Program Summary - Homework_2_Submission.sas

Execution Environment

Author:	u64309835
File:	/home/u64309835/Homework_2/Homework_2_Submission.sas
SAS Platform:	Linux LIN X64 5.14.0-284.30.1.el9_2.x86_64
SAS Host:	ODAWS02-USW2-2.ODA.SAS.COM
SAS Version:	9.04.01M8P02222023
SAS Locale:	en_US
Submission Time:	9/7/2025, 7:03:33 PM
Browser Host:	76-227-92-57.LIGHTSPEED.FTWOTX.SBCGLOBAL.NET
User Agent:	Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/138.0.0.0 Safari/537.36
Application Server:	ODAMID00-USW2-2.ODA.SAS.COM

Code: Homework_2_Submission.sas

```
/*=============================================================================================================*/
/*-------------------------------------------*
| Loading data sheets from LAB_2 demonstration |
*-------------------------------------------*/
/*=============================================================================================================*/
FILENAME REFFILE '/home/u64309835/LAB_2/Datasets.xlsx';
/* Sheet: neg_linear */
proc import datafile=reffile
    dbms=xlsx
    out=neg_linear
    replace;
    sheet='neg_linear';
    getnames=yes;
run;

/* Sheet: no_correlation */
proc import datafile=reffile
    dbms=xlsx
    out=no_correlation
    replace;
    sheet='no_correlation';
    getnames=yes;
run;

/* Sheet: nonlinear */
proc import datafile=reffile
    dbms=xlsx
    out=nonlinear
    replace;
    sheet='nonlinear';
    getnames=yes;
run;
/*=============================================================================================================*/
/*-------------------------------*
| (IPS10-2.6) Make some sketches |
*-------------------------------*/
/*=============================================================================================================*/

/*=================================*
|  (a) Scatter plot of neg linear   |
*=================================*/
title "A WEAK NEGATIVE LINEAR RELATIONSHIP (x,y)";
proc sgplot data=neg_linear;
    loess x=x y=y;
    scatter x=x y=y;
run;

/*=================================*
|  (b) Scatter plot of no correlation   |
*=================================*/
title "NO CORRELATION (x,y)";
proc sgplot data=no_correlation;
    loess x=x y=y;
    scatter x=x y=y;
run;

/*=================================*
|  (c) Scatter plot of nonlinear    |
```

```
/*==================================*/
title "A STRONG CURVILINEAR RELATIONSHIP (x,y)";
proc sgplot data=nonlinear;
    loess x=x y=y;
    scatter x=x y=y;
run;

/*==================================*
|  (d) Scatter plot of nonlinear    |
/*==================================*/
data sample;
    input index $ x $ y;
    datalines;
1 1 1
2 2 2
3 3 3
4 4 4
5 5 3
6 6 2
7 7 1
8 8 2
9 9 3
10 10 4
11 11 3
12 12 2
13 13 1
14 14 2
15 15 3
16 16 4
17 17 3
18 18 2
19 19 1
;
run;

title 'A MORE COMPLICATED RELATIONSHIP (x,y)';
proc sgplot data=sample;
    loess x=x y=y;
    scatter x=x y=y;
run;
/*=============================================================================================*/
/*-------------------------------------------*
|(IPS10-2.8) Blueberries and Anthocyanins     |
*-------------------------------------------*/
/*-Load the blueberry data --------------------*/
FILENAME REFFILE '/home/u64309835/Homework_2/ex02-008berries.csv';
proc import datafile=reffile
    dbms=CSV
    out=bluebar
    replace;
    getnames=yes;
run;

/*=======================================*
|  (a) Scatter plot of Antho4 & Antho3    |
/*=======================================*/
title "Scatter plot of Antho4 & Antho3 ";
proc sgplot data=bluebar;
    scatter x=Antho4 y=Antho3;
run;

/*=================================================*
|  (b) Describing the form, direction, and strength   |
/*=================================================*/

/*================================================*
|  (c) Outliers or unusual observations?          |
/*================================================*/

/*================================================*
|  (d) Scatter plot of Antho4 & Antho3 (w/LSRL)   |
/*================================================*/
title "Scatter plot of Antho4 & Antho3 (w/LSRL) ";
proc sgplot data=bluebar;
    loess x=Antho4 y=Antho3;
    scatter x=Antho4 y=Antho3;
run;

/*=================================================*
|  (e) Scatter plot of Antho4 & Antho3 (w/smoothing)   |
/*=================================================*/
```

```
proc sgplot data=bluebar;
    pbspline x=Antho4 y=Antho3;
    scatter x=Antho4 y=Antho3;
run;

/*=============================================================================================================*/
/*-----------------------------------------------------*
|(IPS10-2.9) Blueberries and Anthocyanins (w/logs)     |
*-----------------------------------------------------*/
/*-Load the blueberry data --------------------*/
FILENAME REFFILE '/home/u64309835/Homework_2/ex02-008berries.csv';
proc import datafile=reffile
    dbms=CSV
    out=bluebar
    replace;
    getnames=yes;
run;

/*-Making a log transformation on the variables----*/
/*=======================================*
|  (a) Scatter plot of Antho4 & Antho3    |
/*=======================================*/
title "Scatter plot of Antho4 & Antho3 (log transformation) ";
proc sgplot data=bluebar;
    scatter x=Antho4 y=Antho3;
    xaxis type=log;
    yaxis type=log;
run;

/*=====================================================*
|  (b) Describing the form, direction, and strength   |
/*=====================================================*/

/*=================================================*
|  (c) Outliers or unusual observations?           |
/*=================================================*/

/*=================================================*
|  (d) Scatter plot of Antho4 & Antho3 (w/LSRL)    |
/*=================================================*/
title "Scatter plot of Antho4 & Antho3 (w/LSRL) ";
proc sgplot data=bluebar;
    loess x=Antho4 y=Antho3;
    scatter x=Antho4 y=Antho3;
    xaxis type=log;
    yaxis type=log;
run;

/*=====================================================*
|  (e) Scatter plot of Antho4 & Antho3 (w/smoothing)   |
/*=====================================================*/
proc sgplot data=bluebar;
    pbspline x=Antho4 y=Antho3;
    scatter x=Antho4 y=Antho3;
    xaxis type=log;
    yaxis type=log;
run;

/*=============================================================================================================*/
/*-------------------------------------------------------*
|(IPS10-2.16) Compare the baseball players with controls |
*-------------------------------------------------------*/
/*--Load the baseball data--------*/
FILENAME REFFILE '/home/u64309835/Homework_2/ex02-016armstr.csv';
proc import datafile=reffile
    dbms=CSV
    out=baseball_data
    replace;
    getnames=yes;
run;


/*=================================================================================*
|  (a) Scatter plot of Baseball Players and Non-Baseball Players (Dominant Hand)   |
/*=================================================================================*/
ods graphics / width=4in height=3in;
title "Scatter plot of Baseball Players and Non-Baseball Players (Dominant Hand)";
proc sgplot data=baseball_data;
    scatter x=Group y=Dom / group=Group;
run;
ods graphics / reset=all;
```

```
/*============================*
|  (b) Summary of observations |
/*============================*/

/*=============================================================================================================*/
/*-----------------------------------------------------*
|(IPS10-2.30) Blueberries and Anthocyanins (correlation)  |
*-----------------------------------------------------*/
/*--Load the blueberry data-----------*/
FILENAME REFFILE '/home/u64309835/Homework_2/ex02-008berries.csv';
proc import datafile=reffile
    dbms=CSV
    out=bluebar
    replace;
    getnames=yes;
run;

/*=========================================*
|  (a) Correlations between Antho4 & Antho3 |
/*=========================================*/
title "Correlation of Antho4 and Antho3";
proc corr data=bluebar;
    var Antho4 Antho3;
run;
/*=========================================*
|  (b) Scatter plot between Antho4 & Antho3 |
/*=========================================*/
title "Scatter plot of Antho4 & Antho3 ";
proc sgplot data=bluebar;
    scatter x=Antho4 y=Antho3;
run;
quit;
/*============================*
|  (c) Summary of observations |
/*============================*/

/*=============================================================================================================*/
/*-------------------------------------------------*
|(IPS10-2.34) Strong association but no correlation |
*-------------------------------------------------*/
data corr_data;
    input x y;
    datalines;
45 30
55 50
65 70
75 50
85 30
;
run;

/*==================*
|  (a) Scatter plot |
/*==================*/
title "Scatter plot of x & y";
proc sgplot data=corr_data;
    scatter x=x y=y;
run;
/*==================================*
|  (b) Description of relationship |
/*==================================*/
proc sgplot data=corr_data;
    scatter x=x y=y;
    loess x=x y=y;
run;
/*==================*
|  (c) Correlation  |
/*==================*/
title "Scatter plot of x & y";
proc corr data=corr_data;
    var x y;
run;
/*=====================================================*
|  (d) The importance of correlation for this exercise |
/*=====================================================*/
```

