**Section 2/Question 2 – Univariate Analysis of Cohort at Risk for Hypertension**

**Variable Name Data Type**

Type Character

Age Numeric

Height Numeric

Weight Numeric

FILENAME F1 'C:\Users\Kimist99\Desktop\RutgersU\BINF5120\BINF5120\_Assign2\Demo11.txt';

DATA Demo11;

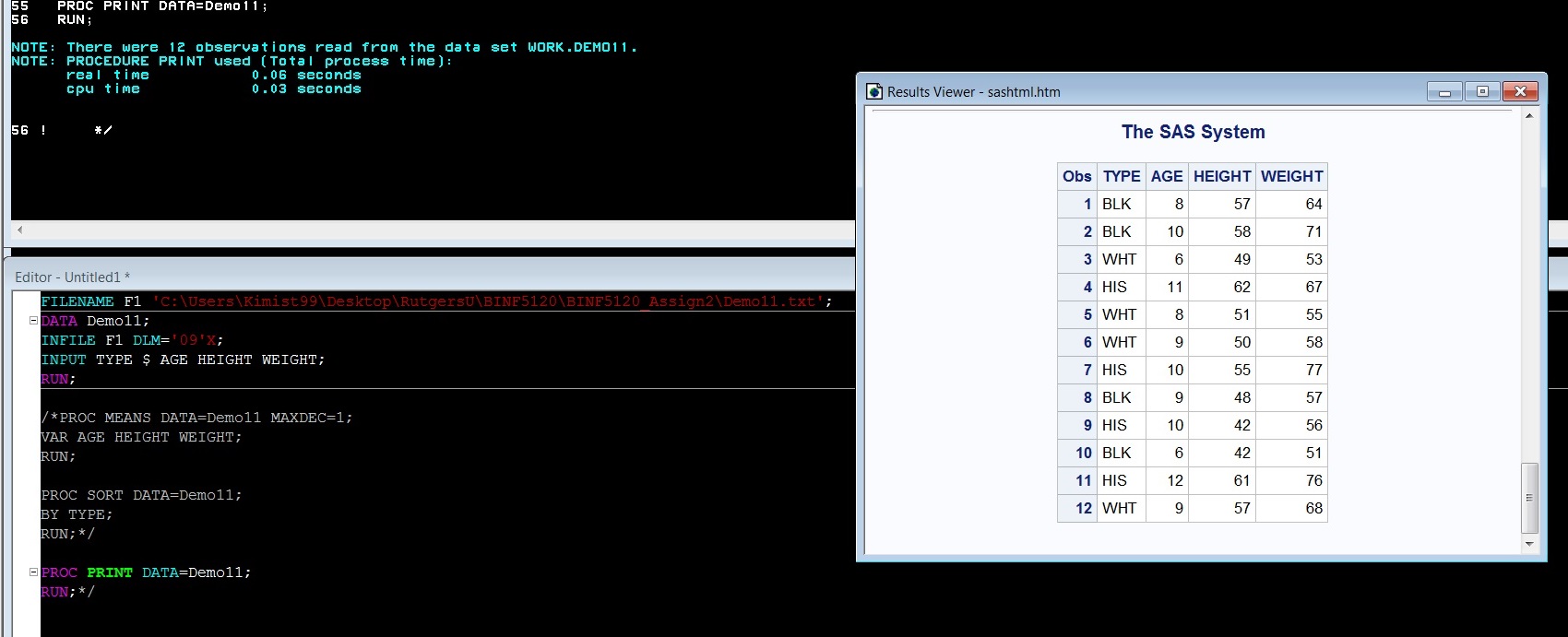
INFILE F1 DLM='09'X;

INPUT TYPE $ AGE HEIGHT WEIGHT;

RUN;

PROC **PRINT** DATA=Demo11;

RUN;

****

**Section 2/Question 3**

FILENAME F1 'C:\Users\Kimist99\Desktop\RutgersU\BINF5120\BINF5120\_Assign2\Demo11.txt';

