

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

**УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«Брестский государственный технический университет»**

Кафедра «Техническая эксплуатация автомобилей»

**КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ
ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА**

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ
к лабораторным занятиям
по дисциплине
«ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ
И ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ»
для студентов специальностей
1 - 37 01 06 «Техническая эксплуатация автомобилей»,
1 – 37 01 07 «Автосервис»**



Брест 2016

УДК 629.331.08

Методические указания к лабораторным занятиям по дисциплине «Основы научных исследований и инновационной деятельности» для студентов специальностей 1-37 01 06 «Техническая эксплуатация автомобилей», 1- 37 01 07 «Автосервис» содержат методику проведения корреляционного и регрессионного анализа, планирования эксперимента применительно к технической эксплуатации автомобилей и могут использоваться при выполнении лабораторных работ и дипломного проекта по вышеназванным специальностям.

Составители: С.В. Монтик, зав. кафедрой ТЭА, доцент, к.т.н.
П.С. Концевич, ст. преподаватель кафедры ТЭА, м.т.н.
С.О. Березуцкая, ассистент кафедры ТЭА

Рецензент: заместитель генерального директора – директор филиала «Автовокзал г. Бреста» ОАО «Брестоблавтотранс» А. М. Сенчук

© Учреждение образования
«Брестский государственный технический университет» 2016

Лабораторная работа №1

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Цель: изучить задачи и методику проведения корреляционного анализа, примеры его практического применения

Краткие сведения о корреляционном анализе

Зависимость между величинами, при которой каждому значению одной величины отвечает с соответствующей вероятностью множество возможных значений другой величины, называют **вероятностной**.

Если при наличии вероятностной зависимости между двумя величинами с изменением значения одной величины изменяется только математическое ожидание второй и наоборот, а дисперсия и тип закона распределения остаются неизменными, то для таких величин характерна **корреляционная зависимость**.

Задача корреляционного анализа – выявление наличия и тесноты (силы) связи между значениями различных случайных величин.

Примеры корреляционной связи: между величиной износа узла трения и пробегом автомобиля; между пределом прочности и пределом текучести стали; между твердостью и износостойкостью стали.

При выполнении корреляционного анализа по результатам эксперимента строят корреляционное поле, по которому по тесноте группирования точек вокруг прямой или кривой линии можно визуально судить о наличии корреляционной зависимости (см. рис.1.1).

Силу линейной статистической связи между случайными величинами X и Y оценивают **коэффициентом корреляции** r_{XY} , который принимает значения в интервале от -1 до $+1$ и не зависит от единиц величин X и Y . Чем больше по абсолютной величине коэффициент корреляции, тем сильнее линейная зависимость между величинами X и Y . Однако обратное не всегда верно. Силу нелинейной статистической связи между случайными величинами оценивают с помощью *корреляционного отношения*.

Задание

Используя изложенную ниже методику проведения корреляционного анализа и табличный процессор MS Excel, определите, существует ли линейная корреляционная зависимость между высотой над уровнем моря и снижением эффективной мощности двигателя автомобиля. Письменно ответить на контрольные вопросы. Исходные данные (таблица А.1), справочные данные (таблица Б.1) и шаблон отчета по лабораторной работе содержатся в файле **Лаб_раб_ОНИИД1.xls**. **Формулы нужно запрограммировать.**

Содержание отчета по лабораторной работе

Тема, цель, исходные данные, распечатка отчета по лабораторной работе (файл **Лаб_раб_ОНИИД1.xls**, письменные ответы на контрольные вопросы.

Контрольные вопросы

1. Перечислите задачи корреляционного анализа?
2. Какая зависимость называется вероятностной?
3. Какая зависимость называется корреляционной?
4. Привести примеры корреляционной зависимости.
5. Что характеризует коэффициент корреляции, его возможные значения?

6. Изобразите примеры корреляционных полей для : а) прямой корреляционной зависимости; б) обратной корреляционной зависимости; в) нелинейной корреляционной зависимости; г) отсутствия корреляционной зависимости.

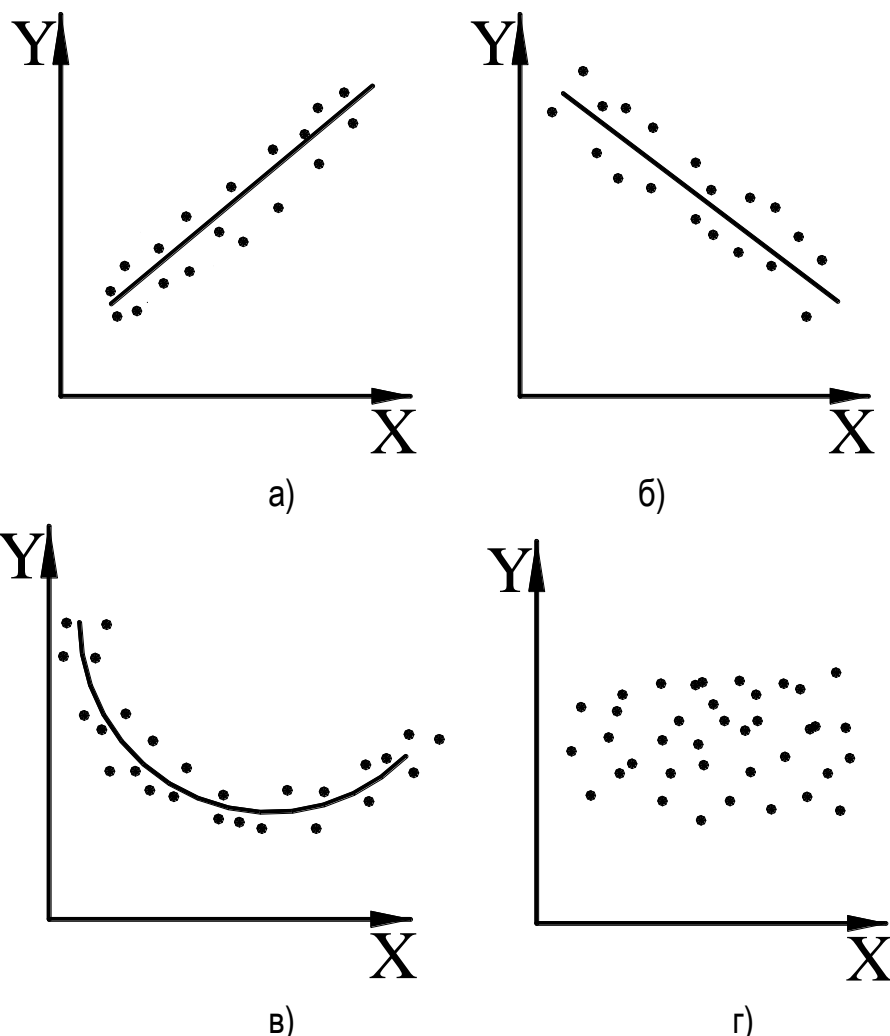


Рисунок 1.1 – Примеры корреляционных полей: а) прямая корреляционная зависимость, $r_{XY} \approx 1$; б) обратная корреляционная зависимость, $r_{XY} \approx -1$; в) нелинейная корреляционная зависимость, $r_{XY} \neq 0$; г) отсутствие корреляционной зависимости, $r_{XY} = 0$

Пример расчета

Исходные данные

Доверительная вероятность $P=0,95$

Таблица 1.1 - Результаты измерений (исходные данные из таблицы А.1)

Номер измерения i	Высота над уровнем моря, м (величина X)	Снижение эффективной мощности двигателя, % (величина Y)
1	0	0,0
2	200	4,7
3	400	6,6
4	600	8,7

Номер измерения i	Высота над уровнем моря, м (величина X)	Снижение эффективной мощности двигателя, % (величина Y)
5	800	11,1
6	1000	7,0
7	1200	9,0
8	1400	10,5
9	1600	11,1
10	1800	14,1

