VCBD: A SAS Macro for Variance Components for Nested Binary Responses (Version 1.0)

Bahjat Qaqish and Habib Moalem (1994)

Department of Biostatistics, CB 7420, University of North Carolina
Chapel Hill, NC 27599-7420, U.S.A.

1. Introduction

VCBD is a SAS macro for fitting a variance-components model to binary responses with three levels of nesting. This document is a brief description of the model, how to use the macro and a sample of the macro's output.

Suppose that there are K medical practices, with n_i providers (or physicians) in the i-th practice and m_{ij} patients seen by the j-th provider in the i-th practice, and let Y_{ijk} denote the outcome of the k-th patient seen by the j-th provider in the i-th practice. The outcome of each patient is a binary outcome where 1 denotes "yes" and 0 denotes "no". Let Y_{ij} denote the number of "yes" outcomes out of the m_{ij} patients.

Let α_i denote a random effect associated with the i-th practice, and assume that

$$E[\alpha_i] = \theta$$

and

$$\text{var}(\alpha_i) = \theta(1-\theta)\lambda_1,\ 0 \le \lambda_1 \le 1.$$

Also let β_{ij}, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$, denote a random effect associated with the j-th provider in the i-th practice and assume that

$$E[\beta_{ij} \mid \alpha_i] = \alpha_i,$$

and

$$\text{var}(\beta_{ij} \mid \alpha_i) = \alpha_i(1-\alpha_i)\lambda_2,\ 0 \le \lambda_2 \le 1.$$

Moreover, assume

$$E[Y_{ijk} \mid \alpha_i, \beta_{ij}] = \beta_{ij}.$$

It is assumed that $\alpha_1, \cdots, \alpha_K$ are independent; that $\beta_{i1}, \cdots, \beta_{in_i}$ are conditionally independent given α_i and that $Y_{ij1}, \cdots, Y_{ijm_{ij}}$ are conditionally independent given β_{ij}.

The interpretation is roughly that θ is the average true performance of all practices, α_i is the unobservable true performance of the i-th practice, β_{ij} is the unobservable true performance of the j-th provider in the i-th practice.

The total variation $\mathrm{var}(Y_{ijk}) = \theta(1-\theta)$ is decomposed into three components:

i- The within-provider component of variation given by

$$\theta(1-\theta)(1-\lambda_1)(1-\lambda_2).$$

ii- The between-provider within-practice component of variation given by

$$\theta(1-\theta)(1-\lambda_1)\lambda_2.$$

iii- The between-practice component of variation given by

$$\theta(1-\theta)\lambda_1.$$

The correlation between two responses within the same provider is given by

$$\mathrm{corr}(Y_{ijk}, Y_{ijk'}) = \lambda_1 + \lambda_2 - \lambda_1\lambda_2.$$

The correlation between providers within practice is given by

$$\mathrm{corr}(Y_{ijk}, Y_{ij'k'}) = \lambda_1.$$

2. Overview of the macro VCBD

The macro uses SAS Summary procedure to calculate sums of squares and a few data steps to estimate θ, λ_1, and λ_2 as well as the three components of variance and the two correlations, by equating the sums of squares to their expected values. Missing observations are allowed, but the analysis will be valid only if they are missing at random.

To facilitate the use of VCBD, the SAS Macro Language has been used for passing parameters. Before this macro can be used in a SAS program, either it has to be directly entered as part of the program or, more conveniently, accessed by an command to call the macro is:

$$\%\mathrm{vcbd}(\mathrm{data} =, \mathrm{i} =, \mathrm{mij} =, \mathrm{yij} =)$$

Parameters may be given in any order and are explained below.

data= This parameter specifies the name of the SAS datset. It is optional. The most recently created SAS dataset will be used if no dataset is specified.

i= This parameter specifies the practice ID. It is required. The practice ID can be a numeric or string variable or a list of variables.

mij= This parameter specifies the number of patients seen by the j-th provider in the i-th practice. It is required and must be a numeric variable.

yij= This parameter specifies the number of "yes" or 1 responses among patients seen by the j-th provider in the i-th practice. It is required and must be a numeric variable. Clearly, y_{ij} must be between 0 and m_{ij}.

Note: If the original dataset contains 0/1 response variables, proc Summary can be used to create the summary variables mij and yij required by the macro. This is illustrated in the example below.

3. Example

```
*****************************************************************;;
filename MAC "vcbd.sas";
%include MAC;
*****************************************************************;;
data A;
    input practice provider response;
 cards;
    1    1     1
    1    1     0
    1    1     0
    1    1     0
    1    2     0
    1    2     0
    1    2     0
    1    2     0
    2    1     0
    2    1     1
```

```
    2    1      1
    2    1      1
    3    3      0
    3    3      1
    3    3      1
    3    4      1
    3    4      1
    3    4      1
    3    7      1
    3    7      0
    3    7      0
    3    7      0
    4    4      1
    4    4      1
    4    4      1
    4    4      1
    4    8      1
    4    8      1
    4    8      1
    5    1      1
    5    1      0
    5    1      1
    5    2      1
    5    3      0
    5    3      1
    5    4      1
   ;
   run;
*******************************************************************;;
proc summary data=A nway;
  class practice provider;
  var  response;
  output out=B sum=y n=m;
  run;
*******************************************************************;;
proc print data=B;      * take a look;
  var practice    provider   y m;
  run;
*******************************************************************;;
%vcbd (data = B, i = practice, mij = m, yij = y);
```

Output:

```
Data set: B
Response: y / m
Practice ID: practice

Total number of practices:  5
Total number of physicians: 12
Total number of patients:   36
Sum of y: 22

Estimates:
    Theta:    0.6111111111
    Lambda_1: 0.3108445033
    Lambda_2: 0.0528424377

Variance components:
    Within provider: 65.27388403 %
    Between provider within practice: 3.6416656409 %
    Between practice: 31.084450329 %

Correlation:
    Within provider: 0.3472611597
    Between provider within practice: 0.3108445033
```

Obs	practice	provider	y	m
1	1	1	1	4
2	1	2	0	4
3	2	1	3	4
4	3	3	2	3
5	3	4	3	3
6	3	7	1	4
7	4	4	4	4
8	4	8	3	3
9	5	1	2	3
10	5	2	1	1
11	5	3	1	2
12	5	4	1	1

REFERENCES

Manton, K. G., Woodbury, M. A. & Stallard, E. (1981). A variance components appraoch to categorical data models with heterogeneous cell populations: analysis of spatial gradients in lung cancer mortality rates in North Carolina counties. *Biometrics* **37**, 259–269.

Han, C-P. (1978). Nonnegative and preliminary test estimators of variance components. *J. Am. Statist. Assoc.* **73**, 855–858.

Anderson, D. A. & Aitkin, M. (1985). Variance component models with binary response. *J. R. Statist. Soc.* B **47**, 203–210.