DATA Demo11;

INFILE F1 DLM='09'X;

INPUT TYPE $ AGE HEIGHT WEIGHT;

RUN;

PROC **MEANS** DATA=Demo11 MAXDEC=1;

VAR AGE HEIGHT WEIGHT;

RUN;

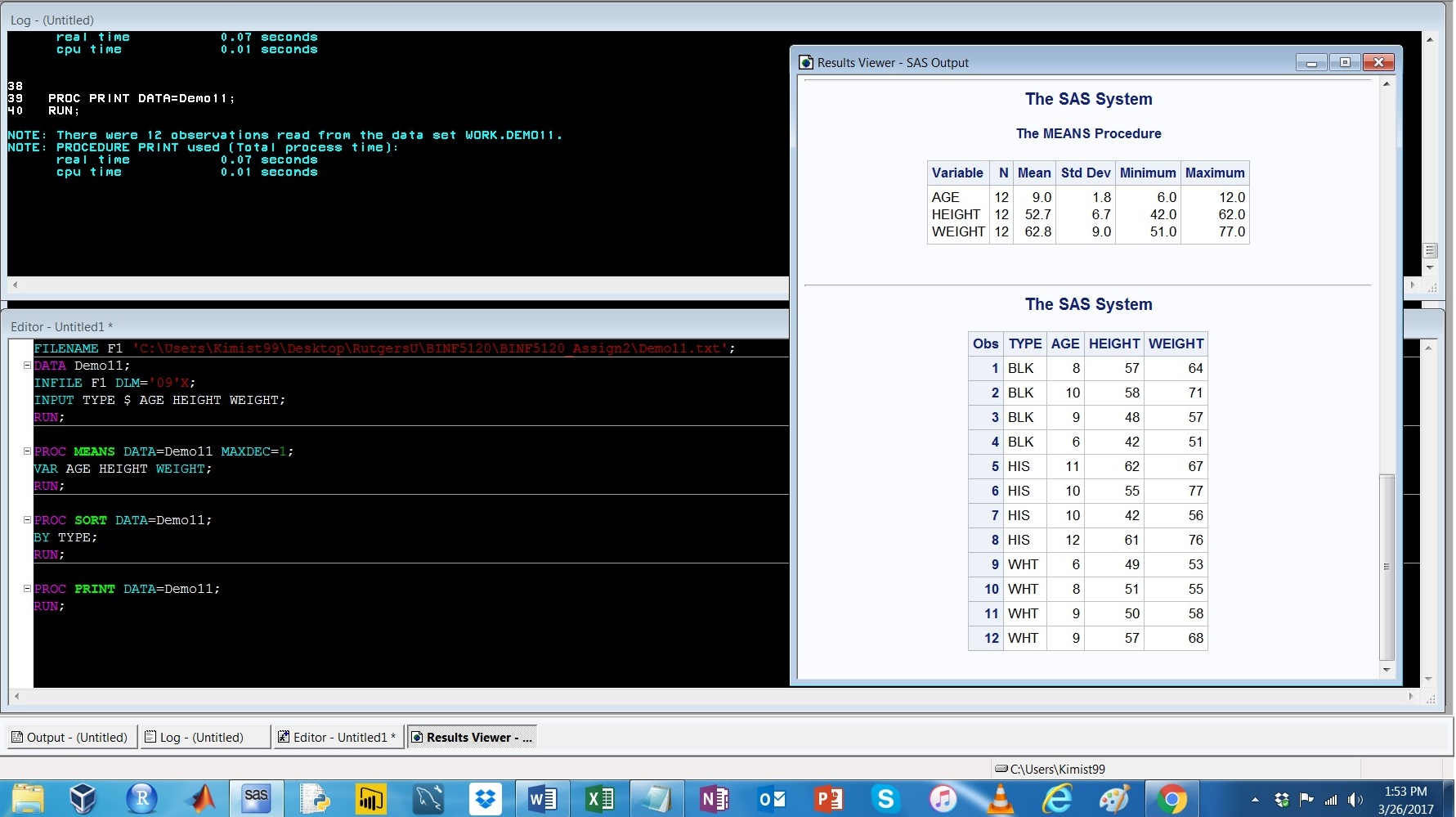
PROC **SORT** DATA=Demo11;

