· Q2(5), 3(5), 5(5),7(5),16(10),19(10), 23(a-d, 20)=60 55/60 = 91.6

Intro to SAS HW Chapter 8

Ian Liu

HW Chapter 8 Q2

2. Which types of plots can be overlaid together in a single graph?

a. box and scatter
b. loess and density
c. scatter and series
d. histogram and bar

C. scatter and series

HW Chapter 8 Q3

3. The STYLE= option can go in which statement in order to specify a style template for a graph?

a. ODS HTML
b. ODS GRAPHICS
c. PROC SGPLOT
d. None of the above

a. ODS HTML

HW Chapter 8 Q5

5. Which statement will draw a normal curve in PROC SGPLOT?

a. `HISTOGRAM Age / DENSITY;`
b. `DENSITY Age;`
c. `HISTOGRAM Age / NORMAL;`
d. `NORMAL Age;`

B. DENSITY Age;

HW Chapter 8 Q7

7. You can produce a fitted line on a scatter plot with which SGPLOT statement?

a. SCATTER
b. SERIES
c. REG
d. All of the above

C. REG

HW Chapter 8 Q16

16. What is the main difference between graphs that are automatically created by ODS Graphics versus graphs created using PROC SGPLOT?

The main difference between graphs from ODS Graphics and graphs from PROC SGPLOT is that graphs for ODS Graphics require the ODS GRAPHICS ON statement to be on, but the SGPLOT procedure always produces graphs so it does not require the statement to be on.

-5

Graphs that are automatically created by ODS Graphics are default graphs that are appropriate to certain procedures and are integrated with the procedure's tabular output. Graphs created with PROC SGPLOT are stand-alone graphs.

HW Chapter 8 Q19

19. When you use a SERIES statement in PROC SGPLOT, the X axis data must be organized in the correct order. Explain what can go wrong if the data are not organized properly and how this can be fixed.

If the data are not organized properly, the dots will not be connected correctly. Instead, they will connect with the next dot based on the sequence of the data. To fix this, simply PROC SORT before sorting your data.

HW Chapter 8 Q23

23. The *World Bank* works internationally with the goal of reducing poverty. Their website tracks population by country. The SAS data set called POPULATION contains data on the estimated number of residents (in 100,000s) of various countries by year.

    a. Examine this SAS data set including the variable labels and attributes. Create a histogram of the most recent population estimates for all countries combined.
    b. Create separate histograms of the most recent population estimates for each continent.
    c. Create a single graph with box plots of the most recent population estimates per continent.
    d. In a comment in your program, describe the main differences between the statistical information that can be visualized in the histograms from part b) compared to the box plots from part c).

23A

```
/* 23A */
DATA POPULATION;
  set '/home/u62223361/Intro to SAS/HW8/population.sas7bdat';
RUN;

PROC CONTENTS DATA= POPULATION;
RUN;

PROC SGPLOT DATA = POPULATION;
        HISTOGRAM Y1;
        DENSITY Y1;
        DENSITY Y1 / TYPE = KERNEL;
RUN;
```

Given the distribution, histograms are a better choice.

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 23A */
70         DATA POPULATION;
71           set '/home/u62223361/Intro to SAS/HW8/population.sas7bdat';
NOTE: Data file WC000001.POPULATION.DATA is in a format that is native to another host, or the file encoding does not match the
      session encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce
      performance.
72         RUN;

NOTE: There were 220 observations read from the data set /home/u62223361/Intro to SAS/HW8/population.sas7bdat.
NOTE: The data set WORK.POPULATION has 220 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              1154.56k
      OS Memory           26788.00k
      Timestamp           04/23/2024 04:17:55 AM
      Step Count                        247  Switch Count  3
      Page Faults                       0
      Page Reclaims                     98
      Page Swaps                        0
      Voluntary Context Switches        25
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264
```

```
73
74          PROC CONTENTS DATA= POPULATION;
75          RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              1846.40k
      OS Memory           27048.00k
      Timestamp           04/23/2024 04:17:55 AM
      Step Count                        248  Switch Count  0
      Page Faults                       0
      Page Reclaims                     95
      Page Swaps                        0
      Voluntary Context Switches        2
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           16


76
77          PROC SGPLOT DATA = POPULATION;
78          HISTOGRAM Y1;
79          DENSITY Y1;
80          DENSITY Y1 / TYPE = KERNEL;
81          RUN;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.08 seconds
      user cpu time       0.04 seconds
      system cpu time     0.01 seconds
      memory              8392.68k
      OS Memory           30508.00k
      Timestamp           04/23/2024 04:17:55 AM
      Step Count                        249  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1126
      Page Swaps                        0
      Voluntary Context Switches        202
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           752

NOTE: There were 220 observations read from the data set WORK.POPULATION.

