

SESUG Paper 148-2023

Systemic Quantitative check to identify if a variable is a confounder in a dataset using SAS® Macro code

Dr. Kannan Deivasigamani, Centene – TSS COA

Dr. Douglas Lunsford – Tutor with the best, LLC.

Abstract

Researchers and statisticians at some point during their analysis tend to think about confounders and if a variable is a confounder in their study. While there are a handful of ways to identify this, I don't recall coming across a tool that would perform a systemic and quantitative check through a simple line of code in SAS. This paper is an attempt to provide that easy option for a researcher to call by simply passing a few parameters as part of the call while the macro code does the rest. Given there are non-quantitative approaches to dissect the essence of the dataset and the study to determine if a certain variable in a dataset is a confounder, this automated quantitative approach is likely to add value by taking away some of the manual steps and tasks by means of the Systemic Quantitative Confounder Check (SQCC) macro. The code obtains the estimate of the coefficient of the predictor variable without the confounder and checks if the value deviates by more than 10% with the presence of the confounder to indicate that the variable is indeed a confounder variable. The researcher can take this information and perform the corresponding activities accordingly. The call to the macro is made simple so it is easy to use and can be called multiple times as needed.

INTRODUCTION

The goal of the author for this paper was to develop a tool that can be called easily with a few parameters that can be passed to identify if a variable is a confounder for that specific study in a systemic quantitative fashion. A literature published by LaMorte & Sullivan (n.d) from Boston university school of public health contains explanation about a quantitative way of identifying a confounder variable. This macro is an effort to systemically perform that operation by a simple macro call that can print out the result indicating whether the suspected variable is a confounder. Choueiry (n.d) also supports a similar method in his literature. Graziano, & Raulin suggest other methods to identify confounder variables. While many methods exist. The scope of this paper is limited to one of the methods presented in the literature by LaMorte & Sullivan (n.d).

MACRO DETAILS

Quantitative Systemic Confounder Check (QSCC)

Name: **QSCC**

Parameters: indsn, iv, dv, confounder variable (all positional parameters)

Indsn = input dataset

iv = independent variable

dv = dependent variable

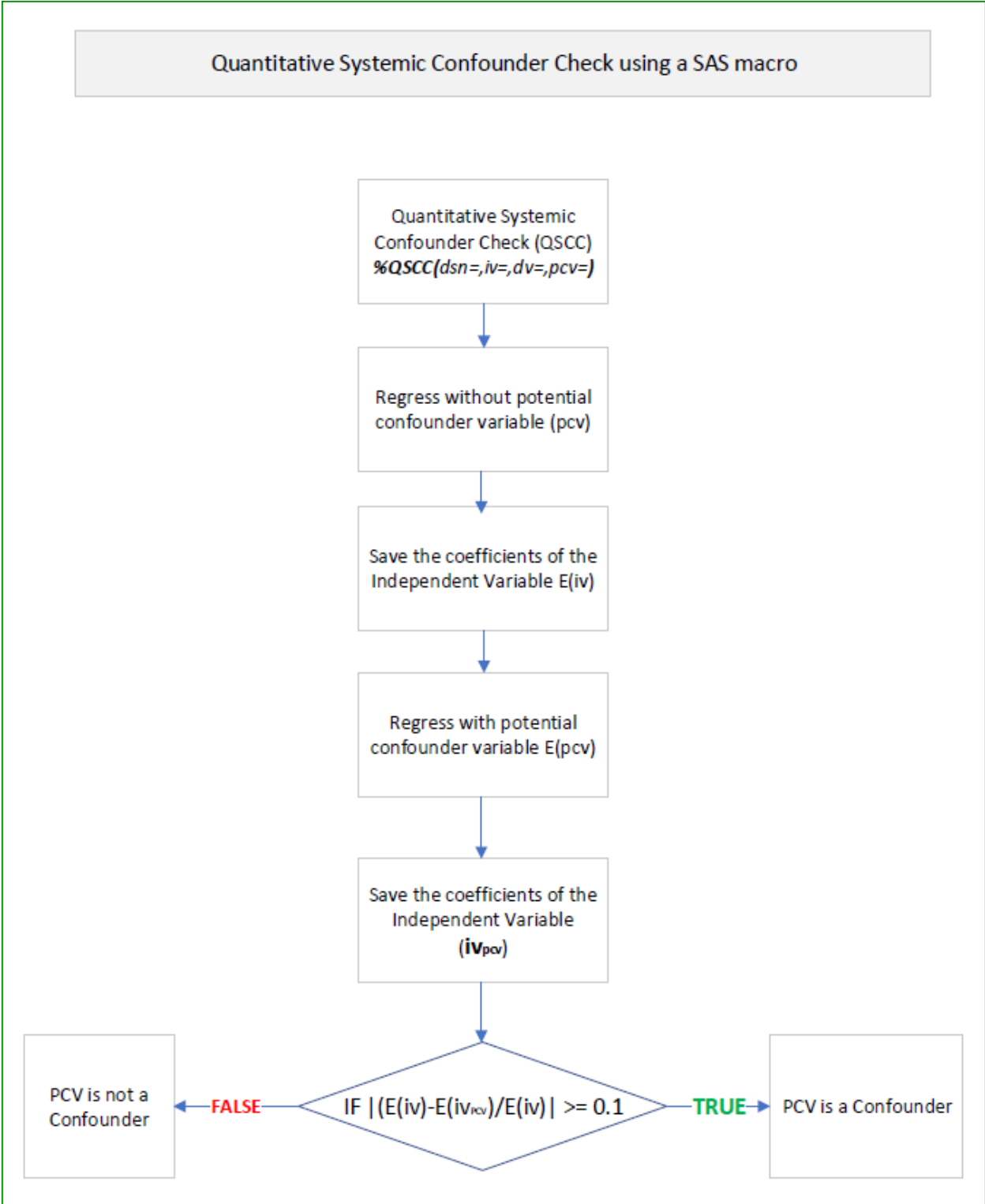
Confounder variable = suspected variable to be the confounder

