Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Институт Информационных Технологий и Анализа Данных

**Название работы** – “ Задача идентификации.”

Отчет по лабораторной работе “Лабораторная работа №2”

по дисциплине Моделирование систем и процессов

Вариант 21

Выполнил

Студент, номер группы ИСМб-19-1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д.Д.Солопов

Дата: \_\_\_\_\_\_\_\_\_\_

Принял

Должность \_\_\_\_\_\_\_\_\_\_\_\_\_\_ О.С.Бучнев

Дата: \_\_\_\_\_\_\_\_\_\_

Иркутск 2021 г.

Содержание

[Содержание 2](#_Toc67747997)

[Цель работы 3](#_Toc67747998)

[Постановка задач 4](#_Toc67747999)

[Задание 4](#_Toc67748000)

[Решение задач 5](#_Toc67748001)

[1. Исходные данные для решения задачи идентификации объекта 5](#_Toc67748002)

[2. Регрессионный анализ полученных данных. 6](#_Toc67748003)

[Вывод 9](#_Toc67748004)

Цель работы

Целью лабораторной работы является изучение одной из задач математического моделирования - задачи идентификации, освоение и закрепление практических навыков решения данной задачи с использованием инструментария регрессионного анализа.

Постановка задач

Задание

1. Получить исходные данные для решения задачи идентификации объекта (файл Варианты\_R.xls, лист 2):

2. Провести регрессионный анализ полученных данных. При этом необходимо получить:

a) Выборочное уравнение регрессии,

b) Проверить гипотезу о значимости коэффициентов регрессии при уровне значимости ∝=0,05.

Составить отчет по лабораторной работе, в который включить все результаты

Решение задач

1. Исходные данные для решения задачи идентификации объекта

x<-c(74.8,57.7,4.71,55.74,8.03,46.29,71.79,7.79,44.76,74.01,23.09,99.99,28.67,44.18,83.48,4.42,45.11,96.47,22.29,94.07,0.87,62.94,79.4,72.87,50.22,98.51,7.17,90.3,62.69,72.42,40,39.31,69.16,43.2,9.23,63.3,2.25,59.87,24.57,81.73,43.93,77.07,33.23,86.23,42.3,11.83,56.43,31.88,18.26,105.76,10.95,28.68,34.55,32.97,51.5,93.06,62.88,7.74,67.94,2.69,16.03,83.97,43.46,77.66,92.38,61.01,77.7,97.17,57.38,40.12,91.31,58.18,35.31,46.23,84.9,38.47,81.35,12.72,69.86,-2.95,39.6,35.11,55.65,68.85,19.1,73.57,24.06,35.95,16.21,26.74,88.37,42.97,76.93,45.15,46.6,104.93,94.31,46.8,3.61,18.9)

y <- c(3.4,53.87,84.3,21.11,31.19,104.7,96.11,81.09,47.58,2.09,101.96,96.23,69.06,40.12,53.56,25.41,17.72,50.99,-0.42,29.99,27.66,86.86,23.13,82.74,23.51,13.95,58.02,50.64,91.59,67.94,64.98,67.9,38.6,32.6,60.91,32.95,66.25,10.14,56.84,33.05,-6.69,70.3,85.62,70.71,17.63,6.12,69.97,104.63,60.81,48.01,72.09,33.64,44.63,46.74,54.68,98.93,21.54,27.92,8.85,29.81,97.81,18.98,16.81,11.17,13.33,7.68,85.03,-0.35,-0.74,64.14,27.42,34.79,12.31,51.48,81.66,47.2,27.23,69.93,69.09,66.79,16.95,62.98,-1.82,82.35,12.16,17.47,7,90.28,74.39,63,7.59,9.54,61.17,29.94,1.29,53.56,7.8,39.38,79.67,11.45)

По данным измеренным величинам (x и y) необходимо найти закон, по которому входной сигнал (x) преобразуется в выходной сигнал (y).

2. Регрессионный анализ полученных данных.

