Sample Size Calculation to Evaluate Mediation Analysis
Rajendra Kadel, University of South Florida, Tampa, Florida

ABSTRACT
Mediation analysis is used to review the comparative change in the amount of strength of association of the primary predictor with the outcome after adjustment for the mediator. Mediation models are very widely used in social sciences and biomedical sciences. Before conducting mediation studies, researchers want to know the sample size (i.e. number of subjects) required for achieving the adequate power when testing for mediation. To the author’s knowledge, there is no any SAS® procedure that produces sample size calculation for mediation analysis. The author presents a macro that implements the methodology for sample size calculation for mediation analysis written by Vittinghoff et al. (2009). It implements the methods to calculate sample sizes for the linear regression models.

INTRODUCTION
Mediation analysis is used to review the comparative change in the amount of strength of association of the primary predictor with the outcome after adjustment for the mediator. Mediation models are very widely used in social sciences, especially in psychological studies. Before conducting mediation studies, researchers want to know the sample size (i.e. number of subjects) required for achieving the adequate power when testing for mediation. This problem of sample size calculation for testing the mediation effect of intermediate variable to the primary predictor is a common issue in epidemiologic and clinical research. To the author’s knowledge, there is no any SAS procedure that produces sample size calculation for mediation analysis. The SAS macro presented in this paper will be used to calculate sample for testing mediation effect based on Vittinghoff et al. (2009). They proposed the method based on variance inflation factor in regression that provide the exact sample size for linear model and approximations for logistic, Poisson, and Cox models. In this paper, I will present sample size calculation for linear models. The following four cases will be considered: (a) continuous primary predictor, continuous mediator (b) binary primary predictor, continuous mediator (c) continuous primary predictor, binary mediator, and (d) binary primary predictor, binary mediator. The authors argued that their method are also applicable to detecting the independent effect of a primary predictor in the presence of substantial confounding. The method of sample size calculation is based on the Wald tests as these tests are more common in practice even though these methods are less reliable in case of small samples and if the alternative hypothesis are far from null.

SAMPLE SIZE CALCULATION FOR LINEAR MODEL
Let $Y$ be the response, $X_1$ a primary predictor variable and $X_2$ is a mediator variable. The simple linear regression model takes the form

$$ y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) $$

(1)

Where $\beta_1$ is the estimator of primary predictor ($X_1$) before the adjustment of mediator ($X_2$), $\beta_2$ is the regression coefficient for the mediator ($X_2$) and $\sigma^2$ is the variance of the random error term in the linear regression. If $\beta_1^*$ is the estimator of the primary predictor ($X_1$) before the adjustment of mediator ($X_2$) in the full model (1), then proportion of treatment effect explained (PTE), which assess the mediation as the relative change in the strength of association of the primary predictor with the outcome after adjustment for the mediator, is of the form

$$ \frac{\beta_1^* - \beta_1}{\beta_1} $$

(2)
Vittinghoff et al. (2009) showed that for the linear regression (1), testing the mediation effect is equivalent to testing the null hypothesis $H_0: \beta_2 = 0$ versus the alternative hypothesis $H_1: \beta_2 \neq 0$. They also showed that, provided the full model is correctly specified, this is equivalent to testing $PTE = 0$, provided the Pearson correlation of $X_1$ and $X_2$, $\rho \neq 0$. This will lead to the following equation to calculate sample size

$$n = \frac{(z_a + z_v)^2 \sigma_e^2}{\beta_2^2 \sigma_z^2 (1 - \rho^2)}$$  \hspace{1cm} (3)

