

**Министерство науки и высшего образования Российской Федерации**  
**Федеральное государственное бюджетное образовательное учреждение**  
**высшего образования**  
**Санкт-Петербургский горный университет**

**Кафедра информатики и компьютерных технологий**

# **ИНФОРМАТИКА**

## **ПОСТРОЕНИЕ ЭМПИРИЧЕСКИХ ФОРМУЛ В ЗАДАЧАХ ПРИКЛАДНОЙ ГЕОЛОГИИ**

*Методические указания к курсовой работе  
для студентов специальности 21.05.02*

**САНКТ-ПЕТЕРБУРГ**  
**2021**

УДК 004.67 (073)

**ИНФОРМАТИКА. Построение эмпирических формул в задачах прикладной геологии:** Методические указания к курсовой работе / Санкт-Петербургский горный университет. Сост.: *В.В. Беляев, Е.Н. Овчинникова*. СПб, 2021. 39 с.

Изложена технология построения эмпирических формул методом наименьших квадратов для аппроксимации экспериментальных данных. Даны общие указания по выполнению и оформлению курсовой работы, приведены примеры построения эмпирических формул в табличном процессоре MS Excel и математическом пакете MathCAD.

Предназначены для студентов специальности 21.05.02 «Прикладная геология», специализация «Геологическая съемка, поиски и разведка твердых полезных ископаемых» и «Геология нефти и газа».

Научный редактор доц. *А.Б. Маховиков*

Рецензент к.т.н. *К.В. Столяров* («Telum Inc»)

© Санкт-Петербургский  
горный университет, 2021

## **ИНФОРМАТИКА**

### **ПОСТРОЕНИЕ ЭМПИРИЧЕСКИХ ФОРМУЛ В ЗАДАЧАХ ПРИКЛАДНОЙ ГЕОЛОГИИ**

*Методические указания к курсовой работе  
для студентов специальности 21.05.02*

Сост.: *В.В. Беляев, Е.Н. Овчинникова*

Печатается с оригинал-макета, подготовленного кафедрой  
информатики и компьютерных технологий

Ответственный за выпуск *В.В. Беляев*

Лицензия ИД № 06517 от 09.01.2002

Подписано к печати 24.05.2021. Формат 60×84/16.

Усл. печ. л. 2,2. Усл.кр.-отт. 2,2. Уч.-изд.л. 2,0. Тираж 75 экз. Заказ 448.

Санкт-Петербургский горный университет  
РИЦ Санкт-Петербургского горного университета  
Адрес университета и РИЦ: 199106 Санкт-Петербург, 21-я линия, 2

## ВВЕДЕНИЕ

В соответствии с действующим учебным планом, студентам специальности 21.05.02 «Прикладная геология» во втором семестре по дисциплине «Информатика» необходимо выполнить курсовую работу.

Целью курсовой работы «Построение эмпирических формул в задачах прикладной геологии» является углубление знаний по информатике, развитие и закрепление навыков работы в табличном процессоре MS Excel, математическом пакете (SMathStudio, MathCAD) и в средах программирования; применение полученных навыков для решения задач из предметной области, связанной с геолого-геофизическими исследованиями.

Отчет по курсовой работе оформляется в виде пояснительной записки. Порядок изложения материала следующий:

- титульный лист;
- задание на курсовую работу;
- аннотация на русском и английском языке;
- оглавление;
- введение;
- теоретические сведения по теме курсовой работы;
- результаты расчетов в табличном процессоре MS Excel с построением линий трендов;
- результаты расчетов в математическом пакете с построением графиков;
- результаты расчетов в среде программирования;
- заключение;
- библиографический список.

При выдаче задания на курсовую работу устанавливаются сроки выполнения ее отдельных этапов, прохождение которых контролируется руководителем. Последовательное выполнение курсовой работы способствует формированию навыков проведения любого научного исследования.

Данные методические указания включают краткие теоретические сведения по теме курсовой работы, подробное описание выполнения заданий, а также варианты заданий для самостоятельного выполнения.

## ПОСТРОЕНИЕ ЭМПИРИЧЕСКИХ ФОРМУЛ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

При анализе эмпирических данных часто возникает необходимость найти в явном виде функциональную зависимость между величинами  $x$  и  $y$ , полученными в результате опытных измерений.

При аналитическом исследовании взаимосвязи между двумя величинами  $x$  и  $y$  производят ряд наблюдений, в результате которых получают таблицу значений:

$x$	$x_1$	$x_2$	$\dots$	$x_i$	$\dots$	$x_n$
$y$	$y_1$	$y_2$	$\dots$	$y_i$	$\dots$	$y_n$

Данная таблица обычно формируется в ходе каких-либо измерений, в которых величины  $x_i$  (независимые величины) задаются исследователем. Значения  $y_i$  получаются в результате эксперимента, поэтому их называют *эмпирическими* или опытными значениями.

Для установления зависимости между величинами  $x$  и  $y$  (аналитический вид ее обычно неизвестен) необходимо решить практически важную задачу – найти эмпирическую формулу:

$$y = f(x; a_1, a_2, \dots, a_m), \quad (1)$$

где  $a_1, a_2, \dots, a_m$  – неизвестные параметры.

Функция (1) обычно выбирается из класса линейных, степенных или показательных функций.

Значения параметров  $a_1, a_2, \dots, a_m$  определяются таким образом, чтобы вычисленные по формуле (1) теоретические значения  $y_i^T = f(x_i; a_1, a_2, \dots, a_m)$  при  $x = x_i$  как можно меньше отличались бы от опытных значений  $y_i$  ( $i = 1, 2, \dots, n$ ).

Согласно *методу наименьших квадратов* (МНК), наилучшими коэффициентами  $a_1, a_2, \dots, a_m$  считаются те, для которых сумма квадратов отклонений найденной эмпирической функции от заданных значений будет минимальной:

$$S(a_1, a_2, \dots, a_m) = \sum_{i=1}^n [f(x_i; a_1, a_2, \dots, a_m) - y_i]^2 \rightarrow \min \quad (2)$$

Поясним геометрический смысл метода наименьших квадратов.

Каждая пара чисел  $(x_i, y_i)$  из исходной таблицы определяет точку  $M_i$  на плоскости  $XOY$ . Используя формулу (1) при различных значениях коэффициентов  $a_1, a_2, \dots, a_m$ , можно построить ряд кривых, которые являются графиками функции (1).

Задача состоит в определении коэффициентов  $a_1, a_2, \dots, a_m$  таким образом, чтобы сумма квадратов расстояний по вертикали от точек  $M_i(x_i, y_i)$  до графика функции (1) была наименьшей.

Разность  $y_i - y_i^T$  называется *отклонением* или *невязкой* в  $i$ -ой точке. Невязка  $d_i$  равна расстоянию по вертикали от точки  $M_i(x_i, y_i)$  до точки  $(x_i, y_i^T)$  на графике эмпирической функции. Геометрический смысл невязки в  $i$ -ой точке показан на рис. 1.

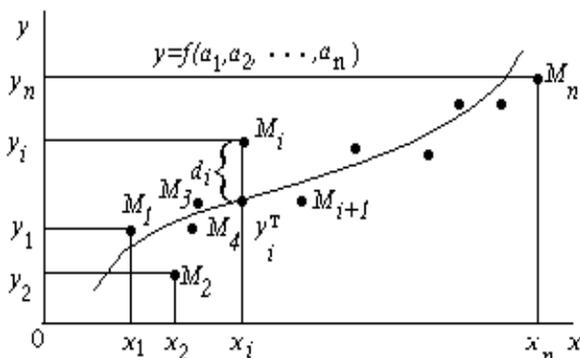


Рис. 1. Фактические данные и график эмпирической функции

Таким образом, построение эмпирической формулы (1) состоит из двух этапов: выяснение общего вида этой формулы и определение ее наилучших параметров  $a_1, a_2, \dots, a_m$ .

На первом этапе выбирается эмпирическая формула аппроксимирующей функции. В общем случае аппроксимация (от латинского «*approximate*» – «приближаться») означает приближенное описание эмпирических данных с помощью аналитических формул. Успешный выбор эмпирической формулы в значительной мере зависит от знаний исследователя в предметной области, используя ко-

торые он может правильно выбрать класс теоретической функции (например, линейный, степенной, показательный или др.).

Далее определяются наилучшие значения коэффициентов  $a_1, a_2, \dots, a_m$ , входящих в эмпирическую формулу. Для этого применяют известные аналитические методы, в частности, метод наименьших квадратов.

Согласно методу наименьших квадратов, для нахождения набора коэффициентов  $a_1, a_2, \dots, a_m$ , которые доставляют минимум функции  $S$ , определяемой формулой (2), используется необходимое условие экстремума функции нескольких переменных – равенство нулю частных производных. В результате получают систему уравнений для определения коэффициентов  $a_i$  ( $i = 1, 2, \dots, m$ ):

$$\frac{\partial S}{\partial a_1} = 0; \frac{\partial S}{\partial a_2} = 0; \dots; \frac{\partial S}{\partial a_m} = 0 \quad (3)$$

Таким образом, нахождение коэффициентов  $a_i$  сводится к решению системы (3).

Конкретный вид системы линейных уравнений для нахождения коэффициентов  $a_i$  зависит от того, из какого класса эмпирических формул мы ищем зависимость (1).