82
83
84          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
94
```

The CONTENTS Procedure

Data Set Name	WORK.POPULATION	**Observations**	220
Member Type	DATA	**Variables**	6
Engine	V9	**Indexes**	0
Created	04/23/2024 00:17:55	**Observation Length**	96
Last Modified	04/23/2024 00:17:55	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

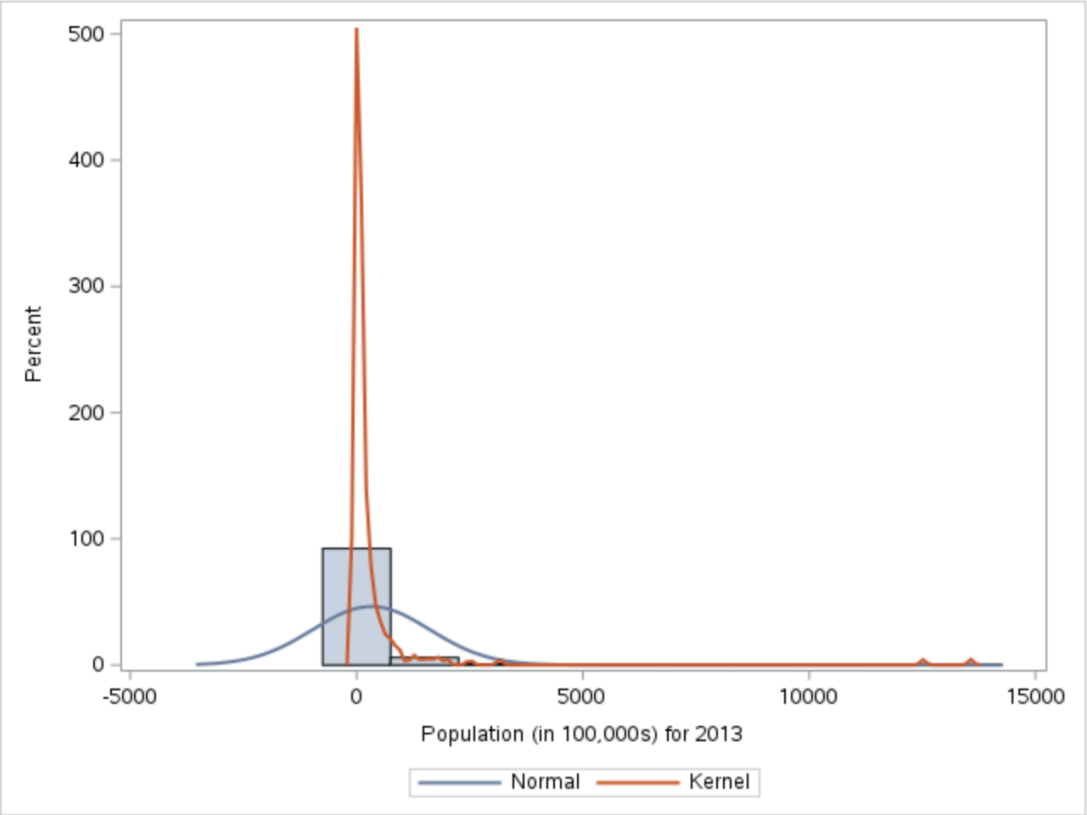
Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1363
Obs in First Data Page	220
Number of Data Set Repairs	0
Filename	/saswork/SAS_workCBE20001910A_odaws02-usw2-2.oda.sas.com/SAS_work8D960001910A_odaws02-usw2-2.oda.sas.com/population.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	1677729574
Access Permission	rw-r--r--
Owner Name	u62223361
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
1	Continent	Char	8	Continent name (AF - Africa, AS - Asia, EU - Europe, NA - North America, SA - South America, OC - Oceania, AN - Antarctica)
2	Country	Char	50	Country Name

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
1	Continent	Char	8	Continent name (AF - Africa, AS - Asia, EU - Europe, NA - North America, SA - South America, OC - Oceania, AN - Antarctica)
2	Country	Char	50	Country Name
3	Y1	Num	8	Population (in 100,000s) for 2013
4	Y2	Num	8	Population (in 100,000s) for 2012
5	Y3	Num	8	Population (in 100,000s) for 2011
