ScienceDirect

Covariance Analysis

Related terms:

Data Assimilation, Meiofauna, Protist, Suspended Sediment, Analysis of Variance, Binomial, Covariance, Dragonfly

Toxicology Testing and Evaluation

S.C. Gad, in Comprehensive Toxicology, 2010

3.13.3.4 Analysis of Covariance

Analysis of covariance (ANCOVA) is a method for comparing sets of data that consist of two variables (treatment and effect, with the effect variable being called the *variate*), when a third variable (called the covariate) exists that can be measured but not controlled and that has a definite effect on the variable of interest. In other words, it provides an indirect type of statistical control, allowing the precision of a study to be increased and potential source of bias to be removed. One common example of this is in the analysis of organ weights in toxicity studies. The interest here is the effect of the dose or exposure level on the specific organ weights; however, most organ weights also increase (in the young, growing animals most commonly used in such studies) in proportion to increases in animal body weight. Because the effect of this covariate (body weight) is not of interest, it is measured only to allow adjustment of the measurement of the variate of concern (the organ weights). ANCOVA allows this adjustment to be made. Care must be taken before using ANCOVA, however, to ensure that the underlying nature of the correspondence between the variate and the covariate is such that it can be relied on as a tool for adjustments (Harris 1975; Litchfield and Wilcoxon 1949).

The underlying assumptions for ANCOVA are fairly rigid and restrictive. These assumptions are as follows:

1. the regression slopes of Y and X are equal from group to group;
2. the relationship between X and Y is linear;
3. the covariate X is measured without error;
4. there are no unmeasured confounding variables;
5. the errors inherent in each variable are independent of each other;
6. the variances of the errors in groups are equal between groups;
7. the measured data that form the groups are normally distributed.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780080468846003201

Statistical Methods in the Atmospheric Sciences

D.S. Wilks, in International Geophysics, 2011

13.4 Maximum Covariance Analysis (MCA)

Maximum covariance analysis (MCA) is a similar technique to CCA, in that it finds pairs of linear combinations of two sets of vector data $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\left.\begin{aligned} v_m &= \boldsymbol{\ell}_m^T \mathbf{x} \\ w_m &= \mathbf{r}_m^T \mathbf{y} \end{aligned}\right\}, m = 1, \cdots, M; \tag{13.39}$$

such that their covariances

$$\mathrm{Cov}(v_m, w_m) = \boldsymbol{\ell}_m^T [S_{x,y}]\, \mathbf{r}_m \tag{13.40}$$

(rather than their correlations, as in CCA) are maximized, subject to the constraint that the vectors $\boldsymbol{\ell}_m$ and $\mathbf{r}_m$ are orthonormal. As in CCA, the number of such pairs $M = \min(I, J)$ is equal to the smaller of the dimensions of the data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, and each succeeding pair of projection vectors is chosen to maximize covariance, subject to the orthonormality constraint. In a typical application to atmospheric data, $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are both time series of spatial fields, and so their projections in Equation 13.39 form time series also.

Computationally, the vectors $\boldsymbol{\ell}_m$ and $\mathbf{r}_m$ are found through a singular value decomposition (Equation 10.68) of the matrix $[S_{x,y}]$ in Equation 13.1, containing the cross-covariances between the elements of $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\underset{(\mathbf{I}\ \mathbf{x}\ \mathbf{J})}{[S_{x,y}]} = \underset{(\mathbf{I}\ \mathbf{x}\ \mathbf{J})}{[L]}\ \underset{(\mathbf{J}\ \mathbf{x}\ \mathbf{J})}{[\Omega]}\ \underset{(\mathbf{J}\ \mathbf{x}\ \mathbf{J})}{[R]^T}. \tag{13.41}$$

The left singular vectors $\boldsymbol{\ell}_m$ are the columns of the matrix $[L]$, and the right singular vectors $\mathbf{r}_m$ are the columns of the matrix $[R]$ (i.e., the rows of $[R]^T$). The elements ω_m of the diagonal matrix $[\Omega]$ of singular values are the maximized covariances (Equation 13.40) between the pairs of linear combinations in Equation 13.39. Because the machinery of the singular value decomposition is used to find the vectors $\boldsymbol{\ell}_m$ and $\mathbf{r}_m$, and the associated covariances ω_m, maximum covariance analysis sometimes unfortunately is known as SVD analysis; although as illustrated earlier in this chapter and elsewhere in this book, the singular value decomposition has a rather broader range of uses. In recognition of the parallels with CCA, the technique is also sometimes called *canonical covariance analysis* and the ω_m are sometimes called the canonical covariances.

There are two main distinctions between CCA and MCA. The first is that CCA maximizes correlation, whereas MCA maximizes covariance. The leading CCA modes may capture relatively little of the corresponding variances (and thus yield small covariances even if the canonical correlations are high). On the other hand, maximum covariance analysis will find linear combinations with large covariances, which may result more from large variances than a large correlation. The second difference is that the vectors $\boldsymbol{\ell}_m$ and $\mathbf{r}_m$ in maximum covariance analysis are orthogonal, and the projections v_m and w_m of the data onto them are in general correlated, whereas the canonical variates in CCA are uncorrelated but the corresponding canonical vectors are not generally orthogonal. However, it is not unusual to find similar results for CCA and MCA applied to the same data sets (e.g., Feddersen et al., 1999; Wilks, 2008). Bretherton et al. (1992), Cherry (1996), Tippett et al. (2008), van den Dool (2007), and Wallace et al. (1992) compare the two methods in greater detail.