**a) Код：**

x<-c(74.8,57.7,4.71,55.74,8.03,46.29,71.79,7.79,44.76,74.01,23.09,99.99,28.67,44.18,83.48,4.42,45.11,96.47,22.29,94.07,0.87,62.94,79.4,72.87,50.22,98.51,7.17,90.3,62.69,72.42,40,39.31,69.16,43.2,9.23,63.3,2.25,59.87,24.57,81.73,43.93,77.07,33.23,86.23,42.3,11.83,56.43,31.88,18.26,105.76,10.95,28.68,34.55,32.97,51.5,93.06,62.88,7.74,67.94,2.69,16.03,83.97,43.46,77.66,92.38,61.01,77.7,97.17,57.38,40.12,91.31,58.18,35.31,46.23,84.9,38.47,81.35,12.72,69.86,-2.95,39.6,35.11,55.65,68.85,19.1,73.57,24.06,35.95,16.21,26.74,88.37,42.97,76.93,45.15,46.6,104.93,94.31,46.8,3.61,18.9)

y <- c(3.4,53.87,84.3,21.11,31.19,104.7,96.11,81.09,47.58,2.09,101.96,96.23,69.06,40.12,53.56,25.41,17.72,50.99,-0.42,29.99,27.66,86.86,23.13,82.74,23.51,13.95,58.02,50.64,91.59,67.94,64.98,67.9,38.6,32.6,60.91,32.95,66.25,10.14,56.84,33.05,-6.69,70.3,85.62,70.71,17.63,6.12,69.97,104.63,60.81,48.01,72.09,33.64,44.63,46.74,54.68,98.93,21.54,27.92,8.85,29.81,97.81,18.98,16.81,11.17,13.33,7.68,85.03,-0.35,-0.74,64.14,27.42,34.79,12.31,51.48,81.66,47.2,27.23,69.93,69.09,66.79,16.95,62.98,-1.82,82.35,12.16,17.47,7,90.28,74.39,63,7.59,9.54,61.17,29.94,1.29,53.56,7.8,39.38,79.67,11.45)

plot(x,y)

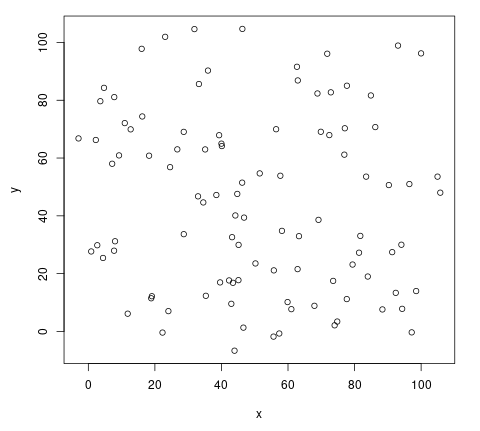


Рисунок 1 – Результат работы кода (визуализация облака точек)

Визуальный анализ графика, представленного на рисунке 1 позволяет предположить нелинейный характер регрессионного уравнения, следовательно, задача идентификации не может быть решена и необходимо построение регрессионного уравнения с полиномом третьего порядка.

**b) Код：**

x<-c(74.8,57.7,4.71,55.74,8.03,46.29,71.79,7.79,44.76,74.01,23.09,99.99,28.67,44.18,83.48,4.42,45.11,96.47,22.29,94.07,0.87,62.94,79.4,72.87,50.22,98.51,7.17,90.3,62.69,72.42,40,39.31,69.16,43.2,9.23,63.3,2.25,59.87,24.57,81.73,43.93,77.07,33.23,86.23,42.3,11.83,56.43,31.88,18.26,105.76,10.95,28.68,34.55,32.97,51.5,93.06,62.88,7.74,67.94,2.69,16.03,83.97,43.46,77.66,92.38,61.01,77.7,97.17,57.38,40.12,91.31,58.18,35.31,46.23,84.9,38.47,81.35,12.72,69.86,-2.95,39.6,35.11,55.65,68.85,19.1,73.57,24.06,35.95,16.21,26.74,88.37,42.97,76.93,45.15,46.6,104.93,94.31,46.8,3.61,18.9)

