

2014 级预防医学专业《卫生统计学》试题（编程测验）

班级：预防医学 1401 班

学号：U201412842

姓名：夏鹏飞

1、学籍编号的编码规则一般是：X-YYYY-ZZZZ，其 X 为学生类别；YYYY 一般为所在班级编码；ZZZZ 为对应的个人识别编号。下面有一组学籍编号，请用程序完成下列要求：

（1）分别列出属于本科生、硕士生、博士生的编号；

（2）分别提取班级编号及个人识别编号。

（提示：列输入方式或者字符串操作）

说明：

本科生：B

硕士生：M

博士生：D

学籍编号	学籍编号
M14141106	B16111038
D16111123	B14111178
D16131098	M16131005
M14141045	B16111075
M14141169	M15121178
D14111020	D15111061
M15121081	B15141086
B16111062	B17111100
D16141192	D14121150
B17111177	D17121158
D16141016	M14141163
M17111045	B16121065
B14131157	B16111023

（提示：请提供相应的程序）

2、试根据 12 名患者的血压测量值，完成以下问题：

表 1 12 名患者血压测量值

学号	收缩压	舒张压	脉压差
135	119	77	42
138	124	77	47
123	131	65	66
125	127	58	69
170	135	83	52
159	127	70	57
169	139	74	65
176	137	80	57
135	130	79	51
180	132	75	57

179	125	75	50
158	142	89	53

问题：

- 1) 创建 SAS 数据集；
- 2) 打印收缩压大于 130 的患者信息；(提示，使用 if 语句)
- 3) 请提供相应的程序、结果及日志文件。

3、某地区 2016 年 14 岁女孩的身高资料（cm）如下：

142.3 148.8 142.7 144.4 144.7 145.1 143.3 154.2 152.3 142.7 156.6
 137.9 143.9 141.2 139.3 145.8 142.2 137.9 141.2 150.6 142.7 151.3
 142.4 141.5 141.9 147.9 125.8 139.9 148.9 154.9 145.7 140.8 139.6
 148.8 147.8 146.7 132.7 149.7 154.0 158.2 138.2 149.8 151.1 140.1
 140.5 143.4 152.9 147.5 147.7 162.6 141.6 143.6 144.0 150.6 138.9
 150.8 147.9 136.9 146.5 130.6 142.5 149.0 145.4 139.5 148.9 144.5
 141.8 148.1 145.4 134.6 130.5 145.2 146.2 146.4 142.4 137.1 141.4
 144.0 129.4 142.8 132.1 141.8 143.3 143.8 134.7 147.1 140.9 137.4
 142.5 146.6 135.5 146.8 156.3 150.0 147.3 142.9 141.4 134.7 138.5
 146.6 134.5 135.1 141.9 142.1 138.1 134.9 146.7 138.5 139.6 139.2
 148.8 150.3 140.7 143.5 140.2 143.6 138.7 138.9 143.5 139.9 134.4
 133.1 145.9 139.2 137.4 142.3 160.9 137.7 142.9 126.8

问题：（1）计算均数、中位数；

（2）计算四分位间距、标准差、变异系数；

（3）观察频数分布情况；

【第一题代码】

```
data xia.stuno;
input sn:$9. @@; /*":表示遇空格或读取结束即读取下一变量*/
deg=substr(sn,1,1); /*substr()函数用于提取或替换字符*/
clsno=substr(sn,2,4);
idno=substr(sn,6,4);
cards;
M14141106 B16111038
D16111123 B14111178
D16131098 M16131005
M14141045 B16111075
M14141169 M15121178
D14111020 D15111061
M15121081 B15141086
B16111062 B17111100
D16141192 D14121150
B17111177 D17121158
D16141016 M14141163
M17111045 B16121065
B14131157 B16111023
;
run;
/*分别列出属于本科生、硕士生、博士生的编号*/
proc print data=xia.stuno label;
var sn;
label sn="本科生学籍编号";
where deg="B";
run;
proc print data=xia.stuno label;
var sn;
label sn="硕士生学籍编号";
where deg="M";
run;
proc print data=xia.stuno label;
var sn;
label sn="博士生学籍编号";
where deg="D";
run;
/*分别提取班级编号及个人识别编号*/
proc print data=xia.stuno label;
var sn clsno;
label sn="学籍编号" clsno="班级编号";
run;
```

```
proc print data=xia.stuno label;  
  var sn idno;  
  label sn="学籍编号" idno="个人识别编号";  
run;
```