Where $z_a$ and $z_v$ are the quintiles of the standard normal distribution corresponding to the specified one-sided type-I and type-II error rates, $\sigma_e^2$ is the residual variance of the outcome, $\sigma_z^2$ is the variance of the mediator ($X_2$). Alternative to formula (3) are provided by Vittinghoff et al. (2009) if values of $PTE, \beta_1^*$ and $\beta_1$ available. The following equations 4-6 are the alternative to equation (3)

$$n = \frac{(z_a + z_v)^2 \rho^2 \sigma_e^2}{(\beta_1^* \sigma_{PTE})^2 (1 - \rho^2)}$$  \hspace{1cm} (4)

$$n = \frac{(z_a + z_v)^2 \sigma_e^2}{(\beta_2 \sigma_z^2)^2 - (\beta_1^* \sigma_{PTE})^2}$$  \hspace{1cm} (5)

$$n = \frac{(z_a + z_v)^2 \sigma_e^2}{(\beta_2 \sigma_z^2)^2 - [(\beta_1^* - \beta_1) \sigma_1]^2}$$  \hspace{1cm} (6)

If either of primary predictor $X_1$ and mediator $X_2$ is binary, its variance will be $\sigma_e^2 = f_j (1 - f_j), j = 1, 2$. Here, $f_j = Pr(X_j = 1)$ is the marginal prevalence. This indicates that formula (3) – (6) provided in Vittinghoff et al. (2009) can be used both for continuous and binary predictors.

**PARAMETERS AND DEFAULT VALUES**

The following are the list of the parameters and their default values that are used to calculate the sample size for mediation analysis in case of simple linear regression. The macro program will ask availability of certain parameter and uses one of the formulas (3) – (6) to calculate sample size. These formulas give user a choice depending upon which parameters can be most easily or readily available what information is available. However, it is not necessary for the user to indentify which formula to use in each of these cases. First three of the parameters listed below are required for each of four formulas and rest of them are based on what parameters are available. For example, if the variance of the predictors ($X_1$ and $X_2$) are available, then equation (5) will be used to calculate sample size. In that case, user will be asked to supply $\beta_1^*$, $\beta_2$ and $PTE$ in addition to the following first three (a-c) required parameters.

(a) Type I Error (e.g. 0.05),
(b) Power desired (e.g. 0.8 for 80%),
(c) Residual Variance of Outcome (e.g. $\sigma_e^2 = 1.00$)
(d) Variance of Primary Predictor $X_1$ (e.g. $\sigma_1^2 = 1$), for continuous primary predictor or marginal prevalence (e.g. $f_1 = 0.5$) if the predictor is binary
(e) Variance of Mediator $X_2$ (e.g. $\sigma_2^2 = 1$), for continuous mediator or marginal prevalence (e.g. $f_2 = 0.5$) of the mediator if it is binary
(f) Pearson Correlation Between $X_1$ and $X_2$ (e.g. $\rho = 0.30$)
(g) Estimate of $X_1$ before adjustment of $X_2$ (e.g. $\beta_1^* = 0.53$)
(h) Estimate of primary predictor $X_1$ after the adjustment of mediator $X_2$, (e.g. $\beta_1 = 0.50$)
(i) Estimate of predictor $X_2$ in full model (e.g. $\beta_2 = 0.10$) and
(j) Proportion of Treatment Effect (e.g $PTE = 5.66$).
SAMPLE SIZE CALCULATION MACRO
When we invoke the macro (attached in the Appendix), a series of windows will pop. The first window will be welcome window (Figure 1) as shown below. Users need to use [ENTER] key to move from one window to another window or one cell to next cell. The second (Figure 2) and third window (not shown here) will ask to choose if the primary predictor and mediators is continuous or binary. Fourth window (Figure 3) will be used to supply required parameters (Type I error, Power and Residual Variance). Rest of the windows will be based on the type of predictor (continuous or binary) chosen in second and third window. Once all the data are supplied, a window will pop up asking to open or save the Microsoft Word document that contains result (Figure 4 below).