Log: Homework_2_Submission.sas

Warnings (2)

Notes (72)

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable some output features.
69
70         /*=================================================================================================*/
71         /*-----------------------------------------*
72         | Loading data sheets from LAB_2 demonstration |
73         *-----------------------------------------*/
74         /*=================================================================================================*/
75         FILENAME REFFILE '/home/u64309835/LAB_2/Datasets.xlsx';
76         /* Sheet: neg_linear */
77         proc import datafile=reffile
78         dbms=xlsx
79         out=neg_linear
80         replace;
81         sheet='neg_linear';
82         getnames=yes;
83         run;

NOTE: The import data set has 20 observations and 2 variables.
NOTE: WORK.NEG_LINEAR data set was successfully created.
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              3340.28k
      OS Memory           29948.00k
      Timestamp           09/08/2025 12:03:31 AM
      Step Count                        84  Switch Count  4
      Page Faults                       0
      Page Reclaims                     703
      Page Swaps                        0
      Voluntary Context Switches        28
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           288


84
85         /* Sheet: no_correlation */
86         proc import datafile=reffile
87         dbms=xlsx
88         out=no_correlation
89         replace;
90         sheet='no_correlation';
91         getnames=yes;
92         run;

NOTE: The import data set has 25 observations and 2 variables.
NOTE: WORK.NO_CORRELATION data set was successfully created.
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              3328.46k
      OS Memory           29948.00k
      Timestamp           09/08/2025 12:03:31 AM
      Step Count                        85  Switch Count  4
      Page Faults                       0
      Page Reclaims                     703
      Page Swaps                        0
      Voluntary Context Switches        28
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           288


93
94         /* Sheet: nonlinear */
95         proc import datafile=reffile
96         dbms=xlsx
97         out=nonlinear
98         replace;
99         sheet='nonlinear';
100        getnames=yes;
101        run;

NOTE: The import data set has 20 observations and 2 variables.
NOTE: WORK.NONLINEAR data set was successfully created.
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              3370.09k
      OS Memory           29948.00k
      Timestamp           09/08/2025 12:03:31 AM
      Step Count                        86  Switch Count  4
```

```
      Page Faults                     0
      Page Reclaims                   703
      Page Swaps                      0
      Voluntary Context Switches      27
      Involuntary Context Switches    2
      Block Input Operations          0
      Block Output Operations         288


102       /*=================================================================================================*/
103       /*------------------------------*
104       | (IPS10-2.6) Make some sketches |
105       *------------------------------*/
106       /*=================================================================================================*/
107
108       /*==================================*
109       |  (a) Scatter plot of neg linear   |
110       *==================================*/
111       title "A WEAK NEGATIVE LINEAR RELATIONSHIP (x,y)";
112       proc sgplot data=neg_linear;
113       loess x=x y=y;
114       scatter x=x y=y;
115       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              8811.31k
      OS Memory           33072.00k
      Timestamp           09/08/2025 12:03:31 AM
      Step Count                      87  Switch Count  1
      Page Faults                     0
      Page Reclaims                   1515
      Page Swaps                      0
      Voluntary Context Switches      171
      Involuntary Context Switches    2
      Block Input Operations          0
      Block Output Operations         776

NOTE: There were 20 observations read from the data set WORK.NEG_LINEAR.

116
117       /*==================================*
118       |  (b) Scatter plot of no correlation   |
119       *==================================*/
120       title "NO CORRELATION (x,y)";
121       proc sgplot data=no_correlation;
122       loess x=x y=y;
123       scatter x=x y=y;
124       run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              2193.37k
      OS Memory           33072.00k
      Timestamp           09/08/2025 12:03:31 AM
      Step Count                      88  Switch Count  1
      Page Faults                     0
      Page Reclaims                   364
      Page Swaps                      0
      Voluntary Context Switches      167
      Involuntary Context Switches    2
      Block Input Operations          0
      Block Output Operations         456

NOTE: There were 25 observations read from the data set WORK.NO_CORRELATION.

125
126       /*==================================*
127       |  (c) Scatter plot of nonlinear    |
128       *==================================*/
129       title "A STRONG CURVILINEAR RELATIONSHIP (x,y)";
130       proc sgplot data=nonlinear;
131       loess x=x y=y;
132       scatter x=x y=y;
133       run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              2322.03k
      OS Memory           33072.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                      89  Switch Count  1
      Page Faults                     0
      Page Reclaims                   293
      Page Swaps                      0
```

```
      Voluntary Context Switches        164
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           480

NOTE: There were 20 observations read from the data set WORK.NONLINEAR.

134
135        /*==================================*
136        |  (d) Scatter plot of nonlinear    |
137        /*==================================*/
138        data sample;
139        input index $ x $ y;
140        datalines;

NOTE: The data set WORK.SAMPLE has 19 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              669.53k
      OS Memory           32168.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        90  Switch Count  2
      Page Faults                       0
      Page Reclaims                     87
      Page Swaps                        0
      Voluntary Context Switches        13
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


160        ;
161        run;
162
163        title 'A MORE COMPLICATED RELATIONSHIP (x,y)';
164        proc sgplot data=sample;
165        loess x=x y=y;
166        scatter x=x y=y;
167        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              2242.31k
      OS Memory           33072.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        91  Switch Count  1
      Page Faults                       0
      Page Reclaims                     293
      Page Swaps                        0
      Voluntary Context Switches        187
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           384

WARNING: Invalid data passed to LOESS. X must be numeric.
WARNING: The LOESSPLOT statement named 'LOESS' will not be drawn because one or more of the required arguments were not supplied.
NOTE: There were 19 observations read from the data set WORK.SAMPLE.