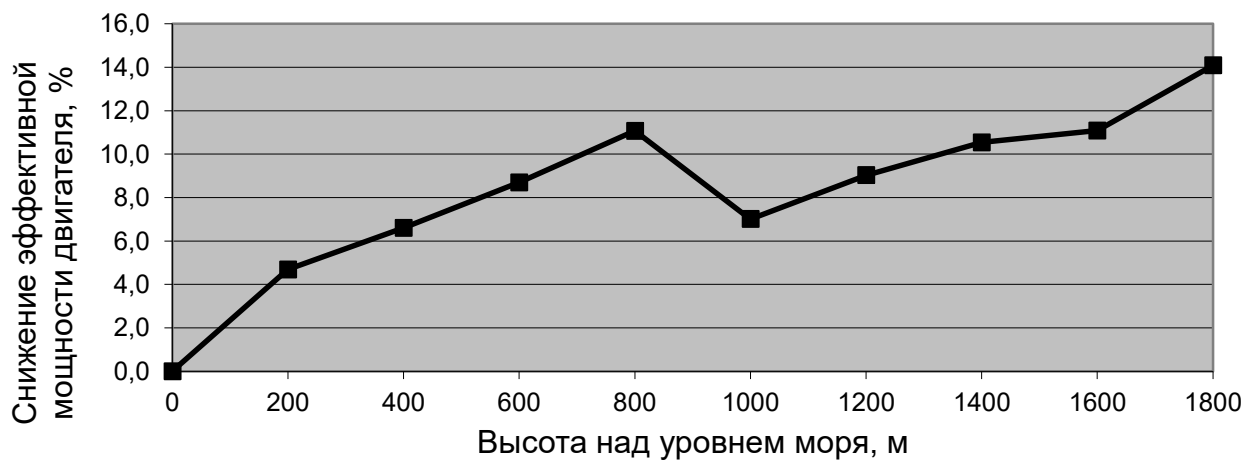


Рисунок 1.2 – Зависимость снижения эффективной мощности двигателя от высоты над уровнем моря

Порядок выполнения

1. Определите коэффициент корреляции

Эмпирический коэффициент корреляции определяется по выражению

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)\sigma_x\sigma_y},$$

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

$$\sigma_x^2 = \frac{1}{(n - 1)} \sum_{i=1}^n (x_i - \bar{x})^2;$$

$$\sigma_y^2 = \frac{1}{(n - 1)} \sum_{i=1}^n (y_i - \bar{y})^2;$$

где x_i, y_i – i -е значения случайных величин X и Y ; $\bar{x}, \bar{y}, \sigma_x, \sigma_y$, – соответственно средние значения и средние квадратические отклонения случайных величин, n – количество измерений, $n=10$.

Для определения коэффициента корреляции в табличном процессоре MS Excel используем функцию **PEARSON**, которая возвращает **коэффициент корреляции Пирсона** – безразмерный индекс в интервале от -1,0 до 1,0 включительно, который отражает степень линейной зависимости между двумя множествами данных.

Синтаксис: **PEARSON(массив1; массив2)**.

Формула имеет вид **=PEARSON(A3:A7;B3:B7)**.

Для примера коэффициент корреляции $r_{xy} = 0,87$

2. Проверьте значимость коэффициента корреляции.

Для этого определяют наблюдаемое значение критерия Стьюдента.

$$t_H = r_{xy} \cdot \sqrt{\frac{n-2}{1-r_{xy}^2}},$$

где n - число измерений. Для примера $t_H = 5,010$.

Критическое значение критерия Стьюдента t_k определяется по таблице Б.1 для заданной доверительной вероятности P и числу степеней свободы $f = n - 2$.

При выборе t_k по таблице Б.1 следует принимать $m = f + 1$, т. е.

$$f = n - 2 = 10 - 2 = 8$$

$$m = f + 1 = 8 + 1 = 9.$$

Для $P=0,95$ и $m = 9$ критическое значение критерия Стьюдента $t_k = 2,306$.

Если $|t_H| < t_k$, то коэффициент корреляции $r_{xy} = 0$ и связи между величинами нет.

Если $|t_H| > t_k$, то коэффициент корреляции значим и между величинами X и Y существует линейная корреляционная зависимость. *Сделайте вывод о наличии зависимости между X и Y .*

Вывод (для примера): Между высотой над уровнем моря (величина X), в метрах, и снижением эффективной мощности двигателя (величина Y), в процентах, существует линейная корреляционная зависимость.

Лабораторная работа №2

ПРОГНОЗИРОВАНИЕ ГРУЗОБОРОТА (ИЛИ ПАССАЖИРОБОРОТА) АВТОТРАНСПОРТНОГО ПРЕДПРИЯТИЯ С ИСПОЛЬЗОВАНИЕМ РЕГРЕССИОННОГО АНАЛИЗА

Цель: освоить методику прогнозирования грузооборота (или пассажирооборота) автотранспортного предприятия с использованием регрессионного анализа

Краткие сведения о регрессионном анализе

Задача регрессионного анализа – установление вида и параметров зависимости от-

клика y от уровней одного или нескольких факторов x_1, x_2, \dots, x_n , т.е. в результате регрессионного анализа определяют вид и параметры функции

$$y = f(x_1, x_2, \dots, x_n; b_0, b_1, \dots, b_m),$$

Данную функцию называют **регрессионной моделью**, а ее параметры b_0, b_1, \dots, b_m – *коэффициентами регрессии*.

Коэффициенты регрессии могут вычисляться с помощью *метода наименьших квадратов*, при котором коэффициенты модели определяются из условия: сумма S квадратов отклонений экспериментальных точек y_i от точек на теоретической кривой должна быть минимальной, т.е.

$$S = \sum_{i=1}^N (y_i - f(x_1, x_2, \dots, x_n; b_0, b_1, \dots, b_m))^2 \rightarrow \min,$$

где N – количество экспериментальных точек.

Регрессионная модель позволяет определять значение отклика y по заданному значению факторов x_1, x_2, \dots, x_n не проводя эксперимент. Например, регрессионный анализ может использоваться для прогнозирования грузооборота или пассажирооборота автотранспортного предприятия на перспективу.

Регрессионные модели могут быть однофакторные и многофакторные, линейные и нелинейные. Примеры регрессионных моделей:

- линейная однофакторная модель $y = b_0 + b_1 \cdot x$, где b_0, b_1 - параметры регрессионной модели (коэффициенты регрессии); x – фактор; y – отклик;
- нелинейная однофакторная модель $y = b_0 x^{b_1}$ (может использоваться для прогнозирования грузооборота или пассажирооборота);
- линейная многофакторная модель $y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3$;
- нелинейная многофакторная модель $y = b_0 + b_1 \cdot \ln x_1 + b_2 \cdot \ln x_2 + b_3 \cdot \ln x_3$.

Задание

На основании статистических данных об изменении пассажирооборота (или грузооборота) автотранспортного предприятия (АТП) за прошедшие семь лет, используя регрессионный анализ, необходимо выполнить прогноз пассажирооборота (или грузооборота) автотранспортного предприятия на заданную перспективу. Письменно ответить на контрольные вопросы. Исходные данные (таблица А.2) и шаблон отчета по лабораторной работе содержатся в файле **Лаб_раб_ОНИИД1.xls**. **Формулы нужно запрограммировать**

Содержание отчета по лабораторной работе

Тема, цель, исходные данные, распечатка отчета по лабораторной работе (файл **Лаб_раб_ОНИИД1.xls**, письменные ответы на контрольные вопросы.

Контрольные вопросы

1. В чем заключается задача регрессионного анализа? Что определяется в результате данного анализа?
2. Как определяются коэффициенты регрессии?

3. Для чего может использоваться регрессионная модель? Приведите примеры из области ТЭА.
4. Запишите вид регрессионной модели при использовании степенной зависимости между X и Y , поясните все обозначения. Запишите формулы для расчета параметров модели.
5. Запишите вид регрессионной модели при использовании полинома второй степени для описания связи между X и Y , поясните все обозначения. Запишите формулы для расчета параметров модели.
6. Что характеризует коэффициент детерминации?

Пример расчета

Для нахождения вида и параметров зависимости грузооборота по годам используем регрессионный анализ.

При выполнении регрессионного анализа будем использовать 2 вида регрессионных моделей: 1) полином второй степени; 2) степенная функция. После нахождения параметров регрессионных моделей, оценки их адекватности, определения коэффициента детерминации выбираем оптимальную модель, которая имеет наибольший коэффициент детерминации, и выполняем прогноз грузооборота (пассажиروборота) на заданный период.

Регрессионная модель в виде полинома второй степени имеет вид

$$Y_{\text{расч}} = b_0 + b_1 X + b_2 X^2,$$

где b_0, b_1, b_2 - неизвестные параметры регрессионной модели, которые нужно определить в результате регрессионного анализа; $Y_{\text{расч}}$ - грузооборот, X - годы.

Для определения параметров модели b_0, b_1, b_2 методом наименьших квадратов решают систему уравнений

$$\begin{cases} b_0 n + b_1 \sum X_i + b_2 \sum X_i^2 = \sum Y_i \\ b_0 \sum X_i + b_1 \sum X_i^2 + b_2 \sum X_i^3 = \sum (YX)_i \\ b_0 \sum X_i^2 + b_1 \sum X_i^3 + b_2 \sum X_i^4 = \sum (YX^2)_i \end{cases},$$

где $i = 1 \dots n$, n - количество лет, за которые известен грузооборот (или пассажируоборот предприятия).

