Program Summary - Homework 2.sas

Execution Environment

Author:	chwang10
File:	/home/chwang10/Homework 2.sas
SAS Platform:	Linux LIN X64 3.10.0-1062.9.1.el7.x86_64
SAS Host:	ODAWS02-USW2.ODA.SAS.COM
SAS Version:	9.04.01M6P11072018
SAS Locale:	en_US
Submission Time:	9/23/2020, 9:47:45 PM
Browser Host:	104.220.37.66
User Agent:	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/85.0.4183.102 Safari/537.36
Application Server:	ODAMID00-USW2.ODA.SAS.COM

Code: Homework 2.sas

```
* Programmed by Charles Hwang   *
* Coded in SAS OnDemand         *
* Wednesday, September 23, 2020 *
* Course: STAT 403              *
* Title: Homework 2             *;

/* 1a */ Data patientinfo;
Input ID$ 1-3 Gender$ 6 Race$ 8-24 CollegeEducated$ 29-31;
Datalines;
005  F Hispanic             Yes
004  M African American     Yes
010  F Asian                No
002  M Asian                Yes
009  F Asian                Yes
001  M White                Yes
003  F Hispanic             Yes
008  F African American     Yes
007  M Hispanic             Yes
006  M African American     No
;

/* 1b */ Data patientvitals;
Set work.patientdata;
BMI=Weight*703/Height**2;
Averagebp=1/3*SBP+2/3*DBP;
Run;

/* 1c */ Proc Means mean median std lclm uclm data=patientvitals;
Title "1c. Statistics for BMI and Average Blood Pressure";
Var BMI Averagebp;
Run; * We are 95 percent confident that the true mean BMI is between 23.4273899 and 27.4236915 and that
the true mean average blood pressure is between 85.3365976 and 102.6634024. *;

/* 1d */ Proc Sort data=patientinfo;
By ID;
Proc Sort data=patientvitals;
By ID;
Data PatientMerge;
Merge patientinfo patientvitals;
By ID;
Proc Sort data=PatientMerge;
By ID;
Proc Print data=PatientMerge;
Title "1d. Merged Data Set";
ID ID;
Run;

/* 1e */ Proc Sort data=PatientMerge;
By Gender;
Proc Boxplot data=PatientMerge;
```

```
Title "1e.";
Plot BMI*Gender;
Run;

/* 2a */ Proc Import out=Boston file="/home/chwang10/Boston.csv" dbms=csv;
Run;
/* 2a(i) */ Proc Means mean median std lclm uclm data=Boston;
Title "2a(i). Student-Teacher Ratio Greater Than 19";
Var crim tax medv;
Where ptratio>19;
Run;
/* 2a(ii) */ Proc Means mean median std lclm uclm data=Boston;
Title "2a(ii). Student-Teacher Ratio Equal To or Below 19";
Var crim tax medv;
Where ptratio<=19;
Run;
/* 2a (iii) */ * The crime rate per capita and property tax rate are significantly lower when
the student-teacher ratio is below 19. The median value of owner-occupied homes is also much higher.
Not only is the mean of these values significantly different, the entire distribution of values is
significantly lower (or higher in the case of median value) when the student-teacher ratio is below 19. *;

/* 2b */ Data RiverSide;
Set Boston;
If CHAS=1;
Data NoRiver;
Set Boston;
IF CHAS=0;
Run;

/* 2c */ Proc Univariate data=RiverSide noprint;
Title "2c. Crime Rate per Capita for Properties Near Charles River";
Histogram crim;
Inset mean="Mean" (5.3) std="Std. Dev." (5.3) skewness="Skewness" (5.3) kurtosis="Kurtosis" (5.3) /pos=NE;
Proc Univariate data=NoRiver noprint;
Title "2c. Crime Rate per Capita for Properties Not Near Charles River";
Histogram crim;
Inset mean="Mean" (5.3) std="Std. Dev." (5.3) skewness="Skewness" (5.3) kurtosis="Kurtosis" (5.3) /pos=NE;
Run; * The skewness and kurtosis of crime rate per capita is significantly less for riverside properties
than for properties not near the Charles River. This means the histogram for crime rate per capita of
non-riverside properties is right-tailed and has a much larger tail than the histogram for crime rate
per capita of riverside properties. *;
```

Log: Homework 2.sas

Warnings (1)

Notes (44)