BY TYPE;

RUN;

PROC **PRINT** DATA=Demo11;

RUN;



**Section 2/Question 4**

FILENAME F1 'C:\Users\Kimist99\Desktop\RutgersU\BINF5120\BINF5120\_Assign2\Demo11.txt';

DATA Demo11;

INFILE F1 DLM='09'X;

INPUT TYPE $ AGE HEIGHT WEIGHT;

RUN;

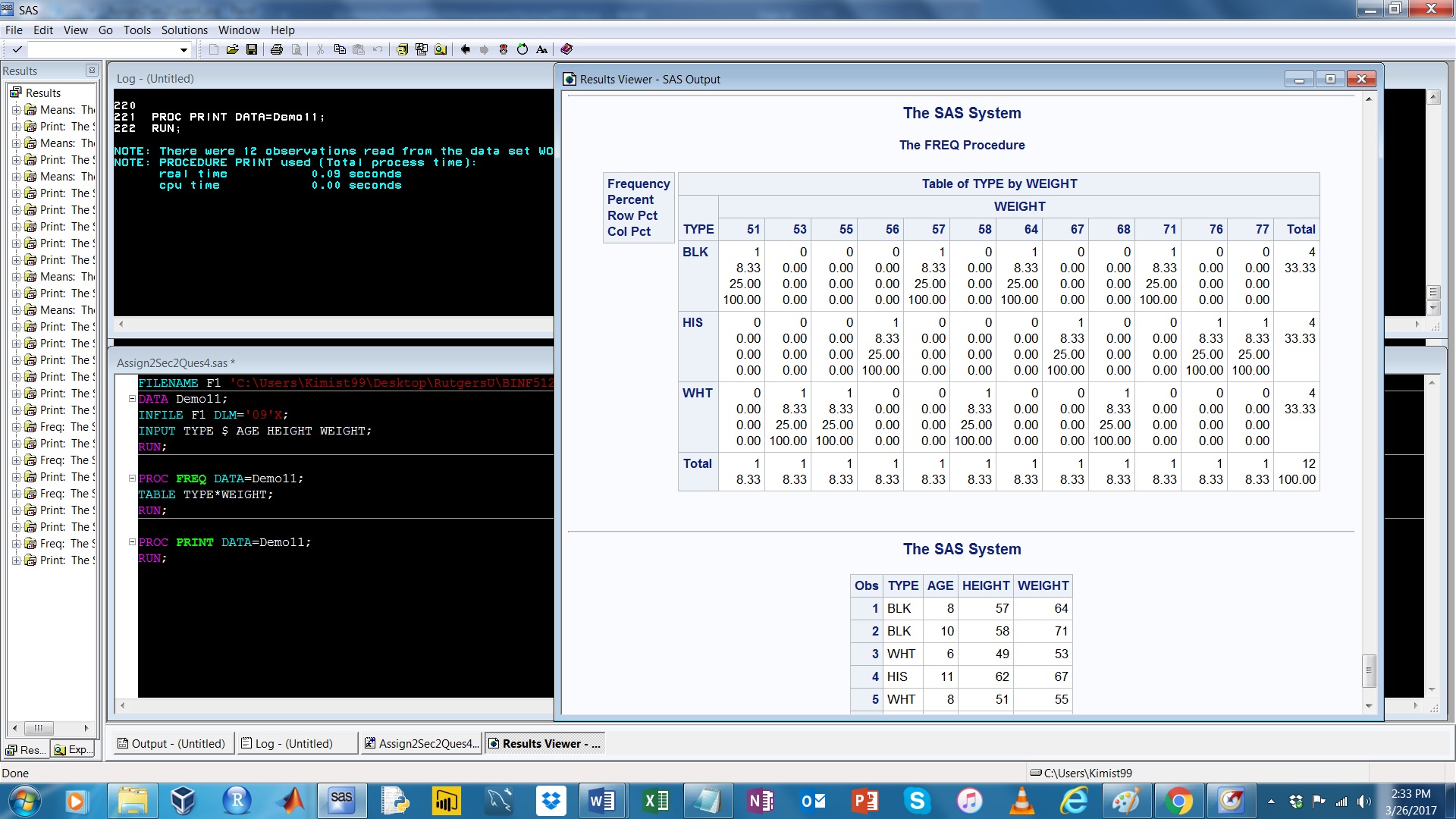
PROC **FREQ** DATA=Demo11;