В случае выбора *линейной аппроксимирующей зависимости* вида  $y = a_1 + a_2x$  система (3) примет следующий вид:

$$\begin{cases} a_1 n + a_2 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases} \quad (4)$$

Данная система линейных уравнений может быть решена любым известным методом (методом обратной матрицы, методом Гаусса, формулами Крамера, методом подстановки и др.).

В случае *квадратичной аппроксимирующей зависимости* вида  $y = a_1 + a_2x + a_3x^2$  система (3) примет вид:

$$\left\{ \begin{array}{l} a_1 n + a_2 \sum_{i=1}^n x_i + a_3 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i \\ a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 + a_3 \sum_{i=1}^n x_i^3 = \sum_{i=1}^n x_i y_i \\ a_1 \sum_{i=1}^n x_i^2 + a_2 \sum_{i=1}^n x_i^3 + a_3 \sum_{i=1}^n x_i^4 = \sum_{i=1}^n x_i^2 y_i \end{array} \right. \quad (5)$$

В случае *экспоненциальной зависимости* аппроксимирующая функция имеет вид:

$$y = a_1 \cdot e^{a_2 x} \quad (6)$$

В этом случае нужно линеаризовать уравнение (6) с помощью логарифмирования, после чего получим соотношение:

$$\ln y = \ln a_1 + a_2 x \quad (7)$$

Обозначим  $\ln y_i$  и  $\ln a_1$  через  $z$  и  $c$  соответственно, тогда зависимость (6) может быть записана в виде  $z = c + a_2 x$ , что позволяет применить систему (4) для определения параметров  $c$  и  $a_2$ :

$$\left\{ \begin{array}{l} nc + a_2 \sum_{i=1}^n x_i = \sum_{i=1}^n z_i \\ c \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i z_i \end{array} \right. \quad (8)$$

Или, возвращаясь к табличным эмпирическим данным, получим систему линейных уравнений:

$$\left\{ \begin{array}{l} nc + a_2 \sum_{i=1}^n x_i = \sum_{i=1}^n \ln y_i \\ c \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i \ln y_i \end{array} \right. \quad (9)$$

## ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИИ

График теоретической функциональной зависимости  $y^T(x)$ , полученный по эмпирическим формулам, называется **кривой регрессии**. Для проверки согласия (справедливости) построенной кривой регрессии с результатами эксперимента, как правило, используют следующие числовые характеристики: коэффициент корреляции и коэффициент детерминированности.

**Коэффициент корреляции** является мерой линейной связи между зависимыми случайными величинами. Он показывает, насколько хорошо, в среднем, может быть представлена (вычислена) одна из величин в виде линейной функции от другой.

Коэффициент корреляции вычисляется по формуле:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (10)$$

где  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$  и  $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$  – среднее арифметическое значение по  $x$  и  $y$  соответственно.

Коэффициент корреляции между случайными величинами по абсолютной величине не превосходит единицу. Чем ближе  $|\rho|$  к 1, тем теснее линейная связь между  $x$  и  $y$ , тем более целесообразна аппроксимация таблично заданной функции линейной зависимостью.

Особо подчеркнем, что если  $|\rho|$  существенно меньше 1, то это не означает отсутствие вообще зависимости между величинами  $x$  и  $y$ . Просто в данном случае линейная аппроксимация не применима, но можно искать аппроксимирующую зависимость среди экспоненциальных, квадратичных и функций других видов.

Вторая числовая характеристика – **коэффициент детерминированности** – позволяет выяснить, насколько точно полученная теоретическая функция описывает взаимосвязь между эмпирическими данными.

Для описания коэффициента детерминированности (детерминации) рассмотрим следующие величины:

$$S_{\text{общ}} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad - \text{общая сумма квадратов отклонений}$$

экспериментальных значений  $y_i$  от  $\bar{y}$  (среднего значения по  $y$ ).

$$S_{\text{ост}} = \sum_{i=1}^n (y_i - y_i^T)^2 \quad - \text{остаточная сумма квадратов (характеризует суммарное отклонение эмпирических данных от теоретических значений, найденных по уравнению регрессии).}$$

квардратов отклонений эмпирических данных от теоретических значений, найденных по уравнению регрессии).

$$S_{\text{рег}} = \sum_{i=1}^n (y_i^T - \bar{y})^2 \quad - \text{регрессионная или факторная сумма}$$

квардратов отклонений (характеризует разброс теоретических значений относительно среднего арифметического значения наблюдаемого значения  $\bar{y}$ ).

Коэффициент детерминации определяется по формуле:

$$R^2 = 1 - \frac{S_{\text{ост}}}{S_{\text{общ}}} \quad (11)$$

Для линейной регрессии можно доказать следующее равенство:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - y_i^T)^2 + \sum_{i=1}^n (y_i^T - \bar{y})^2 \quad (12)$$

Очевидно, что справедливо следующее равенство:

$$S_{\text{общ}} = S_{\text{ост}} + S_{\text{рег}}$$

Поскольку  $S_{\text{ост}} \leq S_{\text{общ}}$ , то  $R^2$  может изменяться в пределах от 0 до 1. Чем меньше остаточная сумма квадратов  $S_{\text{ост}}$  по сравнению с общей суммой квадратов  $S_{\text{общ}}$ , тем больше значение коэффициента детерминированности.

Если  $R^2$  равен единице, то экспериментальные точки лежат на графике эмпирической функции. Если  $R^2$  близок к единице, то уравнение регрессии хорошо описывает фактическую взаимосвязь

между экспериментальными данными и может быть использовано в дальнейшем для анализа и расчетов. В противоположном случае, когда коэффициент детерминированности близок к нулю, выбранная эмпирическая формула неудачна, и уравнение регрессии нецелесообразно использовать в качестве аппроксимирующей функции.

Можно доказать, что в случае линейной зависимости двух переменных коэффициент детерминированности равен квадрату коэффициента корреляции ( $R^2 = \rho^2$ ).

Коэффициент детерминированности служит показателем тесноты связи между переменными. Показателям тесноты связи можно дать качественную оценку по *шкале Чеддока* (таблица 1):

Таблица 1

Шкала Чеддока

Количественная мера тесноты связи	Качественная характеристика силы связи
0,1-0,3	Слабая
0,3-0,5	Умеренная
0,5-0,7	Заметная
0,7-0,9	Высокая
0,9-0,99	Весьма высокая

Для обоснованного применения уравнения регрессии необходимо оценить полученные характеристики уравнения регрессии.

Большое значение имеет установление *статистической значимости* коэффициента детерминированности и параметров полученного уравнения, то есть оценка вероятности того, что данные величины не примут нулевые значения.

Следует помнить, что если уравнение в целом не значимо, то использовать данное уравнение для прогноза не корректно – полученные результаты не заслуживают доверия. При анализе значимости коэффициентов существенным моментом является значимость «старшего» коэффициента ( $a_3$  - для квадратичной зависимости,  $a_2$  - для линейной и экспоненциальной). Если для квадратичной зависимости  $a_3$  не значим, а  $a_2$  – значим, то зависимость вырождается в

линейную. Если для линейной зависимости  $a_2$  не значим, а  $a_1$  - значим, то линейной зависимости фактически нет ( $y(x) = a_1$ ).

Проверка значимости уравнения в целом, то есть гипотезы о наличии линейной зависимости между  $y$  и  $x$ , проводится с помощью **критерия Фишера**.

Использование критерия Фишера предполагает вычисление  $F_{\text{набл}}$  и его сравнение с табличным значением  $F_{\text{табл}}$ , которое зависит от уровня значимости  $\alpha$  (для многих технических приложений эта величина принимается равной 0.05,) и числа степеней свободы для факторной (регрессионная) и остаточной сумм.  $F_{\text{табл}}$  определяется либо с помощью статистических таблиц, либо с использованием специализированных пакетов программ, например, в MS Excel для этого может быть использована функция **F.ОБР.ПХ()**<sup>1</sup>.

Если  $F_{\text{набл}} > F_{\text{табл}}$ , делается вывод о значимости уравнения регрессии (о существенности связи между  $y$  и  $x$ ). Если же величина  $F_{\text{набл}}$  окажется меньше табличной, то есть  $F_{\text{набл}} < F_{\text{табл}}$ , то уравнение регрессии считается статистически незначимым (линейной связи между  $y$  и  $x$  нет).

Для проверки существенности в целом уравнения нелинейной регрессии  $F$ -критерий Фишера может быть вычислен через коэффициент детерминированности по формуле:

$$F_{\text{набл}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}, \quad (13)$$

где  $R^2$  – коэффициент детерминированности;  $n$  – число наблюдений;  $m$  – число параметров при переменной  $x$  (для линейной и экспоненциальной аппроксимации  $m=1$ , для квадратичной аппроксимации  $m=2$ ).

Величина  $m$  характеризует число степеней свободы для регрессионной (факторной) суммы квадратов, а  $(n - m - 1)$  – число степеней свободы для остаточной суммы квадратов.

---

<sup>1</sup>Начиная с MS Office 2010. Для более ранних версий используется функция **FRASПОБР()**.

Для линейной и экспоненциальной функций формула  $F$ -критерия примет вид:

$$F_{набл} = \frac{R^2}{1-R^2} \cdot \frac{n-2}{1}. \quad (14)$$

Формула  $F$ -критерия для квадратичной аппроксимации:

$$F_{набл} = \frac{R^2}{1-R^2} \cdot \frac{n-3}{2}. \quad (15)$$

При выборе уравнения регрессии обычно оценивается значимость не только уравнения в целом, но и отдельных его параметров. Проверка значимости параметров уравнения регрессии (коэффициентов уравнения регрессии) проводится с помощью **критерия Стьюдента**.