168        /*=============================================================================================================*/
169        /*-----------------------------------------*
170        |(IPS10-2.8) Blueberries and Anthocyanins     |
171        *-----------------------------------------*/
172        /*-Load the blueberry data --------------------*/
173        FILENAME REFFILE '/home/u64309835/Homework_2/ex02-008berries.csv';
174        proc import datafile=reffile
175        dbms=CSV
176        out=bluebar
177        replace;
178        getnames=yes;
179        run;

NOTE: Unable to open parameter catalog: SASUSER.PARMS.PARMS.SLIST in update mode. Temporary parameter values will be saved to
WORK.PARMS.PARMS.SLIST.
180         /**********************************************************************
181         *   PRODUCT:   SAS
182         *   VERSION:   9.4
183         *   CREATOR:   External File Interface
184         *   DATE:      07SEP25
185         *   DESC:      Generated SAS Datastep Code
186         *   TEMPLATE SOURCE:  (None Specified.)
187         **********************************************************************/
188            data WORK.BLUEBAR    ;
189            %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
190            infile REFFILE delimiter = ',' MISSOVER DSD  firstobs=2 ;
191               informat ID best32. ;
192               informat Antho1 best32. ;
193               informat Antho2 best32. ;
```

```
194              informat Antho3 best32. ;
195              informat Antho4 best32. ;
196              informat LAntho1 best32. ;
197              informat LAntho2 best32. ;
198              informat LAntho3 best32. ;
199              informat LAntho4 best32. ;
200              format ID best12. ;
201              format Antho1 best12. ;
202              format Antho2 best12. ;
203              format Antho3 best12. ;
204              format Antho4 best12. ;
205              format LAntho1 best12. ;
206              format LAntho2 best12. ;
207              format LAntho3 best12. ;
208              format LAntho4 best12. ;
209           input
210                       ID
211                       Antho1
212                       Antho2
213                       Antho3
214                       Antho4
215                       LAntho1
216                       LAntho2
217                       LAntho3
218                       LAntho4
219           ;
220           if _ERROR_ then call symputx('_EFIERR_',1);  /* set ERROR detection macro variable */
221           run;

NOTE: The infile REFFILE is:
      Filename=/home/u64309835/Homework_2/ex02-008berries.csv,
      Owner Name=u64309835,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=03Sep2025:12:53:28,
      File Size (bytes)=16487

NOTE: 267 records were read from the infile REFFILE.
      The minimum record length was 54.
      The maximum record length was 63.
NOTE: The data set WORK.BLUEBAR has 267 observations and 9 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              9300.56k
      OS Memory           37660.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        92  Switch Count  2
      Page Faults                       0
      Page Reclaims                     136
      Page Swaps                        0
      Voluntary Context Switches        16
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


267 rows created in WORK.BLUEBAR from REFFILE.



NOTE: WORK.BLUEBAR data set was successfully created.
NOTE: The data set WORK.BLUEBAR has 267 observations and 9 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              9300.56k
      OS Memory           37920.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        92  Switch Count  9
      Page Faults                       0
      Page Reclaims                     1157
      Page Swaps                        0
      Voluntary Context Switches        92
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           320


222
223        /*=======================================*
224        |  (a) Scatter plot of Antho4 & Antho3    |
225        /*=======================================*/
226        title "Scatter plot of Antho4 & Antho3 ";
227        proc sgplot data=bluebar;
228        scatter x=Antho4 y=Antho3;
229        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.06 seconds
```

```
      user cpu time       0.04 seconds
      system cpu time     0.01 seconds
      memory              2181.53k
      OS Memory           33584.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        93  Switch Count  1
      Page Faults                       0
      Page Reclaims                     291
      Page Swaps                        0
      Voluntary Context Switches        131
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           400

NOTE: There were 267 observations read from the data set WORK.BLUEBAR.

230
231        /*=================================================*
232        |  (b) Describing the form, direction, and strength   |
233        /*=================================================*/
234
235        /*================================================*
236        |  (c) Outliers or unusual observations?          |
237        /*================================================*/
238
239        /*================================================*
240        |  (d) Scatter plot of Antho4 & Antho3 (w/LSRL)   |
241        /*================================================*/
242        title "Scatter plot of Antho4 & Antho3 (w/LSRL) ";
243        proc sgplot data=bluebar;
244        loess x=Antho4 y=Antho3;
245        scatter x=Antho4 y=Antho3;
246        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.09 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              2488.75k
      OS Memory           33584.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        94  Switch Count  1
      Page Faults                       0
      Page Reclaims                     330
      Page Swaps                        0
      Voluntary Context Switches        165
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           552

NOTE: There were 267 observations read from the data set WORK.BLUEBAR.

247
248        /*=================================================*
249        |  (e) Scatter plot of Antho4 & Antho3 (w/smoothing)   |
250        /*=================================================*/
251        proc sgplot data=bluebar;
252        pbspline x=Antho4 y=Antho3;
253        scatter x=Antho4 y=Antho3;
254        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              2484.18k
      OS Memory           33704.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        95  Switch Count  1
      Page Faults                       0
      Page Reclaims                     431
      Page Swaps                        0
      Voluntary Context Switches        160
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           584

NOTE: There were 267 observations read from the data set WORK.BLUEBAR.

255
256        /*=================================================================================================*/
257        /*-------------------------------------------------*
258        |(IPS10-2.9) Blueberries and Anthocyanins (w/logs)     |
259        *-------------------------------------------------*/
260        /*-Load the blueberry data --------------------*/
261        FILENAME REFFILE '/home/u64309835/Homework_2/ex02-008berries.csv';
262        proc import datafile=reffile
263        dbms=CSV
264        out=bluebar
265        replace;
266        getnames=yes;
```

```
267        run;

NOTE: Unable to open parameter catalog: SASUSER.PARMS.PARMS.SLIST in update mode. Temporary parameter values will be saved to
WORK.PARMS.PARMS.SLIST.
268         /*************************************************************************
269         *   PRODUCT:   SAS
270         *   VERSION:   9.4
271         *   CREATOR:   External File Interface
272         *   DATE:      07SEP25
273         *   DESC:      Generated SAS Datastep Code
274         *   TEMPLATE SOURCE:  (None Specified.)
275         *************************************************************************/
276            data WORK.BLUEBAR    ;
277            %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
278            infile REFFILE delimiter = ',' MISSOVER DSD  firstobs=2 ;
279               informat ID best32. ;
280               informat Antho1 best32. ;
281               informat Antho2 best32. ;
282               informat Antho3 best32. ;
283               informat Antho4 best32. ;
284               informat LAntho1 best32. ;
285               informat LAntho2 best32. ;
286               informat LAntho3 best32. ;
287               informat LAntho4 best32. ;
288               format ID best12. ;
289               format Antho1 best12. ;
290               format Antho2 best12. ;
291               format Antho3 best12. ;
292               format Antho4 best12. ;
293               format LAntho1 best12. ;
294               format LAntho2 best12. ;
295               format LAntho3 best12. ;
296               format LAntho4 best12. ;
297            input
298                        ID
299                        Antho1
300                        Antho2
301                        Antho3
302                        Antho4
303                        LAntho1
304                        LAntho2
305                        LAntho3
306                        LAntho4
307            ;
308            if _ERROR_ then call symputx('_EFIERR_',1);  /* set ERROR detection macro variable */
309            run;

NOTE: The infile REFFILE is:
      Filename=/home/u64309835/Homework_2/ex02-008berries.csv,
      Owner Name=u64309835,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=03Sep2025:12:53:28,
      File Size (bytes)=16487

NOTE: 267 records were read from the infile REFFILE.
      The minimum record length was 54.
      The maximum record length was 63.
NOTE: The data set WORK.BLUEBAR has 267 observations and 9 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              9303.25k
      OS Memory           38940.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        96  Switch Count  2
      Page Faults                       0
      Page Reclaims                     150
      Page Swaps                        0
      Voluntary Context Switches        15
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


267 rows created in WORK.BLUEBAR from REFFILE.