Figure 1: Welcome Window

![Welcome Window](image1.png)

Figure 2: Type of Primary Predictor $X_1$ Window

![Type of Primary Predictor $X_1$ Window](image2.png)
As an example, consider the case where both \( X_1 \) and mediator \( X_2 \) are continuous with type I error rate 0.05, power 80\% (i.e. 0.80), residual variance is 1, variance of primary predictor that is 1 but the variance of mediator \( X_2 \) is not available. In that such case, Pearson's Correlation between \( X_1 \) and \( X_2 \) (\( \rho = 0.30 \) for example), estimate of primary predictor \( X_1 \) before the adjustment of \( X_2 \) (\( \beta_1^* = 0.53 \) for example), Proportion of Treatment Effect (PTE = 5.66 for example) are supplied. Based on this information, the macro will automatically choose the formula (4) to calculate the sample size and the following table will be produced in the Microsoft Word document.

**Figure 4: Microsoft Word Table**

### Sample Size Calculation for Linear Model

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description of Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>type1</td>
<td>TYPE I Error</td>
<td>0.05</td>
</tr>
<tr>
<td>power</td>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>ResVar</td>
<td>Residual Variance of Outcome Y</td>
<td>1.00</td>
</tr>
<tr>
<td>VarX1</td>
<td>Variance of Primary Predictor X1</td>
<td>1.00</td>
</tr>
<tr>
<td>PreEstX1</td>
<td>Estimate of X1 before the adjustment of X2 (B1*)</td>
<td>0.53</td>
</tr>
<tr>
<td>Corr</td>
<td>Pearson's Correlation between X1 and X2</td>
<td>0.30</td>
</tr>
<tr>
<td>PTE</td>
<td>Proportion of Treatment Effect(PTE)</td>
<td>5.66</td>
</tr>
<tr>
<td>n</td>
<td>Sample Size (n)</td>
<td>863.00</td>
</tr>
</tbody>
</table>
CONCLUSION
SAS macro presented in this paper is a simple and user friendly macro that will calculate the sample size for evaluating mediation effect for linear regression. This macro can also be used to calculate sample size in the presence of a common set of confounding variable both in null and full model. Anyone who has basic knowledge of SAS programming and a basic understanding of mediation analysis can use this macro with ease.

REFERENCES

ACKNOWLEDGMENTS
I want to thank Dr. Kevin Kip Ph.D., Associate Professor & Executive Director, College of Nursing Research Center, University of South Florida, Tampa for providing opportunity to work on this project, and advice and support during the preparation of this macro.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at:
   Rajendra Kadel,
   University of South Florida
   13201 Bruce B Downs, MDC56
   Tampa, Florida 33612
   (813) 794-0303
   rkadel@health.usf.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

APPENDIX
/* SAMPLE SIZE CALCULATION MACRO */

/* WINDOW START ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=80 color=GRAY */
#9 @30 'WELCOME TO SAMPLE SIZE CALCULATION PROGRAM' C=RED
#12 @33 'THIS PROGRAM CALCULATES SAMPLE SIZE ' C=BLUE
#14 @47 'FOR' C=BLUE PERSIST=YES
#16 @37 'EVALUATING MEDIATION' C=BLUE
#19 @47 'FOR' C=BLUE
#21 @42 'LINEAR MODEL' C=BLUE
#32 @34 'PRESS [ENTER] TO CONTINUE' C=RED AUTOSKIP=YES;

/* WINDOW X1SELECT ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=80 color=GRAY */
#8 @25 'PRESS [ENTER], THEN INPUT REQUIRED DATA' C=red
#15 @25 'IS PRIMARY PREDICTOR X1 CONTINUOUS OR BINARY ?' C=blue
#17 @40 X1 1 A=REV_VIDEO C=blue REQUIRED=YES
#17 @41 ' A=REV_VIDEO C=blue
#19 @35 '1. CONTINUOUS' C=blue
#20 @35 '2. BINARY' C=blue
#21 @35 'Q. QUIT' C=blue
#32 @35 'Press' C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red ;