TABLE TYPE\*WEIGHT;

RUN;

PROC **PRINT** DATA=Demo11;

RUN;

****

**Section 3/Question 5**

FILENAME F1 'C:\Users\Kimist99\Desktop\RutgersU\BINF5120\BINF5120\_Assign2\Demo12.txt';

DATA Demo12;

INFILE F1 DLM='09'X;

INPUT GENDER $ AGE SBP DBP;

RUN;

PROC **PRINT** DATA=Demo12;

RUN;

**Output Code:**

|  |
| --- |
| The SAS System |

| **Obs** | **GENDER** | **AGE** | **SBP** | **DBP** |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **1** | 1 | 55 | 124 | 88 |
| **2** | 2 | 52 | 128 | 86 |
| **3** | 1 | 63 | 126 | 64 |
| **4** | 1 | 83 | 154 | 54 |
| **5** | 2 | 37 | 104 | 70 |
| **6** | 2 | 50 | 118 | 78 |
| **7** | 2 | 55 | 130 | 86 |
| **8** | 1 | 40 | 126 | 88 |
| **9** | 2 | 37 | 100 | 56 |
| **10** | 2 | 57 | 164 | 70 |
| **11** | 1 | 21 | 122 | 66 |
| **12** | 2 | 85 | 138 | 66 |
| **13** | 2 | 84 | 124 | 54 |
| **14** | 1 | 69 | 124 | 68 |
| **15** | 1 | 55 | 132 | 82 |
| **16** | 2 | 37 | 106 | 76 |
| **17** | 2 | 59 | 110 | 76 |
| **18** | 2 | 73 | 218 | 94 |
| **19** | 1 | 78 | 148 | 84 |
| **20** | 2 | 51 | 128 | 68 |
| **21** | 1 | 71 | 130 | 62 |
| **22** | 1 | 69 | 170 | 92 |
| **23** | 1 | 52 | 116 | 78 |
| **24** | 1 | 25 | 132 | 76 |
| **25** | 2 | 57 | 108 | 72 |
| **26** | 1 | 70 | 130 | 58 |
| **27** | 1 | 85 | 182 | 66 |
| **28** | 2 | 60 | 170 | 86 |
| **29** | 2 | 80 | 202 | 76 |
| **30** | 2 | 57 | 122 | 74 |
| **31** | 1 | 26 | 104 | 58 |
| **32** | 1 | 28 | 114 | 70 |
| **33** | 1 | 56 | 144 | 90 |
| **34** | 2 | 25 | 102 | 56 |
| **35** | 2 | 64 | 132 | 56 |
| **36** | 2 | 68 | 106 | 58 |
| **37** | 1 | 75 | 144 | 70 |
| **38** | 1 | 70 | 116 | 64 |
| **39** | 1 | 26 | 114 | 64 |
| **40** | 1 | 20 | 126 | 74 |
| **41** | 1 | 25 | 124 | 78 |
| **42** | 1 | 76 | 148 | 68 |
| **43** | 1 | 81 | 122 | 54 |
| **44** | 2 | 32 | 104 | 66 |
| **45** | 2 | 38 | 128 | 78 |
| **46** | 2 | 63 | 132 | 72 |
| **47** | 1 | 66 | 140 | 66 |
| **48** | 2 | 60 | 108 | 58 |
| **49** | 2 | 37 | 100 | 74 |
| **50** | 1 | 75 | 158 | 88 |
| **51** | 1 | 55 | 152 | 76 |
| **52** | 1 | 36 | 112 | 68 |
| **53** | 1 | 37 | 114 | 80 |
| **54** | 2 | 20 | 108 | 70 |
| **55** | 1 | 59 | 120 | 74 |
| **56** | 1 | 85 | 136 | 54 |
| **57** | 1 | 72 | 106 | 62 |
| **58** | 2 | 31 | 120 | 72 |
| **59** | 2 | 31 | 126 | 86 |
| **60** | 2 | 63 | 168 | 86 |
| **61** | 1 | 27 | 124 | 82 |
| **62** | 2 | 67 | 172 | 86 |
| **63** | 2 | 71 | 156 | 78 |
| **64** | 2 | 51 | 112 | 74 |
| **65** | 2 | 60 | 146 | 82 |
| **66** | 1 | 65 | 140 | 60 |
| **67** | 2 | 20 | 88 | 50 |
| **68** | 1 | 51 | 114 | 72 |
| **69** | 2 | 69 | 176 | 84 |
| **70** | 2 | 43 | 142 | 100 |
| **71** | 2 | 45 | 118 | 60 |
| **72** | 2 | 61 | 128 | 64 |
| **73** | 1 | 73 | 144 | 50 |
| **74** | 2 | 47 | 162 | 86 |