Example 13.4 Maximum Covariance Analysis of the January 1987 Temperature Data

Singular value decomposition of the cross-covariance submatrix $[S_{x,y}]$ in Equation 13.31c yields

$$\begin{bmatrix} 58.07 & 51.70 \\ 81.63 & 110.8 \end{bmatrix} = \begin{bmatrix} .4876 & .8731 \\ .8731 & -.4876 \end{bmatrix} \begin{bmatrix} 157.4 & 0 \\ 0 & 14.06 \end{bmatrix} \begin{bmatrix} .6325 & .7745 \\ .7745 & -.6325 \end{bmatrix}. \tag{13.42}$$

The results are qualitatively similar to the CCA of the same data in Example 13.1. The first left and right vectors, $\boldsymbol{\ell}_1 = [.4876, .8731]^T$ and $\mathbf{r}_1 = [.6325, .7745]^T$, respectively, resemble the first pair of canonical vectors $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$ in Example 13.1 in that both put positive weights on both variables in both data sets. But here the weights are closer in magnitude and emphasize the minimum temperatures rather than the maximum temperatures. The covariance between the linear combinations defined by these vectors is 157.4, which is larger than the covariance between any other pair of linear combinations for these data, subject to $||\boldsymbol{\ell}_1|| = ||\mathbf{r}_1|| = 1$. The corresponding correlation is

$$\begin{aligned} \mathrm{Corr}(v_1, w_1) &= \frac{\omega_1}{(\mathrm{Var}(v_1)\mathrm{Var}(w_1))^{1/2}} = \frac{\omega_1}{\left(\boldsymbol{\ell}_1^T [S_{x,x}]\, \boldsymbol{\ell}_1\right)^{1/2} \left(\mathbf{r}_1^T [S_{y,y}]\, \mathbf{r}_1\right)^{1/2}} \\ &= \frac{157.44}{(219.8)^{1/2}(126.3)^{1/2}} = 0.945 \end{aligned}, \tag{13.43}$$

which is large, but necessarily smaller than $r_{C1} = 0.969$ for the CCA of the same data.

The second pair of vectors, $\boldsymbol{\ell}_2 = [.8731, -.4876]^T$ and $\mathbf{r}_2 = [.7745. -.6325]^T$, are also similar to the second pair of canonical vectors for the CCA in Example 13.1, in that they also describe a contrast between the maximum and minimum temperatures that can be interpreted as being related to the diurnal temperature ranges. The covariance of the second pair of linear combinations is ω_2, corresponding to a correlation of 0.772. This correlation is slightly larger than the second canonical correlation in Example 13.1, but has not been limited by the CCA constraint that the correlations between v_1 and v_2, and w_1 and w_2 must be zero. ◇

The results of a MCA can be used to forecast one of the fields, say $\mathbf{y}$, using the $\mathbf{x}$ field as the predictor, similarly to the CCA forecasts described in Section 13.2.3. If the projection variables in Equation 13.39 have been computed from anomaly vectors $\mathbf{x}'$ and $\mathbf{y}'$, then the individual regressions will have zero intercept and be of the form

$$\widehat{w}_m = \widehat{\beta}_m\, v_m, \quad m = 1, \ldots, M, \tag{13.44}$$

where the least-squares estimates of the individual regression slopes are

$$\widehat{\beta}_m = \frac{\omega_m}{\boldsymbol{\ell}_m^T [S_{x,x}] \boldsymbol{\ell}_m} \tag{13.45}$$

and the estimated regression error variance is

$$s_e^2 = \mathbf{r}_m^T [S_{y,y}]\, \mathbf{r}_m - \widehat{\beta}_m^2\, \boldsymbol{\ell}_m^T [S_{x,x}]\, \boldsymbol{\ell}_m. \tag{13.46}$$

However, since the since the projections in Equation 13.39 are not uncorrelated for different m, simultaneous application of multiple-m versions of Equation 13.44, as in Equation 13.23 for CCA, will in general not yield optimal predictions. Rather, the framework of multiple linear regression (Section 7.2.8), in which many or all of the $\boldsymbol{\ell}_m$ projections might be used as predictors for any of the $\mathbf{r}_m$ predictands would be more appropriate (Garcia-Morales and Dubus, 2007; Tippett et al., 2008).