y <- c(3.4,53.87,84.3,21.11,31.19,104.7,96.11,81.09,47.58,2.09,101.96,96.23,69.06,40.12,53.56,25.41,17.72,50.99,-0.42,29.99,27.66,86.86,23.13,82.74,23.51,13.95,58.02,50.64,91.59,67.94,64.98,67.9,38.6,32.6,60.91,32.95,66.25,10.14,56.84,33.05,-6.69,70.3,85.62,70.71,17.63,6.12,69.97,104.63,60.81,48.01,72.09,33.64,44.63,46.74,54.68,98.93,21.54,27.92,8.85,29.81,97.81,18.98,16.81,11.17,13.33,7.68,85.03,-0.35,-0.74,64.14,27.42,34.79,12.31,51.48,81.66,47.2,27.23,69.93,69.09,66.79,16.95,62.98,-1.82,82.35,12.16,17.47,7,90.28,74.39,63,7.59,9.54,61.17,29.94,1.29,53.56,7.8,39.38,79.67,11.45)

plot(x,y)

library("lmtest")

fit <- lm(y ~ x + I(x^2) + I(x^3))

print(fit)

plot(x,y)

points(x, fitted(fit), col='red', pch=20)

coeftest(fit)

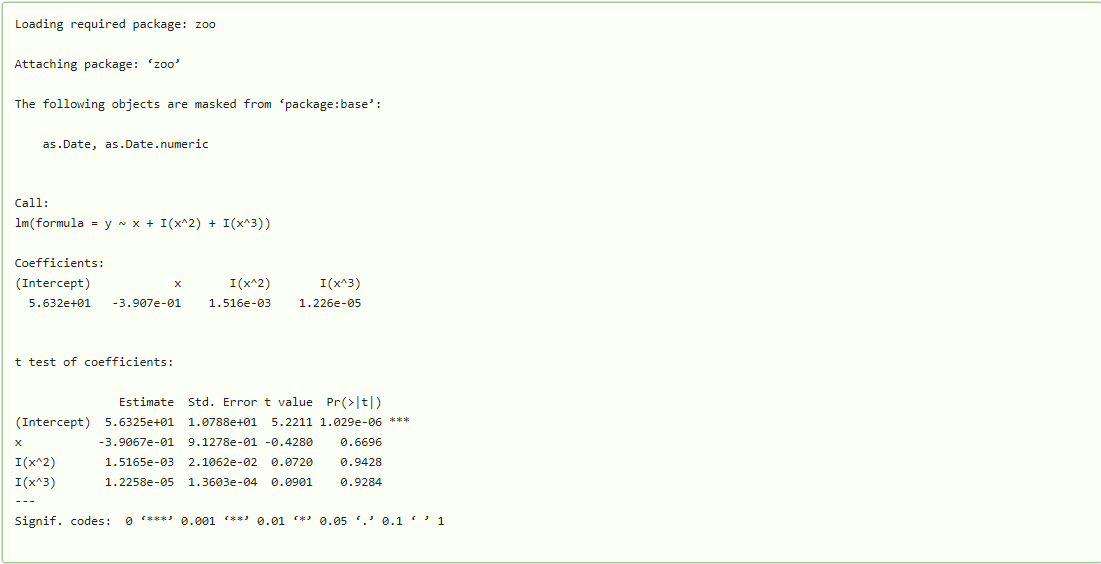


Рисунок 2 – Результат выполнения кода

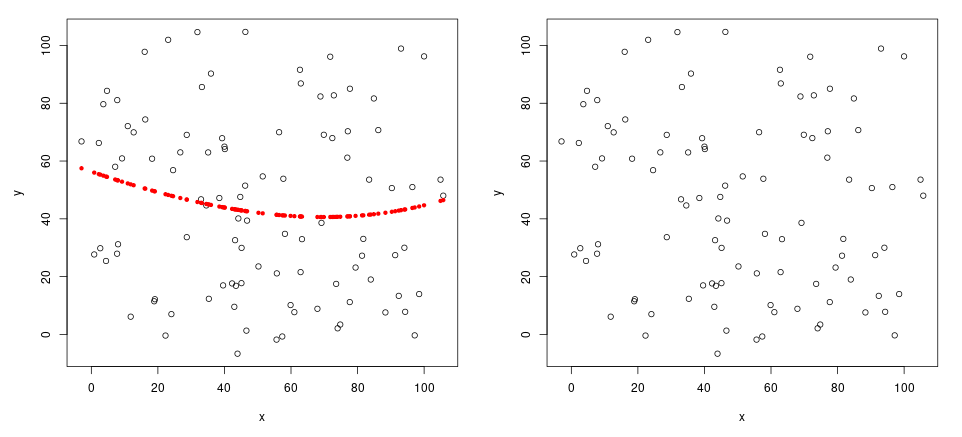


Рисунок 3 – Результат выполнения кода (построенный полином и облако точек)

Анализируя полученные данные из рисунка 2 можно заключить, что значения P-value для коэффициентов регрессионного полинома больше уровня значимости, то есть больше, чем 0,05 (y не зависит от x). Это означает, что мы принимаем гипотезу о равенстве нулю истинных значений коэффициентов регрессионного полинома, что говорит о том, что полученные значения статистически не значимы. Данную картину можно наблюдать и при попытке построения регрессионного уравнения с любым порядком полинома (порядок полинома в таком случае ни на что не влияет, т.к. значения P-value всегда будут больше чем 0,005 (см. рис. 4).

В статистике величину (значение) переменной называют статистически значимой, если мала вероятность случайного возникновения этой или ещё более крайних величин. В данном случае, полученные значения статистически незначимы. Означает это то, что полученные статистические значения достаточно часто будут повторяться (одинаковы или будут более крайними, при одних и тех же тестах будут одни и те же статистические результаты).

В результате было получено следующее регрессионное уравнение (см. результаты рис. 2):

Ответы на контрольные вопросы