| **75** | 1 | 61 | 166 | 70 |
| **76** | 2 | 80 | 180 | 80 |
| **77** | 1 | 40 | 138 | 96 |
| **78** | 2 | 29 | 100 | 66 |
| **79** | 1 | 26 | 118 | 52 |
| **80** | 1 | 85 | 138 | 32 |
| **81** | 1 | 62 | 158 | 88 |
| **82** | 2 | 40 | 166 | 88 |
| **83** | 1 | 20 | 102 | 62 |
| **84** | 2 | 54 | 130 | 76 |
| **85** | 2 | 23 | 124 | 82 |
| **86** | 2 | 51 | 116 | 68 |
| **87** | 1 | 23 | 130 | 88 |
| **88** | 1 | 74 | 148 | 52 |
| **89** | 1 | 21 | 112 | 68 |
| **90** | 2 | 44 | 102 | 60 |
| **91** | 1 | 23 | 114 | 82 |
| **92** | 2 | 34 | 120 | 68 |
| **93** | 2 | 26 | 108 | 40 |
| **94** | 1 | 24 | 128 | 82 |
| **95** | 2 | 49 | 128 | 76 |
| **96** | 1 | 38 | 122 | 76 |
| **97** | 1 | 26 | 124 | 72 |
| **98** | 1 | 31 | 108 | 70 |
| **99** | 2 | 62 | 110 | 72 |
| **100** | 1 | 85 | 160 | 58 |
| **101** | 2 | 67 | 116 | 60 |
| **102** | 1 | 26 | 126 | 82 |
| **103** | 1 | 41 | 122 | 76 |
| **104** | 1 | 43 | 136 | 64 |
| **105** | 2 | 85 | 128 | 44 |
| **106** | 1 | 45 | 120 | 90 |
| **107** | 1 | 44 | 138 | 94 |
| **108** | 2 | 26 | 110 | 74 |
| **109** | 2 | 65 | 142 | 58 |
| **110** | 2 | 22 | 106 | 54 |
| **111** | 1 | 41 | 108 | 70 |
| **112** | 1 | 37 | 132 | 86 |
| **113** | 1 | 63 | 108 | 68 |
| **114** | 2 | 83 | 178 | 82 |
| **115** | 2 | 34 | 100 | 62 |
| **116** | 2 | 64 | 144 | 66 |
| **117** | 1 | 75 | 108 | 68 |
| **118** | 2 | 64 | 122 | 64 |
| **119** | 2 | 72 | 134 | 64 |
| **120** | 1 | 50 | 128 | 84 |
| **121** | 2 | 22 | 108 | 66 |
| **122** | 2 | 22 | 120 | 54 |
| **123** | 2 | 61 | 100 | 58 |
| **124** | 1 | 41 | 136 | 84 |
| **125** | 2 | 32 | 122 | 60 |
| **126** | 2 | 28 | 100 | 74 |
| **127** | 1 | 32 | 134 | 84 |
| **128** | 1 | 51 | 126 | 82 |
| **129** | 2 | 61 | 118 | 62 |
| **130** | 1 | 41 | 100 | 68 |
| **131** | 2 | 43 | 108 | 74 |
| **132** | 1 | 50 | 122 | 66 |
| **133** | 2 | 63 | 152 | 80 |
| **134** | 1 | 42 | 114 | 74 |
| **135** | 2 | 40 | 92 | 70 |
| **136** | 1 | 44 | 110 | 62 |
| **137** | 2 | 43 | 126 | 66 |
| **138** | 2 | 36 | 122 | 78 |
| **139** | 1 | 45 | 140 | 80 |
| **140** | 2 | 44 | 124 | 78 |
| **141** | 1 | 68 | 136 | 54 |
| **142** | 1 | 54 | 144 | 64 |
| **143** | 1 | 21 | 110 | 58 |
| **144** | 2 | 51 | 102 | 62 |
| **145** | 1 | 61 | 142 | 64 |
| **146** | 1 | 77 | 126 | 60 |
| **147** | 1 | 66 | 116 | 70 |
| **148** | 2 | 77 | 116 | 54 |
| **149** | 2 | 61 | 152 | 74 |
| **150** | 2 | 42 | 110 | 80 |
| **151** | 2 | 78 | 140 | 26 |
| **152** | 2 | 45 | 98 | 76 |
| **153** | 1 | 59 | 124 | 74 |
| **154** | 1 | 23 | 122 | 64 |
| **155** | 2 | 58 | 160 | 74 |
| **156** | 1 | 39 | 138 | 90 |
| **157** | 2 | 37 | 120 | 68 |
| **158** | 1 | 35 | 114 | 68 |
| **159** | 1 | 41 | 118 | 78 |
| **160** | 2 | 24 | 108 | 58 |
| **161** | 1 | 30 | 124 | 82 |

