	
$L_i=y_i/y_1$	$l_i=y_i/y_{i-1}$
Коэффициент прироста	
$K_i=(y_i-y_1)/y_1=L_i-1$	$k_i=(y_i-y_{i-1})/y_{i-1}=l_i-1$

Рассмотрим определение среднего абсолютного прироста (цепного). Предположим, что имеется временной ряд $y_1, y_2, \ldots, y_n$.
Тогда $a_2 = y_2 - y_1$, $a_3 = y_3 - y_2$, $a_4 = y_4 - y_3$, …, $a_n = y_n - y_{n-1}$ (цепные приросты).

Средний абсолютный прирост $\bar{a}$ равен

$$\bar{a} = \frac{a_2 + a_3 + a_4 + \cdots + a_n}{n-1} =$$

$$= \frac{(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3) + \cdots + (y_n - y_{n-1})}{n-1} = \frac{y_n - y_1}{n-1}$$

Рассмотрим определение среднего коэффициента роста (цепного)
Предположим, что имеется временной ряд $y_1, y_2, \ldots, y_n$.
Тогда $l_i = \frac{y_i}{y_{i-1}}$ ($i=2,\ldots,n$) – цепные коэффициенты роста.

Средний коэффициент роста $\bar{l}$ равен

$$\bar{i} = \sqrt[n-1]{\frac{y_2}{y_1}\frac{y_3}{y_2}\frac{y_4}{y_3}\dots\frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

3.3. ВЫДЕЛЕНИЕ ТРЕНДА. СГЛАЖИВАНИЕ И ВЫРАВНИВАНИЕ

Временной ряд может быть представлен в виде

$$y_t = f(\vec{a},t) + \varepsilon_t,$$

где $f(\vec{a},t)$ – регулярная составляющая (тренд, основная тенденция);

ε_t – случайная составляющая;

$\vec{a}$ – вектор параметров.

Одним из методов выделения тренда является сглаживание временного ряда с помощью скользящего среднего. Метод состоит в замене уровней ряда динамики y_t средними арифметическими- $\tilde{y}_t$ за определенный интервал (окно сглаживания), длина которого определена заранее. При этом сам выбранный интервал времени «скользит» вдоль ряда.

$$\tilde{y}_t = \frac{\sum_{i=t-k}^{t+k} y_i}{2k+1} \qquad (3.3.1)$$

Например, при *к=2, 2к+1=5* и

$$\tilde{y}_t = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}$$

Получаемый таким образом ряд скользящих средних ведет себя более гладко, чем исходный ряд, из-за усреднения отклонений ряда. Действительно, если индивидуальный разброс значений члена временного ряда y_t около своего среднего значения m характеризуется дисперсией σ^2, то разброс средней из *2к+1* членов временного ряда около того же значения m будет характеризоваться существенно меньшей величиной дисперсии, равной σ^2/*(2к+1)*.

В результате сглаживания получается ряд с меньшим количеством уровней, так как крайние значения теряются.

Пример. Провести сглаживание временного ряда y_t по данным таблицы методом скользящего среднего с интервалом сглаживания 3 года.

t	1	2	3	4	5	6	7	8
y_t	213	171	291	309	317	362	351	361

$$\tilde{y}_t = \frac{\sum_{i=t-1}^{t+1} y_i}{3} \qquad (3.3.2)$$

Например, при $t=2$ по формуле (3.3.2)

$\tilde{y}_2 = \frac{y_1+y_2+y_3}{3} = \frac{213+171+291}{3} = 225{,}0$,

при $t=3$

$\tilde{y}_3 = \frac{y_2+y_3+y_4}{3} = \frac{171+291+309}{3} = 257{,}0$ и т.д.

В результате получим сглаженный ряд

t	1	2	3	4	5	6	7	8
$\tilde{y}_t$	-	225,0	257,0	305,7	329,3	336,3	358,0	-

При аналитическом выравнивании подбирают математическую функцию, значения которой наиболее близки к уровням выравниваемого ряда. Выравнивание ряда сводится к определению параметров ***a*** функции *f*(***a***,*t*). Для этого используется метод наименьших квадратов (МНК).