6	Y4	Num	8	Population (in 100,000s) for 2010



23B

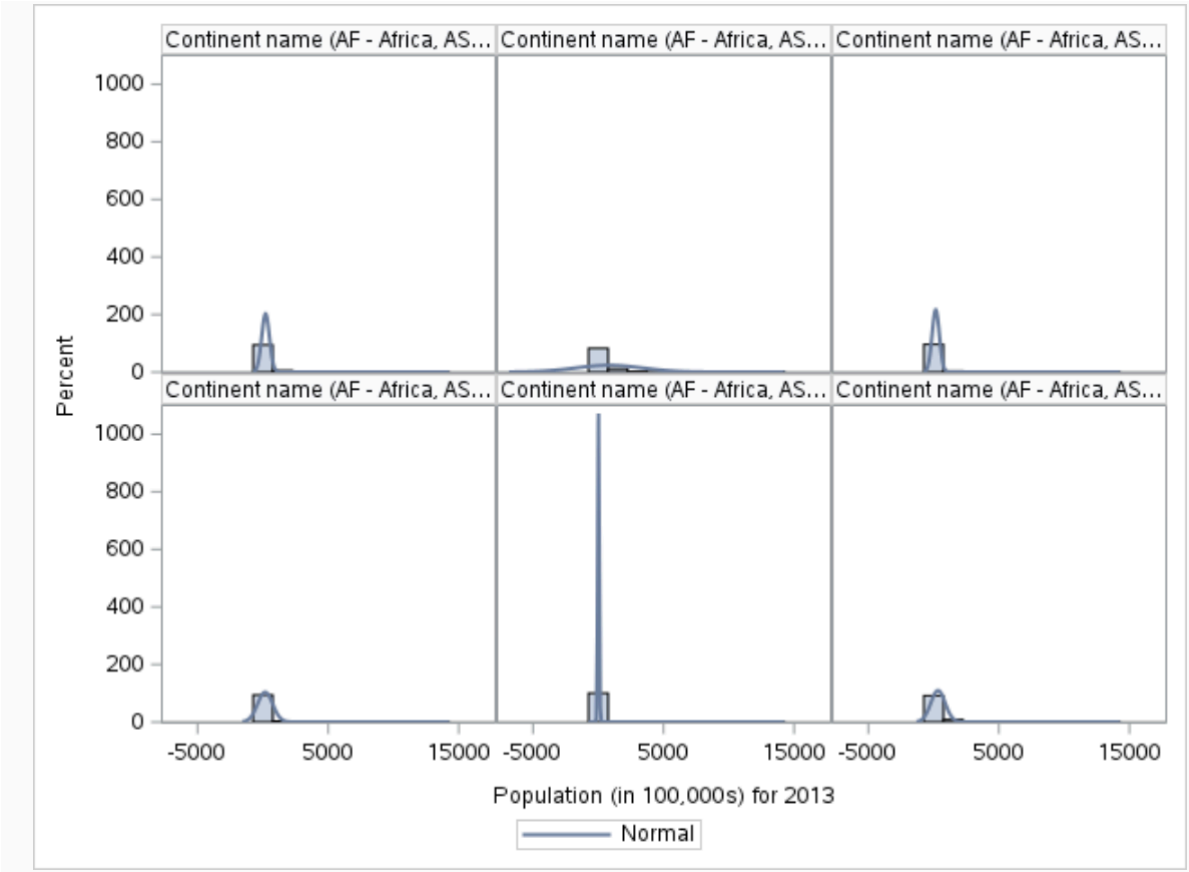
```
/* 23B */
PROC SGPANEL DATA = POPULATION;
        PANELBY Continent;
        HISTOGRAM Y1;
        DENSITY Y1;
RUN;
```

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 23B */
70         PROC SGPANEL DATA = POPULATION;
71         PANELBY Continent;
72         HISTOGRAM Y1;
73         DENSITY Y1;
74         RUN;

NOTE: PROCEDURE SGPANEL used (Total process time):
      real time           0.37 seconds
      user cpu time       0.10 seconds
      system cpu time     0.03 seconds
      memory              9815.31k
      OS Memory           30768.00k
      Timestamp           04/23/2024 04:20:13 AM
      Step Count                        255  Switch Count  40
      Page Faults                       0
      Page Reclaims                     1932
      Page Swaps                        0
      Voluntary Context Switches        3124
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           3176

NOTE: Some of the tick values have been thinned.
NOTE: Some of the tick values have been thinned.
NOTE: Some of the tick values have been thinned.
NOTE: There were 220 observations read from the data set WORK.POPULATION.

75
76         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
86
```