NOTE: WORK.BLUEBAR data set was successfully created.
NOTE: The data set WORK.BLUEBAR has 267 observations and 9 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              9303.25k
      OS Memory           39200.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        96  Switch Count  10
      Page Faults                       0
      Page Reclaims                     1083
```

```
      Page Swaps                       0
      Voluntary Context Switches       92
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          280


310
311       /*-Making a log transformation on the variables----*/
312       /*=======================================*
313       |  (a) Scatter plot of Antho4 & Antho3    |
314       /*=======================================*/
315       title "Scatter plot of Antho4 & Antho3 (log transformation) ";
316       proc sgplot data=bluebar;
317       scatter x=Antho4 y=Antho3;
318       xaxis type=log;
319       yaxis type=log;
320       run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.05 seconds
      user cpu time       0.02 seconds
      system cpu time     0.01 seconds
      memory              1926.21k
      OS Memory           35120.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                       97  Switch Count  1
      Page Faults                      0
      Page Reclaims                    306
      Page Swaps                       0
      Voluntary Context Switches       130
      Involuntary Context Switches     1
      Block Input Operations           0
      Block Output Operations          408

NOTE: There were 267 observations read from the data set WORK.BLUEBAR.

321
322       /*=================================================*
323       |  (b) Describing the form, direction, and strength   |
324       /*=================================================*/
325
326       /*================================================*
327       |  (c) Outliers or unusual observations?          |
328       /*================================================*/
329
330       /*================================================*
331       |  (d) Scatter plot of Antho4 & Antho3 (w/LSRL)   |
332       /*================================================*/
333       title "Scatter plot of Antho4 & Antho3 (w/LSRL) ";
334       proc sgplot data=bluebar;
335       loess x=Antho4 y=Antho3;
336       scatter x=Antho4 y=Antho3;
337       xaxis type=log;
338       yaxis type=log;
339       run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.09 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              2500.43k
      OS Memory           35120.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                       98  Switch Count  1
      Page Faults                      0
      Page Reclaims                    401
      Page Swaps                       0
      Voluntary Context Switches       165
      Involuntary Context Switches     4
      Block Input Operations           0
      Block Output Operations          632

NOTE: There were 267 observations read from the data set WORK.BLUEBAR.

340
341       /*=================================================*
342       |  (e) Scatter plot of Antho4 & Antho3 (w/smoothing)   |
343       /*=================================================*/
344       proc sgplot data=bluebar;
345       pbspline x=Antho4 y=Antho3;
346       scatter x=Antho4 y=Antho3;
347       xaxis type=log;
348       yaxis type=log;
349       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              2385.71k
```

```
      OS Memory           35376.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        99  Switch Count  1
      Page Faults                       0
      Page Reclaims                     411
      Page Swaps                        0
      Voluntary Context Switches        162
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           656

NOTE: There were 267 observations read from the data set WORK.BLUEBAR.

350
351        /*=============================================================================================================*/
352        /*-----------------------------------------------------*
353        |(IPS10-2.16) Compare the baseball players with controls |
354        *-----------------------------------------------------*/
355        /*--Load the baseball data--------*/
356        FILENAME REFFILE '/home/u64309835/Homework_2/ex02-016armstr.csv';
357        proc import datafile=reffile
358        dbms=CSV
359        out=baseball_data
360        replace;
361        getnames=yes;
362        run;

NOTE: Unable to open parameter catalog: SASUSER.PARMS.PARMS.SLIST in update mode. Temporary parameter values will be saved to
WORK.PARMS.PARMS.SLIST.
363         /*************************************************************************
364         *   PRODUCT:   SAS
365         *   VERSION:   9.4
366         *   CREATOR:   External File Interface
367         *   DATE:      07SEP25
368         *   DESC:      Generated SAS Datastep Code
369         *   TEMPLATE SOURCE:  (None Specified.)
370         *************************************************************************/
371            data WORK.BASEBALL_DATA    ;
372            %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
373            infile REFFILE delimiter = ',' MISSOVER DSD  firstobs=2 ;
374               informat ID best32. ;
375               informat Group $8. ;
376               informat NonDom best32. ;
377               informat Dom best32. ;
378               format ID best12. ;
379               format Group $8. ;
380               format NonDom best12. ;
381               format Dom best12. ;
382            input
383                        ID
384                        Group  $
385                        NonDom
386                        Dom
387            ;
388            if _ERROR_ then call symputx('_EFIERR_',1);  /* set ERROR detection macro variable */
389            run;

NOTE: The infile REFFILE is:
      Filename=/home/u64309835/Homework_2/ex02-016armstr.csv,
      Owner Name=u64309835,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=03Sep2025:12:53:28,
      File Size (bytes)=687

NOTE: 30 records were read from the infile REFFILE.
      The minimum record length was 19.
      The maximum record length was 21.
NOTE: The data set WORK.BASEBALL_DATA has 30 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              9293.68k
      OS Memory           38940.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        100  Switch Count  2
      Page Faults                       0
      Page Reclaims                     160
      Page Swaps                        0
      Voluntary Context Switches        15
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           264


30 rows created in WORK.BASEBALL_DATA from REFFILE.



NOTE: WORK.BASEBALL_DATA data set was successfully created.
NOTE: The data set WORK.BASEBALL_DATA has 30 observations and 4 variables.
```

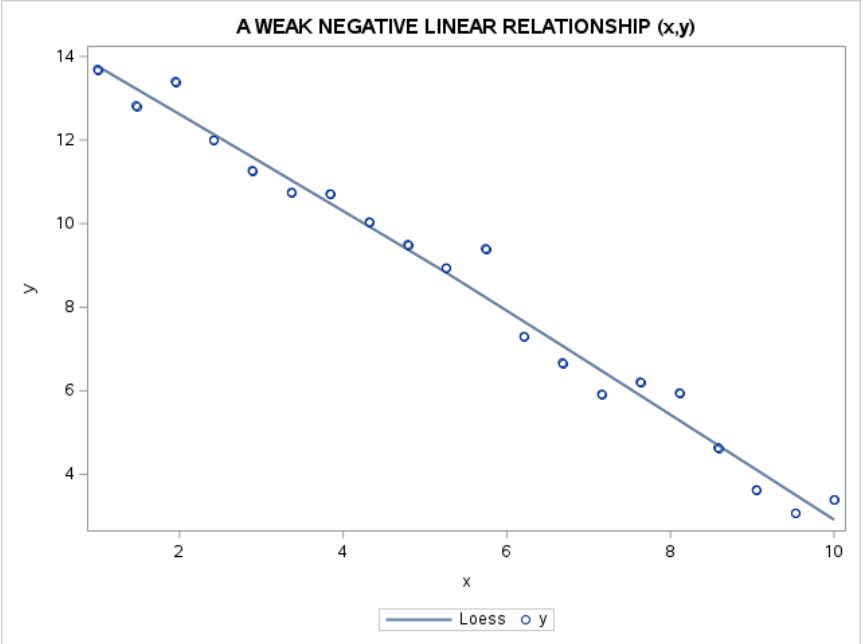
```
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              9293.68k
      OS Memory           39200.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        100  Switch Count  10
      Page Faults                       0
      Page Reclaims                     971
      Page Swaps                        0
      Voluntary Context Switches        92
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           280


390
391
392        /*=============================================================================*
393        |  (a) Scatter plot of Baseball Players and Non-Baseball Players (Dominant Hand)   |
394        /*=============================================================================*/
395        ods graphics / width=4in height=3in;
396        title "Scatter plot of Baseball Players and Non-Baseball Players (Dominant Hand)";
397        proc sgplot data=baseball_data;
398        scatter x=Group y=Dom / group=Group;
399        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              1212.81k
      OS Memory           34800.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        101  Switch Count  1
      Page Faults                       0
      Page Reclaims                     147
      Page Swaps                        0
      Voluntary Context Switches        221
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           360

NOTE: There were 30 observations read from the data set WORK.BASEBALL_DATA.