**Section 3/Question 6**

FILENAME F1 'C:\Users\Kimist99\Desktop\RutgersU\BINF5120\BINF5120\_Assign2\Demo12.txt';

DATA Demo12;

INFILE F1 DLM='09'X;

INPUT GENDER $ AGE SBP DBP;

RUN;

/\*PROC PRINT DATA= Demo12;

RUN;\*/

PROC **SURVEYSELECT** DATA= Demo12

METHOD= SEQ

SAMPRATE= 0.1

SEED= 1224

OUT= Ques6;

;

RUN;

/\*PROC PRINT DATA= Ques6;

RUN;\*/

PROC **UNIVARIATE** DATA= Ques6;

VAR SBP DBP;

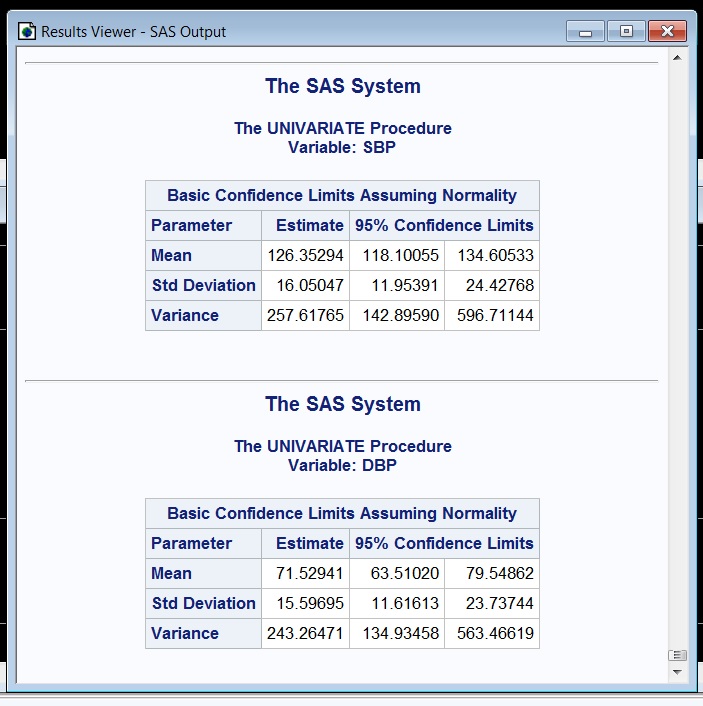
RUN;

ODS SELECT BASICINTERVALS;

