实验一 数据处理编程

姓名：梁家祥 学号:1307200138 班别：电子131

**实验目的**

1、初步掌握工程或科研项目中测量数据处理的理论、步骤、方法和技巧。

2、掌握测量数据处理中常用的计算机语言和软件。

**实验要求**

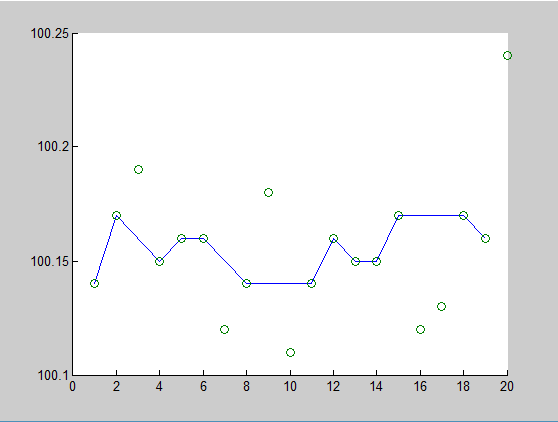
通过此实验要求掌握测量数据的处理方法，并应用Matlab编程实现， 在测量中，按照误差的特点与性质，误差可分为：系统误差，粗大误差和随机误差，借助MATLAB对测量数据进行处理，使处理过程快速、结果可靠。

**实验内容**

程序处理测量数据的过程(参考)如下：

1. 按测量的先后顺序记录下每个测量值Xi，要求可以从数据文件中导入，数据文件格式见附件一，显示导入的测量数据。
2. 判断是否有粗大误差，若有，显示该值并剔除，若无，显示该项检查合格。
3. 判断是否有系统误差，若有，显示依据。将测量数据绘制折线图或点阵图供人工观察。
4. 计算算术平均值、标准差。
5. 给出测量结果。

**实验结果**



空心圆圈代表数据点，用线连起来的数据点表示有效数据点，孤立的数据点为被剔除的粗大误差坏值。

Data为输入数据。arithmetic mean表示算术平均值，Standard deviation为标准差，

NO.3 Value:100.19表示被剔除的坏值是第三次测量的结果，其值为100.19

Matlab的命令行输出结果：

data =

100.1400

100.1700

100.1900

100.1500

100.1600

100.1600

100.1200

100.1400

100.1800

100.1100

100.1400

100.1600

100.1500

100.1500

100.1700

100.1200

100.1300

100.1700

100.1600

100.2400

arithmetic mean :100.1555

Standard deviation :0.028013

bad value found. NO.3 Value:100.19

bad value found. NO.7 Value:100.12

bad value found. NO.9 Value:100.18

bad value found. NO.10 Value:100.11

bad value found. NO.16 Value:100.12

bad value found. NO.17 Value:100.13

bad value found. NO.20 Value:100.24

Marley Kopf criterion delta:-0.04

>>

Marley Kopf criterion delta:-0.04

**实验过程**

**一、读取数据**

要处理数据，就要先解决输入导入的问题。设计要求的数据格式如下：

根据观察，可知第一个数字决定了数值的长度，并起到判断文件的数据是否完整的作用。而数据的排布格式为’%n%s’，每一行上是一个数据加一个字符串。

20 //测量数值的总数

100.3 //第一个测量值

100.5 //第二个测量值

……

因此只需使用textscan函数即可格式化的读入数据。

具体的操作为：

fid = fopen('data.txt');

if fid<0

display('open file failed.');

break;

end

text = textscan(fid,'%n%s');

fclose(fid);

data\_o = text{1,1}(:); %%根据文件的格式，提取出第一列数据

得到数据矩阵之后，需要除去第一个数据，并判断数据是否有效。因此做如下操作：

datasize = data\_o(1,1);

data = data\_o(2:size(data\_o),1);

if datasize ~= size(data) %%如果读入的数据长度与声明的长度不符。退出程序

display('size of data unmatch.');

break;

end

至此得到数据长度datasize与数据data。

**二、剔除粗大误差**

粗大误差可以根据格拉布斯准测判断。

把数学公式转换为matlab实现：

k = 0;

for i=1:datasize

k = k + power((data(i, 1) - x\_mean),2);

end

sigma = sqrt(k / (datasize - 1));

sigma = sigma / sqrt(datasize);

%%求均方差sigma

Gsigma = 2.56 \* sigma; %%根据格布拉斯准则 查表得当n=20 p=0.95时 G=2.56.

即可通过测量值相对平均值的误差绝对值与计算出来的Gsigma比较，判断测量值是否为坏值。

for i=1:datasize

if abs(data(i,1) - x\_mean) > Gsigma

data(i,2) = 1; %%标记出坏值

end

end

cnt = 1;

data\_a = [20,1]; %%好值的缓存

for i=1:datasize %%把标记出来的坏值打印出来，好值则存储到data\_a

if data(i,2) == 0

data\_a(cnt,1) = data(i,1);

cnt = cnt + 1;

else

display(['bad value found. NO.',num2str(i),

' Value:',num2str(data(i,1))]);

end

end

if cnt == datasize %%如果没有坏值

display('No bad value.');

end

**三、判断是否存在系统误差**

由于测量数据为游标卡尺测量木块厚度所得，估计系统主要的误差为累进性误差。

因此对输入值采用马利科夫判据计算。

if rem(datasize\_a,2) == 0

delta = sum(data\_a(1:datasize\_a/2,1)) –

sum(data\_a(datasize\_a/2:datasize\_a,1));

else

delta = sum(data\_a(1:(datasize\_a+1)/2,1)) –

sum(data\_a((datasize\_a+1)/2:datasize\_a,1));

end

**实验心得和体会**

本次实验要求从文档读入数据，再对其进行误差分析，一个首当其冲的难点就是从格式化的文本中提取出数据。在尝试importdata, textread，fread之后，都显得太过复杂。就上网搜索格式化输入的函数有哪些，果然找到了textscan这个方便好用的函数。成功把文本中的信息分类存储到workspace。

至于系统误差和粗大误差的检测没有遇到很大的问题，只要照着数学公式去做matlab实现即可。但依然有些疑问，在对累进性系统误差的判定时，通过采用马利科夫判据计算出的delta近似等于零，则可判定为测量数据中不含有累进性误差。但是这个近似是多少，数值上不好把握。

在读取文件与进行运算时，增加了数值有效性判断，这对本程序实用化非常有意义。可以防止数据格式错误或者输入了非法值时程序有错误输出。

虽然程序简单，但是我感觉这样的作业非常有意义。在实际操作中体会了书中的概念与公式的实际意义。做得过程中反复查看书籍，既获取了书中的知识，又能复习matlab，可谓一举两得。希望能多做些结合真实环境的题目，少一点书本上的习题。

**附录一：数据文本data.txt的内容**

**20 //测量数值的总数**

**100.14 //第1个测量值**

**100.17 //第2个测量值**

**100.19 //第3个测量值**

**100.15 //第4个测量值**

**100.16 //第5个测量值**

**100.16 //第6个测量值**

**100.12 //第7个测量值**

**100.14 //第8个测量值**

**100.18 //第9个测量值**

**100.11 //第10个测量值**

**100.14 //第11个测量值**

**100.16 //第12个测量值**

**100.15 //第13个测量值**

**100.15 //第14个测量值**

**100.17 //第15个测量值**

**100.12 //第16个测量值**

**100.13 //第17个测量值**

**100.17 //第18个测量值**

**100.16 //第19个测量值**

**100.24 //第20个测量值**