23C

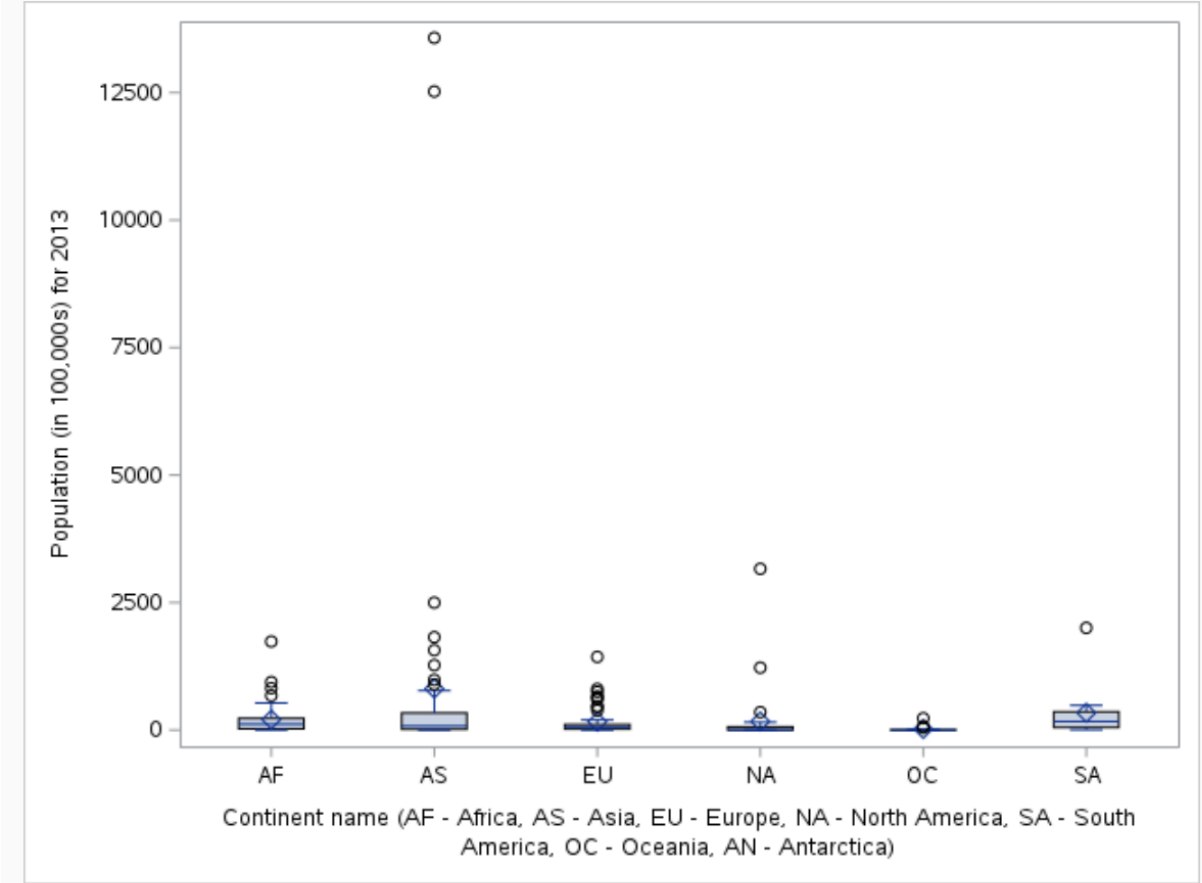
```
/* 23C */
PROC SGPLOT DATA = POPULATION;
        VBOX Y1 / CATEGORY= Continent;
RUN;
```

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 23C */
70         PROC SGPLOT DATA = POPULATION;
71         VBOX Y1 / CATEGORY= Continent;
72         RUN;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.09 seconds
      user cpu time       0.04 seconds
      system cpu time     0.00 seconds
      memory              8876.18k
      OS Memory           29996.00k
      Timestamp           04/23/2024 04:21:09 AM
      Step Count                        261  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1114
      Page Swaps                        0
      Voluntary Context Switches        678
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           760

NOTE: There were 220 observations read from the data set WORK.POPULATION.