PROC **UNIVARIATE** DATA= Ques6 CIBASIC;

VAR SBP DBP;

RUN;



**Section 3/Question 7**

Questions 6 and 7 gave very similar results and this is somewhat predicted due to the way the samples were collected for each question.

FILENAME F1 'C:\Users\Kimist99\Desktop\RutgersU\BINF5120\BINF5120\_Assign2\Demo12.txt';

DATA Demo12;

INFILE F1 DLM='09'X;

INPUT GENDER $ AGE SBP DBP;

RUN;

PROC **SURVEYSELECT** DATA= Demo12

METHOD= SYS

SAMPSIZE= 17

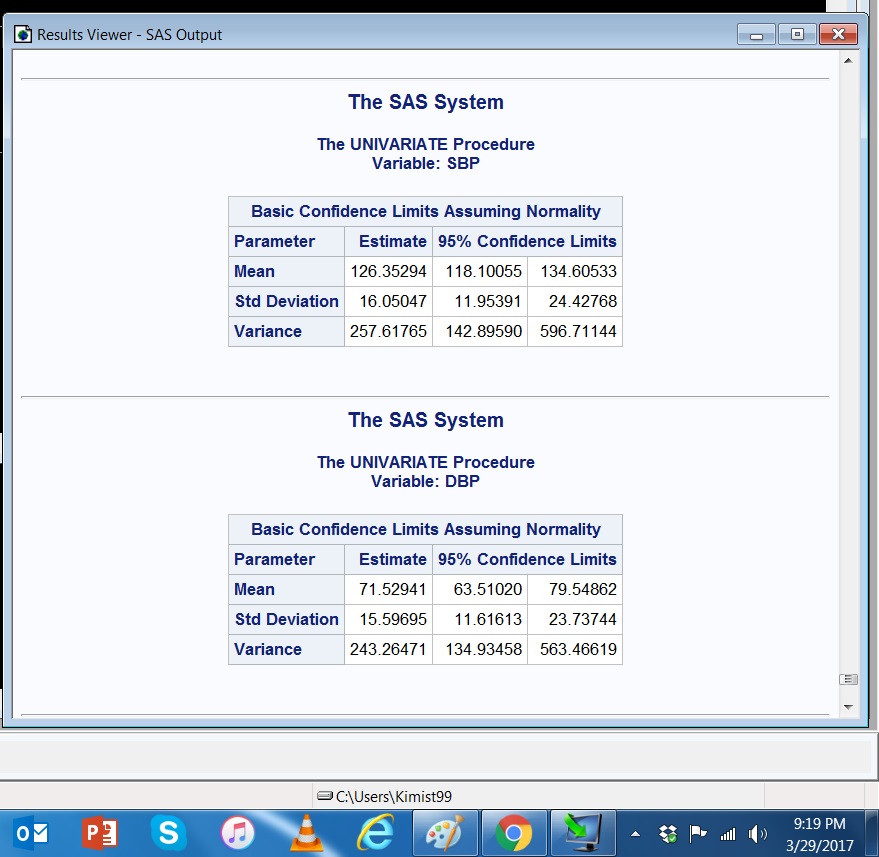
OUT= Ques7;