Так для линейной регрессии с этой целью по каждому из параметров определяется его стандартная ошибка  $s_{a_1}$  и  $s_{a_2}$ .

Стандартная ошибка (средняя квадратичная погрешность) коэффициента регрессии  $a_2$  определяется по формуле:

$$s_{a_2} = \sqrt{\frac{D_{ocm}}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (16)$$

где  $D_{ocm}$  – средний квадрат отклонений  $S_{ocm}$  (не смещенная оценка дисперсии).  $D_{ocm}$  может быть вычислена по формуле:

$$D_{ocm} = \frac{S_{ocm}}{df_{ocm}} = \frac{\sum_{i=1}^n (y_i - y_i^T)^2}{n - m - 1}, \quad (17)$$

где  $m = 1$  для линейной и экспоненциальной зависимостей и  $m = 2$  для квадратичной аппроксимации.

Для оценки существенности коэффициента регрессии  $a_2$  его величина сравнивается с его стандартной ошибкой. Т.е. определяется фактическое значение  $t$ -критерия Стьюдента  $t_{a_2} = \frac{|a_2|}{s_{a_2}}$ , которое затем сравнивается с табличным значением  $t_{табл}$  при определенном уровне значимости  $\alpha$  (например, 0.05) и числе степеней свободы  $n - m - 1$ .

Табличное значение  $t_{табл}$  определяется либо с помощью статистических таблиц, либо с использованием специализированных пакетов программ, например, в MS Excel для этого может быть использована функция **СТЮДЕНТ.ОБР.2X0**<sup>2</sup>.

Если  $t_{a_2} > t_{табл}$ , делается вывод, что коэффициент  $a_2$  значим.

Стандартная ошибка параметра  $a_1$  определяется по формуле:

$$s_{a_1} = \sqrt{D_{осм} \cdot \frac{\sum_{i=1}^n (x_i)^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}. \quad (18)$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии  $a_2$ : вычисляется фактическое значение  $t$ -критерия Стьюдента  $t_{a_1} = \frac{|a_1|}{s_{a_1}}$ , которое затем сравнивается с табличным значением  $t_{табл}$  при определенном уровне значимости  $\alpha$  (например, 0,05) и числе степеней свободы  $n-2$ .

Если  $t_{a_1} > t_{табл}$ , делается вывод, что коэффициент  $a_1$  значим.

Оценку значимости коэффициентов экспоненциальной аппроксимации производят по тем же формулам (16), (18), только при

---

<sup>2</sup> Начиная с MS Office 2010. Для более ранних версий используется функция **СТЮДРАСПОБР()**.

вычислении  $D_{\text{ост}}$  в формуле (17) числитель необходимо заменить на

$$\sum_{i=1}^n [\ln y_i - (\ln y_i)^T]^2.$$

Для оценки значимости квадратичной зависимости используется аналогичный подход. Значения стандартных ошибок вычисляются по формулам:

$$s_{a_1} = \sqrt{D_{\text{ocm}} \left( \frac{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n x_i^4 - \sum_{i=1}^n x_i^3 \cdot \sum_{i=1}^n x_i^3}{|A|} \right)}; \quad (19)$$

$$s_{a_2} = \sqrt{D_{\text{ocm}} \left( \frac{n \cdot \sum_{i=1}^n x_i^4 - \sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n x_i^2}{|A|} \right)}; \quad (20)$$

$$s_{a_3} = \sqrt{D_{\text{ocm}} \left( \frac{n \cdot \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n x_i}{|A|} \right)}, \quad (21)$$

где  $|A| = \det(A)$ ;

$$A = \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{pmatrix}. \quad (22)$$

Заметим, что выражения, стоящие в скобках в формулах (19), (20) и (21) равны диагональным элементам обратной матрицы  $A^{-1}$ .

В прогнозных расчетах предсказываемое значение зависимой величины  $y$  вычисляется путем подстановки в уравнение регрессии  $y = f(x; a_1, a_2, \dots, a_m)$  значения  $x = x_{\text{прогноз}}$ . Вычисленное значение  $y_{\text{прогноз}}$ , равное  $f(x_{\text{прогноз}}; a_1, a_2, \dots, a_m)$ , называется *точечным прогнозом*<sup>3</sup>.

### ПРИМЕР РЕШЕНИЯ ЗАДАЧИ

По результатам лабораторных исследований получены значения величины сцепления грунта ( $c$ ) в зависимости от глубины отбора образца ( $h$ ). Систематизированные значения характеристик сцепления грунта в точках отбора приведены в таблице 2.

Требуется установить тип и параметры аналитической зависимости, аппроксимирующей результаты опытных измерений.

Проверить значимость уравнения регрессии (линейного, квадратичного и экспоненциального) по критерию Фишера при уровне значимости  $\alpha = 0,05$ . Определить значимость коэффициентов регрессии по критерию Стьюдента при уровне значимости  $\alpha = 0,05$ .

Найти точечный прогноз значения величины сцепления грунта ( $c$ ) при условии, что значение глубины отбора образца ( $h$ ) меньше значения максимального значения глубины  $h_{\text{max}}$  на 12% величины размаха переменной  $h$ , т.е.  $h_{\text{прогноз}} = h_{\text{max}} - 0,12 \cdot R_h$ , где  $R_h$  – размах переменной  $h$  (размах выборки), равен разности наибольшего и наименьшего значений  $h$ :

$$R_h = h_{\text{max}} - h_{\text{min}}.$$

---

<sup>3</sup> На практике результат точечного прогноза маловероятен, поэтому часто находят интервальную оценку прогноза.

Таблица 2

## Исходные данные для задачи

Глубина отбора образца $h$ , м	Сцепление $c$ , кг/см <sup>3</sup>
31,0	0,130
32,0	0,139
33,0	0,142
33,5	0,144
34,0	0,145
34,5	0,157
35,0	0,16
36,0	0,158
36,5	0,159
37,0	0,160
38,0	0,165
39,5	0,168
40,0	0,171
41,0	0,180

## РЕШЕНИЕ В ТАБЛИЧНОМ ПРОЦЕССОРЕ MS EXCEL

Для выявления зависимости величины сцепления  $c$  от значения глубины отбора образца  $h$  аппроксимируем эмпирическую зависимость  $c=f(h)$  последовательно линейной, квадратичной и экспоненциальной функциями. Введем обозначения независимых и зависимых величин:  $x$  – глубина отбора образца  $h$  (м);  $y$  – сцепление  $c$  (кг/см<sup>3</sup>).

Для проведения расчетов данные целесообразно расположить в виде, представленном на рис. 2-3, используя средства табличного процессора MS Excel.

▲	A	B	C	D	E	F	G	H	I	J
1	$i$	$x_i$	$y_i$	$x_i^2$	$x_i y_i$	$x_i^3$	$x_i^4$	$x_i^2 y_i$	$\ln(y_i)$	$x_i \ln(y_i)$
2	1	31	0,13	961,00	4,0300	29791,000	923521,000	124,9300	-2,0402	-63,2468
3	2	32	0,139	1024,00	4,4480	32768,000	1048576,000	142,3360	-1,9733	-63,1450
4	3	33	0,142	1089,00	4,6860	35937,000	1185921,000	154,6380	-1,9519	-64,4136
5	4	33,5	0,144	1122,25	4,8240	37595,375	1259445,063	161,6040	-1,9379	-64,9211
6	5	34	0,145	1156,00	4,9300	39304,000	1336336,000	167,6200	-1,9310	-65,6547
7	6	34,5	0,157	1190,25	5,4165	41063,625	1416695,063	186,8693	-1,8515	-63,8771
8	7	35	0,16	1225,00	5,6000	42875,000	1500625,000	196,0000	-1,8326	-64,1404
9	8	36	0,158	1296,00	5,6880	46656,000	1679616,000	204,7680	-1,8452	-66,4258
10	9	36,5	0,159	1332,25	5,8035	48627,125	1774890,063	211,8278	-1,8389	-67,1181
11	10	37	0,16	1369,00	5,9200	50653,000	1874161,000	219,0400	-1,8326	-67,8055
12	11	38	0,165	1444,00	6,2700	54872,000	2085136,000	238,2600	-1,8018	-68,4688
13	12	39,5	0,168	1560,25	6,6360	61629,875	2434380,063	262,1220	-1,7838	-70,4598
14	13	40	0,171	1600,00	6,8400	64000,000	2560000,000	273,6000	-1,7661	-70,6437
15	14	41	0,18	1681,00	7,3800	68921,000	2825761,000	302,5800	-1,7148	-70,3067
16	<b>сумма</b>	<b>501</b>	<b>2,178</b>	<b>18050</b>	<b>78,472</b>	<b>654693</b>	<b>23905063,25</b>	<b>2846,195</b>	<b>-26,1016</b>	<b>-930,627</b>
17	<b>n =</b>	<b>14</b>								