73
74         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
84
```



23D

/* 23D */

/* Histograms can visualize probability distributions

while box plots are good for assessing how the quartiles are distributed and if there are outliers. Boxplots are more concise in describing the spread and can be better when the histogram has lots of ups and downs, which could be confusing. */

HW Chapter 8 Q27

27. Suppose that at a local university the study guidelines for the College of Science and Math are to study two to three hours per unit per week. The instructor of the class, Orientation to the Statistics Major, takes these guidelines very seriously. He asks students to record their study time each week, and at the end of the term he compares their average study time per week to their term GPA. The SAS data set called STUDY_GPA contains student identification information, orientation course-section number, number of units enrolled, average time studied, and term GPA.
    a. Examine this SAS data set including the variable labels and attributes. Create box plots to compare the time studied between the two sections.
    b. Create a simple linear regression plot for time studied and GPA. Turn off the legend.
    c. Create a simple linear regression plot for time studied and GPA with a line for each section. Move the legend to the far right side of the plot.
    d. Add 95% confidence limits for the mean predicted values to your plot from part c). Adjust the transparency so that bands for both sections are visible on the plot.
    e. Add a comment to your program about any potential relationships that you see between the variables included on these three plots.

27A

```
/* 27A */
DATA STUDY_GPA;
  set '/home/u62223361/Intro to SAS/HW8/study_gpa.sas7bdat';
RUN;


PROC CONTENTS DATA= STUDY_GPA;
RUN;
PROC SGPLOT DATA = STUDY_GPA;
        VBOX AveTime / CATEGORY= SECTION;
RUN;
```

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 27A */
70         DATA STUDY_GPA;
71           set '/home/u62223361/Intro to SAS/HW8/study_gpa.sas7bdat';
NOTE: Data file WC000001.STUDY_GPA.DATA is in a format that is native to another host, or the file encoding does not match the
      session encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce
      performance.
72         RUN;

NOTE: There were 122 observations read from the data set /home/u62223361/Intro to SAS/HW8/study_gpa.sas7bdat.
NOTE: The data set WORK.STUDY_GPA has 122 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              1168.37k
      OS Memory           26532.00k
      Timestamp           04/23/2024 04:21:56 AM
      Step Count                        267  Switch Count  3
      Page Faults                       0
      Page Reclaims                     105
      Page Swaps                        0
      Voluntary Context Switches        24
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


73
74          PROC CONTENTS DATA= STUDY_GPA;
75          RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              1860.65k
      OS Memory           26792.00k
      Timestamp           04/23/2024 04:21:56 AM
      Step Count                        268  Switch Count  0
      Page Faults                       0
      Page Reclaims                     94
      Page Swaps                        0
      Voluntary Context Switches        2
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           16


76          PROC SGPLOT DATA = STUDY_GPA;
77          VBOX AveTime / CATEGORY= SECTION;
78          RUN;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.08 seconds
      user cpu time       0.04 seconds
      system cpu time     0.00 seconds
      memory              8424.75k
      OS Memory           30764.00k
      Timestamp           04/23/2024 04:21:56 AM
      Step Count                        269  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1135
      Page Swaps                        0
      Voluntary Context Switches        279
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           720

NOTE: There were 122 observations read from the data set WORK.STUDY_GPA.

79
80          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
90
```