【第一题日志】

```
1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
61  
62      data xia.stuno;  
63      input sn:$9. @@; /* ":"表示遇空格或读取结束即读取下一变量*/  
64      deg=substr(sn,1,1); /*substr()函数用于提取或替换字符*/  
65      clsno=substr(sn,2,4);  
66      idno=substr(sn,6,4);  
67      cards;
```

NOTE: INPUT 语句到达一行的末尾, SAS 已转到新的一行。

NOTE: 数据集 XIA.STUNO 有 26 个观测和 4 个变量。

NOTE: "DATA 语句"所用时间(总处理时间):

实际时间	0.00 秒
CPU 时间	0.01 秒

```
81      ;  
  
82      run;  
83      /*分别列出属于本科生、硕士生、博士生的编号*/  
84      proc print data=xia.stuno label;  
85          var sn;  
86          label sn="本科生学籍编号";  
87          where deg="B";  
88      run;
```

NOTE: 从数据集 XIA.STUNO. 读取了 10 个观测

WHERE deg='B';

NOTE: "PROCEDURE PRINT"所用时间(总处理时间):

实际时间	0.05 秒
CPU 时间	0.04 秒

```
89      proc print data=xia.stuno label;  
90          var sn;  
91          label sn="硕士生学籍编号";  
92          where deg="M";  
93      run;
```

NOTE: 从数据集 XIA.STUNO. 读取了 8 个观测

WHERE deg='M';

NOTE: "PROCEDURE PRINT"所用时间 (总处理时间):

实际时间 0.04 秒

CPU 时间 0.03 秒

```
94      proc print data=xia.stuno label;
95          var sn;
96          label sn="博士生学籍编号";
97          where deg="D";
98      run;
```

NOTE: 从数据集 XIA.STUNO. 读取了 8 个观测

WHERE deg='D';

NOTE: "PROCEDURE PRINT"所用时间 (总处理时间):

实际时间 0.05 秒

CPU 时间 0.05 秒

```
99      /*分别提取班级编号及个人识别编号*/
100      proc print data=xia.stuno label;
101          var sn clsno;
102          label sn="学籍编号" clsno="班级编号";
103      run;
```

NOTE: 从数据集 XIA.STUNO. 读取了 26 个观测

NOTE: "PROCEDURE PRINT"所用时间 (总处理时间):

实际时间 0.09 秒

CPU 时间 0.08 秒

```
104      proc print data=xia.stuno label;
105          var sn idno;
106          label sn="学籍编号" idno="个人识别编号";
107      run;
```

NOTE: 从数据集 XIA.STUNO. 读取了 26 个观测

NOTE: "PROCEDURE PRINT"所用时间 (总处理时间):

实际时间 0.08 秒

CPU 时间 0.07 秒

108

109 OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

122

【第一题结果】

Obs	本科生学籍编号
2	B16111038
4	B14111178
8	B16111075
14	B15141086
15	B16111062
16	B17111100
19	B17111177
24	B16121065
25	B14131157
26	B16111023

Obs	硕士生学籍编号
1	M14141106
6	M16131005
7	M14141045
9	M14141169
10	M15121178
13	M15121081
22	M14141163
23	M17111045

Obs	博士生学籍编号
3	D16111123
5	D16131098
11	D14111020
12	D15111061
17	D16141192
18	D14121150
20	D17121158
21	D16141016

Obs	学籍编号	班级编号
1	M14141106	1414
2	B16111038	1611

3	D16111123	1611
4	B14111178	1411
5	D16131098	1613
6	M16131005	1613
7	M14141045	1414
8	B16111075	1611
9	M14141169	1414
10	M15121178	1512
11	D14111020	1411
12	D15111061	1511
13	M15121081	1512
14	B15141086	1514
15	B16111062	1611
16	B17111100	1711
17	D16141192	1614
18	D14121150	1412
19	B17111177	1711
20	D17121158	1712
21	D16141016	1614
22	M14141163	1414
23	M17111045	1711
24	B16121065	1612
25	B14131157	1413
26	B16111023	1611

