

机械工程学院实验报告

Experiment Report

School of Mechanical Engineering

实验课程名称 (Experiment Course)		数值分析和优化算法					
实验项目名称 (Experiment Item)		描述性分析实验					
姓名 (Name)	宋天佑	学号 (No.)	20211414 10279	系别 (Department)	机械设计制造及其自动化	班级 (Class)	2 班
同组学生姓名 (Accompaniers)							
实验日期 (Date)	2025.3.24	实验地点 (place)	CAD 中心 202	实验学时 (Hour)	3		
指导教师姓名 (Superior)	赵秀粉	成绩 (Grade)		指导教师签名 (Superior Signature)			

(以下是实验报告的参考项目，请学生按照实验指导教师的要求在空白处依次填写，不够加附页)

一、实验目的 (Experiment Objectives)

- 1、熟练掌握利用 MATLAB 软件计算均值、方差、协方差、相关系数、标准差与变异系数、偏度与峰度、中位数、分位数、三均值、四分位极差与极差。
- 2、熟练掌握 jbstest 与 kstest 关于一维数据的正态性检验。
- 3、掌握统计作图方法。
- 4、掌握多维数据的数字特征与相关矩阵的处理方法。

二、实验仪器与设备 (包括名称、型号、规格等) (Main Experiment Apparatus)

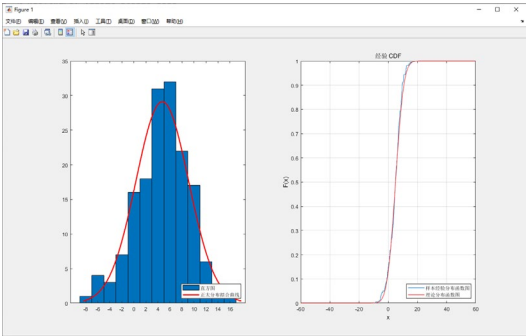
安装有 MATLAB 的计算机。

三、实验内容、步骤与结论 (Experiment Contents, Experiment Steps and Result)

实验 1:

```
1 %  
2 % -*- coding: utf-8 -*-  
3 % @line : 2025/3/23 下午4:05  
4 % @author : Sophomores  
5 % @file : e1.m  
6 % @Software: Visual Studio Code  
7 %  
8 clc, clear  
9 x = normrnd(5, sqrt(20), [1, 100]); % 正态分布随机样本  
10 % x = 5+20*randi(1,100);  
11 x_mean = mean(x);  
12 x_range = range(x);  
13  
14  
15 set(gcf, 'unit', 'normalized', 'position', [0.2, 0.2, 0.5, 0.5]); % 其在屏幕上的显示, 四个参数分别表示的是左边界, 下边界, 高度, 宽度。  
16 % 绘制直方图  
17 subplot(1, 2, 1); histogram(x);  
18 subplot(1, 2, 1); histfit(x);  
19 legend('直方图', '正态分布拟合曲线', 'Location', 'southeast')  
20  
21 % 绘制样本的经验分布函数  
22 subplot(1, 2, 2); [h, stats] = cdfplot(x);  
23 hold on  
24 % 绘制理论分布函数  
25 plot(-60:0.01:60, normcdf(-60:0.01:60, 5, sqrt(20)), 'r');  
26 legend('样本经验分布函数', '理论分布函数图', 'Location', 'southeast')  
27 hold off
```