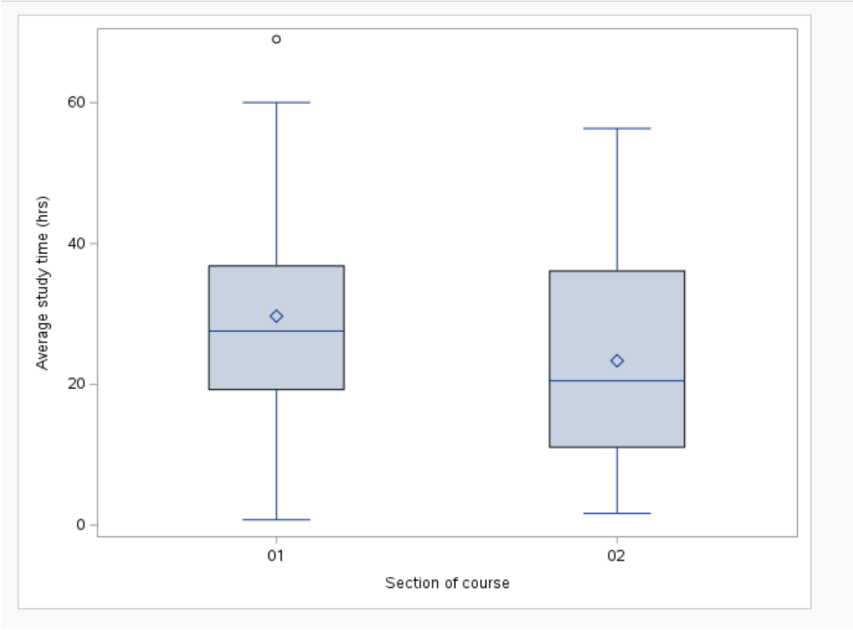
The CONTENTS Procedure

Data Set Name	WORK.STUDY_GPA	**Observations**	122
Member Type	DATA	**Variables**	7
Engine	V9	**Indexes**	0
Created	04/23/2024 00:21:56	**Observation Length**	48
Last Modified	04/23/2024 00:21:56	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	2722
Obs in First Data Page	122
Number of Data Set Repairs	0
Filename	/saswork/SAS_workCBE20001910A_odaws02-usw2-2.oda.sas.com/SAS_work8D960001910A_odaws02-usw2-2.oda.sas.com/study_gpa.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	1677729582
Access Permission	rw-r--r--
Owner Name	u62223361
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
6	AveTime	Num	8	Average study time (hrs)
2	FInitial	Char	1	First name initial
7	GPA	Num	8	GPA
1	ID	Num	8	Student ID
3	LastName	Char	10	Last name
4	Section	Char	2	Section of course
5	Units	Num	8	Number of units enrolled



27B

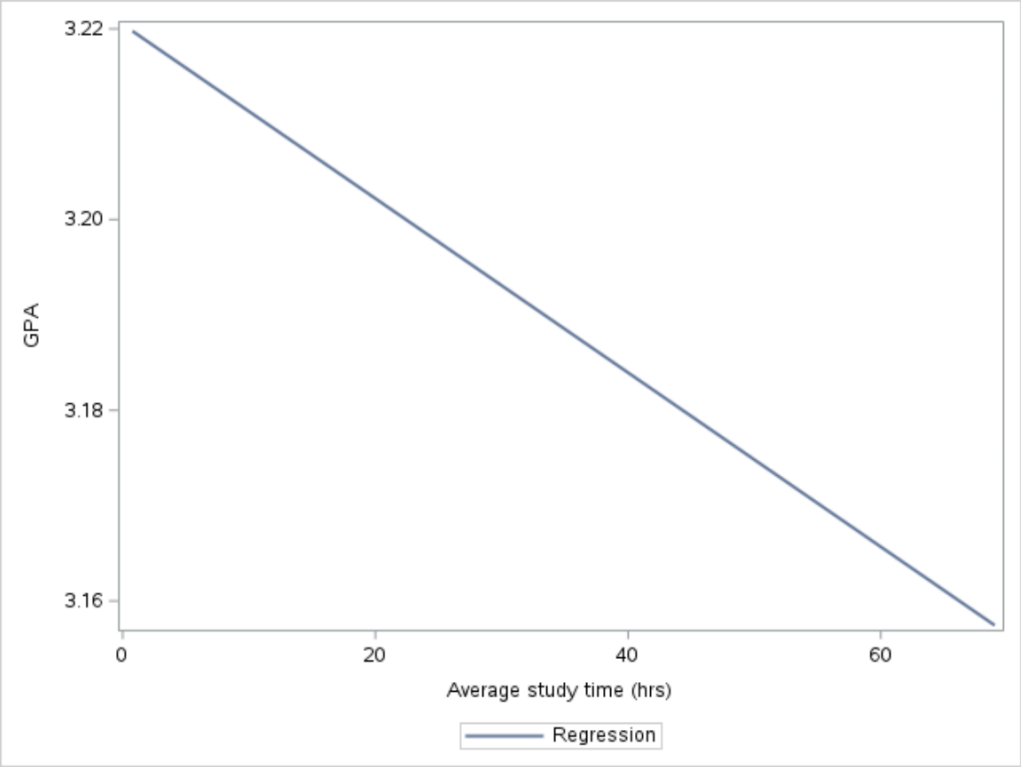
/* 27B */

PROC SGPLOT DATA = STUDY_GPA;

REG X = AveTime Y = GPA / NOMARKERS

NOLEGCLM NOLEGCLI;