Obs	学籍编号	个人识别编号
1	M14141106	1106
2	B16111038	1038
3	D16111123	1123
4	B14111178	1178
5	D16131098	1098
6	M16131005	1005
7	M14141045	1045
8	B16111075	1075
9	M14141169	1169
10	M15121178	1178
11	D14111020	1020
12	D15111061	1061
13	M15121081	1081
14	B15141086	1086
15	B16111062	1062

16	B17111100	1100
17	D16141192	1192
18	D14121150	1150
19	B17111177	1177
20	D17121158	1158
21	D16141016	1016
22	M14141163	1163
23	M17111045	1045
24	B16121065	1065
25	B14131157	1157
26	B16111023	1023

【第二题代码】

```
/*创建SAS数据集*/
data xia.bldpress;
input sn sbp dbp pp;
cards;
135 119 77 42
138 124 77 47
123 131 65 66
125 127 58 69
170 135 83 52
159 127 70 57
169 139 74 65
176 137 80 57
135 130 79 51
180 132 75 57
179 125 75 50
158 142 89 53
;
run;
/*打印收缩压大于130的患者信息*/
proc print data=xia.bldpress label;
var sn sbp dbp pp;
label sn="学号" sbp="收缩压" dbp="舒张压" pp="脉压差";
where sbp>130;
run;
/*请提供相应的程序、结果及日志文件*/
```

【第二题日志】

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
61
```



```
62      /*创建SAS数据集*/
63      data xia.bldpress;
64      input sn sbp dbp pp;
65      cards;
```

NOTE: 数据集 XIA.BLDPRESS 有 12 个观测和 4 个变量。

NOTE: "DATA 语句"所用时间（总处理时间）:

实际时间 0.00 秒
CPU 时间 0.00 秒

```
78      ;

79      run;
80      /*打印收缩压大于130的患者信息*/
81      proc print data=xia.bldpress label;
82          var sn sbp dbp pp;
83          label sn="学号" sbp="收缩压" dbp="舒张压" pp="脉压差";
84          where sbp>130;
85      run;
```

NOTE: 从数据集 XIA.BLDPRESS. 读取了 6 个观测

WHERE sbp>130;

NOTE: "PROCEDURE PRINT"所用时间（总处理时间）:

实际时间 0.06 秒
CPU 时间 0.06 秒

```
86      /*请提供相应的程序、结果及日志文件*/
87
88      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
101
```

【第二题结果】

Obs	学号	收缩压	舒张压	脉压差
3	123	131	65	66
5	170	135	83	52
7	169	139	74	65
8	176	137	80	57
10	180	132	75	57
12	158	142	89	53

【第三题代码】

```
data xia.height;
input ht @@;
low=125;
dis=5;
z=ht-mod(ht-low,dis);/*这三个语句新建了变量z，将原始变量转换为所在组段的下限值，
以z的频数分布代表原始变量各组段的频数分布，可以省略format过程*/
cards;
142.3 148.8 142.7 144.4 144.7 145.1 143.3 154.2 152.3 142.7 156.6
137.9 143.9 141.2 139.3 145.8 142.2 137.9 141.2 150.6 142.7 151.3
142.4 141.5 141.9 147.9 125.8 139.9 148.9 154.9 145.7 140.8 139.6
148.8 147.8 146.7 132.7 149.7 154.0 158.2 138.2 149.8 151.1 140.1
140.5 143.4 152.9 147.5 147.7 162.6 141.6 143.6 144.0 150.6 138.9
150.8 147.9 136.9 146.5 130.6 142.5 149.0 145.4 139.5 148.9 144.5
141.8 148.1 145.4 134.6 130.5 145.2 146.2 146.4 142.4 137.1 141.4
144.0 129.4 142.8 132.1 141.8 143.3 143.8 134.7 147.1 140.9 137.4
142.5 146.6 135.5 146.8 156.3 150.0 147.3 142.9 141.4 134.7 138.5
146.6 134.5 135.1 141.9 142.1 138.1 134.9 146.7 138.5 139.6 139.2
148.8 150.3 140.7 143.5 140.2 143.6 138.7 138.9 143.5 139.9 134.4
133.1 145.9 139.2 137.4 142.3 160.9 137.7 142.9 126.8
;
run;
/*计算均数、中位数*/ /*计算四分位间距、标准差、变异系数*/
proc means data=xia.height mean median qrange std cv;
var ht;
run;
/*观察频数分布情况*/
proc freq data=xia.height; /*频数分布表*/
table z;
run; /*贺佳. SAS统计软件应用[M]. 人民卫生出版社, 2014, 25.*/
proc univariate data=xia.height normal; /*正态性检验 不输出结果*/
var ht;
probplot ht; /*正态概率图*/
output out=normality probn=probnht; /*将正态性检验P值输出为变量probnht */
run;
proc sgplot data=xia.height; /*频数分布直方图*/
histogram ht; /*指定用于绘制直方图的变量*/
density ht; /*指定用于绘制分布图的变量*/
inset "正态性检验" "P=0.4518"; /*在图中插入的注释内容，分两行显示*/
run; /*变量probnht的值无法通过语句实现在图片中显示，可在查看输出数据后手动输入或者
导出图片后用绘图软件编辑。*/
```

【第三题日志】

