ST.310, Spring 2024

Homework #3

Due on Wednesday, 2/14/24

**Alyssa Eddy | Genevieve Torkornoo**

-------------------------------------------------------------------------------------------------------------------------------

**Problem#6.3.** [Page 71]

Generate histograms for nitrate, zinc, and TDS (total dissolved solids) from well#1.dat (see Page 285 for more details), and comment on the shape of the distribution. Use **LABEL** and **TITLE** statements. ***Note***. Use PROC SGPLOT. [15 pts]

**SAS code:** (Direction: Use either [SAS Monospace] font or [Courier New] font for SAS code)

data Well;

infile "\\apporto.com\dfs\LOYOLA\Users\getorkornoo\_loyola\Documents\SAS\_Data/Well#1.dat";

input date $ nitrate zinc tds;

proc sgplot data = Well;

histogram zinc ;

title "Histogram for Zinc";

xaxis label="Zinc";

yaxis label="Frequency";

proc sgplot data = Well;

histogram tds ;

title "Histogram for tds";

xaxis label="Tds";

yaxis label="Frequency";

proc sgplot data = Well;

histogram nitrate ;

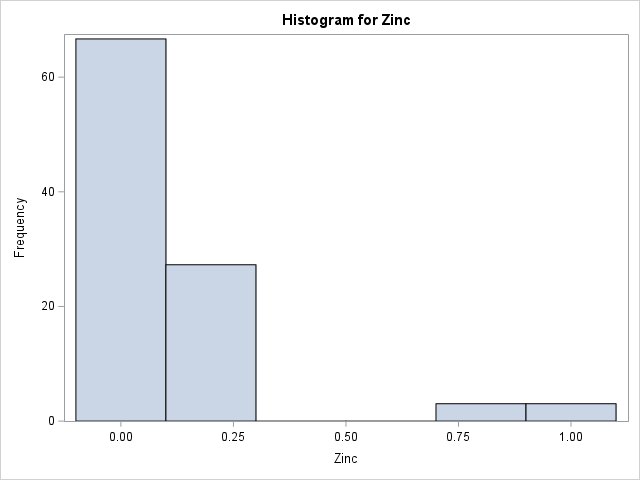
title "Histogram for nitrate";

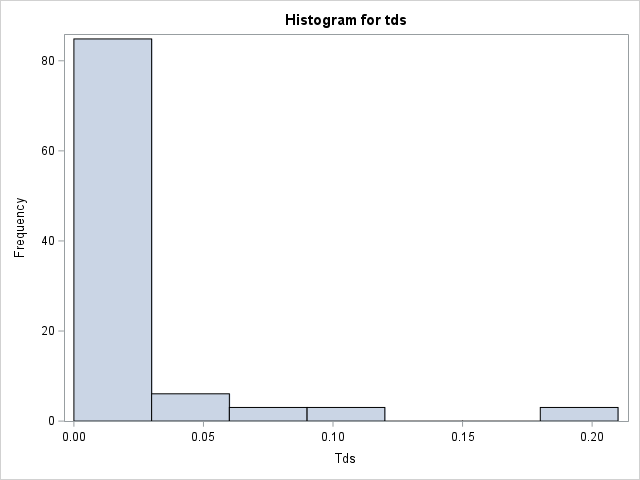
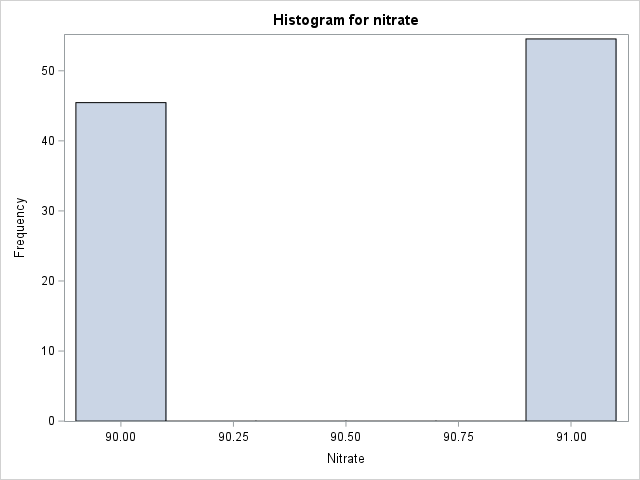
xaxis label="Nitrate";

yaxis label="Frequency";

run;

**SAS output:** (Direction: Paste SA)



-------------------------------------------------------------------------------------------------------------------------------

**Problem#6.8.** [Page 72]

Use the file btt.dat (see Page 262 for more details). Use PROC FORMAT when necessary to make the categories easier to read. Use TITLE and LABEL statements.

