### 第二章作业

##### 10193903446 汤博

*我比较习惯使用 markdown对含有代码的作业进行排版，如果没法看的话可以通过 Visual Studio Code 打开或者 Typora 打开（我比较喜欢 Typora，因为可以清晰地看见排版），如果老师和助教觉得麻烦的话，我会每次都建立一个 docx 存放我的 markdown文档说明截图。*

1. 计算白丁香和紫丁香的展叶期在同一个站点的平均值、中位数，极差，离差、离差平方和、方差和标准差、变异系数；

* 使用table存储打开的资料
* clear;clc;  
  dx = readtable("丁香展叶期数据.xlsx","VariableNamingRule","preserve");
* method 1
* 使用group statistics进行分析，功能有限，不能加入自定义功能

WandP = dx(:,["站点名ID","展叶期"]);   
statarray = grpstats(WandP,"站点名ID",{'mean','median','range','std','var'})

method 2

使用 findgroups 和 splitapply 两个函数，对table进行分类，并且对每一类分类结果运用自定义函数（离差等），并且将结果填入初始表中。

G = findgroups(categorical(table2array(dx(:,"站点名ID"))));   
dx.("分组") = G;   
% 平均数   
meanTime = splitapply(@mean, dx.("展叶期"), G)   
dx.("平均数") = cell2mat(splitapply(@(x) {ones(size(x))\*mean(x)}, dx.("展叶期"), G));   
   
% 中位数 median   
medianTime = splitapply(@median, dx.("展叶期"), G)   
dx.("中位数") = cell2mat(splitapply(@(x) {ones(size(x))\*median(x)}, dx.("展叶期"), G));   
   
% 极差 range   
rangeTime = splitapply(@range, dx.("展叶期"), G)   
dx.("极差") = cell2mat(splitapply(@(x) {ones(size(x))\*range(x)}, dx.("展叶期"), G));   
   
% 离差   
deviationTime = splitapply(@(x){(x-min(x))/(max(x)-min(x))}, dx.("展叶期"), G)   
dx.("离差") = cell2mat(deviationTime);   
   
% 离差平方和   
sumdeviationTime = splitapply(@(x) sum(x-min(x))/(max(x)-min(x)), dx.("展叶期"), G)   
dx.("离差平方和") = cell2mat(splitapply(@(x) {ones(size(x))\*sum(x-min(x))/(max(x)-min(x))}, dx.("展叶期"), G));   
   
% 方差   
varianceTime = splitapply(@var, dx.("展叶期"), G)   
dx.("方差") = cell2mat(splitapply(@(x) {ones(size(x))\*var(x)}, dx.("展叶期"), G));   
   
% 标准差   
standardDivTime = splitapply(@std, dx.("展叶期"), G)   
dx.("标准差") = cell2mat(splitapply(@(x) {ones(size(x))\*std(x)}, dx.("展叶期"), G));   
   
% 变异系数   
coevTime = splitapply(@(x) std(x)./mean(x), dx.("展叶期"), G)   
dx.("变异系数") = cell2mat(splitapply(@(x) {ones(size(x))\*(std(x)./mean(x))}, dx.("展叶期"), G));   
   
writetable(dx,"第一题第一问.csv");

结果已经填入第一题第一问.csv中。

1. 再计算紫丁香在所有站点的偏度系数、峰度系数

dxp = readtable("丁香展叶期数据.xlsx","VariableNamingRule","preserve");   
rows = dxp.("物种") == "紫丁香";   
dxp = dxp(rows,:);   
   
% 偏度系数   
skewTime = skewness(dxp.("展叶期"))   
   
% 峰度系数   
kurtTime = kurtosis(dxp.("展叶期"))

skewTime = 0.2583

kurtTime = 3.1356

均值在峰值的右边，数据分布的集中程度高于正态分布。

1. 计算农户家庭经营性纯收入的集中化指数

nh = readtable("1999年某地区农户家庭经营性纯收入水平及其构成.xlsx","VariableNamingRule","preserve");   
nh = nh(1:9,:);   
nh.("占总收入的比重/%") = nh.("收入/元") ./sum(nh.("收入/元")) \* 100;   
nh.("实际分布") = cumsum(sort(nh.("占总收入的比重/%"),"descend"));   
temp = linspace(0,100,size(nh,1)+1);   
temp = temp(2:size(nh,1)+1);   
nh.("均匀分布") = permute(temp,[2 1]);   
nh.("集中分布") = ones(size(nh,1),1).\*100;   
   
A = sum(nh.("实际分布"))   
R = sum(nh.("均匀分布"))   
M = sum(nh.("集中分布"))   
I = (A - R)/(M - R)