主函数



输出

实验2

```
1 %%%
2 % *- coding: utf-8 -*-
3 % @time : 2025/3/23 下午4:05
4 % @Author : Sophomores
5 % @File : e2.m
6 % @Software: Visual Studio Code
7 %%%
8 clc, clear
9 load data2.mat
10 x = data2(:, 2:end);
11 [r, c] = size(x);
12 % 将标题保存为字符, 放在元胞数组中
13 x_label = {'生铁产量(万吨)', '金属切削机床产量(万台)', '汽车产量(万辆)'};
14
15 %%% 1)
16 disp(['第一题' newline])
17 [mean_x, triple_mean_x, std_x, v_x, sk_x, ku_x] = func2(x);
18 for i = 1:c
19     disp(['第' num2str(i) '个指标 < ' x_label{1, i} ' > 的统计量'])
20     disp(['平均值为: ' num2str(mean_x(1, i))])
21     disp(['三均值为: ' num2str(triple_mean_x(1, i))])
22     disp(['标准差为: ' num2str(std_x(1, i))])
23     disp(['变异系数为: ' num2str(v_x(1, i))])
24     disp(['偏度为: ' num2str(sk_x(1, i))])
25     disp(['峰度为: ' num2str(ku_x(1, i))])
26
27     if i < 3
28         disp(repmat('-', 1, 50))
29     end
30 end
31
32 %%% 2)
33 disp(['第二题' newline])
34 disp(['# ' repmat('#', 1, 50)])
35
36 distribution_test_value = zeros(3, 3);
37 % distribution_test_value 3行3列
38 % 1, 2, 3列分别表示指标 '生铁产量(万吨)', '金属切削机床产量(万台)', '汽车产量(万辆)'
39 % 第1行表示 JB检验的值 0为接受正态分布假设, 1为不接受正态分布假设, 下同
40 % 第2行表示 KS检验的值
41 % 第3行表示 改进KS检验的值
42 f1 = figure('Name', '各项指标的Q-Q图');
43 set(gcf, 'unit', 'normalized', 'position', [0.2, 0.2, 0.5, 0.5])
44 % i表示指标所在的列
45 for i = 1:c
```

```
33 %%% 2)
47 distribution_test_value(1, i) = jbstest(x(:, i));
48 distribution_test_value(2, i) = kstest(x(:, i));
49 distribution_test_value(3, i) = lillietest(x(:, i));
50 figure(f1)
51 subplot(1, 3, i); qqplot(x(:, i))
52 title([x_label{1, i} '的Q-Q图'])
53 ylabel([x_label{1, i} '的分位数'])
54 end
55
56 % 指标 不符合正态分布的次数
57 judge_value = sum(distribution_test_value, 1);
58 % 不符合正态分布的列, 符合为1, 不符合为0
59 not_norm_col = ones(1, 3);
60
61 for i = 1:c
62
63     if judge_value(1, i) > 0
64         disp(['指标 < ' x_label{1, i} ' > 不服从正态分布'])
65         not_norm_col(1, i) = 0;
66     else
67         disp(['指标 < ' x_label{1, i} ' > 服从正态分布'])
68         % 计算95%概率时的值, 不是计算95%分位数的值
69         % 在理论层面, 正态分布的95%分位数即为其CDF达到95%时的值. 两者等同.
70         % 实际样本中, 若数据严格服从正态分布且样本量充分, 两者近似等同
71         disp(['概率为95%时的' x_label{1, i} '为' prctile(x(:, i), 95, 'all')])
72         disp(['概率为95%时的' x_label{1, i} '为' prctile(x(:, i), 95, 'all')])
73     end
74 end
75
76 % 对不符合正态分布的列 进行box-cox变换, 需要提前初始化矩阵
77 not_norm_num = sum(not_norm_col == 0); % 不符合正态分布的指标个数
78
79 % x_b用于保存变换后的指标列
80 x_b = zeros(r, not_norm_num(1, 1));
81
82 f2 = figure('Name', '不符合正态分布的指标的原始数据与变换数据的Q-Q图');
83 set(gcf, 'unit', 'normalized', 'position', [0.2, 0.2, 0.5, 0.5])
84
85 for i = 1:c
86
87     if not_norm_col(1, i) == 0
88
89         disp(['正在使用Box-Cox对指标 < ' x_label{1, i} ' > 进行变换'])
90         [x_b(:, i), t] = boxcox(x(:, i));
```

主函数_1

```
33 %%% 2)
92 figure(f2)
93 subplot(2, not_norm_num(1, 1), i); normplot(x(:, i)) % 原始数据的QQ图
94 title([x_label{1, i} '原始数据的Q-Q图'])
95
96 subplot(2, not_norm_num(1, 1), i + 3); normplot(x_b(:, i)) % 变换数据的QQ图
97 title([x_label{1, i} '变换数据的Q-Q图'])
98 end
99
100 end
101
102 % 打印变换后的指标的统计量数据, 未变换的不打印
103 [mean_x_b, triple_mean_x_b, std_x_b, v_x_b, sk_x_b, ku_x_b] = func2(x_b);
104 disp(repmat('-', 1, 50))
105
106 for i = 1:size(x_b, 2)
107     disp(['第' num2str(i) '个不符合正态分布的指标 < ' x_label{1, i} ' > 变换后的统计量'])
108     disp(['平均值为: ' num2str(mean_x_b(1, i))])
109     disp(['三均值为: ' num2str(triple_mean_x_b(1, i))])
110     disp(['标准差为: ' num2str(std_x_b(1, i))])
111     disp(['变异系数为: ' num2str(v_x_b(1, i))])
112     disp(['偏度为: ' num2str(sk_x_b(1, i))])
113     disp(['峰度为: ' num2str(ku_x_b(1, i))])
114     if i < 3
115         disp(repmat('-', 1, 50))
116     end
117 end
118
119 %%% 3)
120 d = mahal(x, x); % 计算马氏距离
121 d1 = sort(d); % 从小到大排序
122 pt = ((1:r) - 0.5) / r; % 计算分位数
123 x2 = chi2inv(pt, c); % 计算卡方
124 f3 = figure('Name', '3项指标的多维正态检验Q-Q图');
125 set(gcf, 'unit', 'normalized', 'position', [0.2, 0.2, 0.5, 0.5])
126 figure(f3)
127 plot(d1, x2, '*', 0:10, 0:10, '-r')
```

