

数据分析作业

学生姓名:

郑传权

学 号:

2015409020134

学 院:

理学院

班 级:

信计 151

2017 年 10 月

第一章

选题：1.3

1978 年至 1999 年我国居民消费数据表见表 1：

表 1 全国居民消费数据（单位：元）

年份	全国居民	农村居民	城镇居民
1978	184	138	405
1979	270	158	434
1980	236	178	496
1981	262	199	562
1982	284	221	576
1983	311	246	603
1984	354	283	662
1985	437	347	802
1986	485	376	920
1987	550	417	1089
1988	693	508	1431
1989	762	553	1568
1990	803	571	1686
1991	896	621	1925
1992	1070	718	2356
1993	1331	855	3027
1994	1746	1118	3891
1995	2336	1434	4874
1996	2641	1768	5430
1997	2834	1876	5796
1998	2972	1895	6217
1999	3180	1973	6651

- （1） 计算均值、方差、标准差、变异系数、偏度、峰度；
- （2） 计算中位数、上、下四分位数、四分位极差、三均值；
- （3） 作出直方图；
- （4） 作出茎叶图；
- （5） 找出异常值。

解：问题的求解用到了 Matlab2016a，源程序见附录，通过运行程序，得出结果如下：

- （1） 计算均值、方差、标准差、变异系数、偏度、峰度，结果见表 2。

表 2 均值、方差、标准差、变异系数、偏度、峰度求解结果

居民 计算项	全国居民	农村居民	城镇居民
均值	1119.86	747.86	2336.41
方差	979746.12	381506.85	4329948.42
标准差	989.82	617.66	2080.85
变异系数	0.88	0.83	0.89
偏度	0.96	0.94	0.90
峰度	2.38	2.38	2.29

(2) 计算中位数、上、下四分位数、四分位极差、三均值，结果见表 3。

表 3 中位数、上、下四分位数、四分位极差、三均值求解结果

居民 计算项	全国居民	农村居民	城镇居民
中位数	727.50	530.50	1499.50
上四分位数	1746.00	1118.00	3891.00
下四分位数	311.00	246.00	603.00
四分位极差	1435.00	872.00	3288.00
三均值	878.00	606.25	1873.25

(3) 作出直方图，如图 1 所示。

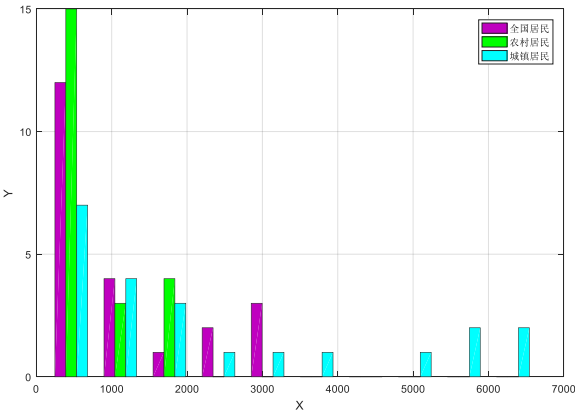


图 1 频数分布直方图

(4) 茎叶图见表 4。

表 4 茎叶图

全国居民	农村居民	城镇居民
180 4	130 8	400 5
230 6	150 8	430 4
260 2	170 8	490 6
270 0	190 9	560 2
280 4	220 1	570 6
310 1	240 6	600 3
350 4	280 3	660 2
430 7	340 7	800 2
480 5	370 6	920 0
550 0	410 7	1080 9
690 3	500 8	1430 1
760 2	550 3	1560 8
800 3	570 1	1680 6
890 6	620 1	1920 5
1070 0	710 8	2350 6
1330 1	850 5	3020 7
1740 6	1110 8	3890 1
2330 6	1430 4	4870 4
2640 1	1760 8	5430 0
2830 4	1870 6	5790 6
2970 2	1890 5	6210 7
3180 0	1970 3	6650 1

(5) 异常值。异常值检测范围：

$$X = \{x | x < Q1 - 1.5d_4, \text{ 或 } x > Q3 + 1.5d_4\}$$

，其中 Q1 为下四分位数，Q3 为上四分位数，d_4 为四分位极差。即：
见表 5。

表 5 异常值界限

	全国居民	农村居民	城镇居民
下界	-1841.50	-1062.00	-4329.00
上届	3898.50	2426.00	8823.00

从表 5 可知，表一中居民消费数据无异常值。

附录

Matlab 源程序:

```
clear;
clc;
data=xlsread('G:\作业\数据分析作业.xlsx','全国居民消费数据','B2:D23');
quanguo=data(:,1);
nongcun=data(:,2);
chengzhen=data(:,3);
%居民平均消费
d_=mean(data);
%居民消费方差
d_var=var(data,1);
%居民消费标准差
d_std=std(data,1);
%变异系数
d_b=d_std./abs(d_);
%偏度系数
d_p=skewness(data);
%峰度系数
d_f=kurtosis(data);
%中位数
d_z=median(data);
%四分位极差
d_4=iqr(data);
%上下四分位数
Q3=prctile(data,75);
Q1=prctile(data,25);
X1=Q1-1.5.*d_4;
X2=Q3+1.5.*d_4;
%三均值
M_=1/4.*Q1+1/2.*d_z+1/4.*Q3;
%4分位标准差
d_4std=d_4./1.349;
%频数分布直方图
hist(data,10);
xlabel('X')
ylabel('Y')
grid on
legend('全国居民','农村居民','城镇居民')
%茎叶图

%全国居民
```

```

leaf = mod(quanguo',10);
stem = quanguo'-leaf;
stem_u = unique(stem);
m = numel(stem_u);
fid = fopen('stemleaf_quanguo.txt','w');
for i = 1:m
    [h,k] = find(stem == stem_u(i));
    n = numel(h);
    fprintf(fid,num2str(stem_u(i)));
    fprintf(fid,'|');
    for j=1:n
        fprintf(fid,num2str(leaf(k(j))));
        if j~= n
            fprintf(fid,',');
        end
    end
    fprintf(fid,'\n');
end
fclose(fid);

```

%农村居民

```

leaf = mod(nongcun',10);
stem = nongcun'-leaf;
stem_u = unique(stem);
m = numel(stem_u);
fid = fopen('stemleaf_nongcun.txt','w');
for i = 1:m
    [h,k] = find(stem == stem_u(i));
    n = numel(h);
    fprintf(fid,num2str(stem_u(i)));
    fprintf(fid,'|');
    for j=1:n
        fprintf(fid,num2str(leaf(k(j))));
        if j~= n
            fprintf(fid,',');
        end
    end
    fprintf(fid,'\n');
end
fclose(fid);

```

%城镇居民

```

leaf = mod(chengzhen',10);
stem = chengzhen'-leaf;

```

```

stem_u = unique(stem);
m = numel(stem_u);
fid = fopen('stemleaf_chengzhen.txt','w');
for i = 1:m
    [h,k] = find(stem == stem_u(i));
    n = numel(h);
    fprintf(fid,num2str(stem_u(i)));
    fprintf(fid,'|');
    for j=1:n
        fprintf(fid,num2str(leaf(k(j))));
        if j~= n
            fprintf(fid,',');
        end
    end
    fprintf(fid,'\n');
end
fclose(fid);
%异常值检测范围:  $X < Q1 - 1.5d_4$ , 或  $X > Q3 + 1.5d_4$ 

```