400        ods graphics / reset=all;
401
402        /*============================*
403        |  (b) Summary of observations |
404        /*============================*/
405
406        /*=============================================================================================================*/
407        /*-----------------------------------------------------*
408        |(IPS10-2.30) Blueberries and Anthocyanins (correlation)  |
409        *-----------------------------------------------------*/
410        /*--Load the blueberry data------------*/
411        FILENAME REFFILE '/home/u64309835/Homework_2/ex02-008berries.csv';
412        proc import datafile=reffile
413        dbms=CSV
414        out=bluebar
415        replace;
416        getnames=yes;
417        run;

NOTE: Unable to open parameter catalog: SASUSER.PARMS.PARMS.SLIST in update mode. Temporary parameter values will be saved to
WORK.PARMS.PARMS.SLIST.
418         /*************************************************************************
419         *   PRODUCT:   SAS
420         *   VERSION:   9.4
421         *   CREATOR:   External File Interface
422         *   DATE:      07SEP25
423         *   DESC:      Generated SAS Datastep Code
424         *   TEMPLATE SOURCE:  (None Specified.)
425         *************************************************************************/
426            data WORK.BLUEBAR    ;
427            %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
428            infile REFFILE delimiter = ',' MISSOVER DSD  firstobs=2 ;
429               informat ID best32. ;
430               informat Antho1 best32. ;
431               informat Antho2 best32. ;
432               informat Antho3 best32. ;
433               informat Antho4 best32. ;
434               informat LAntho1 best32. ;
435               informat LAntho2 best32. ;
436               informat LAntho3 best32. ;
437               informat LAntho4 best32. ;
438               format ID best12. ;
439               format Antho1 best12. ;
440               format Antho2 best12. ;
441               format Antho3 best12. ;
```

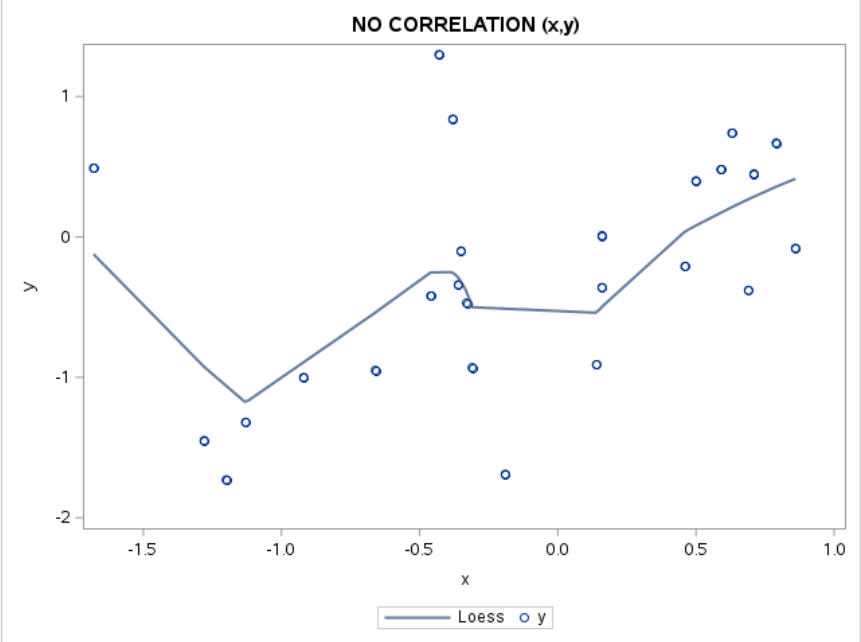
```
442              format Antho4 best12. ;
443              format LAntho1 best12. ;
444              format LAntho2 best12. ;
445              format LAntho3 best12. ;
446              format LAntho4 best12. ;
447           input
448                       ID
449                       Antho1
450                       Antho2
451                       Antho3
452                       Antho4
453                       LAntho1
454                       LAntho2
455                       LAntho3
456                       LAntho4
457           ;
458           if _ERROR_ then call symputx('_EFIERR_',1);  /* set ERROR detection macro variable */
459           run;

NOTE: The infile REFFILE is:
      Filename=/home/u64309835/Homework_2/ex02-008berries.csv,
      Owner Name=u64309835,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=03Sep2025:12:53:28,
      File Size (bytes)=16487

NOTE: 267 records were read from the infile REFFILE.
      The minimum record length was 54.
      The maximum record length was 63.
NOTE: The data set WORK.BLUEBAR has 267 observations and 9 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              9378.71k
      OS Memory           38684.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        102  Switch Count  2
      Page Faults                       0
      Page Reclaims                     107
      Page Swaps                        0
      Voluntary Context Switches        15
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


267 rows created in WORK.BLUEBAR from REFFILE.



NOTE: WORK.BLUEBAR data set was successfully created.
NOTE: The data set WORK.BLUEBAR has 267 observations and 9 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              9378.71k
      OS Memory           38944.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        102  Switch Count  10
      Page Faults                       0
      Page Reclaims                     855
      Page Swaps                        0
      Voluntary Context Switches        88
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           288


460
461        /*=======================================*
462        |  (a) Correlations between Antho4 & Antho3 |
463        /*=======================================*/
464        title "Correlation of Antho4 and Antho3";
465        proc corr data=bluebar;
466        var Antho4 Antho3;
467        run;

NOTE: PROCEDURE CORR used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              1885.18k
      OS Memory           34472.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        103  Switch Count  0
      Page Faults                       0
      Page Reclaims                     52
      Page Swaps                        0
      Voluntary Context Switches        3
```

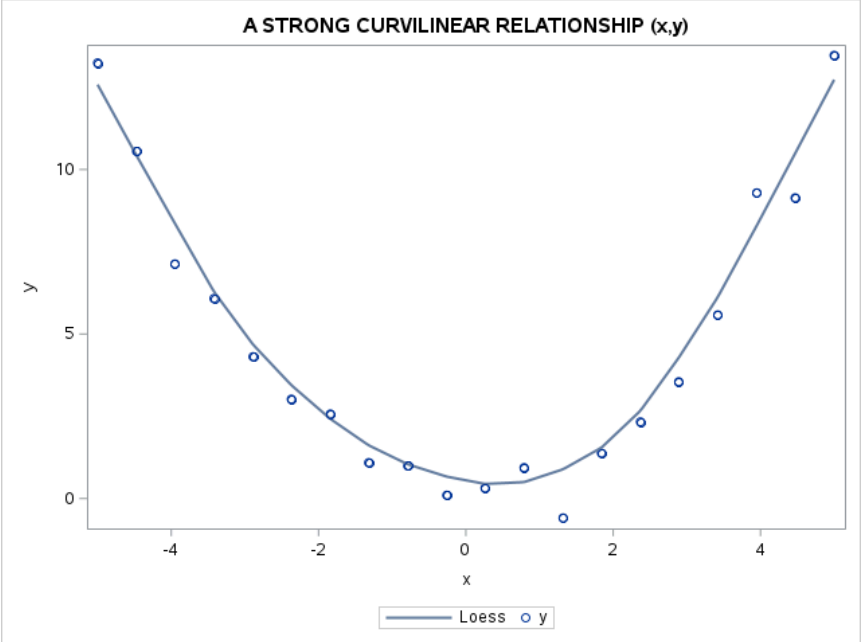
```
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           0


468        /*=========================================*
469        |  (b) Scatter plot between Antho4 & Antho3 |
470        /*=========================================*/
471        title "Scatter plot of Antho4 & Antho3 ";
472        proc sgplot data=bluebar;
473        scatter x=Antho4 y=Antho3;
474        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.05 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              1784.25k
      OS Memory           35376.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        104  Switch Count  1
      Page Faults                       0
      Page Reclaims                     291
      Page Swaps                        0
      Voluntary Context Switches        131
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           424

NOTE: There were 267 observations read from the data set WORK.BLUEBAR.