Invocation: %qsc (dsn, iv, dv, cv)

The example below shows how the macro is called to demonstrate the steps in computing the coefficients of the independent variables and measure if there is a difference of at least 10 % in the value of the coefficient of the independent variable with the inclusion of the potential confounder. The first part of the example shows the sampling of the data from a larger dataset, to allow the computation in SAS Studio®.

The log output is also placed in the sheets that follow revealing the counts and processing info from each step with the logging options of MLOGIC, MPRINT & SYMBOLGEN.

PICTORIAL REPRESENTATION OF THE PROCESS



MACRO OPERATION

The macro QSCC shows 4 parameters that are positional and are required for proper execution. None of the parameters are optional. The first parameter is the source dataset that is used in developing the model with all the necessary variables. The first parameter is DSN and the value expected is a name of a SAS dataset that can be either temporary or permanent, given there are proper access rights. The 2nd parameter is the name of the independent variable (iv). The third parameter is the dependent variable (dv). Both iv and dv are expected to be part of DSN. The final and fourth parameter passed while calling this macro is the name of the suspected or potential confounder variable name (cv).

The first step in the macro is a regression. The code is set up in a way that the output of the regression is saved in a dataset. The coefficient of the independent variable and p-value are of interest for this exercise. If the p-value is over the typical 0.05 for the independent variable, then the coefficient is used for further computation and analysis of confounder. If the basic model does not show validity, the exercise is terminated with an appropriate output message. On the other hand, if the p-value indicates a statistically valid model, then the next step is carried out.

The second step involves another regression with the confounder variable coded. Alike the first regression, the output is saved into another dataset. The interest is towards the coefficient of the independent variable and the confounder variable while we also look at the p-values of both. If the p-values show a statistical significance ($p > 0.05$) then the absolute value of the delta of the coefficients is computed to a percent value. If the difference is equal or over 0.10 or 10% of the initial value, then the confounder is stated as statistically significant and is in fact, a confounder variable. In contrast, if the delta is less than 0.10 or 10%, then the variable is not a confounder. An appropriate message is written to the output as a result. This last part of the percent calculation and checking if the value is meeting the cutoff value is done via a PROC SQL as shown in the code section below.