ID \_all\_;

RUN;

PROC **PRINT** DATA= Ques7;

RUN;



**Section 3/Question 8**

The results for question #8 specifically the standard deviation and variance, were both very different from the results obtained in both #6 and #7 and seeding the sample in question #8, probably contributes to these. Allowing for duplicates potentially skews the data away from a normal distribution. Both the SD and variance, give information about the spread of data from the mean and both are significantly different in question #8 compared to #6 and #7.

FILENAME F1 'C:\Users\Kimist99\Desktop\Demo12.txt';

DATA Demo12;

INFILE F1 DLM='09'X;

INPUT GENDER $ AGE SBP DBP;

RUN;

PROC **SURVEYSELECT** DATA= Demo12

METHOD= URS

SAMPSIZE= 17

SEED= 12345

OUT= Ques8;

ID \_ALL\_;

RUN;

PROC **UNIVARIATE** DATA= Ques8;

VAR SBP DBP;

RUN;

ODS SELECT BASICINTERVALS;

PROC **UNIVARIATE** DATA= Ques8 CIBASIC;

VAR SBP DBP;

RUN;

PROC **PRINT** DATA= Ques8;

RUN;



**Section 3/Question 9**

The variance in question #8 is significantly different from the one obtained in question #9 owing to the differences between a sample using the URS and SRS methods of statistical analysis. As a result, the information contained in the URS run is less reliable or reflective to the population of data being examined than the data collected using the SRS method. As in the previous question comparing the

FILENAME F1 'C:\Users\Kimist99\Desktop\RutgersU\BINF5120\BINF5120\_Assign2\Demo12.txt';

DATA Demo12;

INFILE F1 DLM='09'X;

INPUT GENDER $ AGE SBP DBP;

RUN;

PROC **SURVEYSELECT** DATA= Demo12

METHOD= SRS

SAMPSIZE= 17

OUT= Ques9;

ID \_ALL\_;

RUN;

PROC **UNIVARIATE** DATA= Ques9;

VAR SBP DBP;

RUN;

ODS SELECT BASICINTERVALS;

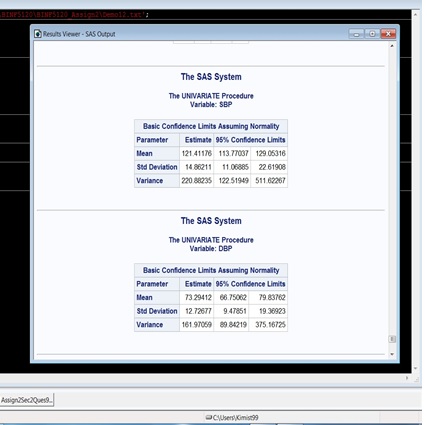
PROC **UNIVARIATE** DATA= Ques9 CIBASIC;

VAR SBP DBP;

RUN;

PROC **PRINT** DATA= Ques9;

RUN;



**Section 3/Question 10**

The data for systolic blood pressure (SBP) is not normally distributed and is skewed slightly to the left (see image below). There are also numerous outliers on the right tail of the histogram which contributes to the skewed/non normalized results. The data for diastolic blood pressure (DBP) is reasonably normally distributed, although there are outliers on the left tail of the histogram (see image below).

FILENAME F1 'C:\Users\Kimist99\Desktop\Demo12.txt';

DATA Demo12;

INFILE F1 DLM='09'X;

INPUT GENDER $ AGE SBP DBP;

RUN;

PROC **SORT** DATA=DEMO12;

BY DESCENDING SBP;

RUN;

PROC **UNIVARIATE** DATA=DEMO12;

VAR SBP DBP;

HISTOGRAM;

RUN;

PROC **PRINT** DATA=Demo12;

RUN;