**1.** Математическая модель — математическое представление реальности, один из вариантов модели как системы, исследование которой позволяет получать информацию о некоторой другой системе. Математическая модель предназначена предсказать поведение реального объекта, но всегда представляет собой ту или иную степень его идеализации.

Задача идентификации заключается в определении модели (т.е. ее структуры и параметров - параметрическая идентификация) или наилучшей аппроксимации характеристик системы (непараметрическая идентификация) по полученному экспериментально набору данных (записанных входных и выходных сигналов).

Задача идентификации сводится, в общем случае, к определению оператора модели, преобразующего входные воздействия объекта в выходные величины.

**2**. Корреляция в широком смысле слова означает связь, соотношение между объективно существующими явлениями и процессами.

Под причинной связью понимается такое соединение явлений и процессов, когда изменение одного из них происходит вследствие изменения другого. Связи между явлениями могут быть различны по силе

Если случайные переменные причинно обусловлены и можно в вероятностном смысле высказаться об их связи, то имеется корреляционная связь, или корреляция.

О функциональной связи говорят в том случае, когда все переменные в причинной связи детерминированы, и, как следствие, предполагается не вероятностная, а детерминированная связь.

Парная линейная корреляция является характеристикой меры линейной зависимости двух величин. При значении коэффициента корреляции зависимость линейная функциональная (одному значению величины соответствует одно значение величины ).

При уменьшении коэффициента корреляции одному значению величины x соответствует уже определенный закон распределения величины y, что, при визуализации выборочных данных, дает разброс значений величины y.

Чем ближе значение коэффициента корреляции к нулю, тем больше рассеяние.

Коэффициент корреляции, всегда будет находиться в интервале от -1 до 1, что обеспечивает универсальность характеристики тесноты связи и ее независимость от масштаба образующих зависимость величин.