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
70
71         * Programmed by Charles Hwang   *
72         * Coded in SAS OnDemand         *
73         * Wednesday, September 23, 2020 *
74         * Course: STAT 403              *
75         * Title: Homework 2             *;
76
77         /* 1a */
77       !          Data patientinfo;
78         Input ID$ 1-3 Gender$ 6 Race$ 8-24 CollegeEducated$ 29-31;
79         Datalines;

NOTE: The data set WORK.PATIENTINFO has 10 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              667.03k
      OS Memory           38316.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        645  Switch Count  2
      Page Faults                       0
      Page Reclaims                     92
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            0
```

```
      Block Output Operations           264


90         ;
91
92         /* 1b */
92       !         Data patientvitals;
93         Set work.patientdata;
94         BMI=Weight*703/Height**2;
95         Averagebp=1/3*SBP+2/3*DBP;
96         Run;

NOTE: There were 10 observations read from the data set WORK.PATIENTDATA.
NOTE: The data set WORK.PATIENTVITALS has 10 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1057.40k
      OS Memory           38576.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        646  Switch Count  2
      Page Faults                       0
      Page Reclaims                     123
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


97
98         /* 1c */
98       !         Proc Means mean median std lclm uclm data=patientvitals;
99         Title "1c. Statistics for BMI and Average Blood Pressure";
100        Var BMI Averagebp;
101        Run;

NOTE: There were 10 observations read from the data set WORK.PATIENTVITALS.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              9071.62k
      OS Memory           43456.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        647  Switch Count  2
      Page Faults                       0
      Page Reclaims                     1352
      Page Swaps                        0
      Voluntary Context Switches        24
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8


101      !      * We are 95 percent confident that the true mean BMI is between 23.4273899 and 27.4236915 and that
102        the true mean average blood pressure is between 85.3365976 and 102.6634024. *;
103
104        /* 1d */
104      !         Proc Sort data=patientinfo;
105        By ID;

NOTE: There were 10 observations read from the data set WORK.PATIENTINFO.
NOTE: The data set WORK.PATIENTINFO has 10 observations and 4 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              938.31k
      OS Memory           38576.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        648  Switch Count  2
      Page Faults                       0
      Page Reclaims                     115
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


106        Proc Sort data=patientvitals;
107        By ID;

NOTE: There were 10 observations read from the data set WORK.PATIENTVITALS.
```

```
NOTE: The data set WORK.PATIENTVITALS has 10 observations and 7 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              927.00k
      OS Memory           38832.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        649  Switch Count  2
      Page Faults                       0
      Page Reclaims                     113
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


108         Data PatientMerge;
109         Merge patientinfo patientvitals;
110         By ID;

WARNING: Multiple lengths were specified for the BY variable ID by input data sets. This might cause unexpected results.
NOTE: There were 10 observations read from the data set WORK.PATIENTINFO.
NOTE: There were 10 observations read from the data set WORK.PATIENTVITALS.
NOTE: The data set WORK.PATIENTMERGE has 10 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1502.03k
      OS Memory           39092.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        650  Switch Count  2
      Page Faults                       0
      Page Reclaims                     163
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


111         Proc Sort data=PatientMerge;
112         By ID;

NOTE: There were 10 observations read from the data set WORK.PATIENTMERGE.
NOTE: The data set WORK.PATIENTMERGE has 10 observations and 10 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              929.06k
      OS Memory           38832.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        651  Switch Count  2
      Page Faults                       0
      Page Reclaims                     113
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


113         Proc Print data=PatientMerge;
114         Title "1d. Merged Data Set";
115         ID ID;
116         Run;

NOTE: There were 10 observations read from the data set WORK.PATIENTMERGE.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.02 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              1051.18k
      OS Memory           38572.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        652  Switch Count  1
      Page Faults                       0
      Page Reclaims                     61
      Page Swaps                        0
      Voluntary Context Switches        6
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0
```

```
117
118        /* 1e */
118      !         Proc Sort data=PatientMerge;
119        By Gender;

NOTE: There were 10 observations read from the data set WORK.PATIENTMERGE.
NOTE: The data set WORK.PATIENTMERGE has 10 observations and 10 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              816.71k
      OS Memory           38832.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        653  Switch Count  2
      Page Faults                       0
      Page Reclaims                     115
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


120        Proc Boxplot data=PatientMerge;
121        Title "1e.";
122        Plot BMI*Gender;
123        Run;

NOTE: Processing beginning for PLOT statement number 1.
NOTE: There were 10 observations read from the data set WORK.PATIENTMERGE.
NOTE: PROCEDURE BOXPLOT used (Total process time):
      real time           0.18 seconds
      user cpu time       0.11 seconds
      system cpu time     0.02 seconds
      memory              20737.87k
      OS Memory           55184.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        654  Switch Count  1
      Page Faults                       0
      Page Reclaims                     4778
      Page Swaps                        0
      Voluntary Context Switches        406
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           800