Рис. 2. Ввод данных и вычисление вспомогательных сумм

▲	K	L	M	N	O	P	Q
1	$(x_i - x_{cp})^2 (y_i - y_{cp})^2$	$(x_i - x_{cp})^2$	$(y_i - y_{cp})^2$	$(y_i - y_i^T_{\text{лин}})^2$	$(y_i - y_i^T_{\text{квадр}})^2$	$(\ln(y_i) - \ln(y_{cp}))^2$	$(\ln(y_i) - \ln(y_i^T_{\text{экс}}))^2$
2	0,1224	22,9031	0,0007	0,0000216	0,000002	0,032246	0,001625
3	0,0627	14,3316	0,0003	0,0000000	0,000002	0,012686	0,000003
4	0,0378	7,7602	0,0002	0,0000019	0,000002	0,008332	0,000075
5	0,0264	5,2245	0,0001	0,0000025	0,000004	0,005974	0,000078
6	0,0189	3,1888	0,0001	0,0000076	0,000013	0,004952	0,000258
7	-0,0018	1,6531	0,0000	0,0000497	0,000035	0,000084	0,002430
8	-0,0035	0,6173	0,0000	0,0000619	0,000042	0,000788	0,002923
9	0,0005	0,0459	0,0000	0,0000022	0,000000	0,000240	0,000173
10	0,0024	0,5102	0,0000	0,0000001	0,000002	0,000475	0,000028
11	0,0054	1,4745	0,0000	0,0000008	0,000005	0,000788	0,000007
12	0,0209	4,9031	0,0001	0,0000001	0,000001	0,003462	0,000000
13	0,0462	13,7959	0,0002	0,0000146	0,000011	0,005907	0,000603
14	0,0650	17,7602	0,0002	0,0000090	0,000003	0,008941	0,000442
15	0,1274	27,1888	0,0006	0,0000026	0,000019	0,021273	0,000004
16	<b>0,53071</b>	<b>121,35714</b>	<b>0,00250</b>	<b>0,00017</b>	<b>0,0001416</b>	<b>0,10615</b>	<b>0,00865</b>
17	<b>x cp.=</b>	<b>35,785714</b>	<b>S общ</b>	<b>S ост.лин</b>	<b>S ост.кв</b>		<b>S ост.экср</b>
18	<b>y cp.=</b>	<b>0,1555714</b>					

Рис. 3. Вычисление вспомогательных сумм и средних значений

Пояснения к расчетам:

- Шаг 1. В ячейки B2:B15 занести значения  $x_i$ .
- Шаг 2. В ячейки C2:C15 занести значения  $y_i$ .
- Шаг 3. В ячейку D2 ввести формулу  $=B2^2$ .
- Шаг 4. В ячейки D3:D15 скопировать эту формулу.
- Шаг 5. В ячейку E2 ввести формулу  $=B2*C2$ .
- Шаг 6. В ячейки E3:E15 скопировать эту формулу.
- Шаг 7. В ячейку F2 ввести формулу  $=B2^3$ .
- Шаг 8. В ячейки F3:F15 скопировать эту формулу.
- Шаг 9. В ячейку G2 ввести формулу  $=B2^4$ .
- Шаг 10. В ячейки G3:G15 скопировать эту формулу.
- Шаг 11. В ячейку H2 ввести формулу  $=B2^2*C2$ .
- Шаг 12. В ячейки H3:H15 скопировать эту формулу.
- Шаг 13. В ячейку I2 ввести формулу  $=LN(C2)$ .
- Шаг 14. В ячейки I3:I15 скопировать эту формулу.
- Шаг 15. В ячейку J2 ввести формулу  $=B2*LN(C2)$ .
- Шаг 16. В ячейки J3:J15 скопировать эту формулу.
- Шаг 17. В ячейку B16 ввести формулу  $=СУММ(B2:B15)$
- Шаг 18. В ячейки C16:J16 скопировать эту формулу.
- Шаг 19. В ячейку B17 вводим формулу  $=СЧЁТ(B2:B15)$ .

Продолжаем заполнение таблицы (рис. 3).

Вычисляем средние значения:

- Шаг 20. В ячейку L17 вводим формулу  $=B16/BS17$ .
- Шаг 21. В ячейку L18 вводим формулу  $=C16/BS17$ .
- Шаг 22. В ячейку K2 вводим формулу  $=(B2-LS17)*(C2-LS18)$ .
- Шаг 23. В ячейки K3:K15 эта формула копируется.
- Шаг 24. В ячейку L2 вводим формулу  $=(B2-LS17)^2$ .
- Шаг 25. В ячейки L3:L15 эта формула копируется.
- Шаг 26. В ячейку M2 вводим формулу  $=(C2-LS18)^2$ .
- Шаг 27. В ячейки M3:M15 эта формула копируется.

Последующие шаги делаем с помощью автосуммирования:

- Шаг 28. В ячейку K16 вводим формулу  $=СУММ(K2:K15)$ .
- Шаг 29. В ячейку L16 вводим формулу  $=СУММ(L2:L15)$ .
- Шаг 30. В ячейку M16 вводим формулу  $=СУММ(M2:M15)$ .

## ЛИНЕЙНАЯ АППРОКСИМАЦИЯ

Аппроксимируем функцию  $c=f(h)$  *линейной функцией* вида  $y(x) = a_1 + a_2x$ .

Решим систему уравнений (4) методом обратной матрицы, используя итоговые суммы, находящиеся в ячейках В16:Е16 (рис. 2), с учетом, что количество измерений  $n=14$ .

Затем вычислим значения коэффициента корреляции и коэффициента детерминированности для линейной регрессии. Далее проверим значимость уравнения, а также установим значимость коэффициентов регрессии при уровне значимости  $\alpha=0,05$  (рис. 4).

Пояснения к расчетам:

Шаг 1. В ячейку А20 вводим формулу = $\$B\$17$ .

Шаг 2. В ячейку В20 и в ячейку А21 вводим формулу =В16.

Шаг 3. В ячейку В21 вводим формулу =D16.

Шаг 4. В ячейку С20 вводим формулу =С16.

Шаг 5. В ячейку С21 вводим формулу =Е16.

Шаг 6. Выделяем ячейки **А24:В25** и вводим формулу =МОБР(А20:В21).

	A	B	C	D	E	F	G	H
19	Матрица А		Столбец В		Линейная			
20	14	501	2,178		аппроксимация			
21	501	18050,00	78,4720					
22								
23	Обратная матрица			Решение системы			$\rho=$	0,9644
24	10,6239	-0,2948793		a1=	-0,0009		$R^2_{\text{лин}}=$	0,9301
25	-0,2949	0,0082401		a2=	0,0044			
26								
27	Гтабл=	4,75	Глин > Гтабл					
28	Глпн=	159,576	Уравнение значимо					
29	Дост=	1,454E-05	гтабл=	2,1604				
30	Sa1=	0,0124304	та1=	0,07443431	та1<гтабл	не значим		
31	Sa2=	0,0003462	та2=	12,6323465	та2>гтабл	значим		

Рис. 4. Фрагмент рабочего листа MS Excel для линейной аппроксимации

- Шаг 7. Выделяем ячейки **E24:E25** и вводим формулу  
 $=\text{МУМНОЖ}(A24:B25;C20:C21)$ .
- Шаг 8. В ячейку H23 вводим формулу  $=K16/((L16^{(1/2)}*M16^{(1/2)})$ .
- Шаг 9. В ячейку N2 вводим формулу  $=(C2-(\$E\$24+\$E\$25*B2))^{2}$ .
- Шаг 10. В ячейки N3:N15 эта формула копируется.
- Шаг 11. В ячейку N16 вводим формулу  $=\text{СУММ}(N2:N15)$ .
- Шаг 12. В ячейку H24 вводим формулу  $=1-N16/M16$ .
- Шаг 13. В ячейке **B27** вычисляем значение критерия Фишера при уровне значимости  $\alpha = 0,05$ ; степень свободы  $df_1 = 1$  и  $df_2 = n - 2 = 12$ .  
Вводим формулу  $=\text{F.ОБР.ПХ}(0,05;1;12)$ .
- Шаг 14. В ячейку B28 вводим формулу  $=H24*(B17-2)/(1-H24)$ .
- Шаг 15. В ячейку B29 вводим формулу  $=N16/(B17-2)$ .
- Шаг 16. В ячейку B30 вводим формулу  
 $=(B29*D16)/(B17*L16)^{(1/2)}$ .
- Шаг 17. В ячейку B31 вводим формулу  $=(B29/L16)^{(1/2)}$ .
- Шаг 18. В ячейку C28 вводим формулу  
 $=\text{ЕСЛИ}(B28>B27;"Уравнение значимо"; "Уравнение не значимо")$ .
- Шаг 19. В ячейке **D29** вычисляем значение критерия Стьюдента по уровню значимости  $\alpha = 0,05$  и степени свободы  $df = n - 1 = 13$ .  
Вводим формулу  $=\text{СТЬЮДЕНТ.ОБР.2X}(0,05;13)$ .
- Шаг 20. В ячейку D30 вводим формулу  $=\text{ABS}(E24)/B30$ .
- Шаг 21. В ячейку D31 вводим формулу  $=\text{ABS}(E25)/B31$ .
- Шаг 22. В ячейку F30 вводим формулу  
 $=\text{ЕСЛИ}(D30>\$D\$29;"значим";"не значим")$ .
- Шаг 23. В ячейку F31 вводим формулу  
 $=\text{ЕСЛИ}(D31>\$D\$29;"значим";"не значим")$ .

Таким образом, уравнение линейной регрессии имеет следующий вид:

$$y(x) = -0,0009 + 0,0044 \cdot x. \quad (23)$$

При этом, согласно критерию Фишера, уравнение линейной регрессии (23) значимо.