Различают два вида зависимостей между явлениями и процессами: а) функциональную зависимость и б) стохастическую зависимость. В случае функциональной зависимости имеется однозначное отображение элементов множества А в элементы множества В. Множество А называют областью определения, а множество В – множеством значения функции. При стохастической зависимости каждому фиксированному значению аргумента соответствует определенное статистическое распределение значений функции. Это обусловлено тем, что зависимая переменная, кроме выделенной переменной, подвержена влиянию ряда неконтролируемых или неучтенных факторов, а так же погрешностью измерений. Поскольку значения зависимой переменной подвержены случайному разбросу, они не могут быть предсказаны с достаточной точностью, а только указаны с определенной вероятностью.

Регрессия – это односторонняя стохастическая зависимость.

Функция регрессии необратима.

Функция регрессии формально устанавливает соответствие между переменными, хотя они могут не состоять в причинно-следственных отношениях.

При анализе зависимости двух переменных строят диаграмму рассеяния (облако точек).

По оси абсцисс отмечают значения независимой переменной, по оси ординат - зависимой. Результат каждого наблюдения отображается точкой на плоскости. Совокупность этих точек образует скопление (облако). Облако точек определяет картину зависимости двух переменных. По ширине разброса точек можно сделать вывод о степени тесноты связи.

Для того, чтобы по диаграмме рассеяния графическим путем определить функцию регрессии, нужно натянуть воображаемую нить так, чтобы по обе стороны от нее оказалось приблизительно одинаковое число точек.

Значения функции регрессии y ̂\_i указывают средние значения зависимой переменной y при заданном значении x\_i объясняющей переменной x.

**3.** Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака у от теоретических минимальна.

Уравнение регрессии - то математическая формула, применяемая к независимым переменным, чтобы лучше спрогнозировать зависимую переменную, которую необходимо смоделировать.

Коэффициенты регрессии (β) - это коэффициенты, которые рассчитываются в результате выполнения регрессионного анализа. Вычисляются величины для каждой независимой переменной, которые представляют силу и тип взаимосвязи независимой переменной по отношению к зависимой.

**4.** Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Нелинейная регрессия — это вид регрессионного анализа, в котором экспериментальные данные моделируются функцией, являющейся нелинейной комбинацией параметров модели и зависящей от одной и более независимых переменных.

Различают два класса нелинейных регрессий. К первому классу можно отнести регрессии, нелинейные относительно включенных в анализ объясняющих переменных x\_k, но линейные по неизвестным, подлежащим оценке параметрам регрессии b\_k. Образующие этот класс нелинейные регрессии называют так же квазилинейными регрессиями. Для них возможно непосредственное применение МНК.

Второй класс регрессий характеризуется нелинейностью по оцениваемым параметрам. Этот класс регрессий часто встречается при исследовании экономических явлений и не допускает применения обычного МНК. Однако некоторые функции с помощью преобразования переменных поддаются линеаризации и последующему применению МНК.

**5.** Проверкой статистической гипотезы о значимости отдельных параметров модели называется проверка предположения о том, что данные параметры значимо отличаются от нуля. Необходимость проверки гипотез о значимости параметров модели вызвана тем, что в дальнейшем построенную модель будут использовать для дальнейших экономических расчётов.

При проверке значимости коэффициентов множественной регрессии выдвигают гипотезы:

H\_0:b\_k=β\_k, т. е. нет существенного различия между оценкой параметра регрессии, полученной по результатам выборки b\_k и истинным значением параметра регрессии генеральной совокупности β\_k;

H\_0:b\_k≠β\_k, т.е. имеется значимая разница между оценкой параметра регрессии и соответствующим значением генеральной совокупности.

Вывод

Была изучена одна из задач математического моделирования – задача идентификации. Были закреплены практические навыки решения задачи идентификации с использованием инструментария регрессионного анализа и пакетами, предоставляемые средой разработки языка R.

Был сделан вывод о том, что задача идентификации не может быть решена, поскольку зависимости y от x не обнаружено. Из этого следует, что значения p-value для коэффициентов регрессионного уравнения будут больше уровня значимости ∝=0,05, что было показано с использованием необходимого инструментария. Данные полученные в результате регрессионного анализа статистически не значимы и все гипотезы о значимости полинома отвергнуты.