475        quit;
476        /*============================*
477        |  (c) Summary of observations |
478        /*============================*/
479
480        /*=============================================================================================================*/
481        /*-------------------------------------------------*
482        |(IPS10-2.34) Strong association but no correlation |
483        *-------------------------------------------------*/
484        data corr_data;
485        input x y;
486        datalines;

NOTE: The data set WORK.CORR_DATA has 5 observations and 2 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              678.31k
      OS Memory           34472.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        105  Switch Count  2
      Page Faults                       0
      Page Reclaims                     87
      Page Swaps                        0
      Voluntary Context Switches        17
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


492        ;
493        run;
494
495        /*==================*
496        |  (a) Scatter plot |
497        /*==================*/
498        title "Scatter plot of x & y";
499        proc sgplot data=corr_data;
500        scatter x=x y=y;
501        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.04 seconds
      system cpu time     0.01 seconds
      memory              2022.09k
      OS Memory           35376.00k
      Timestamp           09/08/2025 12:03:32 AM
      Step Count                        106  Switch Count  1
      Page Faults                       0
      Page Reclaims                     291
      Page Swaps                        0
      Voluntary Context Switches        135
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           344

NOTE: There were 5 observations read from the data set WORK.CORR_DATA.
```

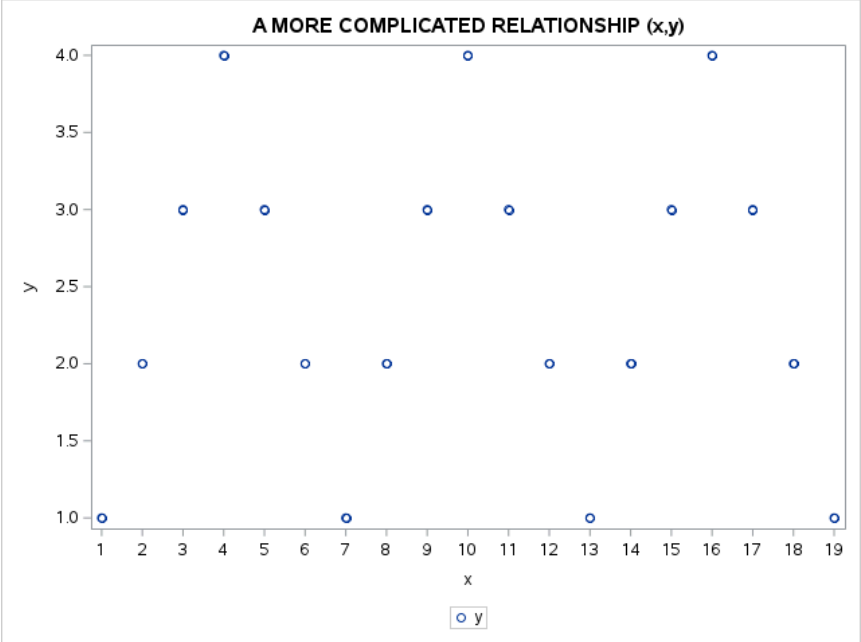
```
502        /*================================*
503        |  (b) Description of relationship |
504        /*================================*/
505        proc sgplot data=corr_data;
506        scatter x=x y=y;
507        loess x=x y=y;
508        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              2488.93k
      OS Memory           35376.00k
      Timestamp           09/08/2025 12:03:33 AM
      Step Count                        107  Switch Count  1
      Page Faults                       0
      Page Reclaims                     293
      Page Swaps                        0
      Voluntary Context Switches        166
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           464

NOTE: There were 5 observations read from the data set WORK.CORR_DATA.

509        /*==================*
510        |  (c) Correlation  |
511        /*==================*/
512        title "Scatter plot of x & y";
513        proc corr data=corr_data;
514        var x y;
515        run;

NOTE: PROCEDURE CORR used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              894.93k
      OS Memory           34472.00k
      Timestamp           09/08/2025 12:03:33 AM
      Step Count                        108  Switch Count  0
      Page Faults                       0
      Page Reclaims                     52
      Page Swaps                        0
      Voluntary Context Switches        4
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0