主函数_3

第一题

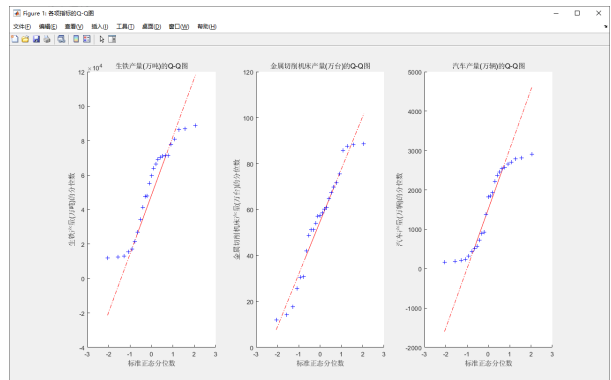
第1个指标 < 生铁产量(万吨) > 的统计量
平均值为: 52357.7544
三均值为: 54074.1837
标准差为: 26363.125
变异系数为: 0.50352
偏度为: -0.30949
峰度为: 1.6764

第2个指标 < 金属切削机床产量(万台) > 的统计量
平均值为: 54.8572
三均值为: 56.03
标准差为: 22.9909
变异系数为: 0.4191
偏度为: -0.32668
峰度为: 2.2433

第3个指标 < 汽车产量(万辆) > 的统计量
平均值为: 1525.8212
三均值为: 1671.8225
标准差为: 1027.0702
变异系数为: 0.67313
偏度为: -0.059964
峰度为: 1.35

第1)题

主函数_2



第1)题 图

```

第二期

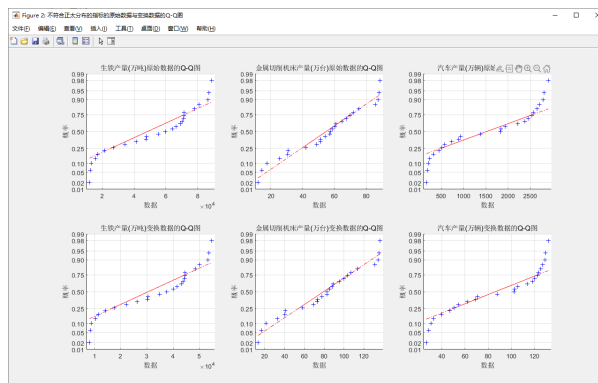
指标 < 生铁产量(万吨)> 不服从正太分布
指标 < 金属切削机床产量(万台)> 不服从正太分布
指标 < 汽车产量(万辆)> 不服从正太分布
正在使用Box-Cox对指标 < 生铁产量(万吨)> 进行变换
正在使用Box-Cox对指标 < 金属切削机床产量(万台)> 进行变换
正在使用Box-Cox对指标 < 汽车产量(万辆)> 进行变换

-----
第1个不符合正太分布的指标 < 生铁产量(万吨)> 变换后的统计量
平均值为: 32889.4019
三均值为: 34032.9742
标准差为: 16015.5237
变异系数为: 0.48788
偏度为: -0.33277
峰度为: 1.6858

-----
第2个不符合正太分布的指标 < 金属切削机床产量(万台)> 变换后的统计量
平均值为: 80.0899
三均值为: 81.215
标准差为: 36.4949
变异系数为: 0.46192
偏度为: -0.22488
峰度为: 2.1917

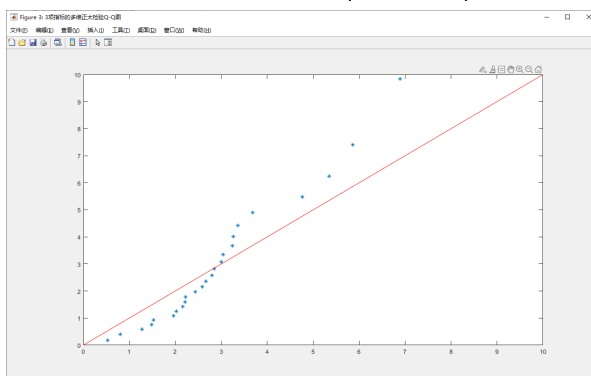
-----
第3个不符合正太分布的指标 < 汽车产量(万辆)> 变换后的统计量
平均值为: 86.091
三均值为: 94.3974
标准差为: 38.4545
变异系数为: 0.44667
偏度为: -0.28595
峰度为: 1.4796