Для решения системы уравнений и нахождения параметров модели b_0, b_1, b_2 можно использовать метод Гаусса, метод Крамера или другие методы.

При использовании степенной зависимости изменения грузооборота (пассажируоборота) по годам уравнение регрессии имеет вид

$$y = b_0 x^{b_1}, \tag{2.1}$$

где y - значение грузооборота (или пассажируоборота), x - годы.

Прологарифмировав выражение (2.1) получаем

$$\lg y = \lg b_0 + b_1 \lg x$$

Заменяя $X = \lg x, Y = \lg y, B_o = \lg b_o$, получаем линейную модель

$$Y = B_o + b_1 X.$$

Значения b_1 и b_o определяются из выражений

$$b_1 = \frac{N \sum_{i=1}^N \lg x_i \lg y_i - \sum_{i=1}^N \lg x_i \sum_{i=1}^N \lg y_i}{N \sum_{i=1}^N (\lg x_i)^2 - \left(\sum_{i=1}^N \lg x_i \right)^2}, \quad (2.2)$$

$$b_o = 10^{\overline{\lg y} - b_1 \overline{\lg x}}, \quad (2.3)$$

где N – количество лет, $N=7$; $\overline{\lg y}$, $\overline{\lg x}$ - среднее арифметические значения логарифмов $\lg y_i$ и $\lg x_i$.

Исходные данные

Таблица 2.1 – Данных об изменении грузооборота автотранспортного предприятия

Год	Значения грузооборота, тыс. тонно-километров,
2004	17282,20
2005	17488,44
2006	17666,13
2007	17732,5
2008	17847,55
2009	17876,96
2010	17950,11

Необходимо выполнить прогноз грузооборота на 3 года вперед.

Порядок расчета

1. Определяем параметры регрессионной модели в виде степенной функции для описания изменения грузооборота.

Для это необходимо определить значения коэффициентов регрессии b_1 и b_o . Для удобства результаты расчетов сводим в таблицу 2.2 и используем в качестве переменной X условный год, т.е. если начальный год изучения пассажирооборота (или грузооборота) 2004, ему присваивается условный год $X=1$, для 2005 года – $X=2$ и т.д. [2].

Для определения параметров регрессионной модели в MS Excel стоим по данным таблицы 2.2 точечную диаграмму (см. рис. 2.1), добавляем линию тренда, параметры линии тренда: степенная функция, показать уравнение на диаграмме, поместить на диаграмму величину достоверности аппроксимации R^2 , прогноз вперед на (указывается на какой сколько лет вперед нужно выполнить прогноз по заданию) период.

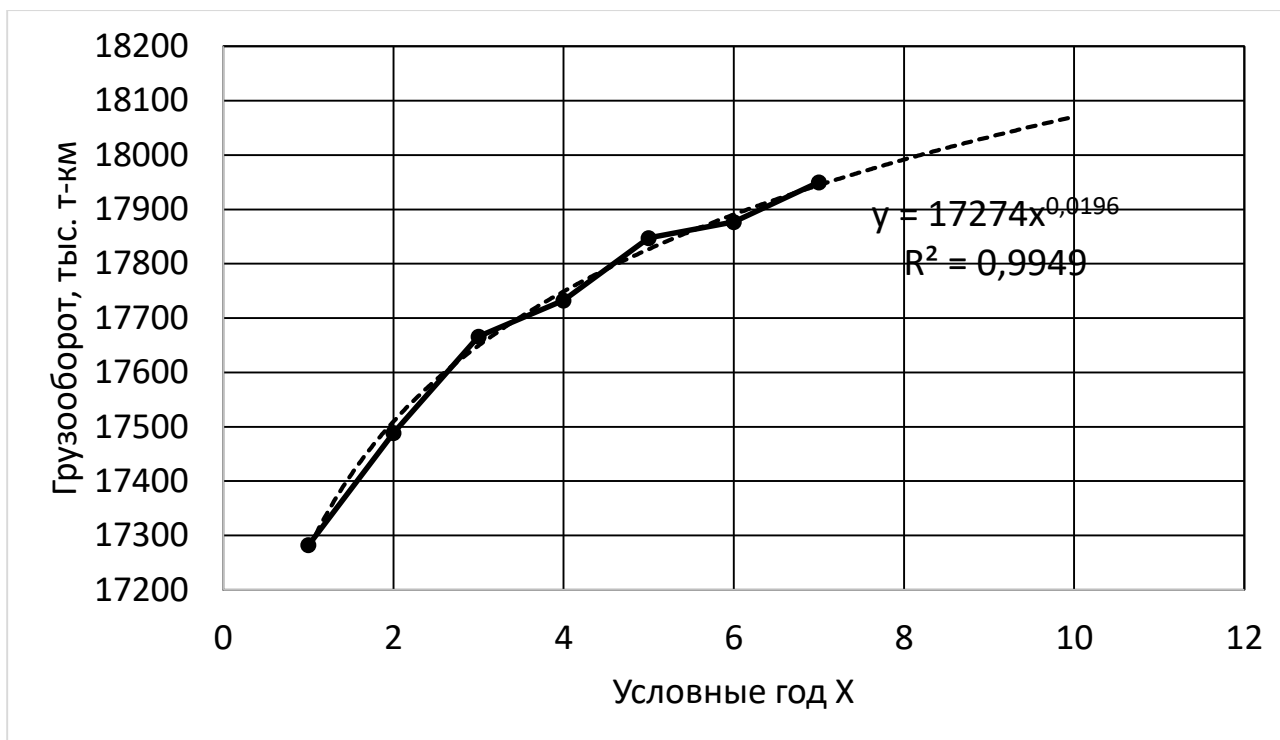


Рисунок 2.1 – Точечная диаграмма с линией тренда, уравнением степенной функции и величиной достоверности аппроксимации R^2 , показывающая зависимость грузооборота Y от условного года X

На диаграмме видно, что полученное уравнение регрессии имеет вид

$$y^T = 17274 \cdot x^{0,0196}, \quad (2.4)$$

где y^T - рассчитанное по модели значение грузооборота, тыс. тонно-километров; x - условный год (см. таблицу 2.2).

2. Проверяем адекватность полученной регрессионной модели

При проверке адекватности оценивается соответствие расчетных данные, полученные по разработанной математической модели, экспериментальным данным и можно ли использовать полученную модель для прогнозирования исследуемого явления или процесса.

Для оценки регрессионной модели используем критерий Фишера. Экспериментальное значение критерия Фишера определяется

$$F_3 = \frac{B \cdot (N - d)}{C},$$

$$B = \sum_{i=1}^N (y_i^T - \bar{y})^2,$$

$$C = \sum_{i=1}^N (y_i - y_i^T)^2,$$

где y_i, y_i^T - экспериментальное и теоретическое значение показателя (значение грузо-

оборота (или пассажирооборота), определяемое по полученной модели), d - число коэффициентов регрессии разработанной регрессионной модели (для примера $d = 2$).

Математическая модель считается адекватной результатам эксперимента и ее можно использовать для решения инженерных задач, если выполняется условие

$$F_{\vartheta} \geq F_{\alpha, \nu_1, \nu_2},$$

где F_{α, ν_1, ν_2} - критическое значение критерия Фишера для уровня значимости α и числа степеней свободы $\nu_1 = N - 1$ и $\nu_2 = N - d$. Для уровня значимости $\alpha = 0,05$; $\nu_1 = N - 1 = 7 - 1 = 6$; $\nu_2 = N - d = 7 - 2 = 5$ критическое значение критерия Фишера по [2] равно $F_{\alpha, \nu_1, \nu_2} = 4,950$. Для определения экспериментального значения критерия Фишера составим таблицу 2.2.

$$\text{Среднее значение } \bar{y} \text{ определяется } \bar{y} = \frac{\sum_{i=1}^N y_i}{N}.$$

В MS Excel нужно использовать функцию **=СРЗНАЧ(Диапазон)**, например, **=СРЗНАЧ(D104:D110)**.