SOURCE CODE

```
1 options mlogic mprint symbolgen;
2
3 %macro prepare_test_data;
4
5     filename reffile '/home/textkannan0/SESUG/2023/Mortgage_Data.csv';
6
7     libname l '/home/textkannan0/SESUG/2023';
8
9
10    proc import datafile=reffile
11
12        dbms=csv
13
14        out=l.Tampa_Mortgage_2023;
15
16        getnames=yes;
17
18    run;
19
20
21    data l.Tampa_Mortgage_2023;
22
23        set l.Tampa_Mortgage_2023;
24
25        rand=rand('Uniform');
26
27    run;
28
29
30    proc sort data=l.Tampa_Mortgage_2023;
31
32        by rand;
33
34    run;
35
36    data l.sample_100;
37
38        set l.Tampa_Mortgage_2023(obs=20000);
39
40    run;
41
42    data sampl;
43
44        set l.sample_100(keep=loan_amount_000s
45                        rate_spread
46                        applicant_income_000s
47                        hud_median_family_income);
48
49        length loan_amount_num
50               rate_spread_num
51               applicant_income_num
52               hud_median_family_inc_num 8.;
53
```

```

54     loan_amount_num = input(trim(loan_amount_000s),8.);
55
56     rate_spread_num = input(trim(rate_spread),4.);
57
58     applicant_income_num = input(trim(applicant_income_000s),8.);
59
60     hud_median_family_inc_num = input(trim(hud_median_family_income),8.);
61
62     run;
63
64 %mend prepare_test_data;
65
66 %macro QSCC(dsn,iv,dv,cv);
67
68     ODS GRAPHICS / RESET IMAGENAME = 'Final'
69     IMAGEFMT =JPEG HEIGHT = 3in WIDTH = 5in;
70
71     ods output ParameterEstimates = parms1;
72
73     proc reg data=sampl OUTEST=EST1/*(DROP=_TYPE_ _MODEL_ _RMSE_ _DEPVAR_) plots(maxpoints=none)*/*;
74
75         model &dv. = &iv.;
76
77         title "Model without confounder var | &dv. = &iv.";
78
79     run;
80
81     ods output ParameterEstimates = parms2;
82
83     proc reg data=sampl OUTEST=EST2/*(DROP=_TYPE_ _MODEL_ _RMSE_ _DEPVAR_) plots(maxpoints=none)*/*;
84
85         model &dv. = &iv. &cv.;
86
87         title "Crude model with confounder var | &dv. = &iv. &cv.";
88
89     run;
90
91     data _null_;
92
93         set parms1;
94
95         where variable="&iv.";
96
97         call symput('iv1_pvalue',probt);
98
99         if probt > 0.05
100         then
101             do;
102                 put '====iv is statistically insignificant====';
103                 abort 99;
104             end;

```

```

105
106 run; proc print data=parms1;title '==1==';run;
107
108 data _null_;
109     set parms2;
110     where variable="&iv.";
111
112     call symput('iv2_pvalue',probt);
113
114     if probt > 0.05
115     then
116     do;
117         put '===iv is statistically insignificant in the presence of cv===';
118         abort 99;
119     end;
120
121 run;
122
123 data _null_;
124     set parms2;
125     where variable="&cv.";
126
127     call symput('cv_pvalue',probt);
128
129 run;
130
131 proc sql;
132     create table chg as
133     select case
134     when (ABS(est2.&iv. - est1.&iv.) / est1.&iv.) ge .1
135     then
136         "&cv. IS A CONFOUNDER"
137     ELSE
138         "&cv. IS NOT A CONFOUNDER"
139     END AS QSCC
140     from est1,
141     est2
142     ;
143
144 quit;
145
146 proc print data=est1;
147     TITLE "ESTIMATES OF IV WITHOUT CONFOUNDER";
148
149 run;
150
151 proc print data=est2;
152     TITLE "ESTIMATES OF IV WITH CONFOUNDER";
153
154 run;
155
156 proc print data=chg;
157     TITLE "RESULT OF CONFOUNDER CHECK";
158
159

```

```

168
169     run;
170
171     %put iv1_pvalue === &iv1_pvalue.;
172     %put iv2_pvalue === &iv2_pvalue.;
173     %put cv_pvalue === &cv_pvalue.;
174
175 %mend QSCC;
176
177 %macro run_pgm;
178
179     %prepare_test_data
180
181     %QSCC(dsn=sampl,
182         dv=loan_amount_num,
183         iv=applicant_income_num,
184         cv=rate_spread_num)
185
186 %mend run_pgm;
187
188 %run_pgm;

```

RESULTS OF MACRO-EXECUTION

The results of the first part of the code, the data preparation is named as sampl. The dataset from where the random sample is created is very large. In order to accommodate the size within SAS Studio®, a smaller sample is taken and used in this example. The first macro called %prepare_test_data contains code to accomplish this. Execution of this macro results in the dataset named as “sampl” which is fed into the next step as a parameter.