Согласно критерию Стьюдента, коэффициент этого уравнения  $a_1 = -0,0009$  не значим, а коэффициент  $a_2 = 0,0044$  значим.

Следовательно, в уравнении линейной регрессии (23) свободным коэффициентом  $a_1$  можно пренебречь, поскольку он практически равен нулю.

### КВАДРАТИЧНАЯ АППРОКСИМАЦИЯ

Аппроксимируем функцию  $c=f(h)$  *квадратичной функцией* вида  $y(x) = a_1 + a_2x + a_3x^2$ .

Решим систему уравнений (5) методом обратной матрицы, используя итоговые суммы, находящиеся в ячейках В16:Н16 (рис. 2), с учетом, что количество измерений  $n=14$ .

Затем вычислим значение коэффициента детерминированности для квадратичной регрессии. Далее проверим значимость уравнения, а также установим значимость коэффициентов регрессии при уровне значимости  $\alpha=0,05$  (рис. 5).

	A	B	C	D	E	F	G
34	<b>Матрица А</b>			<b>Столбец В</b>		<i><b>Квадратичная</b></i>	
35	14	501	18050,0	2,178		<i><b>аппроксимация</b></i>	
36	501	18050	654693,0	78,4720			
37	18050,00	654693	23905063,3	2846,1950		$R^2_{\text{квадр}} =$	0,9433
38							
39	<b>Обратная матрица</b>			<b>Решение системы</b>			
40	1679,18	-93,373777	1,2893475	<b>a1=</b>	-0,2355		
41	-93,374	5,200558	-0,071925	<b>a2=</b>	0,0175		
42	1,28935	-0,071925	0,0009963	<b>a3=</b>	-0,00018		
43							
44	<b>Fтабл=</b>	<b>3,98</b>	<b>Fкв &gt; Fтабл</b>				
45	<b>Fкв=</b>	<b>91,457</b>	<i><b>Уравнение значимо</b></i>				
46	<b>Дост=</b>	<b>1,287E-05</b>	<b>tтабл=</b>	<b>2,1788</b>			
47	<b>Sa1=</b>	<b>0,1469998</b>	<b>ta1=</b>	<b>1,6019452</b>	<b>ta1&lt;tтабл</b>	<i><b>не значим</b></i>	
48	<b>Sa2=</b>	<b>0,0081807</b>	<b>ta2=</b>	<b>2,13402172</b>	<b>ta2&lt;tтабл</b>	<i><b>не значим</b></i>	
49	<b>Sa3=</b>	<b>0,0001132</b>	<b>ta3=</b>	<b>1,60072275</b>	<b>ta3&lt;tтабл</b>	<i><b>не значим</b></i>	

Рис. 5. Фрагмент рабочего листа MS Excel для квадратичной аппроксимации

Пояснения к расчетам:

В ячейках A35:C37 формируем матрицу коэффициентов A:

Шаг 1. В ячейку A35 вводим формулу = $\$B\$17$ .

Шаг 2. В ячейки A36 и B35 вводим формулу = $B16$ .

Шаг 3. В ячейки A37, B36 и C35 вводим формулу = $D16$ .

Шаг 4. В ячейки B37 и C36 вводим формулу = $F16$ .

Шаг 5. В ячейку C37 вводим формулу = $G16$ .

В ячейках D35:D37 формируем столбец коэффициентов B:

Шаг 6. В ячейку D35 вводим формулу = $C16$ .

Шаг 7. В ячейку D36 вводим формулу = $E16$ .

Шаг 8. В ячейку D37 вводим формулу = $H16$ .

Шаг 9. Выделяем ячейки A40:C42 и вводим формулу:  
= $\text{МОБР}(A35:C37)$ .

Шаг 10. Выделяем ячейки E40:E42 и вводим формулу  
= $\text{МУМНОЖ}(A40:C42; D35:D37)$ .

Шаг 11. В ячейку O3 вводим формулу  
= $(C2 - (\$E\$40 + \$E\$41 * B2 + \$F\$42 * B2^2))^2$ .

Шаг 12. В ячейки O3:O15 эта формула копируется.

Шаг 13. В ячейку O16 вводим формулу = $\text{СУММ}(O2:O15)$ .

Шаг 14. В ячейку G37 вводим формулу = $1 - O16/M16$ .

Шаг 15. В ячейке **B44** вычисляем значение критерия Фишера при уровне значимости  $\alpha = 0,05$ ; степень свободы  $df_1 = 2$  и  $df_2 = n - 3 = 11$ .

Вводим формулу = $\text{F.ОБР.ПХ}(0,05; 2; 11)$ .

Шаг 16. В ячейку B45 вводим формулу = $G37 * (B17 - 3) / (2 * (1 - G37))$ .

Шаг 17. В ячейку B46 вводим формулу = $O16 / (B17 - 3)$

Шаг 18. В ячейку B47 вводим формулу = $(\$B\$46 * A40)^{(1/2)}$ .

Шаг 19. В ячейку B48 вводим формулу = $(\$B\$46 * B41)^{(1/2)}$ .

Шаг 20. В ячейку B49 вводим формулу = $(\$B\$46 * C42)^{(1/2)}$ .

Шаг 21. В ячейку C45 вводим формулу  
= $\text{ЕСЛИ}(B45 > B44; \text{"Уравнение значимо"}; \text{"Уравнение не значимо"})$ .

Шаг 22. В ячейке **D46** вычисляем значение критерия Стьюдента по уровню значимости  $\alpha = 0,05$  и степени свободы  $df = n - 2 = 12$ .

Вводим формулу = $\text{СТЮДЕНТ.ОБР.2X}(0,05; 12)$ .

Шаг 23. В ячейку D47 вводим формулу = $\text{ABS}(E40) / B47$ .

Шаг 24. В ячейку D48 вводим формулу = $\text{ABS}(E41) / B48$ .

Шаг 25. В ячейку D49 вводим формулу = $\text{ABS}(E42) / B49$ .

Шаг 26. В ячейку F47 вводим формулу  
=ЕСЛИ(D47>D\$46;"значим";"не значим").

Шаг 27. В ячейку F48 вводим формулу  
=ЕСЛИ(D48>D\$46;"значим";"не значим").

Шаг 28. В ячейку F49 вводим формулу  
=ЕСЛИ(D49>D\$46;"значим";"не значим").

Таким образом, уравнение квадратичной регрессии примет следующий вид:

$$y(x) = -0,2355 + 0,0175 \cdot x - 0,00018 \cdot x^2. \quad (24)$$

Согласно критерию Фишера, уравнение квадратичной регрессии (24) значимо. Все три коэффициента этого уравнения, согласно критерию Стьюдента, не значимы.

Таким образом, в рассмотренном примере уравнение квадратичной регрессии (24) использовать нецелесообразно (не корректно).

## **ЭКСПОНЕНЦИАЛЬНАЯ АППРОКСИМАЦИЯ**

Аппроксимируем функцию  $c=f(h)$  *экспоненциальной функцией* вида  $y(x) = a_1 \cdot e^{a_2 x}$ .

Решим систему уравнений (9) методом обратной матрицы, используя итоговые суммы, находящиеся в ячейках C16, D16, H16, I16 и J16 (рис. 2), с учетом, что количество измерений  $n=14$ .

Затем вычислим значение коэффициента детерминированности для экспоненциальной регрессии. Далее проверим значимость уравнения, а также установим значимость коэффициентов регрессии при уровне значимости  $\alpha=0,05$  (рис. 6).

	A	B	C	D	E	F
50	Матрица А		Столбец В		Экспоненциальная	
51	14	501	-26,1016		аппроксимация	
52	501	18050,00	-930,6270		$R^2_{\text{экс}}$ =	0,9185
53						
54	Обратная матрица			Решение системы		
55	10,6239	-0,2948793		c=	-2,8777	
56	-0,2949	0,0082401		a2=	0,0283	
57						
58					a1=	0,0563
59	Гтабл=	4,75	Гэксп > Гтабл			
60	Гэксп=	135,271	Уравнение значимо			
61	Дост=	0,0007208	гтабл=	2,1604		
62	Са1=	0,0875058	та1=	32,8857676	та1>гтабл	значим
63	Са2=	0,002437	та2=	11,6188944	та2>гтабл	значим

Рис. 6. Фрагмент рабочего листа MS Excel для экспоненциальной аппроксимации

Пояснения к расчетам:

- Шаг 1. В ячейку A51 вводим формулу =B\$17.
- Шаг 2. В ячейку B51 вводим формулу =B16.
- Шаг 3. В ячейку A52 вводим формулу =B16.
- Шаг 4. В ячейку B52 вводим формулу =D16.
- Шаг 5. В ячейку C51 вводим формулу =I16.
- Шаг 6. В ячейку C52 вводим формулу =J16.
- Шаг 7. Выделяем ячейки A55:B56 и вводим формулу =МОБР(A51:B52).
- Шаг 8. Выделяем ячейки E55:E56 и вводим формулу =МУМНОЖ(A55:B56;C51:C52).
- Шаг 9. В ячейку F58 вводим формулу =EXP(E55).
- Шаг 10. В ячейку P2 вводим формулу =(LN(C2)-LN(\$L\$18))^2.
- Шаг 11. В ячейки P3:P15 эта формула копируется.
- Шаг 12. В ячейку Q2 вводим формулу =(LN(C2)-LN(\$F\$58\*EXP(B2\*\$E\$56)))^2.
- Шаг 13. В ячейки Q3:Q15 эта формула копируется.