```



第 2)题

不符合正太分布的指标
原始数据与 Box - Cox 变换数据的 Q - Q 图对比



3)总体指标的多维正态分布检验 Q - Q 图检验

结论: 总体不服从三维正态分布

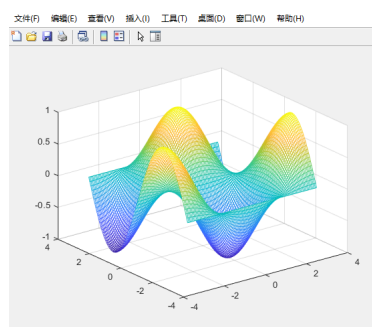
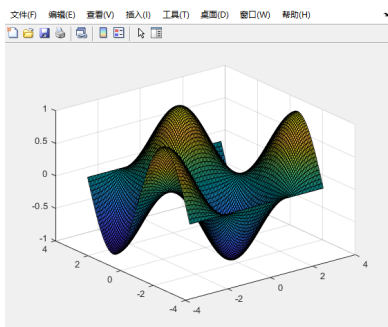
实验 3

```

e3.m >
1  %%%
2  % -*- coding: utf-8 -*-
3  % @time : 2025/3/23 下午4:05
4  % @Author : Sophomores
5  % @File : e3.m
6  % @Software: Visual Studio Code
7  %%%
8  clc, clear
9  n = 100;
10 [x, y] = meshgrid(linspace(-pi, pi, n));
11 z = cos(x) .* sin(y);
12 f1 = figure();
13 figure(f1);
14 mesh(x, y, z)
15
16 f2 = figure();
17 figure(f2);
18 surf(x, y, z)

```

主函数



输出

四、实验思考题

1、各项统计指标反映代表的意义？

统计指标	反映的意义
平均值	数据的中心位置，所有数据点的算术平均，反映数据的集中趋势。易受极端值影响。
三均值	结合四分位数和中位数的稳健中心度量，对异常值不敏感。
标准差	数据离散程度的绝对度量，表示数据偏离平均值的平均距离。值越大，数据越分散。
变异系数	相对离散程度。用于比较不同量纲或均值差异较大的数据集的波动性。
偏度	数据分布的不对称性。正偏（右偏）表示右侧尾部较长，负偏（左偏）表示左侧尾部较长。偏度为 0 时分布对称（如正态分布）。
峰度	数据分布的峰态和尾部厚薄。高峰度 (>3) 表示尖锐峰和厚尾，低峰度 (<3) 表示平缓峰和薄尾。常以正态分布（峰度=3）为基准比较。

2、正态性检验的方法是什么？

检验方法	原理	适用场景
Jarque-Bera 检验	基于样本数据的偏度和峰度，检验其是否符合正态分布的偏度和峰度	适用于大样本数据 ($n>50$) 的正态性检验
Kolmogorov-Smirnov 检验	通过比较样本累积分布函数 (ECDF) 与理论正态分布累积分布函数 (CDF) 的最大垂直距离，判断是否来自正态分布	适用于任意分布的比较
改进 KS 检验	KS 检验的改进版，针对参数未知的正态分布（均值和标准差由样本估计） 调整临界值以提高准确性	适用于正态分布的参数未知且需从样本中估计的场景

3、多维数据如何实现数据变化？如何基于数据变换给出综合评价模型？

多维数据可通过无量纲化、标准化、主成分分析、Box - Cox 等方法进行变换，基于变换后的数据可采用层次分析法、熵权法等方法构建综合评价模型。

4、描述多维数据的正态分布检验的方法有哪些？

多维数据的正态分布检验方法包括 Q - Q 图检验、Mardia 检验、Henze-Zirkler 检验、Royston 检验和能量距离检验等多元正态性检验方法。

五、实验收获与建议

通过本次实验，我系统掌握了多维数据的统计分析方法，包括：

- 1) 各类统计指标（均值、方差、偏度、峰度等）的计算与解读；
- 2) 单变量正态性检验方法（Jarque-Bera 检验、Kolmogorov-Smirnov 检验等）；
- 3) 图形化检验手段（QQ 图、PP 图的绘制与判读）。
