**Практика\_3\_Линейная регрессия**

**Задание:**

1.Разобрать пример.

2. Придумать свою предметную область.

3. Реализовать код.

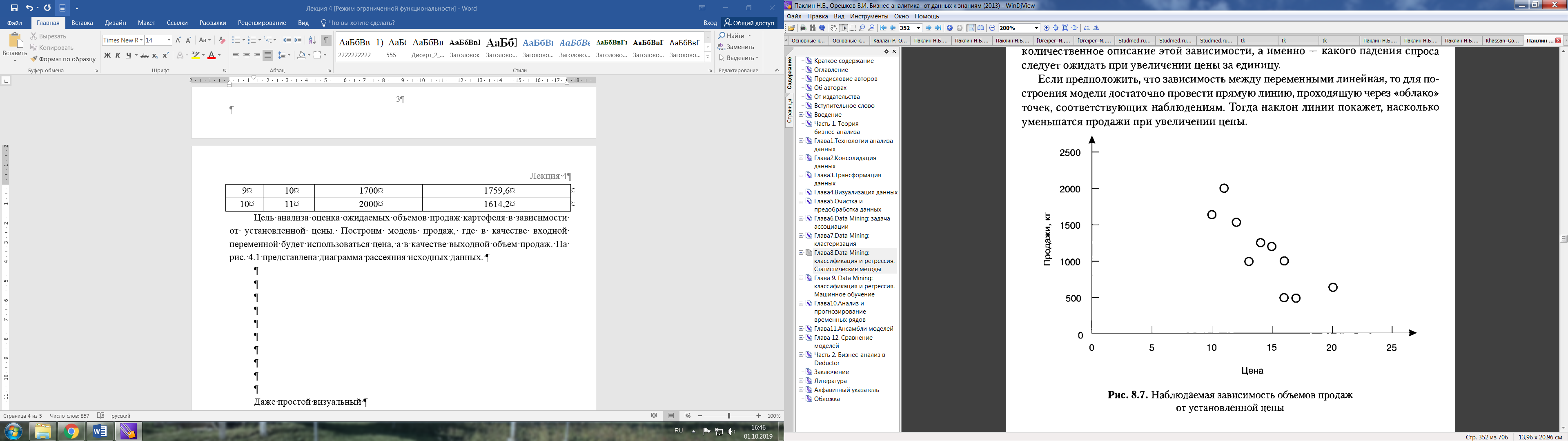
**Пример решения задачи линейной регрессии.**

Результатом собранных наблюдений явилась зависимость ежемесячных продаж картофеля от установленной цены (табл. 1.1).

*Таблица 1.1. Зависимость объема продаж от цены*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № месяца | Цена  за 1 кг, | Количество проданного картофеля, , кг | Количество проданного картофеля, оцененное с помощью регрессии, , кг | Удалось продать картофель |
| 1 | 13 | 1000 | 1323,4 | 0 |
| 2 | 20 | 600 | 305,6 | 1 |
| 3 | 17 | 500 | 741,8 | 0 |
| 4 | 15 | 1200 | 1032,6 | 1 |
| 5 | 16 | 1000 | 887,2 | 1 |
| 6 | 12 | 1500 | 1468,8 | 1 |
| 7 | 16 | 500 | 887,2 | 0 |
| 8 | 14 | 1200 | 1178,0 | 1 |
| 9 | 10 | 1700 | 1759,6 | 0 |
| 10 | 11 | 2000 | 1614,2 | 1 |

Построим модель продаж, где в качестве входной переменной будет использоваться цена, а в качестве выходной объем продаж. На рис. 1.1 представлена диаграмма рассеяния исходных данных.



*Рисунок 1.1. Наблюдаемая зависимость объемов продаж от установленной цены*

Строем прямую (линию) через точки по формуле. На практике линию строят так, чтобы сумма квадратов отклонений наблюдаемых значений от оцененных с помощью данной линейной зависимости была минимальной, то есть:

(1.1)

где – число наблюдений;

– оценка выходного значения для -гo наблюдения, полученная с помощью модели;

– реально наблюдаемое значение объема продаж.