Таблица 2.2 – Данные для определения параметров степенной модели и проверки ее адекватности

№ п/п i	Год	Условный год x_i	Грузооборот y_i , тыс. тонно-километров	Грузооборот по модели (2.4) y_i^T , тыс. тонно-километров	$(y_i - y_i^T)^2$	$(y_i^T - \bar{y})^2$
1	2004	1	17282,21	17273,63	73,6	175035,0
2	2005	2	17488,45	17509,74	453,5	33217,7
3	2006	3	17666,14	17649,35	281,8	1818,6
4	2007	4	17732,58	17749,08	272,5	3258,7
5	2008	5	17847,55	17826,83	429,4	18179,1
6	2009	6	17876,96	17890,60	186,2	39444,0
7	2010	7	17950,11	17944,70	29,2	63859,3
			Сумма	123843,93	1726,2	334812,4

Значения параметров $B = 334\ 812,4$; $C = 1\ 726,2$.

Экспериментальное значение критерия Фишера определяется

$$F_{\vartheta} = \frac{B \cdot (N - d)}{C} = \frac{334812,4 \cdot (7 - 2)}{1726,2} = 387,9$$

Математическая модель считается адекватной, т. к. выполняется условие

$$F_{\vartheta} \geq F_{\alpha, \nu_1, \nu_2} \\ 387,9 > 4,95.$$

3. Определяем коэффициент детерминации

Коэффициент детерминации R^2 (R-квадрат) — это доля дисперсии зависимой переменной Y , объясняемая переменной X . Его рассматривают как универсальную меру зависимости одной случайной величины от множества других. **Чем ближе значение коэффициента к 1, тем сильнее зависимость. При оценке регрессионных моделей это интерпретируется как соответствие модели данным.**

В MS Excel вместо коэффициента детерминации используется термин «**Величина достоверности аппроксимации**».

Как видно из рис. 2.1, коэффициент детерминации (величина достоверности аппроксимации) $R^2 = 0,9949$.

Коэффициент детерминации показывает, что при использовании степенной регрессионной модели 99,49% разброса значений грузооборота Y определяется годом X (т.е. зависит от года), а оставшаяся часть ($1 - 99,79\% = 0,51\%$) разброса значений грузооборота остается необъяснимой (т.е. не учитывается данной моделью). Данная высокая величина R^2 показывает сильную зависимость грузооборота Y от года X , а также хорошее соответствие модели экспериментальным данным (т.е. данным по грузообороту в зависимости от года).

4. Определяем параметры регрессионной модели в виде полинома второй степени для описания изменения грузооборота

Для это необходимо определить значения коэффициентов регрессии b_0 , b_1 и b_2 . Для определения параметров регрессионной модели в MS Excel стоим по данным таблицы 2.3 точечную диаграмму (см. рис. 2.2), добавляем линию тренда, параметры линии тренда: полиномиальная функция (степень 2), показать уравнение на диаграмме, поместить на диаграмму величину достоверности аппроксимации R^2 , прогноз вперед на (указывается на какой сколько лет вперед нужно выполнить прогноз по заданию) период.

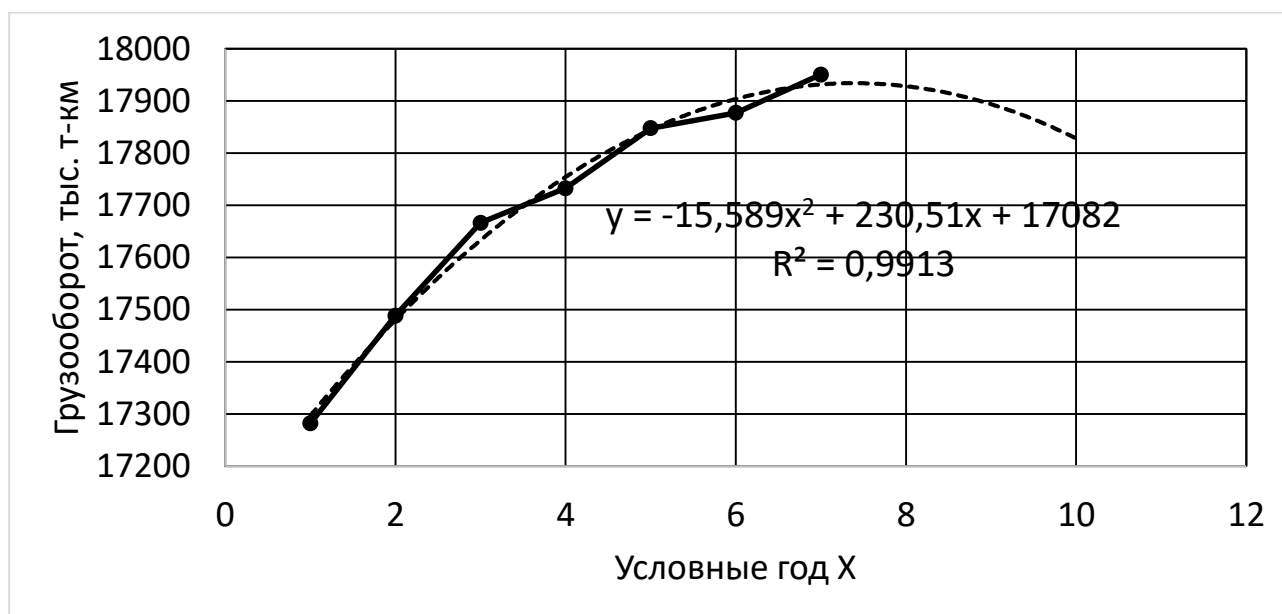


Рисунок 2.2 – Точечная диаграмма с линией тренда, уравнением полиномиальной функции (степень 2) и величиной достоверности аппроксимации R^2 , показывающая зависимость грузооборота Y от условного года X

Полученное уравнение регрессии имеет вид

$$y^T = 17082 + 230,51 \cdot x - 15,589 \cdot x^2 \quad (2.5)$$

где y^T - рассчитанное по модели значение грузооборота, тыс. тонно-километров; x - условный год (см. таблицу 2.3).

5. Проверяем адекватность полученной регрессионной модели

Для оценки регрессионной модели используем критерий Фишера.

Математическая модель считается адекватной, если выполняется условие

$$F_{\vartheta} \geq F_{\alpha, \nu_1, \nu_2},$$

где F_{α, ν_1, ν_2} - критическое значение критерия Фишера для уровня значимости α и числа степеней свободы $\nu_1 = N - 1$ и $\nu_2 = N - d$. Для полинома 2 степени $d=3$. Для уровня значимости $\alpha = 0,05$; $\nu_1 = N - 1 = 7 - 1 = 6$; $\nu_2 = N - d = 7 - 3 = 4$ критическое значение критерия Фишера по [2] равно $F_{\alpha, \nu_1, \nu_2} = 6,163$.

Для определения экспериментального значения критерия Фишера составим таблицу 2.3.

Таблица 2.3 – Данные для расчета критерия Фишера

№ п/п i	Год	Условный год x_i	Грузооборот y_i , тыс. тонно-километров	Грузооборот по модели (2.5) y_i^T , тыс. тонно-километров	$(y_i - y_i^T)^2$	$(y_i^T - \bar{y})^2$
1	2004	1	17282,21	17296,67	208,95	156289,57
2	2005	2	17488,45	17480,42	64,49	44766,40
3	2006	3	17666,14	17632,99	1099,06	3482,43
4	2007	4	17732,58	17754,37	474,85	3890,14
5	2008	5	17847,55	17844,57	8,89	23277,17
6	2009	6	17876,96	17903,58	708,66	44766,40
7	2010	7	17950,11	17931,41	349,79	57315,89
				Сумма	2914,70	333788,00

Значения параметров $C = 2914,70$; $B = 333788,00$.

Экспериментальное значение критерия Фишера определяется

$$F_{\vartheta} = \frac{B \cdot (N - d)}{C} = \frac{333788,00 \cdot (7 - 3)}{2914,70} = 458,08$$

Математическая модель считается адекватной, т. к. выполняется условие

$$F_{\vartheta} \geq F_{\alpha, \nu_1, \nu_2} \\ 458,08 > 6,163.$$

6. Определяем коэффициент детерминации

Как видно из рис. 2.2, коэффициент детерминации (величина достоверности аппроксимации) $R^2 = 0,9913$.

Полученный коэффициент детерминации показывает, что при использовании регрессионной модели в виде полинома второй степени 99,13% разброса значений грузооборота Y определяется годом X (т.е. зависит от года), а оставшаяся часть ($1 - 99,13\% = 0,87\%$) разброса значений грузооборота остается необъяснимой (т.е. не учитывается данной моделью). Данная высокая величина R^2 показывает сильную зависимость грузооборота Y от года X , а также хорошее соответствие модели экспериментальным данным (т.е. данным по грузообороту в зависимости от года). **Однако коэффициент детерминации ниже, чем у степенной регрессионной модели.**

7. Выбор оптимальной регрессионной модели.