516        /*=================================================*
517        |  (d) The importance of correlation for this exercise |
518        /*=================================================*/
519
520
521
522        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
532
```

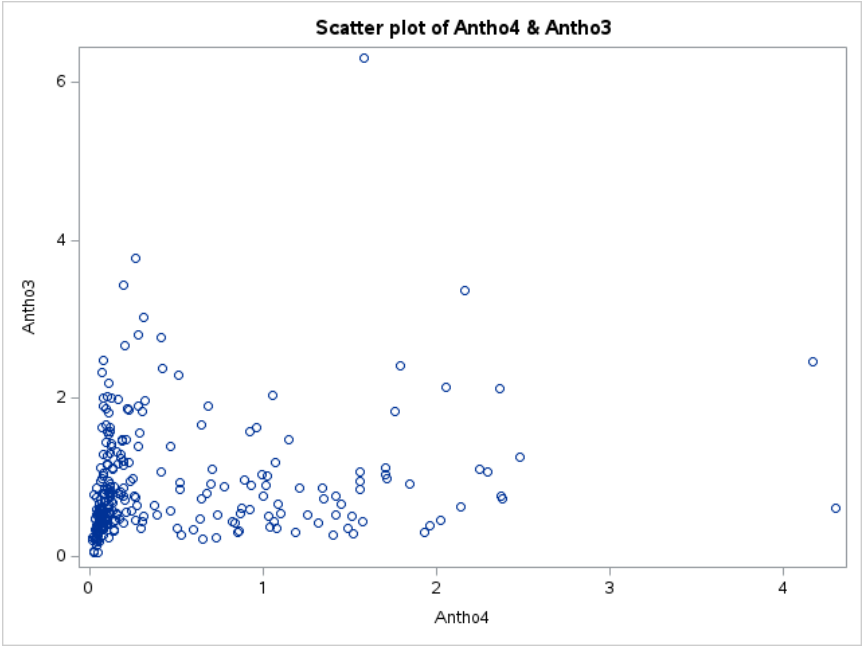
Results: Homework_2_Submission.sas

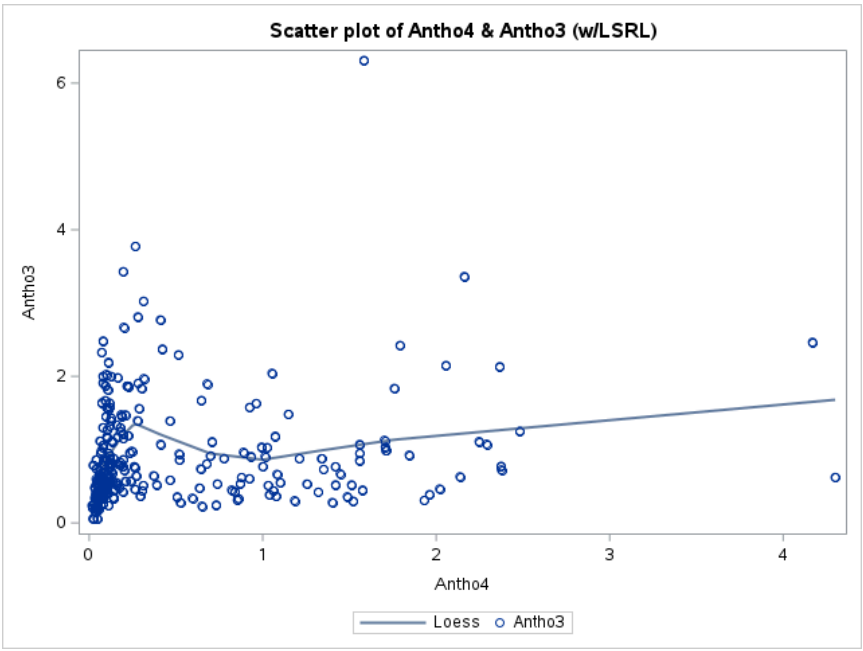

A WEAK NEGATIVE LINEAR RELATIONSHIP (x,y)
14
12
10
8
6
4
y
2
4
6
8
10
x
Loess
y

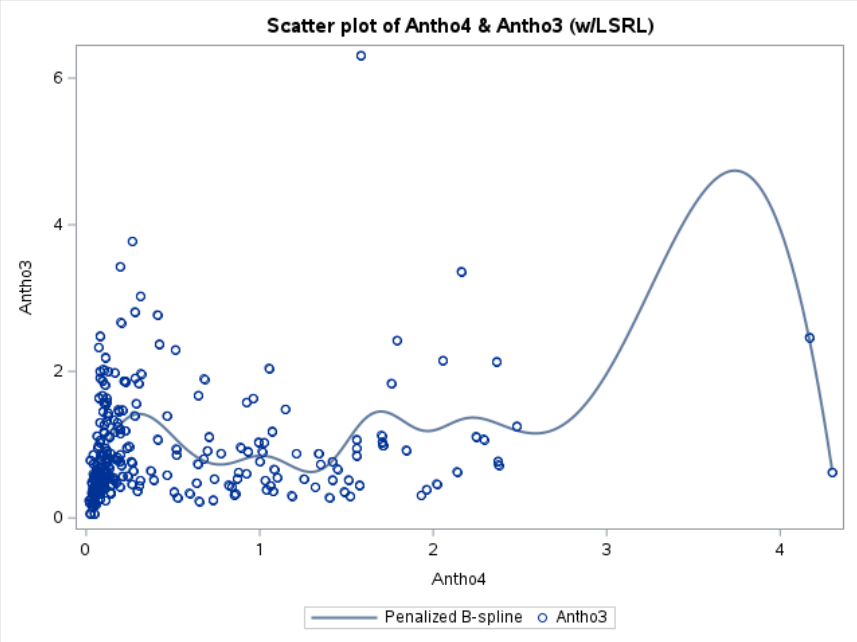

NO CORRELATION (x,y)
1
0
-1
-2
y
-1.5
-1.0
-0.5
0.0
0.5
1.0
x
Loess
y

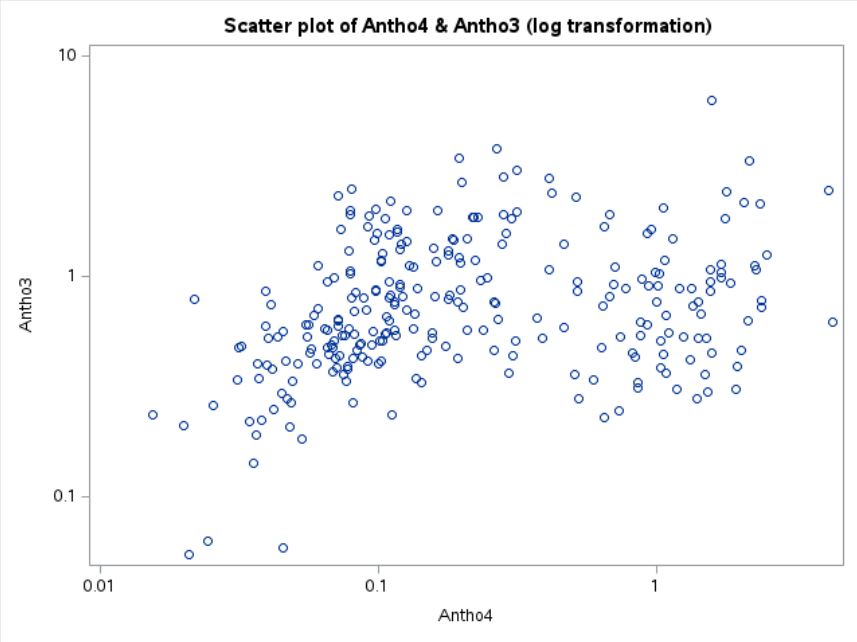

A STRONG CURVILINEAR RELATIONSHIP (x,y)
y
10
5
0
-4
-2
0
2
4
x
Loess
y

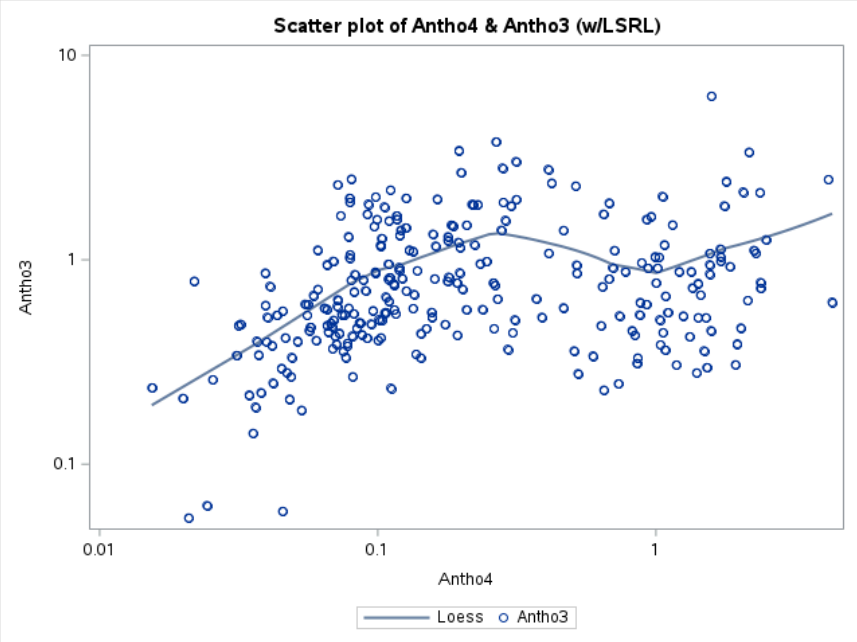

A MORE COMPLICATED RELATIONSHIP (x,y)
y
4.0
3.5
3.0
2.5
2.0
1.5
1.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
x
y

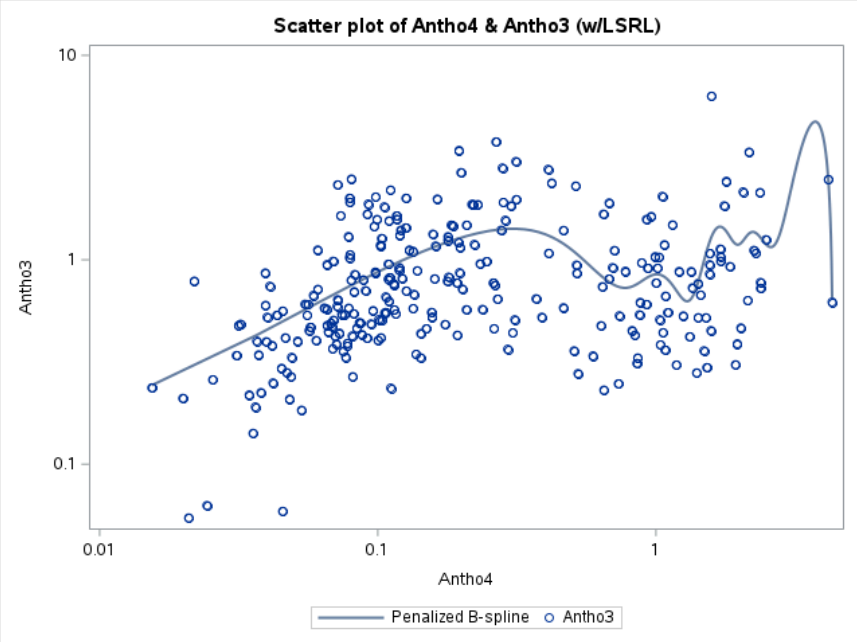

Scatter plot of Antho4 & Antho3
Antho3
Antho4
0
1
2
3
4
6

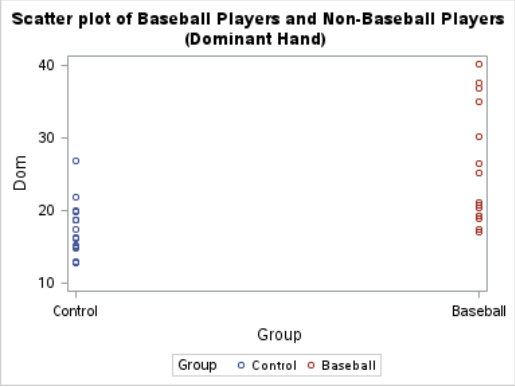

Scatter plot of Antho4 & Antho3 (w/LSRL)
Antho3
Antho4