/* WINDOW X2SELECT ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=80 color=GRAY */
#8 @25 'PRESS [ENTER], THEN INPUT REQUIRED DATA' C=red
#15 @25 'IS MEDIATOR VARIABLE X2 CONTINUOUS OR BINARY?' C=blue
#17 @40 X2 1 A=REV_VIDEO C=blue REQUIRED=YES
#17 @41 ' A=REV_VIDEO C=blue
"Press [ENTER], THEN INPUT REQUIRED DATA." C=red

"ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B1*)' C=BLUE @

"PREVALENCE OF PRIMARY PREDICTOR X1 (f1)' C=BLUE @

"NOTE:" A=UNDERLINE

"ESTIMATE OF PREDICTOR X2 IN FULL MODEL (B2)' C=BLUE @

"VARIANCE OF MEDIATOR X2' C=BLUE @

"PRESS [ENTER], THEN INPUT REQUIRED DATA." C=red

"CORRELATION COEFFICIENT BETWEEN X1 AND X2' C=BLUE @

"VARIANCE OF PRIMARY PREDICTOR X1' C=BLUE @

"NOTE:" A=UNDERLINE

"MARGINAL PREVALENCE OF MEDIATOR X2 IS THE "

"PROPORTION OF YES' IN X2." C=red

"Press" C=red @

"to continue." C=red;

"PRESS [ENTER], THEN INPUT REQUIRED DATA." C=red

"PREVALENCE OF MEDIATOR X2 (f2)' C=BLUE @

"NOTE:" A=UNDERLINE

"ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B1*') C=BLUE @

"VARIANCE OF MEDIATOR X2' C=BLUE @

"NOTE:" A=UNDERLINE

"NOTE:" A=UNDERLINE

"PRESS [ENTER], THEN INPUT REQUIRED DATA." C=red

"PRESS [ENTER], THEN INPUT REQUIRED DATA." C=red

"PRESS [ENTER], THEN INPUT REQUIRED DATA." C=red

"PRESS [ENTER], THEN INPUT REQUIRED DATA." C=red

"PRESS [ENTER], THEN INPUT REQUIRED DATA." C=red
#21 @18 'PROPORTION OF TREATMENT EFFECT (PTE IN %)' C=BLUE @67 PTE 8 A=REV_VIDEO C=blue REQUIRED=YES
#27 @18 "NOTE: " A=UNDERLINE
#27 @24 "PTE WILL BE BETWEEN 0 AND 100, eg 5.66 FOR 5.66%.")
#32 @35 "Press" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

%WINDOW lin_6_4 ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=80 Color=GRAY
#8 @25 "PRESS [ENTER], THEN INPUT REQUIRED DATA" C=red
#13 @18 'VARIANCE OF PRIMARY PREDICTOR X1' C=BLUE @67 VarX1 8 A=REV_VIDEO C=blue REQUIRED=YES C=blue
#15 @18 'VARIANCE OF MEDIATOR X2' C=BLUE @67 VarX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#17 @18 'ESTIMATE OF MEDIATOR X2 IN FULL MODEL (B1)' C=BLUE @67 EstX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#19 @18 'ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B1)*' C=BLUE @67 PreEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#21 @18 'ESTIMATE OF X1 AFTER THE ADJUSTMENT OF X2 (B1)*' C=BLUE @67 PostEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#27 @18 "NOTE:" A=UNDERLINE
#27 @24 "PTE WILL BE BETWEEN 0 AND 100, eg 5.66 FOR 5.66%.")
#32 @35 "Press" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

%WINDOW lin_6_3 ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=80 Color=GRAY
#8 @25 "PRESS [ENTER], THEN INPUT REQUIRED DATA" C=red
#13 @18 'VARIANCE OF PRIMARY PREDICTOR X1' C=BLUE @67 VarX1 8 A=REV_VIDEO C=blue REQUIRED=YES C=blue
#15 @18 'VARIANCE OF MEDIATOR X2' C=BLUE @67 VarX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#17 @18 'ESTIMATE OF MEDIATOR X2 IN FULL MODEL (B2)' C=BLUE @67 EstX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#19 @18 'ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B2)*' C=BLUE @67 PreEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#21 @18 'ESTIMATE OF X1 AFTER THE ADJUSTMENT OF X2 (B2)*' C=BLUE @67 PostEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#27 @18 "NOTE:" A=UNDERLINE
#27 @24 "PTE WILL BE BETWEEN 0 AND 100, eg 5.66 FOR 5.66%.")
#32 @35 "Press" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