Для выбора модели строим графики изменения грузооборота по статистическим данным (см. исходные данные) и по разработанным математическим моделям (2.4) и (2.5). Для удобства построения результаты расчета сведем в таблицу 2.4 и представим в виде графика (см. рис. 2.2).

Таблица 2.8 – Изменение грузооборота по годам

№ п/п	Год	Условный год x_i	Грузооборот y_i , тыс. тонно-километров	Грузооборот по модели (2.4) y_i^T , тыс. тонно-километров	Грузооборот по модели (2.5) y_i^T , тыс. тонно-километров
1	2004	1	17282,21	17273,6	17296,7
2	2005	2	17488,45	17509,7	17480,4
3	2006	3	17666,14	17649,4	17633,0
4	2007	4	17732,58	17749,1	17754,4
5	2008	5	17847,55	17826,8	17844,6
6	2009	6	17876,96	17890,6	17903,6
7	2010	7	17950,11	17944,7	17931,4
8	2011	8	-	17991,7	17928,0
9	2012	9	-	18033,3	17893,5
10	2013	10	-	18070,5	17827,8

Как видно из графиков моделей (см. рис.2.2), при использовании квадратичной регрессионной модели в дальнейшем грузооборот снижается, что не соответствует наблюдаемой ранее тенденции.

Из сравнения коэффициентов детерминации двух моделей следует, что больший коэффициент детерминации имеет регрессионная модель в виде степенной функции, т.е. она лучше соответствует экспериментальным данным.

Исходя из выше сказанного, в качестве регрессионной модели для прогнозирования грузооборота используем степенную функцию.

8. Выполняем прогнозирование грузооборота на заданный период (в соответствии с заданием на 3 года вперед)

Значение грузооборота на заданную перспективу (на 3 года вперед) по (2.4) равно

$$y^* = 17273,626 \cdot x^{0,020} = 17273,626 \cdot 10^{0,020} = 18070,5 \text{ тыс. тонно-километров}$$

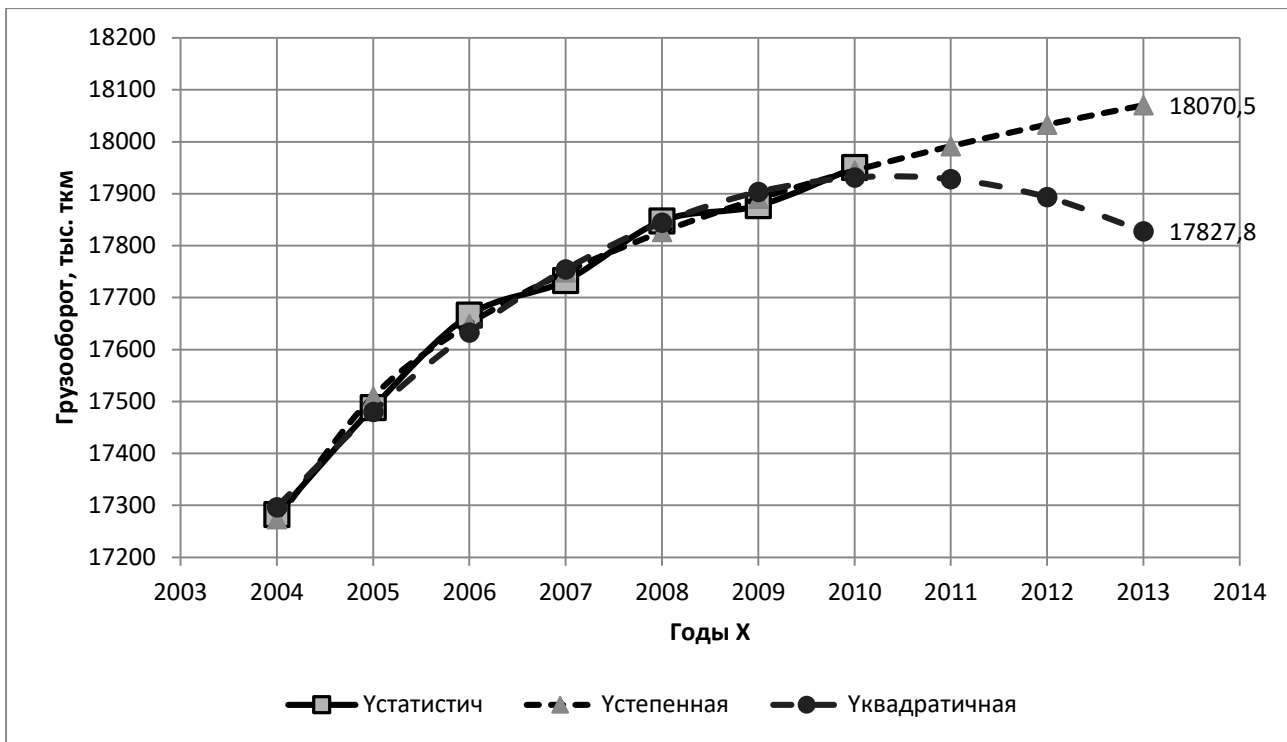


Рисунок 2.2 – Прогнозирование грузооборота предприятия

9. Выполняем оценку точности прогнозирования на основании полученной регрессионной модели [2]

Для оценки точности прогнозирования необходимо определить доверительный интервал для прогнозируемого значения y^* . Для этого определяем несмещенную оценку дисперсии y

$$D(y)_x = \frac{1}{N-2} \sum_{i=1}^N (y_i - y_i^T)^2.$$

Для примера, используя таблицу 2.2, получаем

$$D(y)_x = \frac{1}{7-2} * 1726,2 = 345,2.$$

Затем находим значение половины величины доверительного интервала разброса среднего значения

$$\Delta \bar{y} = t_{\alpha/2, N} \sqrt{\frac{D(y)}{N}},$$

где $t_{\alpha/2, N}$ - значение критерия Стьюдента, для уровня значимости $\alpha = 0,05(5\%)$ и $N=7$ по [2] определяем значение критерия Стьюдента $t_{\alpha/2, N} = 1,8946$.

Для примера $\Delta \bar{y} = 1,8946 \cdot \sqrt{\frac{345,2}{7}} = 13,3$ тонно-километров.

Далее определяем величину периода упреждения (прогноза) Π

$$\Pi = \frac{x^* \cdot N}{x_N} - N,$$

где x_N - максимальное значение фактора x , для примера x - условный год и $x_N = 7$ (см.

табл. 2.3, 2.4); N - количество точек, для примера $N=7$ (т. к. используются данные о грузообороте за 7 лет); x^* - значение фактора x на заданную перспективу для которого выполняется прогноз, для примера $x^*=10$, т.к. выполняется прогноз на 3 года вперед, т.е.
 $x^* = 7 + 3 = 10$.

Для примера величина периода прогноза $\Gamma = \frac{10 \cdot 7}{7} - 7 = 3$.

Значение половины величины доверительного интервала для прогнозируемого значения вычисляется по формуле

$$\Delta y^* = \Delta y \cdot \sqrt{\frac{N+1}{N} + \frac{3 \cdot (N+2\Gamma-1)^2}{N \cdot (N^2-1)}}.$$

Для примера $\Delta y^* = 13,3 \cdot \sqrt{\frac{7+1}{7} + \frac{3 \cdot (7+2 \cdot 3-1)^2}{7 \cdot (7^2-1)}} = 13,3 \cdot 1,56 = 20,7$ тыс. тонно-кило-

метров.

Доверительный интервал для прогнозируемого значения y^*

$$y^* - \Delta y^* < M(y) < y^* + \Delta y^*.$$

Для примера для прогнозируемого значения $y^* = 18070,5$ тыс. тонно-километров доверительный интервал равен

$$18070,5 - 20,7 < M(y) < 18070,5 + 20,7,$$

18049,8 тыс. тонно-километров $< M(y) < 18091,2$ тыс. тонно-километров.

Лабораторная работа №3

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА. ОПРЕДЕЛЕНИЕ ВИДА ЛИНЕЙНОЙ ПО ПАРАМЕТРАМ МНОГОФАКТОРНОЙ МОДЕЛИ (ПОСЛЕДОВАТЕЛЬНОЕ ПЛАНИРОВАНИЕ)

Цель: освоить методику планирования полного факторного эксперимента для построения линейной двухфакторной модели

Теоретические сведения

Эксперимент - это система операций, воздействий и (или) наблюдений, направленных на получение информации об объекте при исследовательских испытаниях. Эксперимент состоит из опытов.