***Note***. Use PROC SGPLOT for the following questions.

1. Construct a bar graph of the mother’s education. [5 pts]

**SAS code:** (Direction: Use either [SAS Monospace] font or [Courier New] font for SAS code)

data btt;

infile "\\apporto.com\dfs\LOYOLA\Users\getorkornoo\_loyola\Documents\SAS\_Data/btt.dat";

input childid $ sex bweight gestage momage parity mdpd msbp momeduc mmedaid socio dbp sbp5 ht5 wt5 hdl5 ldl5 trig5 smoke5 medaid5 socio5;

proc sgplot data = btt;

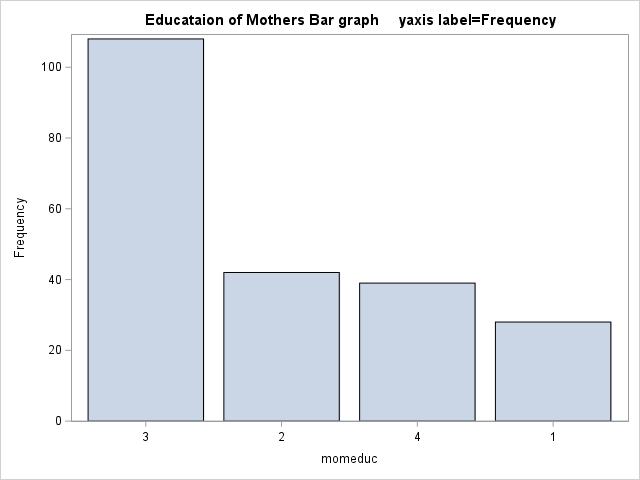
vbar momeduc / categoryorder = respdesc;;

title "Educataion of Mothers Bar graph"

yaxis label="Frequency";

run;

**SAS output:** (Direction: Paste SA)



1. Construct histograms for the discrete variables mother’s socioeconomic status at the child’s birth and at five years. How has the distribution of socioeconomic status changed? [10 pts]

**SAS code:** (Direction: Use either [SAS Monospace] font or [Courier New] font for SAS code)

data btt;

infile "\\apporto.com\dfs\LOYOLA\Users\getorkornoo\_loyola\Documents\SAS\_Data/btt.dat";

input childid $ sex bweight gestage momage parity mdpd msbp momeduc mmedaid socio dbp5 sbp5 ht5 wt5 hdl5 ldl5 trig5 smoke5 medaid5 socio5;

proc sgplot data = btt;

histogram socio5 ;