The results from the second step is written to the log and the output in different ways. This macro is called %QSCC that accepts four parameters as input viz. dsn=, dv=, iv=, and cv writes the final text output to say if the CV is in fact a confounder. The p-values of the coefficient of IV without the CV and with CV are printed along with the significant result whether if the suspected variable is a confounder. This is presented in the results window to the bottom of the results window.

LIMITATIONS

- The macro currently employs just one method of drawing the line to delineate between a confounder or otherwise. It would be more robust if there are additional checks that could be added to this method to confirm more than one way.

CONCLUSION

The macro code was created to help researchers, statisticians, and SAS programmers to save time by simple invocation of this macro and by passing the appropriate parameters, by cutting down the development time, to get results quicker with smaller code making the process efficient. Also, modularizing this may help with easier management and standardization of code if used by multiple users if the macro is in a SAS “autocall” or a %include code to be made available to a group of users or across an organization. The individual has choices and can accomplish the task according to their preferred style as the macro can be called with just a one-line code.

Bibliography

- LaMorte, W. W., & Sullivan, L. (n.d.). *Confounding and Effect Measure Modification*. Retrieved September 10, 2023, from https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704-ep713_confounding-em/bs704-ep713_confounding-em_print.html#:~:text=Identifying%20Confounding,for%20a%20potential%20confounding%20factor
- Choueiry, G. (n.d.). *QUANTIFYING HEALTH*. 4 Simple Ways to Identify Confounding. Retrieved September 10, 2023, from <https://quantifyinghealth.com/identify-confounding/>
- Graziano, & Raulin (n.d.). *Research Methods*. Retrieved September 10, 2023, from <https://graziano-raulin.com/supplements/confvar.htm>

RECOMMENDED READING

- Base SAS® Procedures Guide
- SAS® Macro Language: Reference Guide®
- SAS® Macro Programming

