[线性最小二乘拟合程序]

结题文档

**填写时间**

二零二二 年 12 月

目录

[1 引言 1](#_Toc494135160)

[1.1 背景 1](#_Toc494135161)

[1.2 定义 1](#_Toc494135162)

[2 详细设计 1](#_Toc494135164)

[2.1 概述 1](#_Toc494135165)

[2.2 功能 1](#_Toc494135166)

[2.3 方案设计与实现 1](#_Toc494135167)

[2.4 尚未解决的问题 2](#_Toc494135168)

[3 分工与时间安排 2](#_Toc494135170)

[4 项目小结 2](#_Toc494135171)

[4.1 开发平台及辅助工具 2](#_Toc494135174)

[4.2 开发小结与心得体会 3](#_Toc494135175)

* 1. 引言
     1. 背景

说明：为确定一组具有线性关系的数据对应的具体表达式，编写此程序

待开发软件系统的名称；线性最小二乘拟合程序

本项目开发者：余亚飞 王皓冬 陈思喆 安珂

* + 1. 定义

最小二乘法：是一种在误差估计、不确定度、系统辨识及预测、预报等数据处理诸多学科领域得到广泛应用的数学工具。

离差：离差(dispersion)是随机变数的值(即一组数据)关于某个中心值(通常取为数学期望)偏离或散布的离散程度的一种标志。

相关系数：是研究变量之间线性相关程度的量。

作用域：限定某个名字的可用性的代码范围。

abs：即absolute value 绝对值函数。

* 1. 详细设计

从本章开始，展开软件系统的详细设计。以下给出的提纲仅供参考。

* + 1. 概述

该程序根据预设或输入的两组数据求出其线性最小二乘拟合的方程，输出对应离差，生成并打开excel文件。试图达成检索原数据中是否存在错误数据的目的。

* + 1. 功能

本程序具有以下功能：

计算出给出的自变量及因变量间的相关方程，并对应输出拟合数据及离差；

生成excel文件，便于从图像直观观察拟合情况；

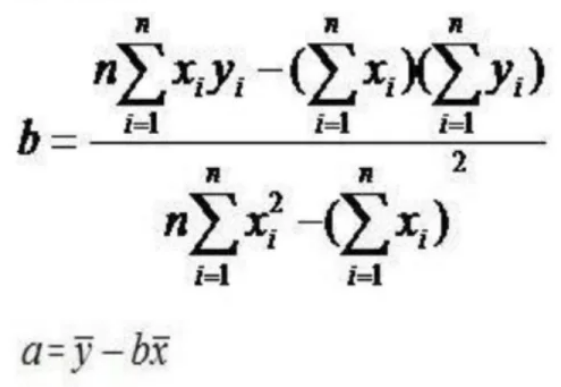
选择数据来源：预设数据或自行输入；

检查数据是否可疑或是否适合线性拟合。

* + 1. 方案设计与实现

·计算相关方程：

主要包含计算环节。首先经由一定程度的数学推导，得出在两组n个数据时的相关方程斜率与相关系数的普适表达形式：



继而将数学形式转换为编程语言。其中循环次数及数据数目的统一与迁移利用了#define，需要更改数据总量时可直接对此数据更改：

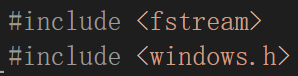


同时，为保证输出数据便于观察，调用iomanip头文件中的setw（）函数，数据具体宽度数值利用#define，以防止空隙过大或数据溢出：

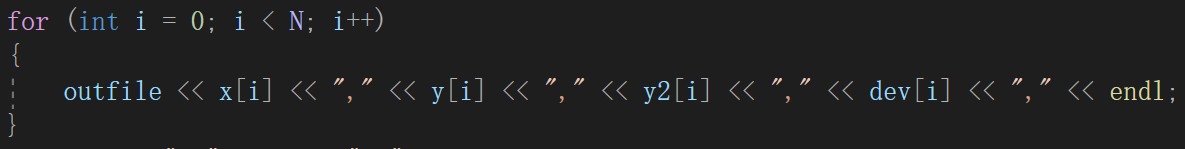
。

·生成Excel文件：

引入头文件：

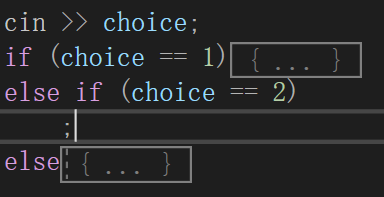


并按语法编写程序。逐个向Excel表格中投入数据时应利用了各数组数据数目相同的特征，采用单个for循环解决：

。

·自主选择数据来源：

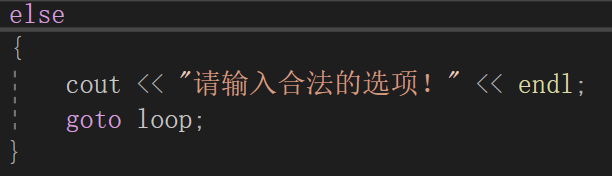
由操作者按照文字提示输入选项，以判断输入选项为架构，采用if-else语句为大框架实现：