A = 763.6237

R = 500

M = 900

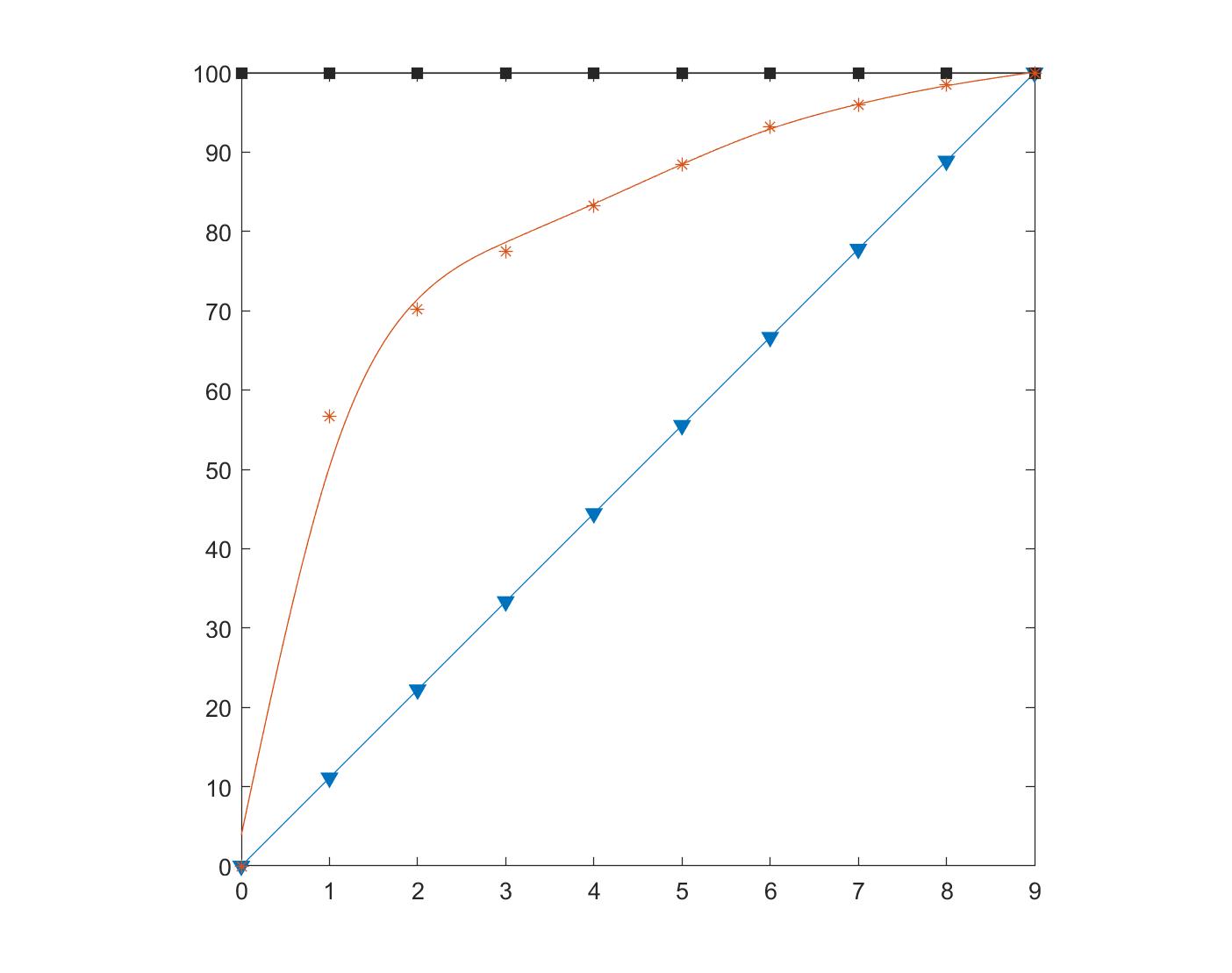
I = 0.6591

由此看来，数据的集中化程度较高。

结果已经填入 1999年某地区农户家庭经营性纯收入水平及其构成.xlsx 中。

1. 绘制洛伦兹曲线

figure;   
x = [0:size(nh,1)];   
y2 = [0; nh.("均匀分布")];   
plot(x',y2,'Marker','v','LineStyle','-','Color',[0 0.447058823529412 0.741176470588235],"MarkerFaceColor",[0 0.447058823529412 0.741176470588235]);   
hold on   
   
y3 = [100; nh.("集中分布")];   
plot(x',y3,'Marker','square','LineStyle','-','Color',[0.149019607843137 0.149019607843137 0.149019607843137],"MarkerFaceColor",[0.149019607843137 0.149019607843137 0.149019607843137]);   
hold on   
   
x2 = 0:0.001:size(nh,1);   
y1 = [0;nh.("实际分布")];   
c = fit(x',y1,'smoothingspline');   
d = c(x2);   
plot(x',y1,'\*',x2,d,'-',"Color",[0.850980392156863 0.325490196078431 0.0980392156862745],"MarkerFaceColor",[0.850980392156863 0.325490196078431 0.0980392156862745]);   
hold on   
xlim([0,9])   
ylim([0,100])   
axis square



洛伦兹曲线

洛伦兹曲线