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Dr. Kannan Deivasigamani

TextDrK@gmail.com

<https://www.linkedin.com/in/kannandeivasigamani>

Dr. Douglas Lunsford

glunsfor@gmail.com

<https://www.linkedin.com/in/g-douglas-lunsford-4a72609/>

Appendix – SAS Log

```
MLOGIC (RUN_PGM): Beginning execution.
MLOGIC (PREPARE_TEST_DATA): Beginning execution.
MPRINT (PREPARE_TEST_DATA): FILENAME REFFILE '/home/textkannan0/SESUG/2023/Mortgage_Data.csv';
MPRINT (PREPARE_TEST_DATA): libname l '/home/textkannan0/SESUG/2023';
NOTE: Libref L was successfully assigned as follows:
      Engine:          V9
      Physical Name:   /home/textkannan0/SESUG/2023
MPRINT (PREPARE_TEST_DATA): PROC IMPORT DATAFILE=REFFILE DBMS=CSV OUT=l.Tampa_Mortgage_2023;
MPRINT (PREPARE_TEST_DATA): ADLM;
MPRINT (PREPARE_TEST_DATA): GETNAMES=YES;
MPRINT (PREPARE_TEST_DATA): RUN;

NOTE: Import cancelled. Output dataset L.TAMPA_MORTGAGE_2023 already exists. Specify REPLACE option to
overwrite it.
NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time          0.00 seconds
      user cpu time      0.00 seconds
      system cpu time    0.00 seconds
      memory             416.46k
      OS Memory         29088.00k
      Timestamp          09/11/2023 01:41:47 AM
      Step Count         360  Switch Count  0
      Page Faults        0
      Page Reclaims      15
      Page Swaps         0
      Voluntary Context Switches  2
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 0

MPRINT (PREPARE_TEST_DATA): ;

MPRINT (PREPARE_TEST_DATA): data l.Tampa_Mortgage_2023;
MPRINT (PREPARE_TEST_DATA): set l.Tampa_Mortgage_2023;
MPRINT (PREPARE_TEST_DATA): rand=rand('Uniform');
MPRINT (PREPARE_TEST_DATA): run;
NOTE: There were 365896 observations read from the data set L.TAMPA_MORTGAGE_2023.
NOTE: The data set L.TAMPA_MORTGAGE_2023 has 365896 observations and 79 variables.
NOTE: DATA statement used (Total process time):
      real time          0.98 seconds
      user cpu time      0.06 seconds
      system cpu time    0.58 seconds
      memory             3866.46k
      OS Memory         31912.00k
      Timestamp          09/11/2023 01:41:48 AM
      Step Count         361  Switch Count  4
      Page Faults        0
      Page Reclaims      497
      Page Swaps         0
      Voluntary Context Switches 4746
      Involuntary Context Switches 1
      Block Input Operations 712960
      Block Output Operations 755464

MPRINT (PREPARE_TEST_DATA): proc sort data=l.Tampa_Mortgage_2023;
MPRINT (PREPARE_TEST_DATA): by rand;
MPRINT (PREPARE_TEST_DATA): run;

NOTE: There were 365896 observations read from the data set L.TAMPA_MORTGAGE_2023.
NOTE: The data set L.TAMPA_MORTGAGE_2023 has 365896 observations and 79 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time          1.36 seconds
      user cpu time      0.35 seconds
      system cpu time    0.85 seconds
      memory             422626.57k
      OS Memory         449816.00k
      Timestamp          09/11/2023 01:41:50 AM
      Step Count         362  Switch Count  7
      Page Faults        0
      Page Reclaims     98861
      Page Swaps         0
      Voluntary Context Switches 6707
      Involuntary Context Switches 17
      Block Input Operations 755488
      Block Output Operations 755480
```

```

MPRINT(PREPARE_TEST_DATA):  data l.sample_100;
MPRINT(PREPARE_TEST_DATA):  set l.Tampa_Mortgage_2023(obs=20000);
MPRINT(PREPARE_TEST_DATA):  run;
NOTE: There were 20000 observations read from the data set L.TAMPA_MORTGAGE_2023.
NOTE: The data set L.SAMPLE_100 has 20000 observations and 79 variables.
NOTE: DATA statement used (Total process time):
      real time           0.13 seconds
      user cpu time       0.00 seconds
      system cpu time     0.08 seconds
      memory              3756.00k
      OS Memory          31912.00k
      Timestamp           09/11/2023 01:41:50 AM
      Step Count          363  Switch Count  4
      Page Faults         0
      Page Reclaims       502
      Page Swaps          0
      Voluntary Context Switches 320
      Involuntary Context Switches 0
      Block Input Operations 42528
      Block Output Operations 41488

MPRINT(PREPARE_TEST_DATA):  data sampl;
MPRINT(PREPARE_TEST_DATA):  set l.sample_100(keep=loan_amount_000s rate_spread applicant_income_000s
hud_median_family_income);
MPRINT(PREPARE_TEST_DATA):  length loan_amount_num rate_spread_num applicant_income_num
hud_median_family_inc_num 8.;
MPRINT(PREPARE_TEST_DATA):  loan_amount_num = input(trim(loan_amount_000s),8.);
MPRINT(PREPARE_TEST_DATA):  rate_spread_num = input(trim(rate_spread),4.);
MPRINT(PREPARE_TEST_DATA):  applicant_income_num = input(trim(applicant_income_000s),8.);
MPRINT(PREPARE_TEST_DATA):  hud_median_family_inc_num = input(trim(hud_median_family_income),8.);
MPRINT(PREPARE_TEST_DATA):  run;

NOTE: There were 20000 observations read from the data set L.SAMPLE_100.
NOTE: The data set WORK.SAMPL has 20000 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      user cpu time       0.00 seconds
      system cpu time     0.02 seconds
      memory              3196.09k
      OS Memory          31656.00k
      Timestamp           09/11/2023 01:41:50 AM
      Step Count          364  Switch Count  5
      Page Faults         0
      Page Reclaims       477
      Page Swaps          0
      Voluntary Context Switches 239
      Involuntary Context Switches 0
      Block Input Operations 41504
      Block Output Operations 2312

MLOGIC(PREPARE_TEST_DATA):  Ending execution.
MLOGIC(QSCC):  Beginning execution.
MLOGIC(QSCC):  Parameter DSN has value sampl
MLOGIC(QSCC):  Parameter DV has value loan_amount_num
MLOGIC(QSCC):  Parameter IV has value applicant_income_num
MLOGIC(QSCC):  Parameter CV has value rate_spread_num
MPRINT(QSCC):  ODS GRAPHICS / RESET IMAGENAME = 'Final' IMAGEFMT =JPEG HEIGHT = 3in WIDTH = 5in;
MPRINT(QSCC):  ods output ParameterEstimates = parms;
MPRINT(QSCC):  proc reg data=sampl OUTEST=EST1(DROP=_TYPE_ _MODEL_ _RMSE_ _DEPVAR_) plots(maxpoints=none);
SYMBOLGEN:  Macro variable DV resolves to loan_amount_num
SYMBOLGEN:  Macro variable IV resolves to applicant_income_num
MPRINT(QSCC):  model loan_amount_num = applicant_income_num;
SYMBOLGEN:  Macro variable DV resolves to loan_amount_num
SYMBOLGEN:  Macro variable IV resolves to applicant_income_num
MPRINT(QSCC):  title "Model without confounder var | loan_amount_num = applicant_income_num";
MPRINT(QSCC):  run;
NOTE: The data set WORK.PARMS has 2 observations and 8 variables.
NOTE: The data set WORK.EST1 has 1 observations and 3 variables.
NOTE: PROCEDURE REG used (Total process time):
      real time           8.13 seconds
      user cpu time       7.46 seconds
      system cpu time     0.18 seconds
      memory              23201.75k
      OS Memory          48072.00k
      Timestamp           09/11/2023 01:41:58 AM
      Step Count          365  Switch Count  75
      Page Faults         0
      Page Reclaims       30417