而考虑到if语句中定义的数组作用域仅在if内，无法利用if完成不同情况下两种数组的定义。因此采取数值传递的思路完成该目标。

具体操作如下，首先在if语句前完成预设值数组的初始化，同时另定义两个未初始化的数组。当判定为“自行输入数据”时，则利用for循环cin完成对未初始化数组的初始化。完成该过程后，再次采用for循环将原数组数据对应一一覆盖。当判定为“采用预设数据”时，则执行空语句。

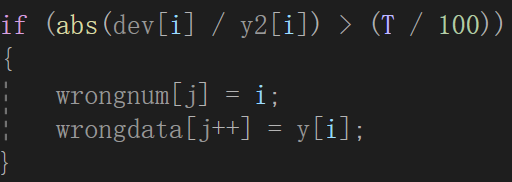
考虑到输入的选项可能不符合规格，当判定输入值不为1或2时，将会进行文字提示，并采用goto语句返回选项输入的语句：



。

·检查数据是否可疑：

计算出离差后，希望通过离差判定原数据偏离拟合数据幅度是否过大，从而一定程度上判断数据不适合线性拟合或输入数据有误。对原数组对应离差占拟合数据比重依次审查，若超出一定幅度则判定为“可疑”，并将数据对应的自变量及因变量分别储存在两个数组中：



利用了cmath头文件中的abs函数，避免离差及拟合数据的正负影响结果。

* + 1. 尚未解决的问题

本程序中检查可疑数据模块可以一定程度上提示存在可疑数据，但无法精确检查，原因是当至少有一个数据大幅偏离时，其回归方程也会随之偏离，导致可能存在多个数据均满足偏离拟合数据幅度足够大而被判定为错误数据。可能的解决方案是：其一，对a个判定为可疑的数据依次删除后判定剩余数据形成的回归方程是否存在可疑数据，a的个数随循环减少，直到判定为仅有1个数据可疑；其二，依次比较所有可疑数据的偏离幅度并求均值，超出均值一定幅度者判定为可疑。

* 1. 分工与时间安排

1. 程序结构设计及框架编写
   陈思喆

根据要求完成程序的初步设计，包括需调用的头文件、初步需编写的函数、程序总体流程设计等方面的内容；编写程序的大致框架

1. 函数编程
   余亚飞

根据所需编写所有定义的函数，同时保证函数对不同数组数据的普适性；并在main函数编程的过程中通过函数简化部分编程

1. main函数编程
   王皓冬

完成程序主体功能的编程，调用统合所引用头文件中的函数及已定义的函数，通过语句尽量简化地实现程序的所有功能

1. 汇总修正及调整

安珂

根据预定的目标与设计及时反馈各板块人员进度及效果，并在初版程序完成后对不必要的参数等进行删改，修正语句、格式及文字提示

* 1. 项目小结
     1. 开发平台及辅助工具

开发平台：Microsoft Visual.Studio.2022

辅助工具：Excel

两款实用的软件，且我们对此熟悉程度高。Microsoft Visual.Studio.2022具备完善的开发环境与调试手段，为程序编写提供便利。Excel中的表格与散点图可以更直观地让开发者对结果进行验证。

* + 1. 开发小结与心得体会

在本次项目实践中，我们小组应用课堂知识与课外查阅资料编写程序，并对其进行不断改进，实现了线性最小二乘拟合程序的编写。在方案论证方面，我们对最小二乘法对应的几个公式进行比较，在充分考虑精确性与可行性的基础上进行方案论证，最终得出切实可行的方案。这不仅锻炼了我们的分析能力，同时也使得看问题的角度更加全面。在开发平台的选择上，我们采用较为熟悉的Microsoft Visual.Studio.2022。好的开发平台既有利于程序编写，也为后续的调试与改进提供便利。在团队分工方面，本小组四人经商讨合理分配工作，主动把握自己擅长的部分，最后圆满完成本次作业。总而言之，经过此次任务，我们对程序编写的掌握更加熟练，对将课本知识投入实践有了更进一步的认识。同时，大量课外资料的查阅，也令我们接触到新鲜的课外知识，这对丰富我们的头脑大有裨益。组内合作，锻炼了我们的沟通能力与团队协作能力。本次软件开发，确实是一次各方面能力的锻炼，我们从中受益匪浅。