

# **Chapter 36**

## **Variance Components for Assessing the Magnitude of Random Effects (40 Patients)**

### **General Purpose**

If we have reasons to believe that in a study certain patients due to co-morbidity, co-medication and other factors will respond differently from others, then the spread in the data is caused not only by residual effect, but also by some subgroup property, otherwise called some random effect. Variance components analysis is able to assess the magnitudes of random effects as compared to that of the residual error of a study.

### **Primary Scientific Question**

Can a variance components analysis by including the random effect in the analysis reduce the unexplained variance in a study, and, thus, increase the accuracy of the analysis model as used.

---

This chapter was previously published in “Machine learning in medicine-cookbook 3” as Chap. 3, 2014.

### Example

Variables			
PAT	treat	gender	cad
52,00	,00	,00	2,00
48,00	,00	,00	2,00
43,00	,00	,00	1,00
50,00	,00	,00	2,00
43,00	,00	,00	2,00
44,00	,00	,00	1,00
46,00	,00	,00	2,00
46,00	,00	,00	2,00
43,00	,00	,00	1,00
49,00	,00	,00	2,00
28,00	1,00	,00	1,00
35,00	1,00	,00	2,00

PAT=episodes of paroxysmal atrial tachycardias  
 treat=treatment modality (0=placebo treatment, 1=active treatment)  
 gender=gender (0=female)  
 cad=presence of coronary artery disease (1 no, 2=yes)

The first 12 of a 40 patient parallel-group study of the treatment of paroxysmal tachycardia with numbers of episodes of PAT as outcome is given above. The entire data file is in “variancecomponents”, and is available at [extras.springer.com](http://extras.springer.com). We had reason to believe that the presence of coronary artery disease would affect the outcome, and, therefore, used this variable as a random rather than fixed variable. SPSS statistical software was used for data analysis. Start by opening the data file in SPSS.

**Command:**

Analyze...General Linear Model...Variance Components...Dependent Variable: enter "paroxtachyc"...Fixed Factor(s): enter "treat, gender"...Random Factor(s): enter "corartdisease"...Model: mark Custom...Model: enter "treat, gender, cad"...click Continue...click Options...mark ANOVA...mark Type III...mark Sums of squares...mark Expected mean squares...click Continue...click OK.

The output sheets are given underneath. The Variance Estimate table gives the magnitude of the Variance due to cad, and that due to residual error (unexplained variance, otherwise called Error). The ratio of the  $\text{Var}(\text{cad}) / [\text{Var}(\text{Error}) + \text{Var}(\text{cad})]$  gives the proportion of variance in the data due to the random cad effect ( $5.844 / (28.426 + 5.844) = 0.206 = 20.6\%$ ). This means that 79.4 % instead of 100 % of the error is now unexplained.

Variance estimates	
Component	Estimate
Var(cad)	5,844
Var(Error)	28,426

Dependent variable: paroxtach  
 Method: ANOVA (Type III sum of squares)

The underneath ANOVA table gives the sums of squares and mean squares of different effects. E.g. the mean square of cad=139.469, and that of residual effect=28.426.

ANOVA			
Source	Type III sum of squares	df	Mean square
Corrected model	727,069	3	242,356
Intercept	57153,600	1	57153,600
treat	515,403	1	515,403
gender	,524	1	,524
cad	139,469	1	139,469
Error	1023,331	36	28,426
Total	58904,000	40	
Corrected total	1750,400	39	

Dependent variable: paroxtach

The underneath Expected Mean Squares table gives the results of a special procedure, whereby variances of best fit quadratic functions of the variables are minimized to obtain the best unbiased estimate of the variance components. A little mental arithmetic is now required.

Expected mean squares			
Source	Variance component		
	Var(cad)	Var(Error)	Quadratic term
Intercept	20,000	1,000	Intercept, treat, gender
treat	,000	1,000	treat
gender	,000	1,000	gender
cad	19,000	1,000	
Error	,000	1,000	

Dependent variable: paroxtach

Expected mean squares are based on Type III sums of squares

For each source, the expected mean square equals the sum of the coefficients in the cells times the variance components, plus a quadratic term involving effects in the Quadratic Term cell

$$\begin{aligned} \text{EMS (expected mean square) of cad (the random effect)} \\ &= 19 \times \text{Variance (cad)} + \text{Variance (Error)} \\ &= 139.469 \end{aligned}$$

$$\begin{aligned} \text{EMS of Error (the residual effect)} \\ &= 0 + \text{Variance (Error)} \\ &= 28.426 \end{aligned}$$

$$\begin{aligned} \text{EMS of cad - Variance (Error)} \\ &= 19 \times \text{Variance (cad)} \\ &= 139.469 - 28.426 \\ &= 110.043 \end{aligned}$$

$$\begin{aligned} \text{Variance (cad)} \\ &= 110.043 / 19 \\ &= 5.844 \text{ (compare with the results of the above Variance Estimates table)} \end{aligned}$$

It can, thus, be concluded that around 20 % of the uncertainty is in the data is caused by the random effect.

## Conclusion

If we have reasons to believe that in a study certain patients due to co-morbidity, co-medication and other factors will respond differently from others, then the spread in the data will be caused, not only by the residual effect, but also by the subgroup property, otherwise called the random effect. Variance components analysis, by including the random effect in the analysis, reduces the unexplained variance in a study, and, thus, increases the accuracy of the analysis model used.

## Note

More background, theoretical and mathematical information of random effects models are given in Machine learning in medicine part three, Chap. 9, Random effects, pp 81–94, 2013, Springer Heidelberg Germany, from the same authors.