- Шаг 14. В ячейку P16 вводим формулу =СУММ(P2:P15).
- Шаг 15. В ячейку Q16 вводим формулу =СУММ(Q2:Q15).
- Шаг 16. В ячейку F52 вводим формулу =1-Q16/P16.
- Шаг 17. В ячейку **B59** вводим формулу =B27 (табличные значения критерия Фишера для экспоненциальной и линейной зависимости при одинаковом числе экспериментальных данных  $n$  совпадают).
- Шаг 18. В ячейку B60 вводим формулу =F52\*(B17-2)/(1-F52).
- Шаг 19. В ячейку B61 вводим формулу =Q16/(B17-2)
- Шаг 20. В ячейку B62 вводим формулу  

$$=((B61*D16)/(B17*L16))^{(1/2)}$$
- Шаг 21. В ячейку B63 вводим формулу =(B61/L16)^(1/2).
- Шаг 22. В ячейку C59 вводим формулу  
 =ЕСЛИ(B60>B59;"Уравнение значимо";"Уравнение не значимо").
- Шаг 23. В ячейку **D61** вводим формулу =D29 (табличные значения коэффициента Стьюдента для экспоненциальной и линейной зависимости при одинаковом числе экспериментальных данных  $n$  совпадают).
- Шаг 24. В ячейку D62 вводим формулу =ABS(E55)/B62.
- Шаг 25. В ячейку D63 вводим формулу =ABS(E56)/B63.
- Шаг 26. В ячейку F62 вводим формулу  
 =ЕСЛИ(D62>D\$61;"значим";"не значим").
- Шаг 27. В ячейку F63 вводим формулу  
 =ЕСЛИ(D63>D\$61;"значим";"не значим").

Таким образом, уравнение экспоненциальной регрессии имеет следующий вид:

$$y(x) = 0,0563e^{0,0283 \cdot x} \quad (25)$$

Согласно критерию Фишера, уравнение экспоненциальной регрессии (25) значимо. Оба коэффициента этого уравнения, согласно критерию Стьюдента, также значимы.

### **Вывод.**

Представим результаты расчетов в виде таблицы (табл. 3).

В уравнении квадратичной регрессии все три коэффициента не значимы, поэтому использовать такую зависимость для прогноза нельзя.

## Анализ результатов расчетов

№ п/п	Зависимость	$R^2$	Значимость уравнения в целом	Значимость коэффициентов	Возможность использования для прогноза
1	Линейная	0,9301	Да	Старший коэффициент $a_2$ значим	Да
2	Квадратичная	0,9433	Да	Не значимы	Нет
3	Экспоненциальная	0,9185	Да	Значимы	Да

Для нахождения лучшей эмпирической формулы сравним коэффициенты детерминации для зависимостей, которые могут быть использованы для прогнозирования. В нашем случае это линейная и экспоненциальная зависимости.

Поскольку справедливо соотношение  $R^2_{\text{лин}} = 0,9301 > R^2_{\text{экспон}} = 0,9185$ , то приходим к выводу, что лучше всего результаты испытаний аппроксимирует линейная функция вида  $y(x) = -0,0009 + 0,0044 \cdot x$ .

Следовательно, за аппроксимирующую функцию, отражающую зависимость сцепления грунта ( $c$ ) от глубины отбора образца ( $h$ ) в пределах предварительно выделенного ИГЭ, следует принять функцию:

$$c(h) = -0,0009 + 0,0044 \cdot h. \quad (26)$$

Учитывая то, что коэффициент  $a_1 = -0,0009$  не значим, уравнение (26) может быть записано в виде

$$c(h) = 0,0044 \cdot h. \quad (27)$$

Такая форма записи позволяет избежать «парадокса», состоящего в том, что при малых глубинах сцепление, вычисленное по формуле (26), отрицательно.

## ПОСТРОЕНИЕ ЛИНИИ ТРЕНДА В MS EXCEL

Представим результаты расчетов, полученные выше, графически: исследуем характер зависимости  $x$  и  $y$  с помощью построения линий тренда в MS Excel (рис. 7 - 9).

Для этого необходимо:

1) Выделить ячейки **B2:C15** с эмпирическими данными и построить точечную диаграмму зависимости  $y(x)$ .

2) Построить линию тренда: щелкнуть по одной из точек графика правой кнопкой мыши, выбрать команду **Добавить линию тренда**, выбрать тип аппроксимирующей кривой.

3) Получить числовые характеристики коэффициентов уравнения: активировать опции «показывать уравнение на диаграмме» и «поместить на диаграмму величину  $R^2$ ».

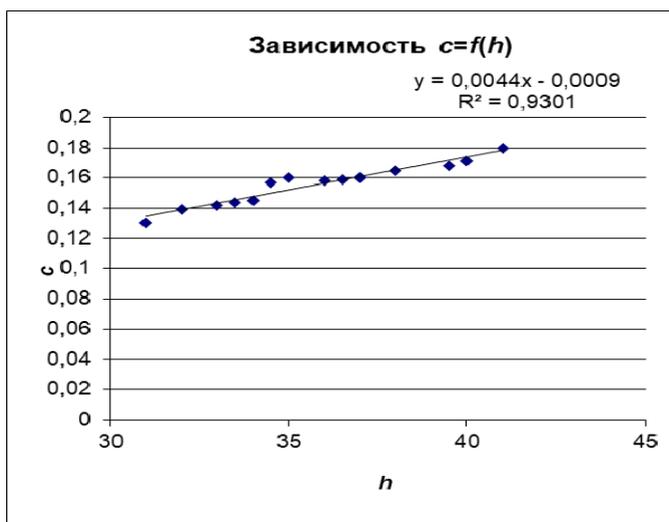


Рис. 7. Эмпирические данные и линия тренда  $c=f(h)$  для линейной аппроксимации

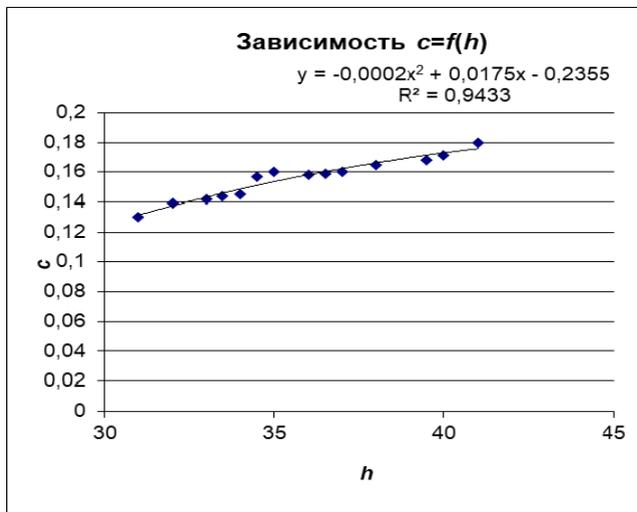


Рис. 8. Эмпирические данные и линия тренда  $c=f(h)$  для квадратичной аппроксимации

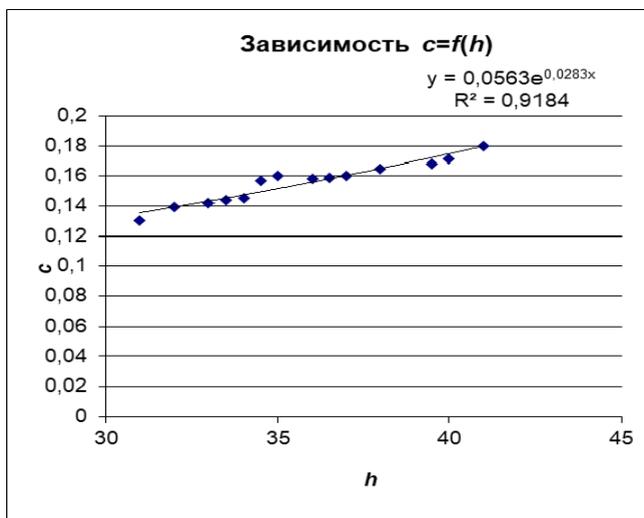


Рис. 9. Эмпирические данные и линия тренда  $c=f(h)$  для экспоненциальной аппроксимации

Полученные коэффициенты уравнений и значения коэффициентов детерминации для линейного, квадратичного и экспоненциального трендов полностью совпадают с коэффициентами, рассчитанными по методу наименьших квадратов с помощью матричных функций MS Excel.

Следовательно, зависимость  $c=f(h)$  может быть описана линейной функцией:  $c(h) = -0,0009 + 0,0044 \cdot h$ .

## ВЫЧИСЛЕНИЕ ТОЧЕЧНОГО ПРОГНОЗА

Прогнозное (точечное) значение сцепления при заданной глубине отбора образца вычисляем с помощью функции, которая наиболее качественно аппроксимирует табличные данные, т.е. по формуле (26).

Результаты вычисления прогнозного (точечного) значения представлены на рис. 10.

Значение глубины отбора образца, для которой будет построен прогноз, приближенно равно 39,8, вычислено по формуле  $x_{\max} - 0,12 \cdot R$  и находится в ячейке I31. В ячейке I32 находится вычисленное с помощью эмпирической формулы (26) прогножное значение величины сцепления.