title "Mother’s Socioeconomic status (scale

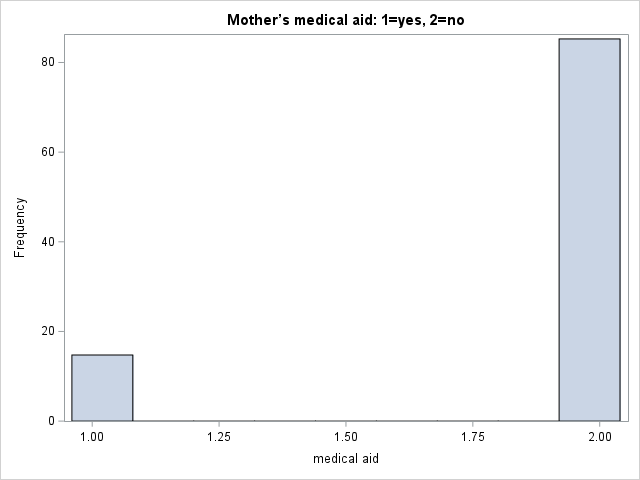
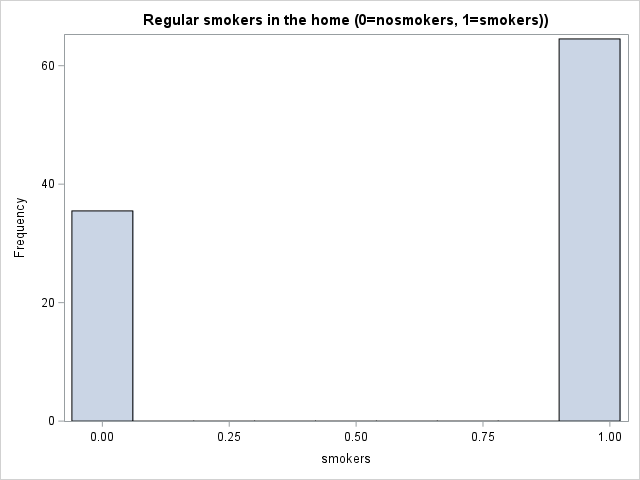
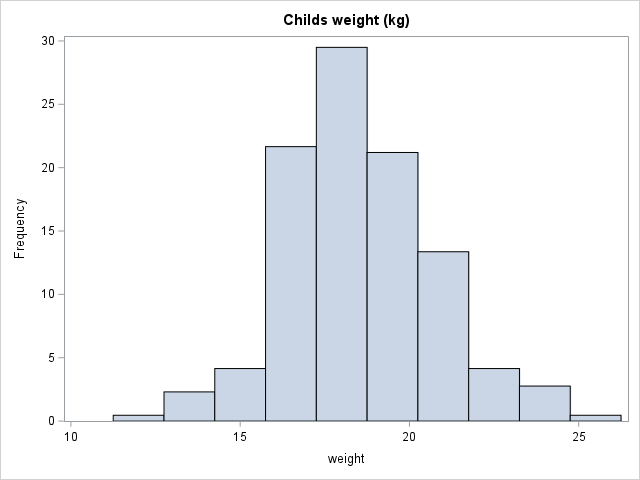
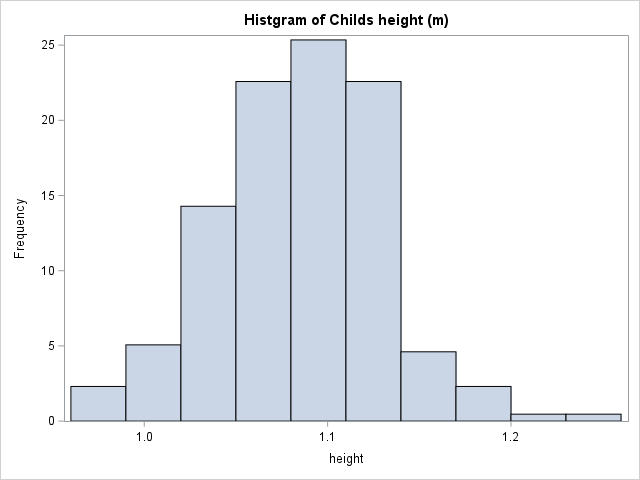
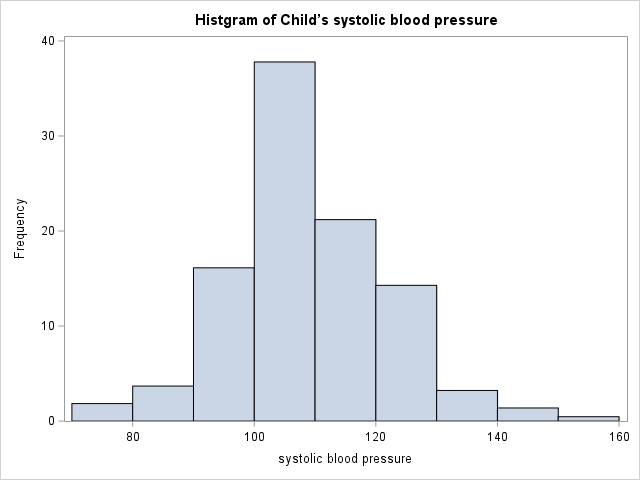