%WINDOW lin_6_2 ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=80 Color=GRAY
#8 @25 "PRESS [ENTER], THEN INPUT REQUIRED DATA" C=red
#13 @18 'VARIANCE OF PRIMARY PREDICTOR X1' C=BLUE @67 VarX1 8 A=REV_VIDEO C=blue REQUIRED=YES C=blue
#15 @18 'VARIANCE OF MEDIATOR X2' C=BLUE @67 VarX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#17 @18 'ESTIMATE OF MEDIATOR X2 IN FULL MODEL (B2)' C=BLUE @67 EstX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#19 @18 'ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B2)*' C=BLUE @67 PreEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#21 @18 'ESTIMATE OF X1 AFTER THE ADJUSTMENT OF X2 (B2)*' C=BLUE @67 PostEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#27 @18 "NOTE:" A=UNDERLINE
#27 @24 "PTE WILL BE BETWEEN 0 AND 100, eg 5.66 FOR 5.66%.")
#32 @35 "Press" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

%WINDOW lin_6_1 ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=80 Color=GRAY
#8 @25 "PRESS [ENTER], THEN INPUT REQUIRED DATA" C=red
#13 @18 'VARIANCE OF PRIMARY PREDICTOR X1' C=BLUE @67 VarX1 8 A=REV_VIDEO C=blue REQUIRED=YES C=blue
#15 @18 'VARIANCE OF MEDIATOR X2' C=BLUE @67 VarX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#17 @18 'ESTIMATE OF MEDIATOR X2 IN FULL MODEL (B2)' C=BLUE @67 EstX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#19 @18 'ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B1)*' C=BLUE @67 PreEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#21 @18 'ESTIMATE OF X1 AFTER THE ADJUSTMENT OF X2 (B1)*' C=BLUE @67 PostEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#27 @18 "NOTE:" A=UNDERLINE
#27 @24 "PTE WILL BE BETWEEN 0 AND 100, eg 5.66 FOR 5.66%.")
#32 @35 "Press" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

%WINDOW lin_6 ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=80 Color=GRAY
#8 @25 "PRESS [ENTER], THEN INPUT REQUIRED DATA" C=red
#13 @18 'VARIANCE OF PRIMARY PREDICTOR X1' C=BLUE @67 VarX1 8 A=REV_VIDEO C=blue REQUIRED=YES C=blue
#15 @18 'VARIANCE OF MEDIATOR X2' C=BLUE @67 VarX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#17 @18 'ESTIMATE OF MEDIATOR X2 IN FULL MODEL (B2)' C=BLUE @67 EstX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#19 @18 'ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B2)*' C=BLUE @67 PreEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#21 @18 'ESTIMATE OF X1 AFTER THE ADJUSTMENT OF X2 (B2)*' C=BLUE @67 PostEstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#27 @18 "NOTE:" A=UNDERLINE
#27 @24 "PTE WILL BE BETWEEN 0 AND 100, eg 5.66 FOR 5.66%.")
#32 @35 "Press" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;
"PRESS [ENTER], THEN INPUT REQUIRED DATA" C=red

"MARGINAL PREVALENCE OF PRIMARY PREDICTOR X1 (f1)" C=BLUE @67 f1 8 A=REV_VIDEO
C=blue REQUIRED=YES C=blue

"MARGINAL PREVALENCE OF MEDIATOR X2 (f2)" C=BLUE @67 f2 8 A=REV_VIDEO C=blue
REQUIRED=YES

"ESTIMATE OF PREDICTOR X2 IN FULL MODEL (B2)" C=BLUE @67 EstX2 8 A=REV_VIDEO C=blue
REQUIRED=YES

"ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B1)" C=BLUE @67 PreEstX1 8 A=REV_VIDEO
C=blue REQUIRED=YES

"PROPORTION OF TREATMENT EFFECT (PTE IN %)" C=BLUE @67 PTE 8 A=REV_VIDEO C=blue
REQUIRED=YES

"NOTE:" A=UNDERLINE

"PTE WILL BE BETWEEN 0 AND 100, eg 5.66 FOR 5.66%.