	Н	l
27	<b>Прогнозное значение</b>	
28	<b>Xmax=</b>	<b>41</b>
29	<b>Xmin=</b>	<b>31</b>
30	<b>R=</b>	<b>10</b>
31	<b>x=</b>	<b>39,80</b>
32	<b>y=</b>	<b>0,1731</b>

Рис. 10. Фрагмент листа MS Excel с вычислением точечного прогноза

Пояснения к расчетам:

В ячейку в ячейку I28 введена формула =МАКС(B2:B15).

В ячейку I29 введена формула =МИН(B2:B15).

В ячейку I30 введена формула =I28-I29.

В ячейку I31 введена формула = I28-0,12\*I30.

В ячейку I32 введена формула =E24+E25\*I31.

Таким образом, при глубине отбора образца, равной 39,8 м точечное значение прогноза сцепления равно 0,173 кг/см<sup>3</sup>.

### РЕШЕНИЕ ЗАДАЧИ В ПАКЕТЕ MATHCAD

На рисунках 11 - 14 приведено решение задачи и графическое представление результатов расчетов в математическом пакете MathCAD.

x :=	( 31 )	y :=	( 0.13 )
	( 32 )		( 0.139 )
	( 33 )		( 0.142 )
	( 33.5 )		( 0.144 )
	( 34 )		( 0.145 )
	( 34.5 )		( 0.157 )
	( 35 )		( 0.16 )
	( 36 )		( 0.158 )
	( 36.5 )		( 0.159 )
	( 37 )		( 0.160 )
	( 38 )		( 0.165 )
	( 39.5 )		( 0.168 )
	( 40 )		( 0.171 )
	( 41 )		( 0.18 )

Рис. 11. Эмпирические данные

**Примечание.** 1. Данные в MathCAD можно экспортировать из таблицы MS Excel, скопировав их и вставив затем в пустую матрицу (вектор) MathCAD (предварительно заменив в таблице данных Excel знак десятичной запятой на точку).

2. Параметры линейной, квадратичной и экспоненциальной регрессий в математическом пакете MathCAD можно определить методом обратной матрицы, а также с помощью встроенной функции  $lsolve()$ , реализующей метод Гаусса.

Аппроксимация линейной функцией

$$\sum x = 501 \quad \sum y = 2.178 \quad \sum x^2 = 18050 \quad \sum (x \cdot y) = 78.472$$

Решение системы методом обратной матрицы:

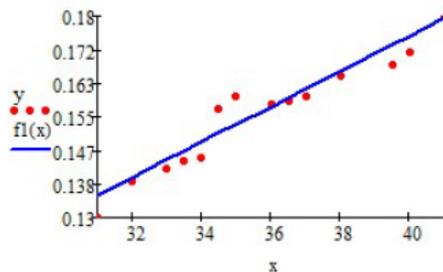
$$A := \begin{pmatrix} 14 & 501 \\ 501 & 18050 \end{pmatrix} \quad B := \begin{pmatrix} 2.178 \\ 78.472 \end{pmatrix} \quad A^{-1} \cdot B = \begin{pmatrix} -9.2525 \times 10^{-4} \\ 4.3732 \times 10^{-3} \end{pmatrix}$$

Коэффициенты линейной аппроксимации:

$$a1 := -0.00092525 \quad a2 := 0.0043732$$

Уравнение линейной функции:

$$f1(x) := 0.0044 \cdot x - 0.0009$$



Коэффициент корреляции:

$$\text{corr}(x, y) = 0.9643963$$

Коэффициент детерминированности:

$$1 - \frac{\sum [(a1 + a2 \cdot x) - y]^2}{\sum (y - \text{mean}(y))^2} = 0.93006$$

Рис. 12. Аппроксимация линейной функцией в MathCAD

**Примечание.** Вычисление параметров линейной регрессии в системе MathCAD можно также выполнить с применением встроенной функции  $line(x, y)$ .

### Аппроксимация квадратичной функцией

$$\sum x = 501 \quad \sum y = 2.178 \quad \sum x^2 = 18050 \quad \sum \overrightarrow{(x \cdot y)} = 78.472$$

$$\sum x^3 = 654693 \quad \sum x^4 = 23905063.25 \quad \sum \overrightarrow{(x^2 \cdot y)} = 2846.195$$

Решение системы методом обратной матрицы:

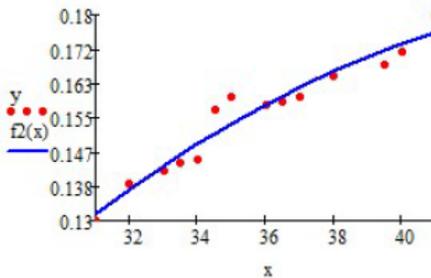
$$A := \begin{pmatrix} 14 & 501 & 18050 \\ 501 & 18050 & 654693 \\ 18050 & 654693 & 23905063.25 \end{pmatrix} \quad B := \begin{pmatrix} 2.178 \\ 78.472 \\ 2846.195 \end{pmatrix} \quad \underline{K}_{\text{min}} := A^{-1} \cdot B$$

Коэффициенты квадратичной аппроксимации:

$$\underline{a1} := K_0 \quad \underline{a2} := K_1 \quad \underline{a3} := K_2 \quad K = \begin{pmatrix} -0.23548556 \\ 0.01745788 \\ -0.00018125 \end{pmatrix}$$

$$a1 = -0.23548556 \quad a2 = 0.01745788 \quad a3 = -0.00018125$$

Уравнение квадратичной функции:  $f2(x) := a3 \cdot x^2 + a2 \cdot x + a1$



Коэффициент детерминированности:

$$1 - \frac{\sum [(a1 + a2 \cdot x + a3 \cdot x^2) - y]^2}{\sum (y - \text{mean}(y))^2} = 0.9433$$

Рис. 13. Аппроксимация квадратичной функцией в MathCAD

**Примечание.** Одномерную полиномиальную регрессию с произвольной степенью полинома  $n$  в MathCAD можно выполнить также с использованием встроенной функцией  $regress(x,y,n)$ .

### Аппроксимация экспоненциальной функцией

$$\sum x = 501 \quad \sum x^2 = 18050 \quad \sum \ln(y) = -26.1016 \quad \sum (x \cdot \ln(y)) = -930.627$$

Решение системы методом обратной матрицы:

$$A_{\text{exp}} := \begin{pmatrix} 14 & 501 \\ 501 & 18050 \end{pmatrix} \quad B_{\text{exp}} := \begin{pmatrix} -26.1016 \\ -930.627 \end{pmatrix} \quad K_{\text{exp}} := A_{\text{exp}}^{-1} \cdot B_{\text{exp}}$$

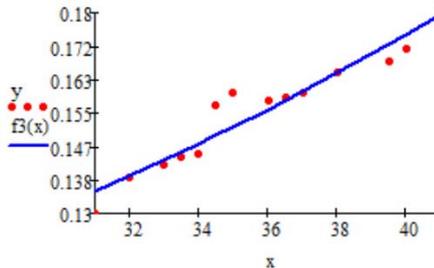
Коэффициенты экспоненциальной аппроксимации:

$$K_{\text{exp}} = \begin{pmatrix} -2.87801825 \\ 0.02832466 \end{pmatrix}$$

$$a1 := e^{(K_{\text{exp}})_0} \quad a1 = 0.0562 \quad a2 := K_{\text{exp}}_1 \quad a2 = 0.02832466$$

Уравнение экспоненциальной функции:

$$f3(x) := a1 \cdot e^{a2 \cdot x}$$



Коэффициент детерминированности:

$$1 - \frac{\sum (a1 \cdot e^{a2 \cdot x} - y)^2}{\sum (y - \text{mean}(y))^2} = 0.91936$$

Рис. 14. Аппроксимация экспоненциальной функцией в MathCAD

## ВЫВОД ПО РЕШЕНИЮ ЗАДАЧИ

Сравнивая результаты расчетов, полученных средствами табличного процессора MS Excel и математического пакета MathCAD, видим, что они практически совпадают. Расхождения обусловлены тем, что в MathCAD и MS Excel для вычисления значений встроенных функций используются различные алгоритмы и пределы точности.

Учитывая высокое значение коэффициента детерминированности для линейной аппроксимации ( $R^2=0,9301$ ) и значимость коэффициентов, за аппроксимирующую функцию, отражающую зависи-

мость сцепления грунта ( $c$ ) от глубины отбора образца ( $h$ ), следует принять квадратичную функцию вида:  $c(h) = -0,0009 + 0,0044 \cdot h$ .

Используя полученную зависимость при глубине отбора образца ( $h$ ), равной 39,8 м, получен точечный прогноз для величины сцепления ( $c$ ) равный 0,173 кз/см<sup>3</sup>.

## ВАРИАНТЫ ЗАДАНИЙ

По результатам исследований получены систематизированные значения различных характеристик грунта в зависимости от глубины отбора образца.

Во всех вариантах требуется:

1. Установить тип аналитической зависимости (линейный, квадратичный и экспоненциальный), аппроксимирующей результаты опытных измерений.

2. Вычислить коэффициент корреляции (только для линейной зависимости).

3. Для каждого вида зависимости вычислить коэффициент детерминированности.

4. Проверить значимость уравнения регрессии (линейного, квадратичного и экспоненциального) по критерию Фишера-Снедекора при уровне значимости  $\alpha = 0,05$ .

5. Определить значимость коэффициентов регрессии (линейной, квадратичной и экспоненциальной) по критерию Стьюдента при уровне значимости  $\alpha = 0,05$ .