Опыт - это воспроизведение исследуемого явления в определенных условиях проведения эксперимента при возможности регистрации его результатов. Условия опытов определяются **уровнями (значениями) факторов**.

Фактор - это переменная величина, по предположению влияющая на результаты эксперимента. **Уровень фактора** - это фиксированное значение фактора относительно начала отсчета.

Отклик - это наблюдаемая случайная переменная, по предположению, зависящая от факторов.

Пассивный эксперимент – это эксперимент, при котором уровни факторов в каждом опыте регистрируются исследователем, но не задаются. Пример пассивного эксперимента - это **подконтрольная эксплуатация автомобилей**. В этом случае в автотранспортном предприятии (АТП) выделяется специальная группа подконтрольных автомобилей (выборка), выполняющая обычную транспортную работу. На каждый автомобиль заводится специальный журнал, где фиксируется и накапливается информация о всех отказах и неисправностях, на каком пробеге они произошли или выявлены, данные о нагрузках, виде перевозимого груза, среднесуточных пробегах, пробегах до ТО и между ремонтами и т.п. Основным недостатком пассивного эксперимента: информация слишком «запаздывает», т.е. время обратной связи очень значительно. Например, от разработки какого-либо узла до момента поступления информации о его надежности из сферы эксплуатации проходит несколько лет.

Активный эксперимент - это эксперимент, в котором уровни факторов в каждом опыте задаются исследователем. Активный эксперимент проводится в соответствии с планом эксперимента.

Этапы проведения экспериментальных исследований

1. Формирование целей исследований
2. Выдвижение гипотезы об исследуемом объекте
3. Планирование эксперимента
4. Проведение эксперимента
5. Обработка и анализ результатов
6. Проверка правильности выдвинутой гипотезы
7. Выдвижение новой гипотезы, если необходимо
8. Проверка условий окончания эксперимента
9. Планирование нового эксперимента (при необходимости)

Планирование эксперимента – это выбор плана эксперимента, удовлетворяющего заданным требованиям. Основным требованием при планировании эксперимента является минимизация времени и числа опытов при сохранении требуемой достоверности результатов. Изменение уровней факторов происходит в соответствии с планом эксперимента.

План эксперимента - совокупность данных, определяющих число, условия и порядок реализации опытов. Исследователь в каждом опыте изменяет *уровни управляемых факторов* x_1, x_2, \dots, x_n , воздействующих на объект исследования, и регистрирует *отклик (выходной параметр)* y . В результате проведения эксперимента необходимо определить **функцию отклика** y - зависимость математического ожидания отклика от факторов x_1, x_2, \dots, x_n :

$$y = f(x_1, x_2, \dots, x_n; b_0, b_1, \dots, b_m).$$

где b_0, b_1, \dots, b_m – параметры модели.

Функцию отклика называют также *моделью регрессионного анализа или регрессионной моделью*, а b_0, b_1, \dots, b_m – коэффициентами регрессии.

Функция отклика должна иметь числовое выражение, четкий физический или экономический смысл, быть информативной (т.е. характеризовать свойства объекта) и измеряться с требуемой точностью. Факторы должны быть управляемыми, совместимыми, независимыми, однозначными и непосредственно влиять на отклик [1,2].

Планирование эксперимента обладает рядом *преимуществ* [2]:

- резко сокращается число испытаний;
- схема исследования объекта становится формализованной.
- проводится **последовательный эксперимент**, т.е. эксперимент, реализуемый в виде серий опытов, при этом условия проведения каждой последующей серии определяются результатами предыдущих
- процедура разработки математических моделей упрощается;
- точность математических моделей (их адекватность результатам эксперимента) повышается.

Планирование эксперимента в основном сводится к выбору числа уровней факторов и определению уровня каждого фактора в опыте.

План, содержащий все возможные комбинации всех факторов на определенном числе уровней, называется **полным факторным планом**, а эксперимент, проводимый в соответствии с данным планом, – **полным факторным экспериментом (ПФЭ)**. Выбранное число уровней p в сочетании с числом факторов k определяет число возможных опытов N , которое равно $N = p^k$.

Для обработки результатов эксперимента факторы **нормализуют**, т.е. преобразуют натуральные значения факторов в безразмерные. Для определения параметров линейной модели достаточно каждый фактор фиксировать на одном из двух уровней: *верхнем и нижнем* (*верхний уровень* – большее значение, *нижний* – меньшее значение). Верхний уровень нормализованного фактора обозначают «+1», нижний «-1», среднее значение – «0». Нормализация факторов выполняется по следующей формуле

$$x_i = \frac{X_i - X_{CPI}}{\Delta X_i}, \quad \Delta X_i = \frac{X_{MAXi} - X_{MINi}}{2}, \quad X_{CPI} = \frac{X_{MAXi} + X_{MINi}}{2}, \quad (3.1)$$

где x_i - нормализованное значение i -го фактора, X_i - натуральное значение i -го фактора; X_{CPI} - среднее значение i -го фактора, ΔX_i - интервал варьирования i -го фактора.

Рассмотрим планирование эксперимента на примере анализа зависимости отклика Y от двух факторов X_1 и X_2 . При последовательном эксперименте порядок модели до опыта неизвестен. На первом этапе предполагается, что модель линейна и имеет виде

$$Y = b_0 + b_1 X_1 + b_2 X_2, \quad (3.2)$$

где X_1 и X_2 – два контролируемых фактора, b_0, b_1, b_2 – параметры модели (коэффициенты регрессии).

Для двух факторов количество опытов полного факторного плана равно $N = p^k = 2^2 = 4$. Составляют **матрицу плана** – стандартную форму записи условий проведения эксперимента в виде прямоугольной таблицы, строки которой отвечают опытам, столбцы – факторам (см. табл. 3.1).

Таблица 3.1 – Матрица полного факторного эксперимента для двух факторов

№ опыта u	Нормализованные значения факторов		Значение функции отклика, полученное в ходе эксперимента, y_u
	x_1	x_2	
1	+1	+1	y_1
2	-1	+1	y_2
3	-1	-1	y_3
4	+1	-1	y_4

Графический план ПФЭ для двух факторов в факторном пространстве можно представить в виде квадрата (см. рис. 3.1).

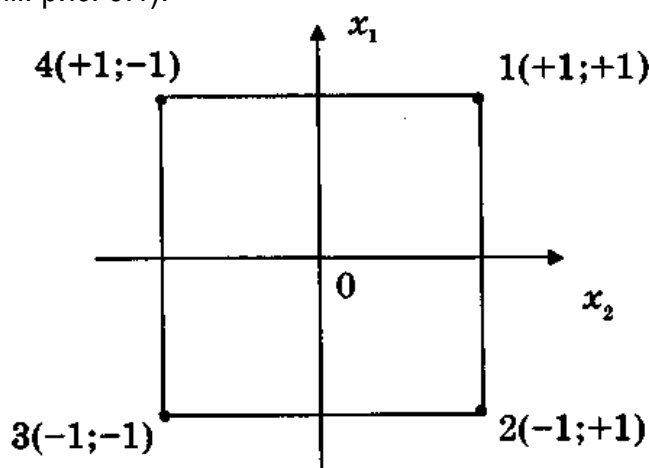


Рисунок 3.1 – Схема плана ПФЭ типа 2^2 [2]

Проводят эксперимент, определяют параметры b_0, b_1, b_2 и проверяют адекватность модели. Если модель адекватна, то заканчивают эксперимент. Адекватность модели проверяют с помощью критерия Фишера. В противном случае модель предполагается в виде

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2, \quad (3.3)$$

т. е. учитывается эффект взаимодействия факторов $X_1 X_2$, вычисляют параметры модели и проверяют ее адекватность. Коэффициент b_{12} оценивает эффект парного взаимодействия факторов X_1 и X_2 , и показывает силу влияния одного фактора в зависимости от уровня другого. Матрица планирования эксперимента с учетом эффектов взаимодействия называется расширенной матрицей планирования (см. табл. 3.2).