3.4. ЛИНЕЙНЫЕ МОДЕЛИ ТРЕНДА

Предположим, что имеет место линейная зависимость $f(t) = a + bt$ т. е.

$$y_t = a + bt + \varepsilon_t\,. \tag{3.4.1}$$

Найдем оценки коэффициентов *a* и *b* по фактическим данным об уровнях ряда (t_i; y_i) ($i=1,\ldots,n$) так, чтобы сумма квадратов отклонений теоретической кривой от реальных данных была минимальной

$$\sum_{i=1}^{n} \varepsilon_i^{\,2} = \sum_{i=1}^{n} (y_i - f(t_i))^2 \to \min \tag{3.4.2}$$

или

$$Q = \sum_{i=1}^{n} (y_i - a - bt_i)^2 \to \min\,. \tag{3.4.2a}$$

Возьмем частные производные Q по параметрам a и b и приравняем их нулю

$$\begin{cases} \dfrac{\partial Q}{\partial a} = 2\sum_{i=1}^{n} (y_i - a - bt_i)(-1) = 0\,, \\ \dfrac{\partial Q}{\partial b} = 2\sum_{i=1}^{n} -t_i (y_i - a - bt_i) = 0\,. \end{cases} \tag{3.4.3}$$

Первое уравнение системы (3.4.3) можно преобразовать к виду

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bt_i = 0$$

или

$$na + b\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} y_i \, .$$

Второе уравнение можно преобразовать к виду

$$a\sum_{i=1}^{n} t_i + b\sum_{i=1}^{n} t_i^{\,2} = \sum_{i=1}^{n} t_i y_i \, .$$

Таким образом, мы имеем систему уравнений

$$\begin{cases} na + b\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} y_i \, , \\ a\sum_{i=1}^{n} t_i + b\sum_{i=1}^{n} t_i^{\,2} = \sum_{i=1}^{n} t_i y_i \, . \end{cases} \qquad (3.4.4)$$

Ее решение позволяет найти оценки параметров *a* и *b*.

Для упрощения расчетов (при нечетном количестве точек ряда – *2к+1*) будем считать, что ряд образуется для моментов времени *–к, -к+1, ... 0, 1, 2, ... к* .

Тогда

$$\sum_{i=1}^{n} t_i = 0$$

и система уравнений имеет решение

$$\begin{cases} \hat{a} = \dfrac{\sum_{i=1}^{n} y_i}{n} \, , \\ \hat{b} = \dfrac{\sum_{i=1}^{n} t_i y_i}{\sum_{i=1}^{n} t_i^{\,2}} \, . \end{cases} \qquad (3.4.5)$$

Полученная модель используется для прогноза показателя на момент времени t_L

$$\hat{y}(t_L) = \hat{a} + \hat{b} t_L \qquad (3.4.6)$$

Ошибки в оценке параметров приводят к ошибке в оценке тренда (среднего уровня), т. е. величина $\hat{y}\,(t_L)$ является случайной.

Дисперсия ошибки прогноза оценивается по формуле

$$\widehat{D}(\hat{y}(t_L) - y(t_L)) = s^2 \left(\frac{n+1}{n} + \frac{3(n+2L-1)^2}{n(n^2-1)}\right), \qquad (3.4.7)$$

где $y(t_L)$– истинное значение величины:

$$y(t_L) = a + bt_L + \varepsilon_L;$$

n – количество точек имеющегося временного ряда;

L – количество точек периода упреждения;

$s^2 = \dfrac{\sum_{i=1}^{n} \varepsilon_i^2}{n-2}$ - оценка остаточной дисперсии.

Из формулы (3.4.7) следует, что дисперсия ошибки прогноза увеличивается с увеличением периода упреждения (L) и уменьшается с увеличением числа точек временного ряда (n).

ПРИМЕР. Опишем динамику добычи угля в Англии за ряд лет (*табл.* 5) линейной зависимостью.

Т а б л и ц а 5. Динамика добычи угля в Англии

t_i	y_i	t_i^2	$y_i t_i$	$\hat{y}_i$	ε_i	ε_i^2
-2	227	4	-454	227	0	0
-1	219	1	-219	218	1	1
0	209	0	0	209	0	0
1	197	1	197	200	-3	9
2	193	4	386	191	2	4
Итого	1045	10	-90		0	14

$$\hat{a} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{1045}{5} = 209;$$

$$\hat{b} = \frac{\sum_{i=1}^{n} t_i y_i}{\sum_{i=1}^{n} t_i^2} = \frac{-90}{10} = -9,$$

т. е. линейная модель имеет вид

$$\hat{y}(t) = 209 - 9t.$$

При этом

$$\hat{y}_i = \hat{y}(t_i) = 209 - 9t_i$$

Результаты расчетов приведены в пятом столбце таблицы

При прогнозировании на 3 года (t_L=5) прогноз добычи угля по модели составит

$$\widehat{y}(t_L) = 209 - 9 \times 5 = 164 \ .$$

Определим дисперсию ошибки прогноза по формуле (3.4.7), оценив предварительно остаточную дисперсию.