```
1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
61
62      data xia.height;
63          input ht @@;
64          low=125;
65          dis=5;
66          z=ht-mod(ht-low,dis);
66      ! /*这三个语句新建了变量z，将原始变量转换为所在组段的下限值，以z的频数分
布代表原始变
66      ! 量各组段的频数分布，可以省略format过程*/
67      cards;
```

NOTE: INPUT 语句到达一行的末尾，SAS 已转到新的一行。

NOTE: 数据集 XIA.HEIGHT 有 130 个观测和 4 个变量。

NOTE: “DATA 语句”所用时间（总处理时间）：

实际时间	0.01 秒
CPU 时间	0.01 秒

```
80      ;

81      run;
82      /*计算均数、中位数*/ /*计算四分位间距、标准差、变异系数*/
83      proc means data=xia.height mean median qrange std cv;
84          var ht;
85      run;
```

NOTE: 从数据集 XIA.HEIGHT. 读取了 130 个观测

NOTE: “PROCEDURE MEANS”所用时间（总处理时间）：

实际时间	0.07 秒
CPU 时间	0.07 秒

```
86      /*观察频数分布情况*/
87      proc freq data=xia.height; /*频数分布表*/
88          table z;
89      run;
```

NOTE: 从数据集 XIA.HEIGHT. 读取了 130 个观测

NOTE: “PROCEDURE FREQ”所用时间（总处理时间）：

实际时间	0.05 秒
CPU 时间	0.05 秒

```
89      !      /*贺佳. SAS统计软件应用[M]. 人民卫生出版社, 2014, 25.*/

90      proc univariate data=xia.height normal; /*正态性检验 不输出结果
*/
91      var ht;
92      probplot ht; /*正态概率图*/
93      output out=normality probn=probnht; /*将正态性检验P值输出为变
量probnht */
94      run;

WARNING: The font CSongGB18030C-Light is not available. Arial Unicode
MS will be used.
NOTE: 数据集 WORK.NORMALITY 有 1 个观测和 1 个变量。
NOTE: "PROCEDURE UNIVARIATE"所用时间（总处理时间）:
      实际时间      0.79 秒
      CPU 时间      0.38 秒

95      proc sgplot data=xia.height; /*频数分布直方图*/
96      histogram ht; /*指定用于绘制直方图的变量*/
97      density ht; /*指定用于绘制分布图的变量*/
98      inset "正态性检验" "P=0.4518"; /*在图中插入的注释内容，分两行显示
*/
99      run;

99      !      /*变量probnht的值无法通过语句实现在图片中显示，可在查看输出数据
后手动输入或者导
99      ! 出图片后用绘图软件编辑。*/
NOTE: "PROCEDURE SGPLOT"所用时间（总处理时间）:
      实际时间      0.64 秒
      CPU 时间      0.20 秒

WARNING: The font CSongGB18030C-Light is not available. Arial Unicode
MS will be used.
NOTE: 从数据集 XIA.HEIGHT. 读取了 130 个观测