Таблица 3.2 – Расширенная матрица полного факторного эксперимента для двух факторов

№ опыта u	Нормализованные значения факторов			Значение функции отклика, полученное в ходе эксперимента, y_u
	x_1	x_2	$x_1 \cdot x_2$	
1	+1	+1	+1	y_1

№ опыта u	Нормализованные значения факторов			Значение функции отклика, полученное в ходе эксперимента, y_u
	x_1	x_2	$x_1 \cdot x_2$	
2	-1	+1	-1	y_2
3	-1	-1	+1	y_3
4	+1	-1	-1	y_4

Параметры нормализованной регрессионной модели определяются по формулам:

$$\begin{aligned}
 b_0 &= \frac{1}{N} \sum_{u=1}^N Y_u, \\
 b_i &= \frac{1}{N} \sum_{u=1}^N (x_i Y)_u, \\
 b_{ij} &= \frac{1}{N} \sum_{u=1}^N (x_i x_j Y)_u,
 \end{aligned} \tag{3.4}$$

где N – число опытов; i, j – номера факторов; x_i, x_j – нормализованные значения факторов; Y_u – измеренное значение отклика в u -м опыте. Например, значение коэффициента b_1 определяется

$$b_1 = \frac{(+1) \cdot y_1 + (-1) \cdot y_2 + (-1) \cdot y_3 + (+1) \cdot y_4}{4}.$$

Графическим представлением функции отклика является поверхность отклика (см. рис. 3.2).

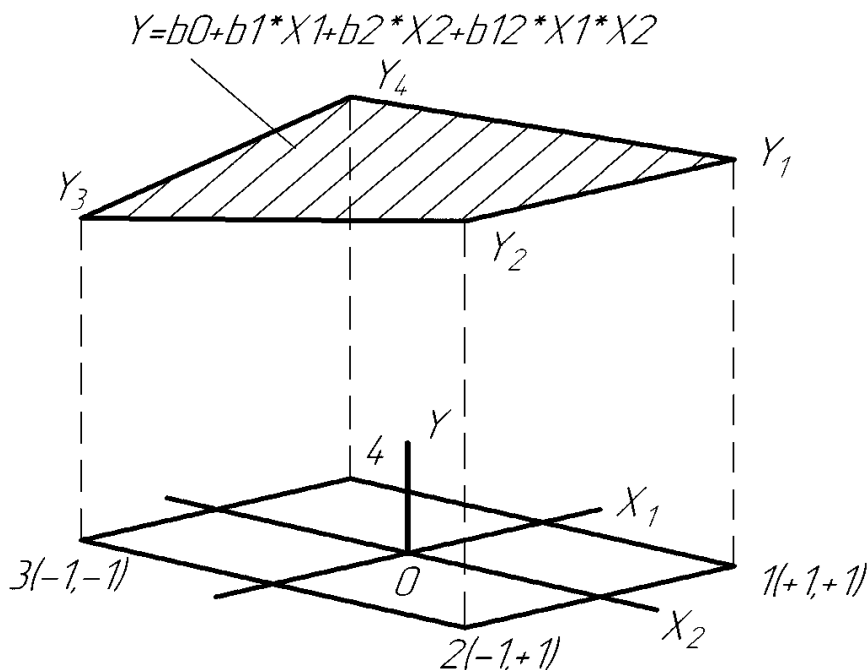


Рисунок 3.2 – Поверхность отклика $y = f(x_1, x_2)$ в факторном пространстве

Если модель адекватна, то заканчивают эксперимент. Если модель не адекватна, то модель предполагается квадратичной в виде

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2 + b_{11} X_1^2 + b_{22} X_2^2 \quad (3.5)$$

Далее составляется план для определения параметров квадратичной модели, проводятся недостающие опыты, определяются параметры модели и проверяется ее адекватность.

После определения параметров нормализованной модели переходят к натуральной модели, с помощью которой можно определить значение отклика Y по значениям факторов X_1, X_2 не проводя эксперимент.

Задание

Используя изложенную ниже методику планирования полного факторного эксперимента и табличный процессор MS Excel, определите параметры модели зависимости средней толщины h (отклик Y), осажденного на катоде металла при нанесении гальванического покрытия, от плотности тока на катоде D_k (фактор X_1) и времени нанесения покрытия t (фактор X_2), а также спрогнозировать значение толщины гальванического покрытия для заданных плотности тока и времени нанесения гальванического покрытия. Модель ищем в виде (3.3). Письменно ответить на контрольные вопросы. Исходные данные (таблица А.3) и шаблон отчета по лабораторной работе содержатся в файле **Лаб_раб_ОНИИД1.xls. Формулы нужно запрограммировать.**

Пример расчета

Таблица 3.3 - Исходные данные

Наименование процесса	<i>Износостойкое хромирование</i>
Плотность металла покрытия, г/см ³	6,9
Электрохимический эквивалент C , г/А*ч	0,324
Выход металла по току, %	13
Пределы изменения плотности тока D_k , А/дм ²	50 - 75
Пределы изменения времени нанесения покрытия t , ч	4 – 6
Дисперсия воспроизводимости S_v , мм ²	0,0001
Доверительная вероятность P	0,95
Уровень значимости $\alpha = 1 - P$	0,05
Определите толщину наносимого гальванического покрытия для	
Время нанесения t , ч	6,5
Плотности тока, А/дм ²	65

Порядок выполнения

1 Нормализуем натуральные значения факторов по формулам (3.1) и сводим в таблицу 3.4

Для фактора X_1 (плотности тока на катоде):

$$X_{CP1} = \frac{X_{MAX1} + X_{MIN1}}{2} = \frac{75 + 50}{2} = 62.5; \Delta X_1 = \frac{X_{MAX1} - X_{MIN1}}{2} = \frac{75 - 50}{2} = 12.5$$

$$x_{1MIN} = \frac{X_{1MIN} - X_{CP1}}{\Delta X_1} = \frac{50 - 62.5}{12.5} = -1; \quad x_{1MAX} = \frac{X_{1MAX} - X_{CP1}}{\Delta X_1} = \frac{75 - 62.5}{12.5} = 1$$

Таблица 3.6 – Натуральные и нормализованные значения факторов

Обозначение фактора	Минимальное значение фактора		Максимальное значение фактора		Среднее значение фактора		Интервал варьирования фактора ΔX (натуральное значение)
	Натуральное	Нормализованное	Натуральное	Нормализованное	Натуральное	Нормализованное	
X_1	50	-1	75	+1	62,5	0	12,5
X_2	4	-1	6	+1	5	0	1

2 Составляем матрицу планирования полного факторного эксперимента (ПФЭ) для 2 факторов (по аналогии с табл. 3.2).

После составления матрицы планирования ПФЭ по составленному плану проводится эксперимент, при этом в каждом опыте фактор фиксируется на заданном уровне (минимальном или максимальном). В результате проведения каждого опыта определяется значение отклика Y , т.е. значение средней толщины h (мм) гальванического покрытия. Данные значения также заносятся в матрицу планирования (см. табл.3.7).

Таблица 3.7 – План полного факторного эксперимента (факторный план)

№ опыта u	X_1 (плотность тока D_k)		X_2 (времена нанесения покрытия t)		$X_1 \cdot X_2$	Значение отклика Y (Толщина покрытия h , мм), полученное в результате опыта
	натуральное значение, A/dm^2	Нормализованное значение	натуральное значение, ч	Нормализованное значение		
1	75	1	6	1	1	0,2747
2	75	1	4	-1	-1	0,1831
3	50	-1	6	1	-1	0,1831
4	50	-1	4	-1	1	0,1221

3 Определяем коэффициенты регрессии (параметры регрессионной модели) b_0, b_1, b_2, b_3 по формулам (3.4) и их значимость.

Для коэффициента b_0 :

$$b_0 = \frac{1}{N} \sum_{u=1}^N Y_u = \frac{0.275+0.183+0.183+0.122}{4} = 0.1908$$

Для коэффициента b_1 :

$$b_1 = \frac{1}{4} \sum_{u=1}^4 (x_1 Y)_u = \frac{(+1) \cdot y_1 + (+1) \cdot y_2 + (-1) \cdot y_3 + (-1) \cdot y_4}{4} =$$

$$= \frac{(+1) \cdot 0.275 + (+1) \cdot 0.183 + (-1) \cdot 0.183 + (+1) \cdot 0.122}{4} = 0.0382.$$

Аналогично определяем значения для b_2, b_{12} : $b_2 = 0,0382$; $b_{12} = 0,0076$.

Статистическая оценка значимости коэффициентов регрессии предназначена для исключения второстепенных факторов, не влияющих на функцию отклика.

Для оценки коэффициентов регрессии сначала определяют *дисперсию воспроизводимости* эксперимента, которая характеризует разброс значений отклика при проведении параллельных (дублирующих) опытов:

$$S_y^2 = \frac{1}{N} \sum_{u=1}^N S_u^2,$$

где N - количество основных опытов (серий); S_u^2 - дисперсия, полученная при проведении параллельных опытов в каждой серии:

$$S_u^2 = \frac{\sum_{k=1}^m (y_{uk} - \bar{y}_u)^2}{m-1}, (k = 1, 2, \dots, m);$$

где y_{uk} - значение отклика в k -м опыте; \bar{y}_u - среднее значение отклика в u -ой серии; количество параллельных опытов в серии ($m=3$).