6. Для каждой зависимости построить линию тренда.

7. Используя лучшую зависимость, найти точечный прогноз значения зависимой переменной ( $y$ ) при условии, что значение независимой переменной ( $x$ ) больше ее минимального значения на 20% от величины размаха, т.е.  $x = x_{min} + 20\% \cdot R$ , где  $R$  – размах независимой переменной  $x$  (размах выборки), равен разности наибольшего и наименьшего значений  $x$ :  $R = x_{max} - x_{min}$ .

8. Выполнить решение задачи в математическом пакете (п.п. 1-3, п. 6).

9. В целях получения максимальной оценки за выполнение курсовой работы, произвести расчеты (п.п. 1-3) в среде программирования (VBA, Python и др.).

Варианты заданий представлены на рис. 15-17.

Вариант 1		Вариант 2		Вариант 3	
Глубина отбора образца $h$ , м	Естественная плотность $\rho$ , г/см <sup>3</sup>	Глубина отбора образца $h$ , м	Естественная плотность $\rho$ , г/см <sup>3</sup>	Глубина отбора образца $h$ , м	Естественная плотность $\rho$ , г/см <sup>3</sup>
1	2,01	4	1,95	2	2,05
1,5	2,02	4,2	1,98	2,4	2,06
2	2,04	5	1,97	3	2,07
2,2	2,04	5,7	1,96	3,2	2,08
2,5	2,05	7	1,96	3,5	2,09
2,8	2,05	7,3	1,97	3,8	2,09
3	2,06	8	1,97	4	2,08
3,1	2,07	8,1	1,98	4,2	2,09
3,5	2,08	8,4	1,99	4,5	2,1
4	2,07	9	2,02	5	2,12
4,2	2,08	9,6	2,03	5,3	2,11
4,5	2,09	10	2,05	5,5	2,11
5	2,1	10,5	2,04	6	2,13
5,5	2,09	11	2,07	6,5	2,12

Вариант 4		Вариант 5		Вариант 6	
Глубина отбора образца $h$ , м	Показатель консистенции $I_L$ , д.е.	Глубина отбора образца $h$ , м	Показатель консистенции $I_L$ , д.е.	Глубина отбора образца $h$ , м	Показатель консистенции $I_L$ , д.е.
2	0,2	1	0,11	4	0,64
2,4	0,3	1,5	0,14	4,2	0,58
2,7	0,31	1,7	0,15	4,5	0,61
3	0,34	2	0,2	5	0,59
3,5	0,33	2,5	0,15	6	0,72
3,8	0,35	2,6	0,2	7,5	0,71
4	0,4	3	0,23	8	0,7
4,5	0,44	3,5	0,28	8,4	0,67
4,8	0,43	3,8	0,29	8,7	0,69
5	0,43	4	0,3	9	0,7
5	0,41	4,2	0,32	9,6	0,71
5,5	0,37	4,5	0,33	10	0,71
6	0,44	5	0,34	10,5	0,73
6,5	0,46	5,5	0,37	11	0,75

Рис. 15. Варианты заданий (начало)

Вариант 7		Вариант 8		Вариант 9	
Глубина отбора образца $h$ , м	Коэффициент пористости $e$ , д.е.	Глубина отбора образца $h$ , м	Коэффициент пористости $e$ , д.е.	Глубина отбора образца $h$ , м	Коэффициент пористости $e$ , д.е.
1	0,53	10	0,4	11	0,49
2	0,52	10,6	0,42	11,5	0,5
3	0,51	11	0,46	11,6	0,47
3,2	0,49	12	0,43	12	0,44
3,7	0,5	13,1	0,44	13,2	0,44
4	0,5	13,5	0,45	13,5	0,45
4,6	0,48	14	0,48	14	0,42
4,9	0,44	15	0,47	14,5	0,43
5,5	0,45	15,5	0,49	14,8	0,42
6	0,47	16	0,52	15	0,41
6,6	0,45	16,7	0,53	15,6	0,4
7,2	0,46	17	0,54	16	0,4
7,5	0,49	18	0,58	16,4	0,38
7,9	0,51	19	0,61	16,9	0,37

Вариант 10		Вариант 11		Вариант 12	
Глубина отбора образца $h$ , м	Влажность $W$ , д.е.	Глубина отбора образца $h$ , м	Влажность $W$ , д.е.	Глубина отбора образца $h$ , м	Влажность $W$ , д.е.
11	0,35	1,8	0,27	1	0,31
11,5	0,35	2	0,26	1,4	0,3
12	0,34	2,8	0,25	1,8	0,3
13	0,33	3	0,21	2	0,28
13,4	0,32	3,2	0,21	2,2	0,28
13,5	0,31	3,8	0,22	2,4	0,29
13,8	0,31	3,9	0,21	2,5	0,28
14	0,3	4	0,2	2,6	0,27
14,5	0,29	5	0,19	3,1	0,26
15	0,27	5,8	0,17	3,6	0,25
15,7	0,27	5,5	0,16	3,8	0,24
16	0,26	6	0,15	4	0,23
17	0,24	6,5	0,14	4,3	0,22
18	0,21	6,8	0,11	4,7	0,2

Рис. 16. Варианты заданий (продолжение)

Вариант 13		Вариант 14		Вариант 15	
Глубина отбора образца $h$ , м	Сцепление $c$ , кг/см <sup>2</sup>	Глубина отбора образца $h$ , м	Сцепление $c$ , кг/см <sup>2</sup>	Глубина отбора образца $h$ , м	Сцепление $c$ , кг/см <sup>2</sup>
1	0,328	3	0,031	1	0,07
1,4	0,33	4	0,04	2	0,071
2	0,31	4,1	0,04	2,5	0,071
3	0,329	4,6	0,041	3,2	0,072
3,3	0,3	4,8	0,042	3,8	0,074
3,5	0,333	5	0,044	4	0,075
3,8	0,34	5,5	0,045	4,4	0,078
4	0,35	6	0,046	4,6	0,08
5	0,378	6,5	0,047	4,9	0,097
5,3	0,388	6,7	0,048	5,7	0,099
6	0,39	7	0,05	6	0,1
7	0,401	7,2	0,049	7,2	0,11
8	0,409	7,5	0,052	7,5	0,115
9	0,413	7,9	0,055	7,7	0,12

Вариант 16		Вариант 17		Вариант 18	
Глубина отбора образца $h$ , м	Показатель консистенции $I_L$ , д.е.	Глубина отбора образца $h$ , м	Сцепление $c$ , кг/см <sup>2</sup>	Глубина отбора образца $h$ , м	Коэффициент пористости $e$ , д.е.
2	0,16	1	0,13	3	1
2,4	0,21	1,2	0,139	3,5	0,99
2,6	0,21	2	0,14	4	0,99
3	0,22	2,7	0,142	4,1	0,98
3,3	0,26	3	0,143	4,8	0,96
3,5	0,33	3,5	0,144	5	0,95
4	0,29	4	0,145	5,6	0,97
4,2	0,33	4,2	0,15	6	0,96
4,5	0,36	4,5	0,157	6,5	0,94
5	0,4	5	0,16	7	0,92
5,2	0,42	5,7	0,161	7,1	0,93
5,5	0,44	6	0,162	7,2	0,92
6	0,43	7	0,167	7,5	0,91
6,5	0,47	8	0,17	8	0,9

Рис. 17. Варианты заданий (окончание)

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Бондарик Г.К.* Инженерная геология. Вопросы теории и практики. Философские и методологические основы геологии: Учеб. пособие / Г.К. Бондарик, Л.А. Ярг. – М.: КДУ, 2015, 296 с.
2. *Гарнаев А.Ю.* Microsoft Office Excel 2010: разработка приложений / А.Ю. Гарнаев, Л.В. Рудикова. – СПб.: БВХ-Петербург, 2011. 528 с.
3. ГОСТ 20522-2012. Грунты. Методы статистической обработки результатов испытаний.
4. *Грошев А.С.* Информатика: учебник для вузов/ А.С. Грошев. – Москва, Берлин: Директ-Медиа, 2015. 484 с.
5. Инженерная геодинамика: Учебник / Г.К. Бондарик, В.В. Пендин, Л.А. Ярг. – М.: Книжный дом «Университет», 2015, 472 с.
6. *Макаров Е.Г.* Инженерные расчеты в Mathcad 15: учебный курс. – СПб.: Питер, 2011, 399 с.
7. *Цытович Н.А.* Механика грунтов. Краткий курс: учебник для вузов. – М.: Ленард, 2014. 288 с.
8. Эконометрика. Парная линейная регрессия. Методические указания к лабораторным работам для студентов направлений подготовки бакалавриата 21.03.02 и 38.03.01. Сост.: В.В. Беляев, Т.Р. Косовцева. СПб, 2017, 50 с.

## СОДЕРЖАНИЕ

Введение .....	3
Построение эмпирических формул методом наименьших квадратов	4
Элементы теории корреляции .....	8
Пример решения задачи .....	15
Решение в табличном процессоре MS Excel .....	16
Линейная аппроксимация .....	19
Квадратичная аппроксимация .....	21
Экспоненциальная аппроксимация .....	23
Построение линии тренда в MS Excel .....	27
Вычисление точечного прогноза .....	29
Решение задачи в пакете MathCAD .....	30
Вывод по решению задачи.....	33
Варианты заданий.....	33
Библиографический список.....	38