100
101
102
103
104
105      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
118
```

【第三题结果】

MEANS PROCEDURE

分析变量: ht				
均值	中位数	四分位数极差	标准差	变异系数
143.2161538	142.7500000	7.8000000	6.4317864	4.4909644

FREQ 过程

ht	频数	百分比	累积频数	累积百分比
125-130	3	2.31	3	2.31
130-135	11	8.46	14	10.77
135-140	24	18.46	38	29.23
140-145	44	33.85	82	63.08
145-150	32	24.62	114	87.69
150-155	11	8.46	125	96.15
155-160	3	2.31	128	98.46
160-165	2	1.54	130	100.00

UNIVARIATE PROCEDURE

变量: ht

矩			
N	130	权重总和	130
均值	143.216154	观测总和	18618.1
标准差	6.43178642	方差	41.3678766
偏度	0.14519373	峰度	0.69472443
未校平方和	2671749.13	校正平方和	5336.45608
变异系数	4.49096436	标准误差均值	0.56410499

基本统计测度			
位置		变异性	
均值	143.2162	标准差	6.43179
中位数	142.7500	方差	41.36788
众数	142.7000	极差	36.80000
		四分位间距	7.80000

注意: 显示的众数是 2 个众数中最小的众数, 其计数为 3。

位置检验: Mu0=0				
检验	统计量	p 值		
Student t	t	253.8821	Pr > t	<.0001
符号检验	M	65	Pr >= M	<.0001

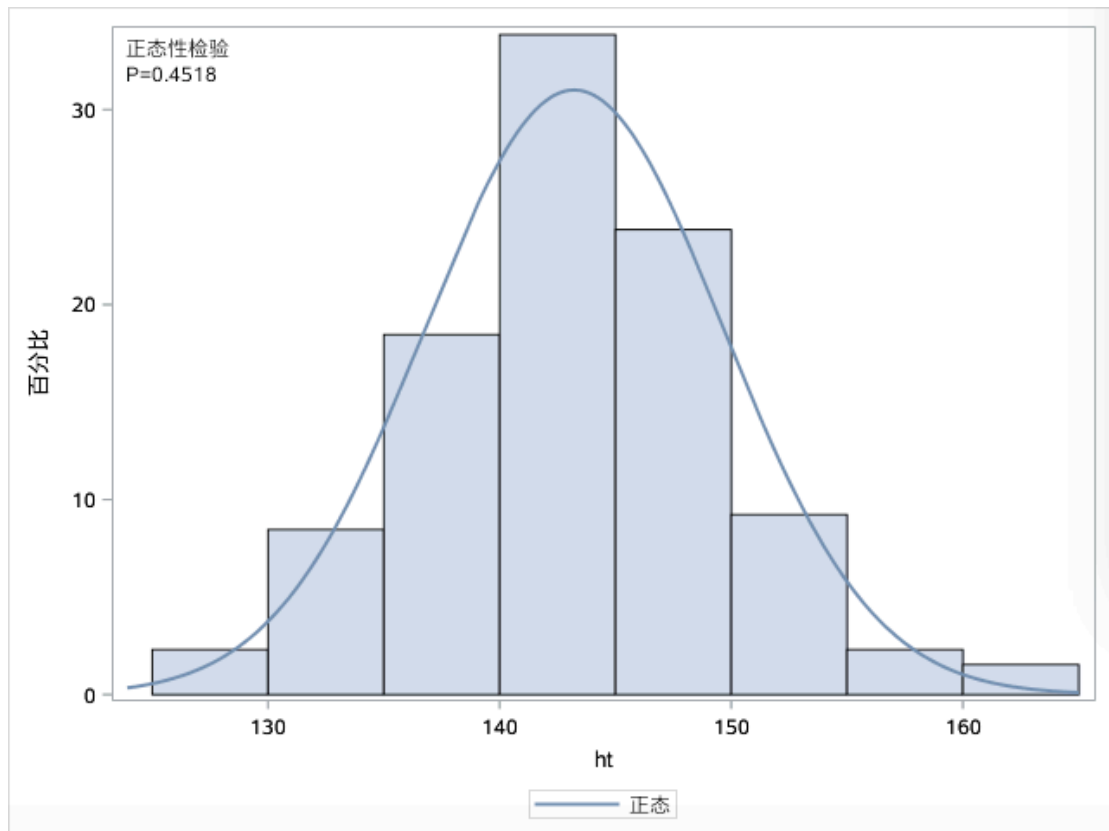
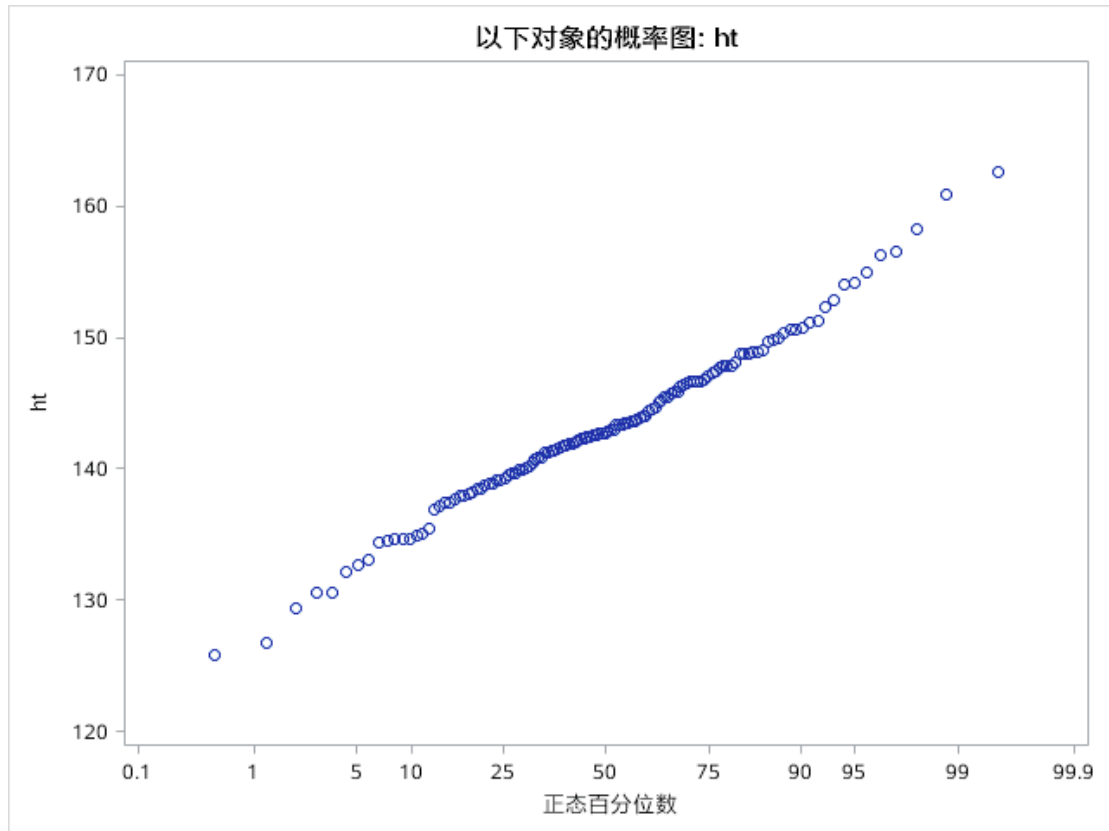
符号秩检验	S	4257.5	Pr >= S	<.0001
-------	---	--------	----------	--------

正态性检验				
检验	统计量		p 值	
Shapiro-Wilk	W	0.989753	Pr < W	0.4518
Kolmogorov-Smirnov	D	0.059193	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.075278	Pr > W-Sq	0.2408
Anderson-Darling	A-Sq	0.449233	Pr > A-Sq	>0.2500

分位数 (定义 5)	
水平	分位数
100% 最大值	162.60
99%	160.90
95%	154.20
90%	150.70
75% Q3	147.10
50% 中位数	142.75
25% Q1	139.30
10%	134.80
5%	132.70
1%	126.80
0% 最小值	125.80

极值观测			
最小值		最大值	
值	观测	值	观测
125.8	27	156.3	93
126.8	130	156.6	11
129.4	79	158.2	40
130.5	71	160.9	127
130.6	60	162.6	50

UNIVARIATE PROCEDURE



【附：三道题的另一种实现方式代码】

【第一题】