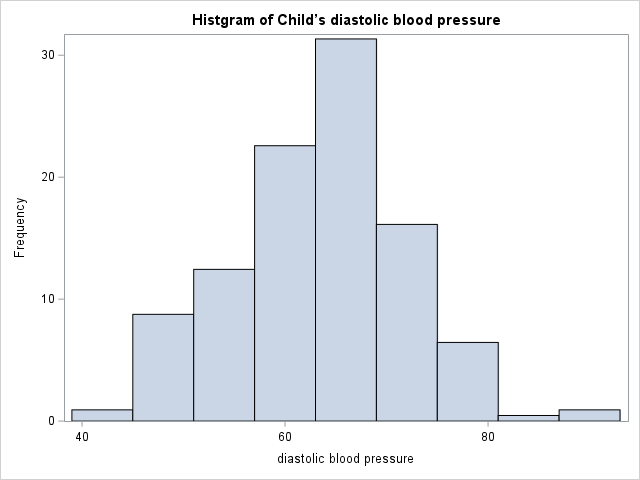
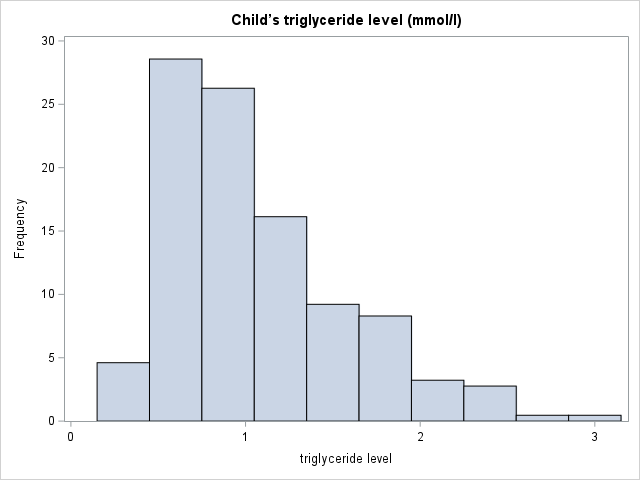
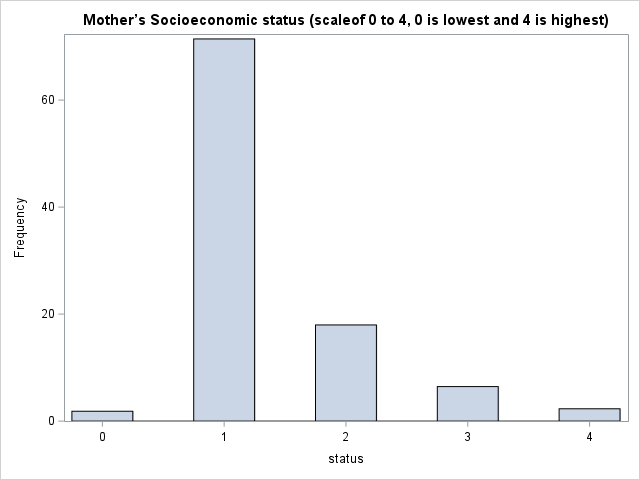
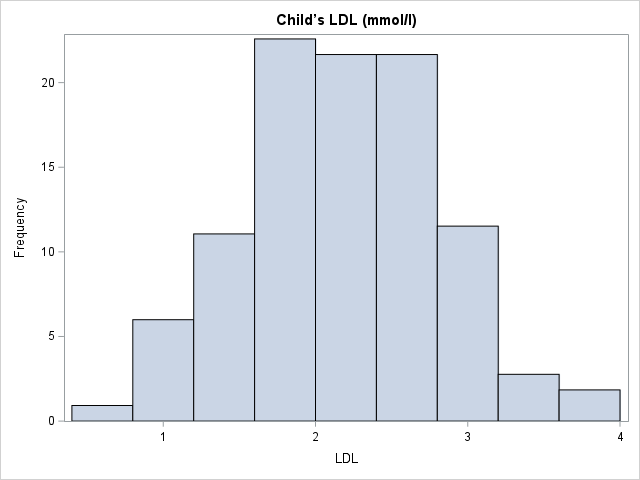
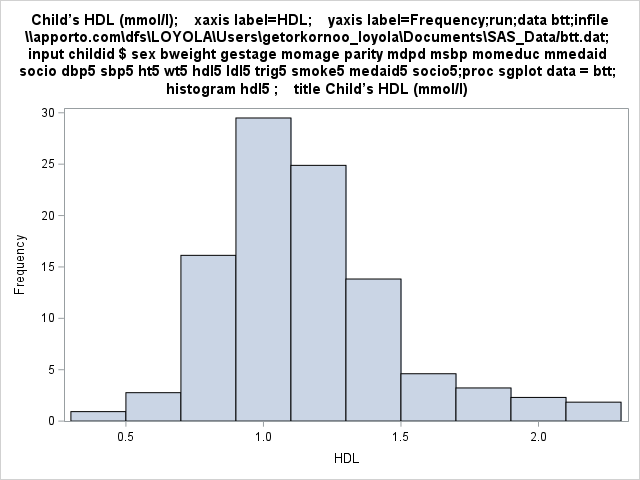
of 0 to 4, 0 is lowest and 4 is highest)";