124
125        /* 2a */
125      !         Proc Import out=Boston file="/home/chwang10/Boston.csv" dbms=csv;
126        Run;

NOTE: Import cancelled.  Output dataset WORK.BOSTON already exists.  Specify REPLACE option to overwrite it.
NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              192.46k
      OS Memory           53672.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        655  Switch Count  0
      Page Faults                       0
      Page Reclaims                     15
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0

127        /* 2a(i) */


127      !            Proc Means mean median std lclm uclm data=Boston;
128        Title "2a(i). Student-Teacher Ratio Greater Than 19";
129        Var crim tax medv;
130        Where ptratio>19;
131        Run;

NOTE: There were 253 observations read from the data set WORK.BOSTON.
      WHERE ptratio>19;
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.02 seconds
```

```
      user cpu time       0.02 seconds
      system cpu time     0.01 seconds
      memory              6408.06k
      OS Memory           59072.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        656  Switch Count  4
      Page Faults                       0
      Page Reclaims                     1380
      Page Swaps                        0
      Voluntary Context Switches        29
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0


132         /* 2a(ii) */
132       !             Proc Means mean median std lclm uclm data=Boston;
133         Title "2a(ii). Student-Teacher Ratio Equal To or Below 19";
134         Var crim tax medv;
135         Where ptratio<=19;
136         Run;

NOTE: There were 253 observations read from the data set WORK.BOSTON.
      WHERE ptratio<=19;
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.02 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              6515.25k
      OS Memory           59072.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        657  Switch Count  4
      Page Faults                       0
      Page Reclaims                     1357
      Page Swaps                        0
      Voluntary Context Switches        32
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0


137         /* 2a (iii) */ * The crime rate per capita and property tax rate are significantly lower when
138         the student-teacher ratio is below 19. The median value of owner-occupied homes is also much higher.
139         Not only is the mean of these values significantly different, the entire distribution of values is
140         significantly lower (or higher in the case of median value) when the student-teacher ratio is below 19. *;
141
142         /* 2b */
142       !           Data RiverSide;
143         Set Boston;
144         If CHAS=1;

NOTE: There were 506 observations read from the data set WORK.BOSTON.
NOTE: The data set WORK.RIVERSIDE has 35 observations and 15 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              845.34k
      OS Memory           54192.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        658  Switch Count  2
      Page Faults                       0
      Page Reclaims                     123
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


145         Data NoRiver;
146         Set Boston;
147         IF CHAS=0;
148         Run;

NOTE: There were 506 observations read from the data set WORK.BOSTON.
NOTE: The data set WORK.NORIVER has 471 observations and 15 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1022.00k
      OS Memory           54192.00k
      Timestamp           09/24/2020 04:47:44 AM
      Step Count                        659  Switch Count  2
      Page Faults                       0
```

```
      Page Reclaims                     109
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


149
150       /* 2c */
150     !         Proc Univariate data=RiverSide noprint;
151       Title "2c. Crime Rate per Capita for Properties Near Charles River";
152       Histogram crim;
153       Inset mean="Mean" (5.3) std="Std. Dev." (5.3) skewness="Skewness" (5.3) kurtosis="Kurtosis" (5.3) /pos=NE;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.15 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              7411.40k
      OS Memory           55304.00k
      Timestamp           09/24/2020 04:47:45 AM
      Step Count                        660  Switch Count  1
      Page Faults                       0
      Page Reclaims                     670
      Page Swaps                        0
      Voluntary Context Switches        189
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           440


154       Proc Univariate data=NoRiver noprint;
155       Title "2c. Crime Rate per Capita for Properties Not Near Charles River";
156       Histogram crim;
157       Inset mean="Mean" (5.3) std="Std. Dev." (5.3) skewness="Skewness" (5.3) kurtosis="Kurtosis" (5.3) /pos=NE;
158       Run;

NOTE: At least one W.D format was too small for the number to be printed. The decimal may be shifted by the "BEST" format.
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.16 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              7349.15k
      OS Memory           55304.00k
      Timestamp           09/24/2020 04:47:45 AM
      Step Count                        661  Switch Count  1
      Page Faults                       0
      Page Reclaims                     599
      Page Swaps                        0
      Voluntary Context Switches        193
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           416