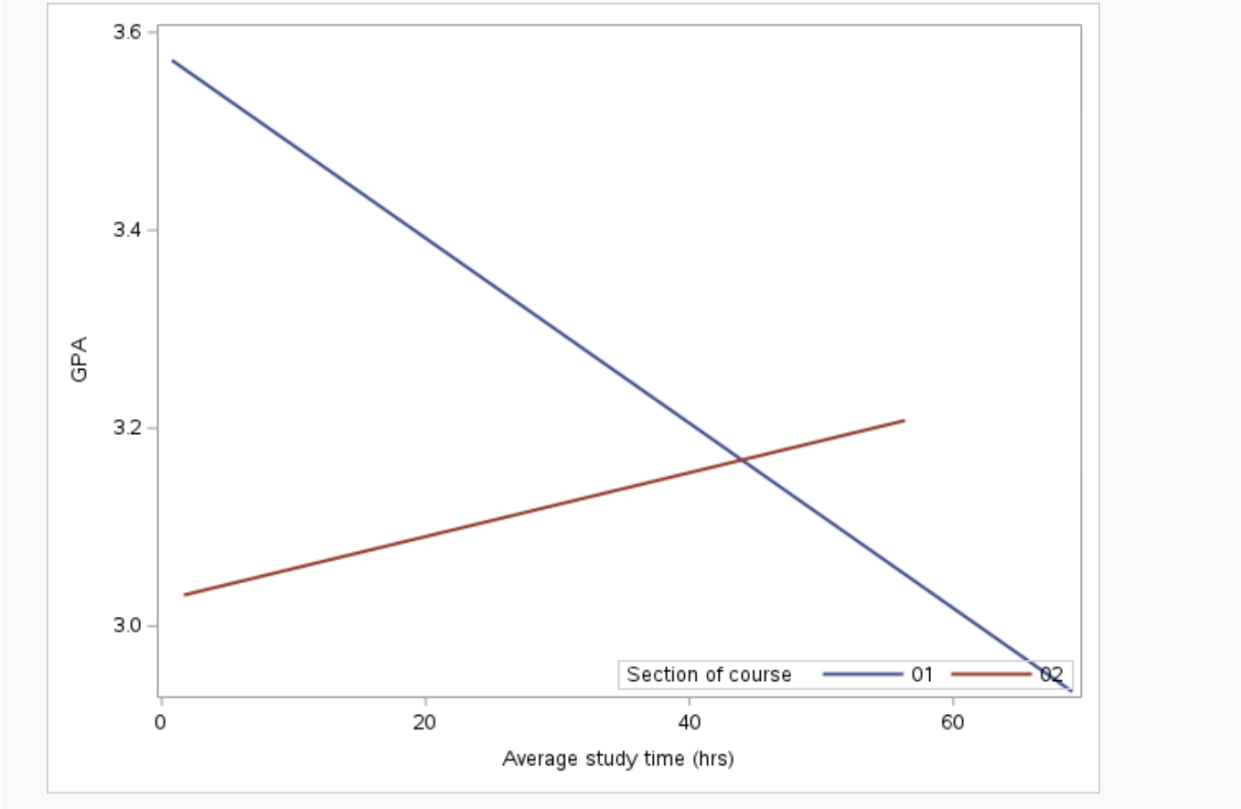
```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 27B */
70         PROC SGPLOT DATA = STUDY_GPA;
71         REG X = AveTime Y = GPA / NOMARKERS
72         NOLEGCLM NOLEGCLI;
73
74         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
84
```



27C

```
/* 27C */
PROC SGPLOT DATA = STUDY_GPA;
        REG X = AveTime Y = GPA / NOMARKERS
        ALPHA=0.05 CLMTRANSPARENCY=0.5 GROUP=SECTION;
        KEYLEGEND / LOCATION = INSIDE POSITION = BOTTOMRIGHT;
```