xaxis label="status";

yaxis label="Frequency";

( swapped the variables out for each column)

**SAS output:** (Direction: Paste SA)



For socialeconomic many people are with the range of one meaning they are not entirely poor but they sometimes don’t have enough to provide for themselves and their family. While 2nd most filled is two, which the mother has enough to take care of the family and herself but less money for pleasures and wants. Th histogram shows us there's more people struggling then well off.

-------------------------------------------------------------------------------------------------------------------------------

**Problem#8.6.** [Page 90]

Generate 1000 random samples of size 10 from an exponential distribution with *λ* = 7. Create a new variable to find the average of these ten values. Use PROC SGPLOT to look at a histogram (with a normal curve overlaid) of the distribution of the mean. Describe the shape of the distribution. [20 pts]

**SAS code:** (Direction: Use either [SAS Monospace] font or [Courier New] font for SAS code)

data eight;

do i = 1 to 1000;

e = ranexp(11765);

x = e/7;

output;

end;

run;

proc means data=eight;

var x;

run;

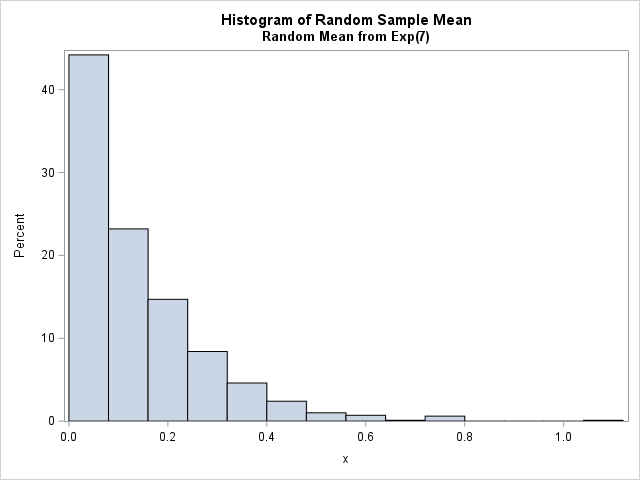
proc sgplot data = eight;

histogram x;

title2 'Random Mean from Exp(7)';

run;

**SAS output:** (Direction: Paste SA)



-------------------------------------------------------------------------------------------------------------------------------

**Problem#8.7.** [Page 90]

Generate 1000 random samples of size 10 from a uniform distribution on (10*,* 20). Create a new variable to find the average of these ten values. Use PROC SGPLOT to look at a histogram (with a normal curve overlaid) of the distribution of the mean. Describe the shape of the distribution. [20 pts]

**SAS code:** (Direction: Use either [SAS Monospace] font or [Courier New] font for SAS code)

data three;

do i = 1 to 1000;

x01 = 10 + (20-10) \* ranuni(55223);

x02 = 10 + (20-10) \* ranuni(55223);

x03 = 10 + (20-10) \* ranuni(55223);

x04 = 10 + (20-10) \* ranuni(55223);

x05 = 10 + (20-10) \* ranuni(55223);

x06 = 10 + (20-10) \* ranuni(55223);

x07 = 10 + (20-10) \* ranuni(55223);

x08 = 10 + (20-10) \* ranuni(55223);

x09 = 10 + (20-10) \* ranuni(55223);

x10 = 10 + (20-10) \* ranuni(55223);

ave\_x = (x01 + x02 + x03 + x04 + x05 +

x06 + x07 + x08 + x09 + x10) / 10;

output;

end;

run;

proc sgplot data = three;