```
data xia.stuno;
input sn:$9. @@; /*:"表示遇空格或读取结束即读取下一变量*/
deg=substr(sn,1,1); /*substr()函数用于提取或替换字符*/
clsno=substr(sn,2,4);
idno=substr(sn,6,4);
cards;
M14141106 B16111038
D16111123 B14111178
D16131098 M16131005
M14141045 B16111075
M14141169 M15121178
D14111020 D15111061
M15121081 B15141086
B16111062 B17111100
D16141192 D14121150
B17111177 D17121158
D16141016 M14141163
M17111045 B16121065
B14131157 B16111023
;
run;
/*分别列出属于本科生、硕士生、博士生的编号*/
data xia.bachelor xia.master xia.doctor;
  set xia.stuno;
  if (deg="B") then output xia.bachelor;
  else if (deg="M") then output xia.master;
  else if (deg="D") then output xia.doctor;
run;
proc print data=xia.bachelor;
run;
proc print data=xia.master;
run;
proc print data=xia.doctor;
run;
/*分别提取班级编号及个人识别编号*/
data xia.class;
  set xia.stuno;
  keep sn clsno;
run;
proc print;
run;
data xia.id;
```

```
set xia.stuno;
keep sn idno;
run;
proc print;
run;
```

【第二题】

```
/*创建SAS数据集*/
data xia.bldpress;
input sn sbp dbp pp;
cards;
135 119 77 42
138 124 77 47
123 131 65 66
125 127 58 69
170 135 83 52
159 127 70 57
169 139 74 65
176 137 80 57
135 130 79 51
180 132 75 57
179 125 75 50
158 142 89 53
;
run;
proc print;
run;
/*打印收缩压大于130的患者信息; (提示, 使用if 语句)*/
data xia.htn;
set xia.bldpress;
if sbp>130 then output xia.htn;
run;
proc print;
run; /*请提供相应的程序、结果及日志文件*/
```

【第三题】

```
data xia.height;
input ht @@;
cards;
142.3 148.8 142.7 144.4 144.7 145.1 143.3 154.2 152.3 142.7 156.6
137.9 143.9 141.2 139.3 145.8 142.2 137.9 141.2 150.6 142.7 151.3
142.4 141.5 141.9 147.9 125.8 139.9 148.9 154.9 145.7 140.8 139.6
148.8 147.8 146.7 132.7 149.7 154.0 158.2 138.2 149.8 151.1 140.1
140.5 143.4 152.9 147.5 147.7 162.6 141.6 143.6 144.0 150.6 138.9
150.8 147.9 136.9 146.5 130.6 142.5 149.0 145.4 139.5 148.9 144.5
141.8 148.1 145.4 134.6 130.5 145.2 146.2 146.4 142.4 137.1 141.4
```

```

144.0 129.4 142.8 132.1 141.8 143.3 143.8 134.7 147.1 140.9 137.4
142.5 146.6 135.5 146.8 156.3 150.0 147.3 142.9 141.4 134.7 138.5
146.6 134.5 135.1 141.9 142.1 138.1 134.9 146.7 138.5 139.6 139.2
148.8 150.3 140.7 143.5 140.2 143.6 138.7 138.9 143.5 139.9 134.4
133.1 145.9 139.2 137.4 142.3 160.9 137.7 142.9 126.8
;
run;
/*计算均数、中位数*/
proc means data=xia.height mean median;
run;
/*计算四分位间距、标准差、变异系数*/
proc means data=xia.height qrange std cv;
run;
/*观察频数分布情况*/
proc format;
    value ht_range 125-130 = '125-130'
                  130-135 = '130-135'
                  135-140 = '135-140'
                  140-145 = '140-145'
                  145-150 = '145-150'
                  150-155 = '150-155'
                  155-160 = '155-160'
                  160-165 = '160-165';

run;
proc freq data=xia.height; /*频数分布表*/
    format ht ht_range.;
    table ht;
run;
/*http://blog.sina.com.cn/s/blog_58ea0d1f0101d272.html*/
proc univariate data=xia.height normal; /*正态性检验*/
    var ht;
    histogram ht; /*绘制直方图*/
    probplot ht; /*频数概率图*/
run;
proc sgplot data=xia.height; /*频数分布图*/
    histogram ht; /*指定用于绘制直方图的变量*/
    density ht; /*指定用于绘制分布图的变量*/
    inset "正态性检验" "P=0.4518"/textattrs=(size=12) position=topright; /*在
图中插入的注释内容(已知P=0.4518)，分两行显示；指定插入注释的字体大小和位置*/
    xaxis valueattrs=(size=12) labelattrs=(size=12); /*设定x轴刻度和标签字体大小
*/
    yaxis valueattrs=(size=12) labelattrs=(size=12); /*设定y轴刻度和标签字体大小
(指定y轴刻度值字体大小为12，默认单位是pixels；指定标签字体大小为12) */
run;

```