```

```

Page Swaps                                0
Voluntary Context Switches                3706
Involuntary Context Switches              18
Block Input Operations                     0
Block Output Operations                    126432

```

```

MPRINT(QSCC):  data parms1;
MPRINT(QSCC):  set parms;
MPRINT(QSCC):  run;
NOTE: There were 2 observations read from the data set WORK.PARMS.
NOTE: The data set WORK.PARMS1 has 2 observations and 8 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             965.56k
OS Memory          39848.00k
Timestamp          09/11/2023 01:41:58 AM
Step Count         366  Switch Count  2
Page Faults        0
Page Reclaims     127
Page Swaps         0
Voluntary Context Switches  11
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  264

```

```

MPRINT(QSCC):  ods output ParameterEstimates = parms;
MPRINT(QSCC):  proc reg data=sampl OUTEST=EST2(DROP=_TYPE_ _MODEL_ _RMSE_ _DEPVAR_) plots(maxpoints=none);
SYMBOLGEN: Macro variable DV resolves to loan_amount_num
SYMBOLGEN: Macro variable IV resolves to applicant_income_num
SYMBOLGEN: Macro variable CV resolves to rate_spread_num
MPRINT(QSCC):  model loan_amount_num = applicant_income_num rate_spread_num;
SYMBOLGEN: Macro variable DV resolves to loan_amount_num
SYMBOLGEN: Macro variable IV resolves to applicant_income_num
SYMBOLGEN: Macro variable CV resolves to rate_spread_num
MPRINT(QSCC):  title "Crude model with confounder var | loan_amount_num = applicant_income_num
rate_spread_num";

```

```

MPRINT(QSCC):  run;
NOTE: The data set WORK.PARMS has 3 observations and 8 variables.
NOTE: The data set WORK.EST2 has 1 observations and 4 variables.
NOTE: PROCEDURE REG used (Total process time):

```

```

real time          0.56 seconds
user cpu time      0.38 seconds
system cpu time    0.05 seconds
memory             14625.87k
OS Memory          50152.00k
Timestamp          09/11/2023 01:41:58 AM
Step Count         367  Switch Count  27
Page Faults        0
Page Reclaims     13398
Page Swaps         0
Voluntary Context Switches  855
Involuntary Context Switches 1
Block Input Operations  0
Block Output Operations  10576

```

```

MPRINT(QSCC):  data parms2;
MPRINT(QSCC):  set parms;
MPRINT(QSCC):  run;
NOTE: There were 3 observations read from the data set WORK.PARMS.
NOTE: The data set WORK.PARMS2 has 3 observations and 8 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             949.40k
OS Memory          40360.00k
Timestamp          09/11/2023 01:41:58 AM
Step Count         368  Switch Count  2
Page Faults        0
Page Reclaims     127
Page Swaps         0
Voluntary Context Switches  15
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  264

```

```

MPRINT(QSCC):  data _null_;
MPRINT(QSCC):  set parms1;
SYMBOLGEN:    Macro variable IV resolves to applicant_income_num
MPRINT(QSCC):  where variable="applicant_income_num";
MPRINT(QSCC):  call symput('iv1_pvalue',probt);
MPRINT(QSCC):  if probt > 0.05 then do;
MPRINT(QSCC):  put '====iv is statistically insignificant====';
MPRINT(QSCC):  abort 99;
MPRINT(QSCC):  end;
MPRINT(QSCC):  run;

```

NOTE: Numeric values have been converted to character values at the places given by: (Line):(Column).
9559:213

NOTE: There were 1 observations read from the data set WORK.PARMS1.
WHERE variable='applicant_income_num';

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 | 867.56k |
| OS Memory | 40100.00k |
| Timestamp | 09/11/2023 01:41:58 AM |
| Step Count | 369 |
| Switch Count | 0 |
| Page Faults | 0 |
| Page Reclaims | 61 |
| Page Swaps | 0 |
| Voluntary Context Switches | 0 |
| Involuntary Context Switches | 0 |
| Block Input Operations | 0 |
| Block Output Operations | 0 |

```

MPRINT(QSCC):  proc print data=parms1;
MPRINT(QSCC):  title '==1==';
MPRINT(QSCC):  run;

```

NOTE: There were 2 observations read from the data set WORK.PARMS1.