$$s^2 = \frac{\sum_{i=1}^{5} \varepsilon_i^{\ 2}}{5-2} = \frac{14}{3} = 4{,}67\,;$$

$$\widehat{D}(\widehat{y}(t_L) - y(t_L)) = 4{,}67 \times \left(\frac{5+1}{5} + \frac{3 \times (5 + 2 \times 3 - 1)^2}{5 \times (5^2 - 1)}\right) = 17{,}28\,.$$

Задание № 2

Выберите из *табл.*6 временной ряд в соответствии с номером Вашего варианта (по последней цифре шифра зачетной книжки)

Таблица 6. Исходные данные по вариантам

Номер варианта	Временной ряд				
	1	2	3	4	5
1	26,7	27,3	28,8	30,9	31,2
2	616	635	657	707	716
3	85,4	87,2	93,4	97,1	97,2
4	865	867	910	999	1025
5	304	325	344	359	386
6	212,3	216,2	219,8	223,2	226,4
7	145,0	152,9	164,6	168,8	181,3
8	59,1	56,1	58,9	58,4	57,5
9	78,5	81,1	87,3	91,7	96,4
10	292,3	327,6	369,3	412,4	458,9

1. Рассчитать показатели динамики – абсолютный прирост, коэффициент роста, коэффициент прироста (цепные и базисные).
2. Найти средний абсолютный прирост и средний коэффициент роста.
3. Подобрать линейную зависимость вида $f(t) = a + bt$. Найти оценки коэффициентов $\widehat{a}$ и $\widehat{b}$ по методу наименьших квадратов.
4. Сделать прогноз показателя по математической модели тренда на 3 года вперед и найти оценку дисперсии ошибки прогноза

ВОПРОСЫ К ЗАЧЕТУ И ЭКЗАМЕНУ

1. Функциональная и статистическая зависимость. Исследование статистической зависимости.
2. Парный коэффициент корреляции.
3. Проверка значимости коэффициента корреляции.
4. Математическая модель парной линейной регрессии.
5. Метод наименьших квадратов (МНК) в оценке параметров парной линейной регрессии.
6. Свойства оценок параметров парной линейной регрессии, полученных МНК.
7. Проверка статистической значимости оценок коэффициентов парной линейной регрессии.
8. Прогноз и дисперсия ошибки прогноза на основе парной линейной регрессии.
9. Проверка значимости уравнения парной линейной регрессии по критерию Фишера.
10. Показатели качества парной линейной регрессии. Средняя ошибка аппроксимации.
11. Понятие временного ряда**.** Этапы анализа временных рядов.
12. Трендовая, сезонная, циклическая и случайная компоненты временных рядов.
13. Показатели динамики временных рядов.
14. Выделение тренда. Сглаживание и выравнивание.
15. Аналитическое выравнивание временного ряда с помощью линейной модели.
16. Методы прогнозирования.
17. Источники ошибок прогнозирования.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Ивченко Ю.С. Эконометрика [Электронный ресурс]: курс лекций/ Ивченко Ю.С.— Электрон. текстовые данные.— Саратов: Вузовское образование, 2018.— 121 с.— Режим доступа: http://www.iprbookshop.ru/73609.html.— ЭБС «IPRbooks»
2. Кремер Н.Ш. Эконометрика [Электронный ресурс]: учебник для студентов вузов/ Кремер Н.Ш., Путко Б.А.— Электрон. текстовые данные.— М.: ЮНИТИ-ДАНА, 2017.— 328 с.— Режим доступа: http://www.iprbookshop.ru/71071.html.— ЭБС «IPRbooks»
3. Ивченко Ю.С. Эконометрика в MS EXCEL [Электронный ресурс]: лабораторный практикум/ Ивченко Ю.С.— Электрон. текстовые данные.— Саратов: Ай Пи Эр Медиа, 2018.— 94 с.— Режим доступа: http://www.iprbookshop.ru/70785.html.— ЭБС «IPRbooks»
4. Яковлев В.П. Эконометрика [Электронный ресурс]: учебник для бакалавров/ Яковлев В.П.— Электрон. текстовые данные.— М.: Дашков и К, 2016.— 384 с.— Режим доступа: http://www.iprbookshop.ru/60631.html.— ЭБС «IPRbooks»
5. Богданов А. И. Математические модели прогнозирования: монография.- СПб.: СПГУТД, 2007.- 128 с.
6. Богданов А.И. Эконометрика: учеб. пособие. – СПб: СПГУТД, 2010.- 105 с.
7. Дубина И. Н. Математико-статистические методы в эмпирических социально-экономических исследованиях: учеб. пособие по дисциплине «Эконометрика». –М. : Финансы и статистика, 2010. – 413 с.