"PRESS" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

"MARGINAL PREVALENCE PRIMARY PREDICTOR X1 (f1)" C=BLUE @67 PreEstX1 8 A=REV_VIDEO
C=blue REQUIRED=YES C=blue

"ESTIMATE OF X1 AFTER THE ADJUSTMENT OF X2 (B1)" C=BLUE @67 EstX1 8 A=REV_VIDEO
C=blue REQUIRED=YES

"ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B1)" C=BLUE @67 PreEstX1 8 A=REV_VIDEO
C=blue REQUIRED=YES

"PRESS" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

"MARGINAL PREVALENCE MEDIATOR X2 (f2)" C=BLUE @67 VarX2 8 A=REV_VIDEO C=blue
REQUIRED=YES C=blue

"VARIANCE OF MEDIATOR X2" C=BLUE @67 VarX2 8 A=REV_VIDEO C=blue
REQUIRED=YES

"VARIANCE OF PRIMARY PREDICTOR X1" C=BLUE @67 VarX1 8 A=REV_VIDEO C=blue
REQUIRED=YES C=blue

"PRESS" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

"PRESS" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

"PRESS" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

"PRESS" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;
#17 @18 'ESTIMATE OF PREDICTOR X2 IN FULL MODEL (B2)' C=BLUE @67 EstX2 8 A=REV_VIDEO C=blue REQUIRED=YES
#19 @18 'ESTIMATE OF X1 AFTER THE ADJUSTMENT OF X2 (B1)' C=BLUE @67 EstX1 8 A=REV_VIDEO C=blue REQUIRED=YES
#21 @18 'ESTIMATE OF X1 BEFORE THE ADJUSTMENT OF X2 (B1*)' C=BLUE @67 PreEstX1 8 A=REV_VIDEO AUTOSkip=NO C=blue REQUIRED=YES
#32 @35 "Press" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

%WINDOW ERRORWIN_2 ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=80 color=GRAY
#8 @40 "INCORRECT INPUT!!" C=red

#15 @18 "PLEASE ENTER 1 FOR 'CONTINUOUS' , 2 FOR 'BINARY' OR Q TO 'QUIT'. " C= BLUE
#32 @35 "Press" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

%WINDOW ERRORWIN_3 ICOLUMN=5 IROW=5 ROWS=50 COLUMNS=83 color=GRAY
#8 @40 "INCORRECT INPUT!!" C=red

#17 @20 "PLEASE ENTER Y FOR 'YES' , N FOR 'NO' or Q FOR QUIT. " C= BLUE
#32 @35 "Press" C=red @43 "ENTER" A=underline C=red @50 "to continue." C=red;

%let X1=; %let X2=; %let p=0.5; %let type1=0.05; %let power=0.80;
%let ResVar=1.00; %let VarX1=1.00; %let VarX2=1.00;
%let VarX1YN=; %let VarX2YN=; %let EstX1YN=;%let f1=0.50; %let f2=0.50;
%let EstX1=; %let EstX2=; %let PTE=;%let PreEstX1=; %let Corr=;