В каждом проводимом основном опыте (серии) опыт повторялся 3 раза, а затем вычислялась дисперсия воспроизводимости. По данным эксперимента дисперсия воспроизводимости равна

$$S_y^2 = 0,0001.$$

Далее находим дисперсию оценок коэффициентов регрессии

$$S_b^2 = \frac{S_y^2}{N \cdot m},$$

где N - количество опытов, m – количество параллельных опытов ($m=3$).

Для проводимого эксперимента дисперсия оценок коэффициентов регрессии

$$S_b^2 = \frac{S_y^2}{N \cdot m} = \frac{0,0001}{4 \cdot 3} = 0,0000083.$$

Доверительный интервал рассеяния коэффициентов уравнения регрессии определяют

$$\delta = S_b \cdot t(\alpha, k_2),$$

где $t(\alpha, k_2)$ - критерий Стьюдента, его принимают исходя из заданного уровня значимости $\alpha=0,05$ (т.е. вероятность ошибки 5%) и числу степеней свободы $k_2 = N \cdot (m-1)$.

Для проводимого эксперимента $k_2 = 4 \cdot (3-1) = 8$, а критерий Стьюдента $t(0,05; 8) = 2,306$.

Доверительный интервал рассеяния коэффициентов уравнения регрессии

$$\delta = t(\alpha, k_2) \cdot S_b = t(\alpha, k_2) \cdot \sqrt{S_b^2} = 2,306 \cdot \sqrt{0,0000083} = 0,0067$$

Если $|b_j| \geq \delta$, то коэффициент регрессии значим. Если $|b_j| < \delta$, то коэффициент регрессии не значим, т.е. $b_j = 0$.

По результатам эксперимент все коэффициенты значимы:

$$|0,1908| > 0,0067$$

$$|0,0382| > 0,0067$$

$$|0,0382| > 0,0067$$

$$|0,0076| > 0,0067.$$

Нормализованная регрессионная модель имеет вид

$$Y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_{12} \cdot x_{12} = 0,1908 - 0,0382 \cdot x_1 + 0,0382 \cdot x_2 + 0,0076 \cdot x_{12}.$$

4 Проверяем полученную регрессионную модель на адекватность для того, чтобы определить насколько расчетные значения соответствуют эмпирическим и можно ли использовать полученную регрессионную модель для расчета и прогнозирования значений отклика.

Находим опытное значение критерия Фишера $F_{оп}$, если $F_{оп} < F_{ТЕОР}$, то модель адекватна, и наоборот. Критерий Фишера вычисляется [1]:

$$F_{оп} = \frac{S_{yHA}^2}{S_y^2},$$

где S_{yHA}^2 - дисперсия неадекватности:

$$S_{yHA}^2 = \frac{m \sum_{u=1}^N (y_{uPACЧ} - \overline{y_{uОПЫТ}})^2}{N - n},$$

где $y_{uPACЧ}$ - значение отклика, рассчитанное по регрессионной модели для u -й комбинации факторов, $\overline{y_{uОПЫТ}}$ - среднее значение отклика, полученное в u -м опыте, n - количество факторов.

Теоретическое значение критерия Фишера $F_{ТЕОР}$ определяется по справочным таблицам [1, 2] в зависимости от уровня значимости α и числа степеней свободы $k_1 = N - n$ и k_2 .

Для расчета критерия Фишера сведем в табл. 3.8 значения отклика, полученные по регрессионной модели и полученные в ходе эксперимента. Из данной таблицы видно, что значения совпадают, поэтому считаем, что модель адекватна и расчет критерия Фишера не выполняем.

Таблица 3.8 – Сравнение значений экспериментальных и расчетных значений отклика

№ опыта, и	Значение отклика $y_{\text{ОПЫТ}}$ (толщина покрытия h , мм), полученное в результате опыта	Значение отклика $y_{\text{РАСЧ}}$ (толщина покрытия h , мм), рассчитанное по регресси- онной модели
1	0,2747	0,2747
2	0,1831	0,1831
3	0,1831	0,1831
4	0,1221	0,1221

5 Переходим к натуральной модели и выполняем прогнозирование значения отклика для заданных значений факторов, которые лежат в области экспериментирования.

Для перехода к натуральной модели в нормализованной модели нормализованные факторы заменяем на натуральные с помощью формул (3.1):

$$\begin{aligned}
 Y &= b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_{12} \cdot x_{12} = \\
 &= b_0 + b_1 \cdot \frac{X_1 - X_{CP1}}{\Delta X_1} + b_2 \cdot \frac{X_2 - X_{CP2}}{\Delta X_2} + b_{12} \cdot \frac{X_1 - X_{CP1}}{\Delta X_1} \cdot \frac{X_2 - X_{CP2}}{\Delta X_2}
 \end{aligned}$$

Для примера натуральная модель имеет вид:

$$\begin{aligned}
 Y &= 0.1908 - 0.0382 \cdot x_1 + 0.0382 \cdot x_2 + 0.0076 \cdot x_{12} = \\
 &= 0.1908 - 0.0382 \cdot \frac{X_1 - 62.5}{12.5} + 0.0382 \cdot \frac{X_2 - 5}{1} + 0.0076 \cdot \frac{X_1 - 62.5}{12.5} \cdot \frac{X_2 - 5}{1}
 \end{aligned}$$

Выполняем прогнозирование значения отклика Y (толщина покрытия h , мм) для заданных значений факторов X_1, X_2 в соответствии с исходными данными.

Выполняем расчет с использованием натуральной модели

$$\begin{aligned}
 Y &= 0.1908 - 0.0382 \cdot \frac{X_1 - 62.5}{12.5} + 0.0382 \cdot \frac{X_2 - 5}{1} + 0.0076 \cdot \frac{X_1 - 62.5}{12.5} \cdot \frac{X_2 - 5}{1} = \\
 &= 0.1908 - 0.0382 \cdot \frac{65 - 62.5}{12.5} + 0.0382 \cdot \frac{6.5 - 5}{1} + 0.0076 \cdot \frac{65 - 62.5}{12.5} \cdot \frac{6.5 - 5}{1} = \\
 &= 0.2579 \text{ мм}
 \end{aligned}$$

Содержание отчета по лабораторной работе

Тема, цель, исходные данные, распечатка отчета по лабораторной работе (файл Лаб_раб_ОНИиИД1.xls, письменные ответы на контрольные вопросы.

Контрольные вопросы

1. Дайте определение понятий: эксперимент, опыт, фактор, отклик, уровень фактора, план эксперимента, активный эксперимент, полный факторный эксперимент, матрица плана эксперимента.
2. Что такое подконтрольная эксплуатация автомобилей? К какому виду экспериментов она относится? Укажите его достоинства и недостатки.
3. Что определяется в ходе эксперимента?
4. Назовите основные требования при планировании эксперимента?
5. В чем заключается нормализация факторов при планировании эксперимента? С какой целью она проводится?
6. Запишите расширенную матрицу полного факторного эксперимента для двух факторов и регрессионную модель, параметры которой определяются по данной матрице.
7. Изобразите поверхность отклика для 2-х факторов.
8. Сколько и какие уровни факторов используются для определения параметров линейной модели? Назовите нормализованные значения уровней факторов?

Список используемых источников

1. Коваленко, Н. А. Научные исследования и решение инженерных задач в сфере автомобильного транспорта: учеб. пособие / Н. А. Коваленко. – Минск: Новое знание; М.: ИНФА-М, 2011 – 271 с.: ил.
2. Научные исследования и решение инженерных задач: Учебн. пособие/ С. С. Кучур, М. М. Болбас, В. К. Ярошевич. – Мн.: Адукацыя і выхаванне, 2003.

Учебное издание

Составители: Монтик Сергей Владимирович
Концевич Павел Сергеевич
Березуцкая Светлана Олеговна

**КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ
ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА**

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ
к лабораторным занятиям
по дисциплине
«ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ
И ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ»**

для студентов специальностей
1 - 37 01 06 «Техническая эксплуатация автомобилей»,
1 – 37 01 07 «Автосервис»

Ответственный за выпуск Монтик С.В.
Редактор

Подписано к печати ____ .2016 г. Формат 60x84/16 Бумага писчая N 1. Усл. п.л.
____. Уч. изд. л. . Заказ N ____ . Тираж 40 экз. Отпечатано на ризографе Учреждения
образования «Брестский государственный технический университет». 224017, г.
Брест, ул. Московская, 267.