NOTE: PROCEDURE PRINT used (Total process time):

| | |
|------------------------------|------------------------|
| real time | 0.00 seconds |
| user cpu time | 0.00 seconds |
| system cpu time | 0.00 seconds |
| memory | 673.28k |
| OS Memory | 40100.00k |
| Timestamp | 09/11/2023 01:41:58 AM |
| Step Count | 370 |
| Switch Count | 1 |
| Page Faults | 0 |
| Page Reclaims | 70 |
| Page Swaps | 0 |
| Voluntary Context Switches | 8 |
| Involuntary Context Switches | 0 |
| Block Input Operations | 0 |
| Block Output Operations | 0 |

```

MPRINT(QSCC):  data _null_;
MPRINT(QSCC):  set parms2;
SYMBOLGEN:    Macro variable IV resolves to applicant_income_num
MPRINT(QSCC):  where variable="applicant_income_num";
MPRINT(QSCC):  call symput('iv2_pvalue',probt);
MPRINT(QSCC):  if probt > 0.05 then do;
MPRINT(QSCC):  put '====iv is statistically insignificant in the presence of cv====';
MPRINT(QSCC):  abort 99;
MPRINT(QSCC):  end;
MPRINT(QSCC):  run;

```

NOTE: Numeric values have been converted to character values at the places given by: (Line):(Column).
9568:219

NOTE: There were 1 observations read from the data set WORK.PARMS2.
WHERE variable='applicant_income_num';

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 | 755.43k |
| OS Memory | 40100.00k |
| Timestamp | 09/11/2023 01:41:58 AM |
| Step Count | 371 |
| Switch Count | 0 |
| Page Faults | 0 |
| Page Reclaims | 61 |
| Page Swaps | 0 |
| Voluntary Context Switches | 0 |

```

Involuntary Context Switches      0
Block Input Operations             0
Block Output Operations            8

```

```

MPRINT(QSCC):  data _null_;
MPRINT(QSCC):  set parms2;
SYMBOLGEN:    Macro variable CV resolves to rate_spread_num
MPRINT(QSCC):  where variable="rate_spread_num";
MPRINT(QSCC):  call symput('cv_pvalue',probt);
MPRINT(QSCC):  run;

```

NOTE: Numeric values have been converted to character values at the places given by: (Line):(Column).
265:205

NOTE: There were 1 observations read from the data set WORK.PARMS2.
WHERE variable='rate_spread_num';

```

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              866.53k
OS Memory           40100.00k
Timestamp           09/11/2023 01:41:58 AM
Step Count          372  Switch Count  0
Page Faults         0
Page Reclaims       61
Page Swaps          0
Voluntary Context Switches  0
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

```

MPRINT(QSCC):  proc sql;
SYMBOLGEN:    Macro variable IV resolves to applicant_income_num
SYMBOLGEN:    Macro variable IV resolves to applicant_income_num
SYMBOLGEN:    Macro variable IV resolves to applicant_income_num
SYMBOLGEN:    Macro variable CV resolves to rate_spread_num
SYMBOLGEN:    Macro variable CV resolves to rate_spread_num
MPRINT(QSCC):  create table chg as select case when (ABS (est2.applicant_income_num -
est1.applicant_income_num) /
est1.applicant_income_num) ge .1 then "rate_spread_num IS A CONFOUNDER" ELSE "rate_spread_num IS NOT A
CONFOUNDER" END AS QSCC from
est1, est2;

```

NOTE: The execution of this query involves performing one or more Cartesian product joins that cannot be optimized.

NOTE: Table WORK.CHG created, with 1 rows and 1 columns.