/* inner macro */

%macro lin_samp;

data sampsize;type1=&type1;power=&power;zalpha=quantile('NORMAL',1-type1/2);
zbeta=quantile('NORMAL',power);ResVar=&ResVar;
%if &X1=1 and &X2=1 %then %do;VarX1=&VarX1;VarX2=&VarX2;
label VarX1="Variance of Primary Predictor X1";
%end;
%else %if &X1=2 and &X2=1 %then %do;VarX1=&f1*(1-&f1);VarX2=&VarX2;
label f1="Marginal Prevalance of Primary Predictor X1 (f1)";label VarX2="Variance of Mediator X2";
%end;
%else %if &X1=1 and &X2=2 %then %do;f2=&f2;VarX1=&VarX1;VarX2=f2*(1-f2);
label VarX1="Variance of Primary Predictor X1";
label f2="Marginal Prevalance of Mediator X2(f2)";%end;
%else %if &X1=2 and &X2=2 %then %do;
f1=&f1;f2=&f2;VarX1=f1*(1-f1);VarX2=f2*(1-f2);
label f1="Marginal Prevalance of Primary Predictor X1 (f1)";
label f2="Marginal Prevalance Mediator X2(f2)";%end;

/* method1 */
%if %upcase(&VarX1YN)=N %then %do; method='method1';EstX2=&EstX2;Corr=&Corr;
bottom=((EstX2)**2*(VarX2)*((1-(Corr)**2)));
label Corr="Pearson's Correlation between X1 and X2";
label EstX2="Estimate of Predictor X2 (B2)";drop varX1;%goto sampsize;%end;

/* method2: */
%if %upcase(&VarX1YN)=Y and %upcase(&VarX2YN)=N %then %do;
method='method2';PreEstX1=&PreEstX1;Corr=&Corr;PTE=&PTE;
Bottom=(PreEstX1*PTE/100*(1-Corr)**2)*2*VarX1*(1-(Corr)**2);
label Corr="Pearson's Correlation between X1 and X2";
lable PreEstX1="Estimate of X1 before the adjustment of X2 (B1*)";
lable PTE="Proportion of Treatment Effect(PTE)";drop varX2 ;%goto sampsize;
/*method 3*/
%if %upcase(&VarX1YN)=Y and %upcase(&VarX2YN)=Y and %upcase(&EstX1YN)=N %then %do;
method='method3';EstX2=&EstX2;PreEstX1=&PreEstX1;PTE=&PTE;
Bottom=(EstX2)**2*(VarX2)-(PreEstX1*PTE/100)**2*VarX1;
lbl PreEstX1="Estimate of X1 before the adjustment of X2 (B1*)";
lbl EstX2="Estimate of Predictor X2 (B2)"
label PTE="Proportion of Treatment Effect(PTE)";
%goto sampsize;
%end;

/*method 4*/
%if %upcase(&VarX1YN)=Y and %upcase(&VarX2YN)=Y and %upcase(&EstX1YN)=Y %then %do;
method='method4';EstX1=&EstX1;EstX2=&EstX2;PreEstX1=&PreEstX1;
Bottom=(EstX2)**2*(VarX2)-(PreEstX1-EstX1)**2*VarX1;
lbl EstX1="Estimate of Predictor X2 (B1)";
lbl EstX2="Estimate of Predictor X2 (B2)"
label PreEstX1="Estimate of X1 before the adjustment of X2 (B1*)";
%goto sampsize;
%end;

%sampsize:
n=ceil(((zalpha+zbeta)**2*(ResVar))/Bottom); /* round to next integer */
label ResVar= "Residual Variance of Outcome Y";
lbl type1="TYPE I Error";lbl power="Power";lbl n="Sample Size (n)"
run;

proc transpose data=sampsize(drop=Bottom zalpha zbeta)out=sampsize2;run;
ods rtf  style=SASWeb  bodytitle; title 'Sample Size Calculation for Linear Model';
footnote1 j=l 'Source : Vittinghoff E, Sen S, McCulloch CE . Sample size calculation for evaluating mediation.';
footnote2 j=l 'Statistics in Medicine 2009;28: 541-557';
proc print data=sampsize2 label noobs;label _NAME_="Variable Name";
label _LABEL_="Description of Variable";label COL1="Value";run;
ods rtf close;
%mend lin_samp;