Loess
Antho3
0
1
2
3
4
6


Scatter plot of Antho4 & Antho3 (w/LSRL)
Antho3
Antho4
0
2
4
6
0
1
2
3
4
Penalized B-spline
Antho3


Scatter plot of Antho4 & Antho3 (log transformation)
Antho3
Antho4
10
1
0.1
0.01
0.1
1







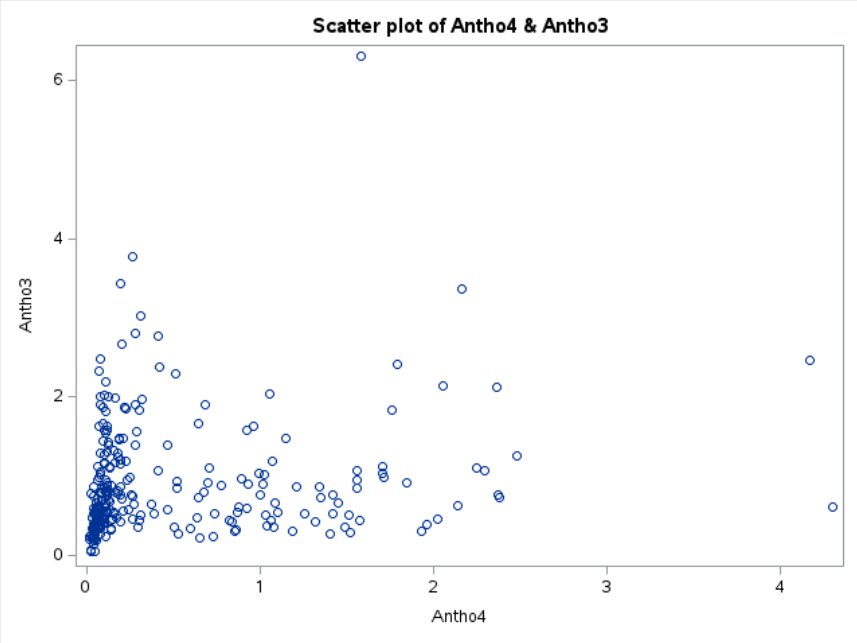
Correlation of Antho4 and Antho3

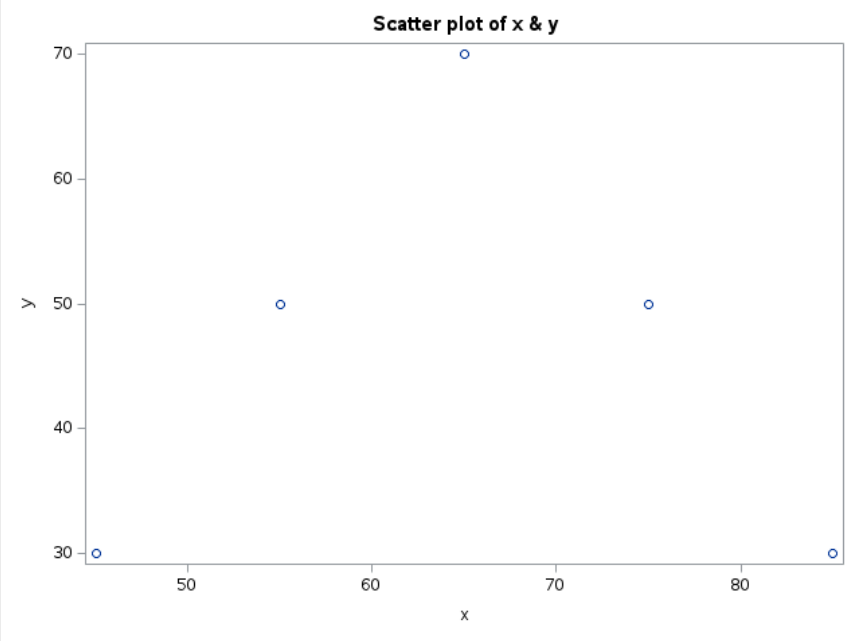
The CORR Procedure

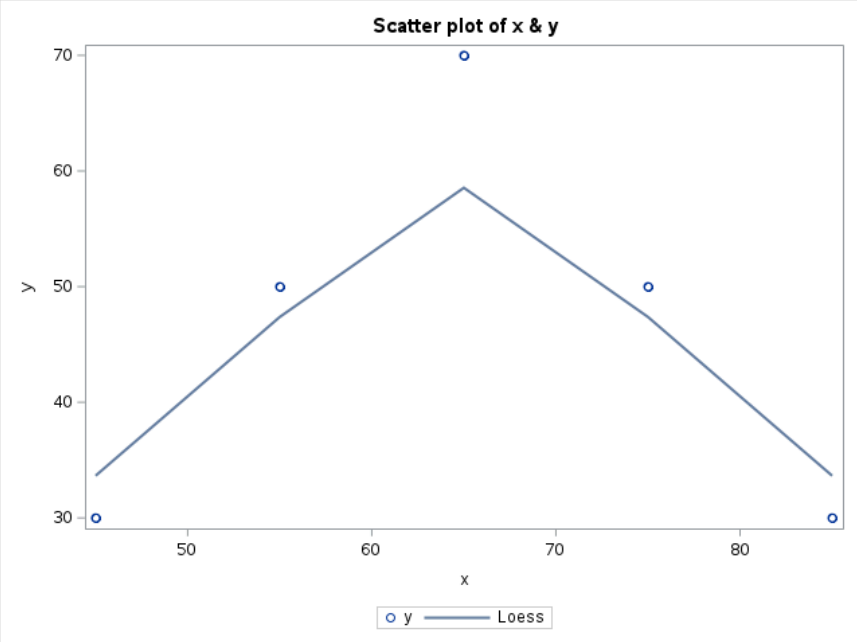
2 Variables:	Antho4 Antho3

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Antho4	267	0.48370	0.67708	129.14800	0.01540	4.29950
Antho3	267	0.91505	0.72788	244.31900	0.05460	6.31100

Pearson Correlation Coefficients, N = 267 Prob > |r| under H0: Rho=0		
	Antho4	Antho3
Antho4	1.00000	0.16324 0.0075
Antho3	0.16324 0.0075	1.00000







Scatter plot of x & y

The CORR Procedure

2 Variables:	x y

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
x	5	65.00000	15.81139	325.00000	45.00000	85.00000
y	5	46.00000	16.73320	230.00000	30.00000	70.00000

Pearson Correlation Coefficients, N = 5
Prob > |r| under H0: Rho=0

	x	y
x	1.00000	0.00000 1.0000
y	0.00000 1.0000	1.00000