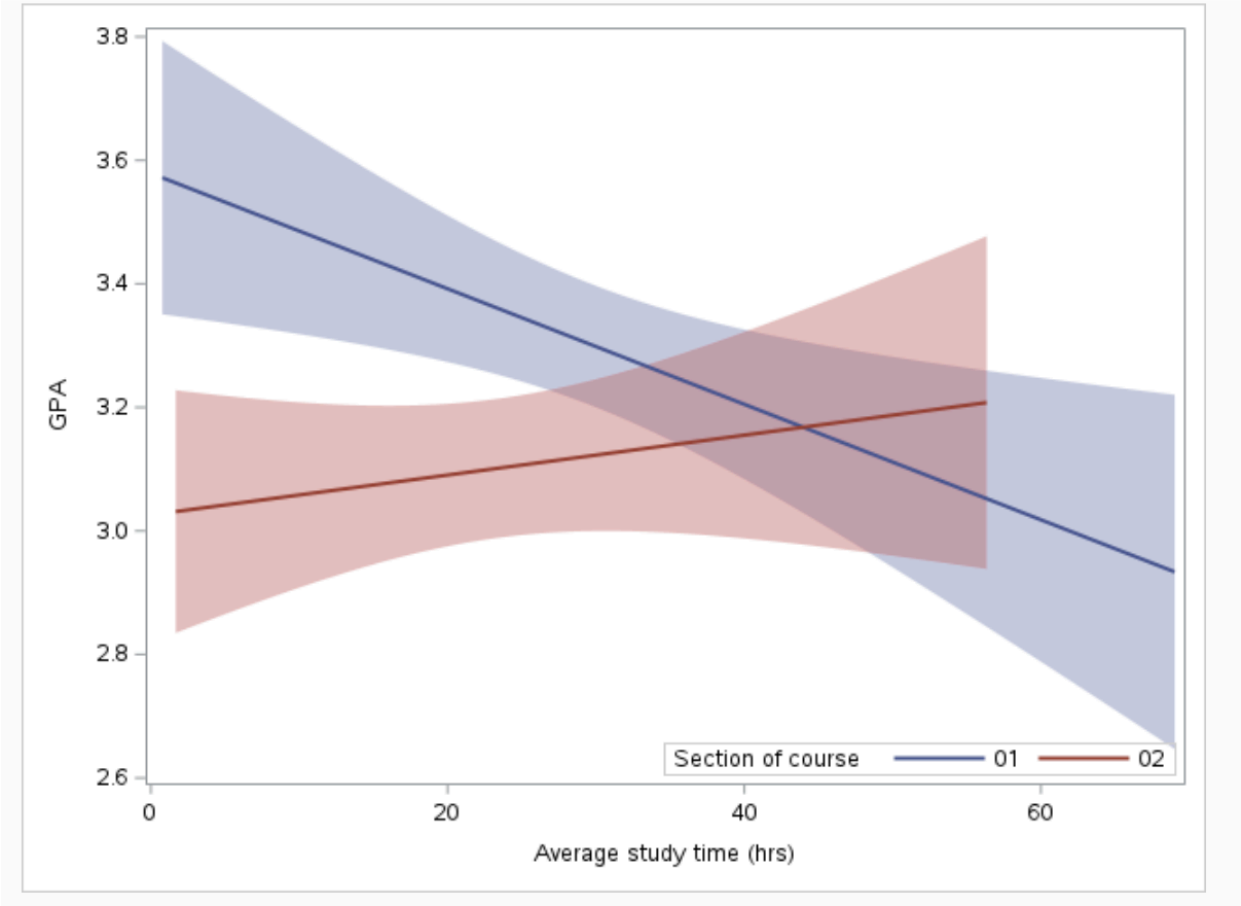
```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 27C */
70         PROC SGPLOT DATA = STUDY_GPA;
71         REG X = AveTime Y = GPA / NOMARKERS
72         ALPHA=0.05 CLMTRANSPARENCY=0.5 GROUP=SECTION;
73         KEYLEGEND / LOCATION = INSIDE POSITION = BOTTOMRIGHT;
74
75         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
85
```



27D

```
/* 27D */
PROC SGPLOT DATA = STUDY_GPA;
        REG X = AveTime Y = GPA / NOMARKERS CLM GROUP=SECTION
        NOLEGCLM NOLEGCLI ALPHA=0.05 CLMTRANSPARENCY=0.6;
        KEYLEGEND / LOCATION = INSIDE POSITION = BOTTOMRIGHT;
```

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 27D */
70         PROC SGPLOT DATA = STUDY_GPA;
71         REG X = AveTime Y = GPA / NOMARKERS CLM GROUP=SECTION
72         NOLEGCLM NOLEGCLI ALPHA=0.05 CLMTRANSPARENCY=0.6;
73         KEYLEGEND / LOCATION = INSIDE POSITION = BOTTOMRIGHT;
74
75         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
85
```



27E

/* 27E */

/* As average study time increases in section 1, GPA tends to go down.

As average study time increases in section 2, GPA tends to go up.

Average study time and GPA have a inverse correlation in section 1,

and average study time and GPA have a positive correlation in section 2. */