/* outer macro */
%macro med_samplesize;
%display start ;

%first:
%let X1=; %let X2=; %let VarX1YN=; %let VarX2YN=; %let EstX1YN=

%next1:
%display X1SELECT;
%if %upcase(&X1)=Q %then %do; %abort;%end;
%if %upcase(&X1)^=1 AND %upcase(&X1)^=2 AND %upcase(&X1)^=Q %then %do;
%display ERRORWIN_2;%let X1=;%goto next1;%end;

%next2:
%display X2SELECT;
%if %upcase(&X2)=Q %then %do; %abort;%end;
%if %upcase(&X2)^=1 AND %upcase(&X2)^=2 AND %upcase(&X2)^=Q %then %do;
%display ERRORWIN_2;%let X2=;%goto next2;%end;

%let ResVar=1.00; %let PreEstX1=0.53; %let EstX1=0.50; %let EstX2=0.10;
%let Corr=0.30; %let f1=0.50; %let f2=0.50; %let PTE=5.66;
%if &X1=2 %then %do; %let PTE=11.30; %end;
%if &X2=2 %then %do;%let EstX2=0.20;%end;
%display lin_1;
%second:
%let VarX1YN=;
%if &X1=1 %then %do; %display lin_2; %end;
%else %if &X1=2 %then %do; %display lin_2_1; %end;
%end;
%if %upcase(&VarX1YN)=N %then %do;
%if &X2=1 %then %do; %display lin_3; %end;
%else %if &X2=2 %then %do; %display lin_3_1; %end;
%end;
%else %if %upcase(&VarX1YN)=Y %then %do;
%third:
%let VarX2YN=;
%if &X2=1 %then %do; %display lin_4; %end;
%else %if &X2=2 %then %do; %display lin_4_1; %end;
%if %upcase(&VarX2YN)=N %then %do;
%if &X1=1 %then %do; %display lin_5; %end;
%else %if &X1=2 %then %do; %display lin_5_2; %end;
%end;
%else %if %upcase(&VarX2YN)=Y %then %do;
%fourth:
%let EstX1YN=; %display lin_6;
%if %upcase(&EstX1YN)=N %then %do;
%if &X1=1 and &X2=1 %then %do; %display lin_6_1; %end;
%else %if &X1=2 and &X2=1 %then %do; %display lin_6_2; %end;
%else %if &X1=1 and &X2=2 %then %do; %display lin_6_3; %end;
%else %if &X1=2 and &X2=2 %then %do; %display lin_6_4; %end;
%end;
%else %if %upcase(&EstX1YN)=Y %then %do;
%if &X1=1 and &X2=1 %then %do; %display lin_8_1; %end;
%else %if &X1=2 and &X2=1 %then %do; %display lin_8_2; %end;
%else %if &X1=1 and &X2=2 %then %do; %display lin_8_3; %end;
%else %if &X1=2 and &X2=2 %then %do; %display lin_8_4; %end;
%end;
%else %if %upcase(&EstX1YN)=Q %then %do; %abort; %end;
%else %if %upcase(&EstX1YN)^=Y AND %upcase(&EstX1YN)^=N AND %upcase(&EstX1YN)^=Q %then %do;
%display ERRORWIN_3; %goto fourth; %end;
%end;
%else %if %upcase(&VarX2YN)=Q %then %do; %abort; %end;
%else %if %upcase(&VarX2YN)^=Y AND %upcase(&VarX2YN)^=N AND %upcase(&VarX2YN)^=Q %then %do;
%display ERRORWIN_3; %goto third; %end;
%end;
%else %if %upcase(&VarX1YN)=Q %then %do; %abort; %end;
%else %if %upcase(&VarX1YN)^=Y AND %upcase(&VarX1YN)^=N AND %upcase(&VarX1YN)^=Q %then %do;
%display ERRORWIN_3; %goto second; %end;
%lin samp
%mend med_samplesize;
%med_samplesize