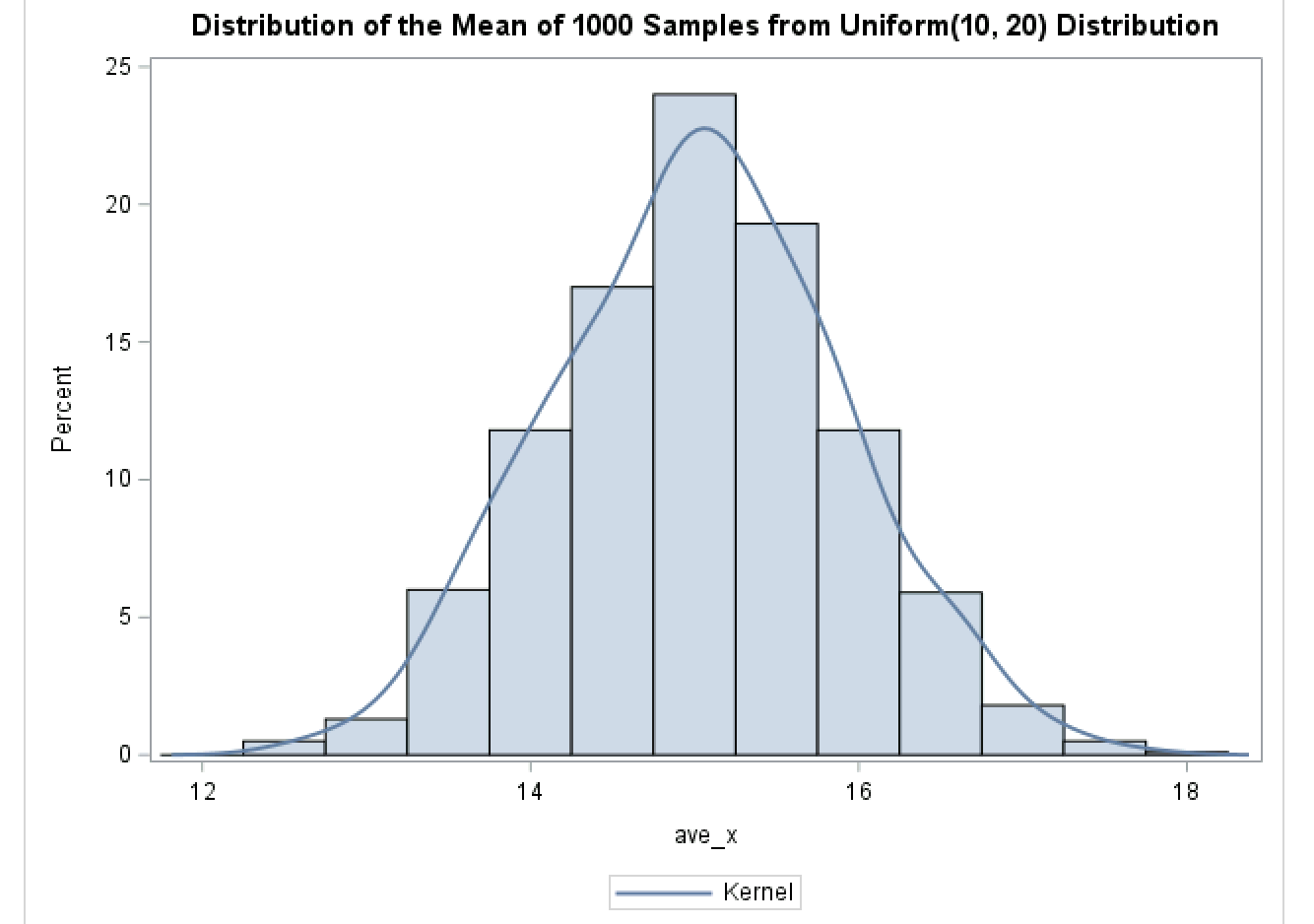
histogram ave\_x / binwidth=0.5;

density ave\_x / type=kernel;

title1 'Distribution of the Mean of 1000 Samples from Uniform(10, 20) Distribution ';

run;

**SAS output:** (Direction: Paste SA)



This distribution is bell-shaped with the averages perfectly following the distribution line. This shows that the data has a normal distribution.

-------------------------------------------------------------------------------------------------------------------------------

**Problem#9.4.** [Page 98]

Use handinj.dat (see Page270 for more details) to create an X-Y plot of days lost and cost. Describe the relationship between the two variables. [15 pts]

***Note***. Use PROC SGPLOT.

**SAS code:** (Direction: Use either [SAS Monospace] font or [Courier New] font for SAS code)

data hand;

infile "\\apporto.com\dfs\LOYOLA\Users\aneddy\_loyola\Desktop\SasData/Handinj.dat";

input id $ activity $ days\_lost cost;

run;