**Линия регрессия** – это прямая наилучшего приближения для множества пар значений входной и выходной переменной (), выбираемая таким образом, чтобы сумма квадратов расстояний от точек () до этой прямой, измеренных вертикально (то есть вдоль оси ), была минимальна, Уравнение, описывающее линию регрессии, называется уравнением регрессии :

+e (1.2)

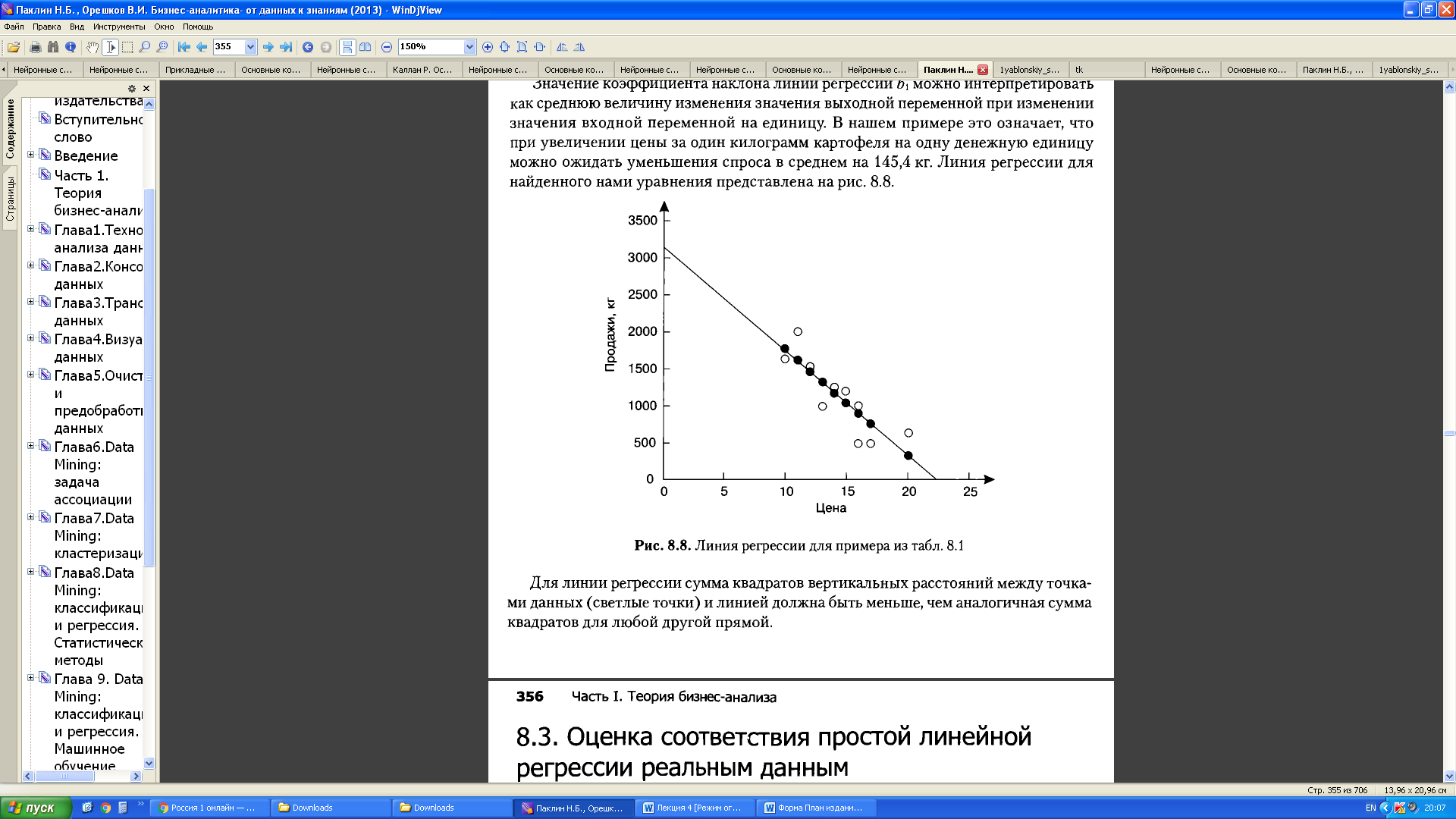
где – оценка значения выходной переменной;