```

MPRINT(QSCC):  quit;
NOTE: PROCEDURE SQL used (Total process time):
real time           0.00 seconds
user cpu time       0.00 seconds
system cpu time     0.00 seconds
memory              6019.15k
OS Memory           45484.00k
Timestamp           09/11/2023 01:41:58 AM
Step Count          373  Switch Count  2
Page Faults         0
Page Reclaims       167
Page Swaps          0
Voluntary Context Switches  12
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  264

```

```

MPRINT(QSCC):  proc print data=est1;
MPRINT(QSCC):  TITLE "ESTIMATES OF IV WITHOUT CONFOUNDER";
MPRINT(QSCC):  run;

```

NOTE: There were 1 observations read from the data set WORK.EST1.

```

NOTE: PROCEDURE PRINT used (Total process time):
real time           0.00 seconds
user cpu time       0.01 seconds
system cpu time     0.00 seconds
memory              607.18k
OS Memory           40100.00k
Timestamp           09/11/2023 01:41:58 AM
Step Count          374  Switch Count  1
Page Faults         0
Page Reclaims       65
Page Swaps          0

```

```

Voluntary Context Switches      8
Involuntary Context Switches    0
Block Input Operations           0
Block Output Operations          16

```

```

MPRINT(QSCC):  proc print data=est2;
MPRINT(QSCC):  TITLE "ESTIMATES OF IV WITH CONFOUNDER";
MPRINT(QSCC):  run;
NOTE: There were 1 observations read from the data set WORK.EST2.
NOTE: PROCEDURE PRINT used (Total process time):
  real time          0.00 seconds
  user cpu time      0.01 seconds
  system cpu time    0.00 seconds
  memory             611.71k
  OS Memory          40100.00k
  Timestamp          09/11/2023 01:41:58 AM
  Step Count         375  Switch Count  1
  Page Faults        0
  Page Reclaims      65
  Page Swaps         0
  Voluntary Context Switches  11
  Involuntary Context Switches 0
  Block Input Operations 0
  Block Output Operations 0

```

```

MPRINT(QSCC):  proc print data=chg;
MPRINT(QSCC):  TITLE "RESULT OF CONFOUNDER CHECK";
MPRINT(QSCC):  run;
NOTE: There were 1 observations read from the data set WORK.CHG.
NOTE: PROCEDURE PRINT used (Total process time):
  real time          0.00 seconds
  user cpu time      0.00 seconds
  system cpu time    0.00 seconds
  memory             600.75k
  OS Memory          40100.00k
  Timestamp          09/11/2023 01:41:58 AM
  Step Count         376  Switch Count  1
  Page Faults        0
  Page Reclaims      63
  Page Swaps         0
  Voluntary Context Switches  9
  Involuntary Context Switches 0
  Block Input Operations 0
  Block Output Operations 0

```

```

MLOGIC(QSCC):  %PUT iv1_pvalue === &iv1_pvalue.
SYMBOLGEN:    Macro variable IV1_PVALUE resolves to          0
iv1_pvalue ===          0
MLOGIC(QSCC):  %PUT iv2_pvalue === &iv2_pvalue.
SYMBOLGEN:    Macro variable IV2_PVALUE resolves to 2.03582E-149
iv2_pvalue === 2.03582E-149
MLOGIC(QSCC):  %PUT cv_pvalue === &cv_pvalue.
SYMBOLGEN:    Macro variable CV_PVALUE resolves to 0.0001212325
cv_pvalue === 0.0001212325
MLOGIC(QSCC):  Ending execution.
MLOGIC(RUN_PGM): Ending execution.
266
267      OPTIONS NOSYNTAXCHECK;
268      ODS HTML CLOSE;
SYMBOLGEN:    Macro variable GRAPHTERM resolves to GOPTIONS NOACCESSIBLE;
269      &GRAPHTERM; ;*';*";*/;RUN;QUIT;
270      QUIT;RUN;
271      ODS HTML5 (ID=WEB) CLOSE;
272
273      FILENAME _GSFNAME;
NOTE: Fileref _GSFNAME has been deassigned.
274      DATA _NULL_;
275      RUN;

```

```

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             474.75k
  OS Memory          29096.00k
  Timestamp          09/11/2023 01:41:58 AM
  Step Count         377  Switch Count  0
  Page Faults        0

```


| | |
|------------------------------|----|
| Page Reclaims | 25 |
| Page Swaps | 0 |
| Voluntary Context Switches | 1 |
| Involuntary Context Switches | 0 |
| Block Input Operations | 0 |
| Block Output Operations | 0 |

276 OPTIONS NOTES STIMER SOURCE SYNTAXCHECK;
277