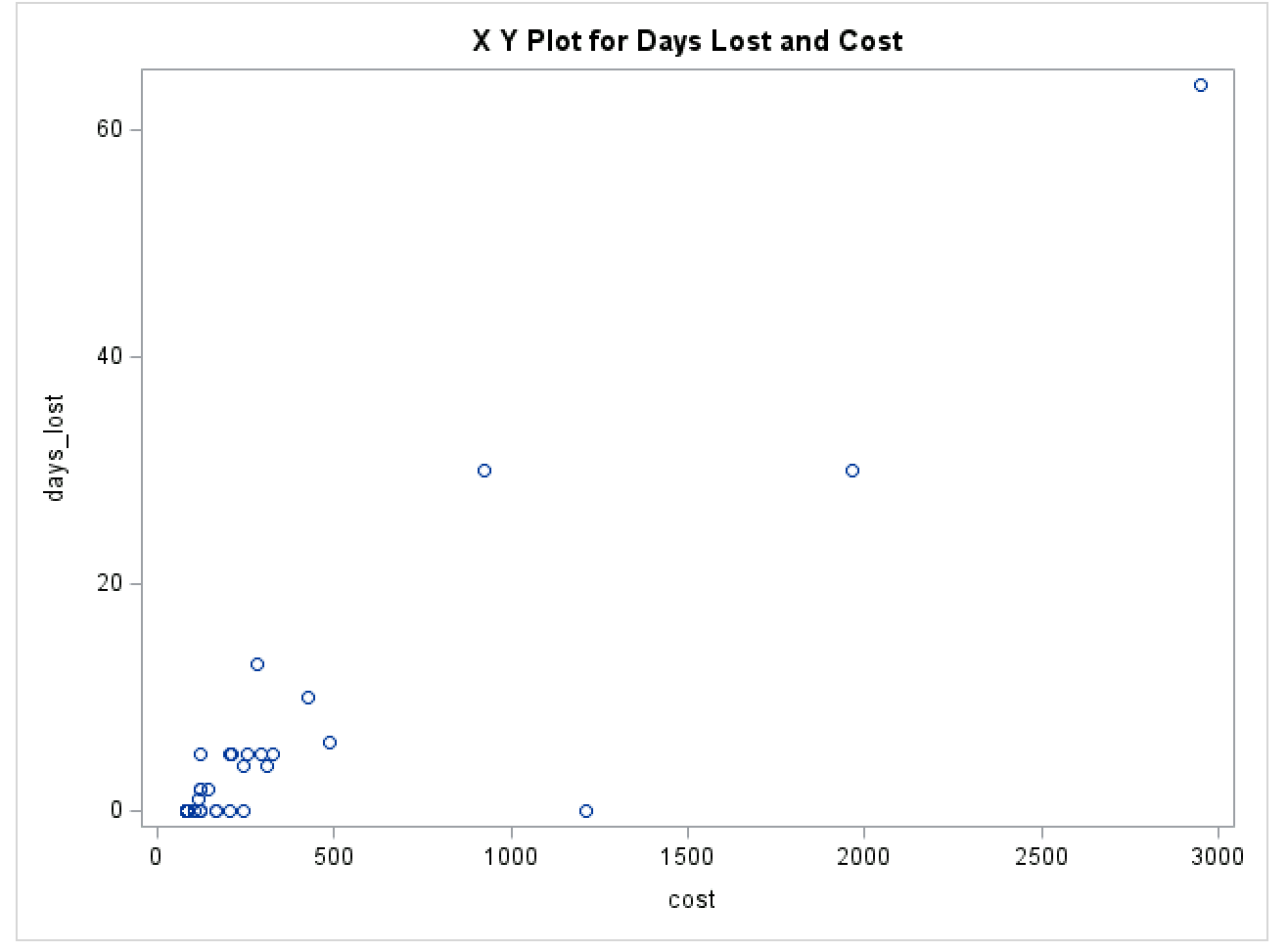
proc sgplot data = hand;

scatter y = days\_lost x = cost;

title "X Y Plot for Days Lost and Cost";

run;

**SAS output:** (Direction: Paste SA)



There seems to be no linear relationship or correlation between the variable days\_lost and cost. There is no trend in the graph which indicates no connection between the variables.

-------------------------------------------------------------------------------------------------------------------------------

**Problem#9.6.** [Page 99]

Use the file athlete.dat (see Page 260 for more details) to plot systolic blood pressure versus diastolic blood pressure. Describe the relationship between the two variables. [15 pts]

***Note***. Use PROC SGPLOT.

**SAS code:** (Direction: Use either [SAS Monospace] font or [Courier New] font for SAS code)

data athlete;

infile "\\apporto.com\dfs\LOYOLA\Users\aneddy\_loyola\Desktop\SASDATA/athlete.dat";

input Systolic Diastolic Gender $ AthorSed;

label Diastolic = 'Diastolic Blood Pressure'