- коэффициент, определяющий точку пересечения линии с осью , называемый также свободным членом Коэффициент определяет наклон линии относительно оси (иногда ero называют *угловым коэффициентом*).

– это величина, на которую изменяется значение выходной переменной при изменении входной переменной на единицу.

e-ошибка.

Коэффициенты линейного уравнения и называются *коэффициентами регрессии*



Система нормальных уравнений.

Для наших данных система уравнений имеет вид

10 b0+144·b1=11200 (1.3)

144·d0 + 2156·b1 = 149300 (1.4)

Решим систему уравнений выразим b0 из формулы (1.3):

b0 = (1.5)

Подставим в формулу (1.4) выражение b0 =

Получим:

161280 – 2073,6 b1 + 2156 b1 + 149300

-2073,6 b1 + 2156 b1 = 149300 – 161280

82,4 b1 = -11980

b1 =

Подставим коэффициент b1 в формулу (1.5) чтобы получить коэффициент b0

b0 =

Получили коэффициенты регрессии:

b0 = 3213,58

b1 = -145,388

Подставим коэффициенты b0 и b1 в уравнение регрессии в формулы (1.2), получим:

(1.6)

Оценку значения получаем из выражения (1.6).

*А также коэффициенты b0 и b1 можно найти другим способом. Они минимизируются , путем дифференцирования уравнения (1.2) по b0 и b1. Можно использовать метод градиентного спуска или метод Ньютона.*

Стандартная ошибка равна корню квадратному среднеквадратической ошибки (), то есть сумме квадратов разностей между реальным и оцененным значениями, вычисленной по всем наблюдениям и отнесенной к числу степеней свободы выборки [1]:

(1.7)

где – количество независимых переменных, которое для простой линейной регрессии равно 1.

можно рассматривать как меру изменчивости выходной переменной, объясняемую регрессией.

В то время стандартная ошибка оценивания ориентируются следующим способом [1]:

(1.8)

Найдем из формулы (1.7) значения:

= + + + + + + + + + = 593 989,28

= 74248,66

Найдем стандартную ошибку из формулы (1.8)

= = 272,4860730

Простая линейная регрессионная модель задается следующим образом. Пускай существует подборка сведений, включающая исследований, в любом из которых значению самостоятельной величины соответствует зависимой величине , сопряженных с помощью линейной связи:

(1.9)

где и – параметры модели, определяющие точку пересечения линии регрессии с осью и наклон линии регрессии соответственно;

– член, определяющий ошибку отклонения реального наблюдения от оценки, полученной с помощью данной модели.

Подставим наши значение b0, b1 и ошибку отклонения в формулу (1.9)

Получим:

)

*Найдем коэффициент корреляции*

Еще одной мерой, используемой для количественного описания линейной зависимости между двумя числовыми переменными, является коэффициент корреляции, который определяется следующим образом:

(1.9)

где и стандартные отклонения соответствующих переменных. Значение коэффициента корреляции всеrда расположено в диапазоне от до

Вычисление коэффициента корреляции по формуле (1.9)

Если коэффициент корреляции близок к 1, то между переменными имеет место сильная положительная корреляция. Иными словами, наблюдается высокая степень зависимости входной и выходной переменных (если значения входной переменной возрастают, то и значения выходной переменной также будут увеличиваться).