158     !      * The skewness and kurtosis of crime rate per capita is significantly less for riverside properties
159       than for properties not near the Charles River. This means the histogram for crime rate per capita of
160       non-riverside properties is right-tailed and has a much larger tail than the histogram for crime rate
161       per capita of riverside properties. *;
162
163       OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
174
```

Results: Homework 2.sas

1c. Statistics for BMI and Average Blood Pressure

The MEANS Procedure

Variable	Mean	Median	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
BMI	25.4255407	25.1756984	2.7932222	23.4273899	27.4236915
Averagebp	94.0000000	96.6666667	12.1106014	85.3365976	102.6634024

1d. Merged Data Set

ID	Gender	Race	CollegeEducated	SBP	DBP	Weight	Height	BMI	Averagebp
001	M	White	Yes	108	98	152	70	21.8073	101.333
002	M	Asian	Yes	128	78	171	69	25.2495	94.667
003	F	Hispanic	Yes	154	84	154	62	28.1639	107.333
004	M	African American	Yes	102	86	173	66	27.9199	91.333
005	F	Hispanic	Yes	126	100	132	55	30.6764	108.667

ID	Gender	Race	CollegeEducated	SBP	DBP	Weight	Height	BMI	Averagebp
006	M	African American	No	104	54	170	69	25.1019	70.667
007	M	Hispanic	Yes	152	74	186	74	23.8784	100.000
008	F	African American	Yes	96	70	169	72	22.9180	78.667
009	F	Asian	Yes	114	76	137	65	22.7955	88.667
010	F	Asian	No	116	90	150	64	25.7446	98.667

1e.

2a(i). Student-Teacher Ratio Greater Than 19

The MEANS Procedure

Variable	Mean	Median	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
crim	6.8248508	3.6736700	11.2716642	5.4292325	8.2204690
tax	509.0079051	666.0000000	179.2284609	486.8164602	531.1993500
medv	18.1102767	18.2000000	7.2157492	17.2168477	19.0037057

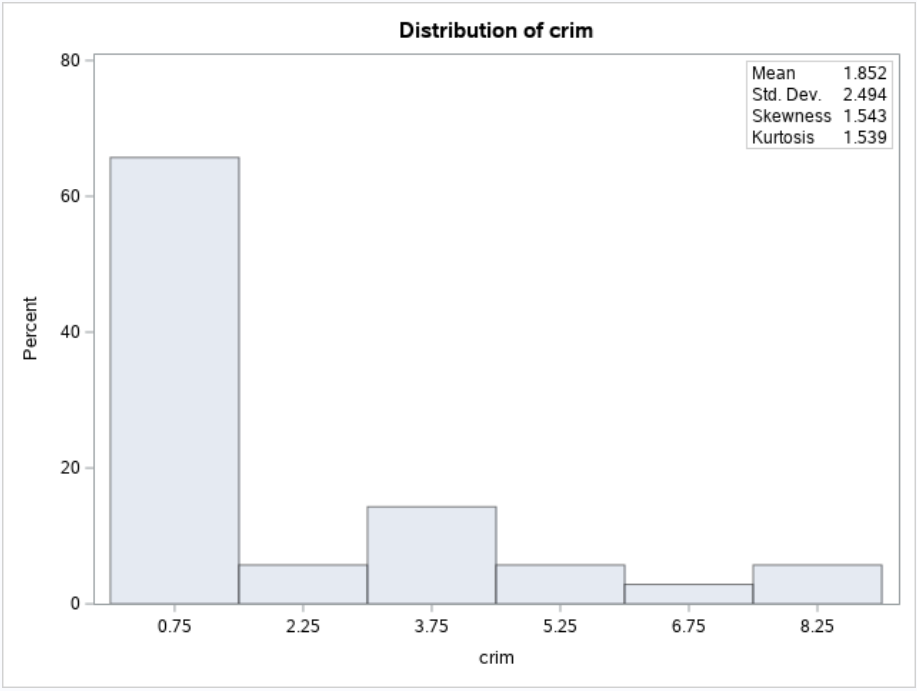
2a(ii). Student-Teacher Ratio Equal To or Below 19

The MEANS Procedure

Variable	Mean	Median	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
crim	0.4021964	0.1015300	0.7136175	0.3138387	0.4905540
tax	307.4664032	300.0000000	66.4026607	299.2446566	315.6881497
medv	26.9553360	24.0000000	8.8413408	25.8606319	28.0500400

2c. Crime Rate per Capita for Properties Near Charles River

The UNIVARIATE Procedure



2c. Crime Rate per Capita for Properties Not Near Charles River

The UNIVARIATE Procedure