Systolic = 'Systolic Blood Pressure'

run;

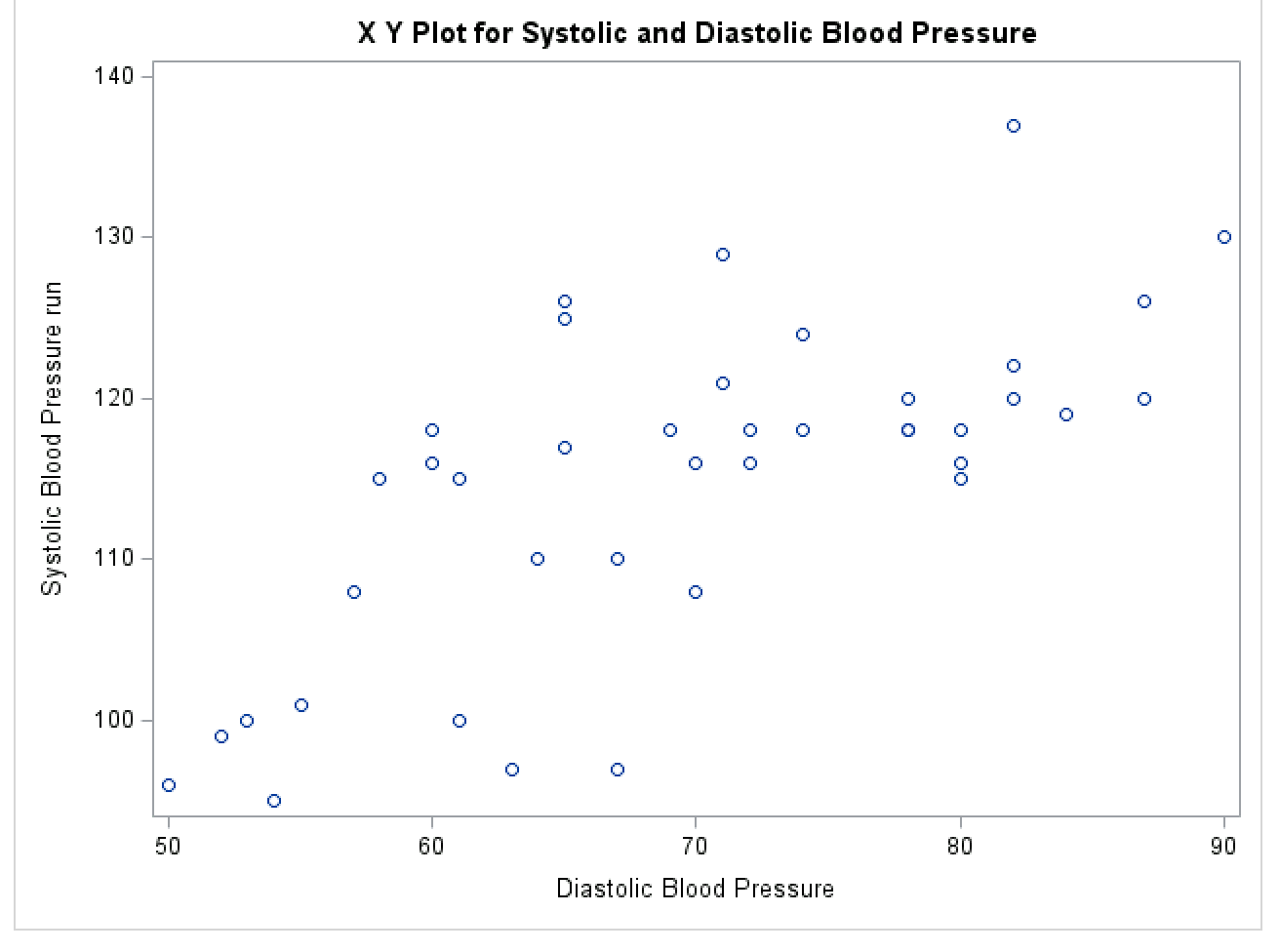
proc sgplot data = athlete;

scatter y = Systolic x = Diastolic;

title "X Y Plot for Systolic and Diastolic Blood Pressure";

run;

**SAS output:** (Direction: Paste SA)



In this graph, there seems to be somewhat of a positive correlation between Systolic and Diastolic Blood Pressure with some outliers. I believe so because there is a small trend of the data increasing going to the right. There are a few outliers at the beginning.

-------------------------------------------------------------------------------------------------------------------------------