The papers of Bretherton et al. (1992) and Wallace et al. (1992) have been influential advocates for the use of maximum covariance analysis. One advantage over CCA that sometimes is cited is that no matrix inversions are required, so that a maximum covariance analysis can be computed even if $n < \max(I, J)$. However,

both techniques are subject to similar sampling problems in limited-data situations, so it is not clear that this advantage is of practical importance. Some cautions regarding maximum covariance analysis have been offered by Cherry (1997) and Hu (1997); and Newman and Sardeshmukh (1995) emphasize that the $\boldsymbol{\ell}_m$ and $\boldsymbol{r}_m$ vectors may not represent physical modes of their respective fields, just as the eigenvectors in PCA do not necessarily represent physically meaningful modes.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780123850225000130

Multivariate Analysis of Vector Pairs

Daniel S. Wilks, in Statistical Methods in the Atmospheric Sciences (Fourth Edition), 2019

14.3.1 Definition of MCA

Maximum covariance analysis (MCA) is a similar technique to CCA, in that it finds pairs of linear combinations of two sets of vector data $\boldsymbol{x}$ and $\boldsymbol{y}$ (Equation 14.2) such that the squares of their covariances

$$\mathrm{Cov}(v_m, w_m) = \boldsymbol{a}_m^T [S_{x,y}]\, \boldsymbol{b}_m \tag{14.44}$$

(rather than their correlations, as in CCA) are maximized, subject to the constraint that the vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ are orthonormal. Maximization of squared covariance allows for the possibility that a pair of vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ may yield a negative covariance in Equation 14.44. As in CCA, the number of such pairs $M = \min(I, J)$ is equal to the smaller of the dimensions of the data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, and each succeeding pair of projection vectors are chosen according to the maximization criterion, subject to the orthonormality constraint. In a typical application to atmospheric data, $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are both time series of spatial fields, or the leading principal components of these fields, and so their projections in Equation 14.2 form time series also.

Computationally, the vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ are found through a singular value decomposition (Equation 11.72) of the matrix $[S_{x,y}]$ in Equation 14.1, containing the cross-covariances between the elements of $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\underset{(\mathbf{I}\times\mathbf{J})}{[\boldsymbol{S}_{\boldsymbol{x},\boldsymbol{y}}]} = \underset{(\mathbf{I}\times\mathbf{J})}{[\boldsymbol{A}]}\ \underset{(\mathbf{J}\times\mathbf{J})}{[\boldsymbol{\Omega}]}\ \underset{(\mathbf{J}\times\mathbf{J})}{[\boldsymbol{B}]^{\boldsymbol{T}}}. \tag{14.45}$$

The left singular vectors $\boldsymbol{a}_m$ are the columns of the matrix $[A]$ and the right singular vectors $\boldsymbol{b}_m$ are the columns of the matrix $[B]$ (i.e., the rows of $[B]^{\mathrm{T}}$). The elements ω_m of the diagonal matrix $[\Omega]$ of singular values are the covariances (Equation 14.44) between the pairs of linear combinations in Equation 14.2. Comparison of Equation 14.45 with its CCA counterpart in Equation 14.35 shows that MCA is computed on the basis of the unwhitened (not subjected to Mahalanobis transformations) data vectors $\boldsymbol{x}'$ and $\boldsymbol{y}'$, and that the projection vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ in MCA are not scaled subsequent to the SVD (cf. Equation 14.32 for CCA).

The proportions of the variances of the underlying variables represented by the projections v_m and w_m are

$$R_m^2(\boldsymbol{x}) = \frac{\boldsymbol{a}_m^T [S_{x,x}]\, \boldsymbol{a}_m}{tr([S_{x,x}])} \tag{14.46a}$$

and

$$R_m^2(\boldsymbol{y}) = \frac{\boldsymbol{b}_m^T[S_{y,y}]\,\boldsymbol{b}_m}{\mathrm{tr}([S_{y,y}])}, \tag{14.46b}$$

the numerators of which are Var(v_m) and Var(w_m), respectively. The homogeneous correlations are

$$\mathrm{Corr}(v_m, \boldsymbol{x}) = \frac{\boldsymbol{a}_m^T[S_{x,x}][D_x]^{-1}}{\left(\boldsymbol{a}_m^T[S_{x,x}]\,\boldsymbol{a}_m\right)^{1/2}} \tag{14.47a}$$

and

$$\mathrm{Corr}(w_m, \boldsymbol{y}) = \frac{\boldsymbol{b}_m^T[S_{y,y}][D_y]^{-1}}{\left(\boldsymbol{b}_m^T[S_{y,y}]\,\boldsymbol{b}_m\right)^{1/2}}, \tag{14.47b}$$

which differ from their counterparts for CCA in Equation 14.7 because the square roots of the variances of the projection variables v_m and w_m in the denominators of Equations 14.47 are not equal to 1. Similarly, the heterogeneous correlations are

$$\mathrm{Corr}(v_m, \boldsymbol{y}) = \frac{\boldsymbol{a}_m^T[S_{x,y}][D_y]^{-1}}{\left(\boldsymbol{a}_m^T[S_{x,x}]\,\boldsymbol{a}_m\right)^{1/2}} \tag{14.48a}$$

and

$$\mathrm{Corr}(w_m, \boldsymbol{x}) = \frac{\boldsymbol{b}_m^T[S_{y,x}][D_x]^{-1}}{\left(\boldsymbol{b}_m^T[S_{y,y}]\boldsymbol{b}_m\right)^{1/2}}, \tag{14.48b}$$

which correspond to Equations 14.8.

Because the machinery of the singular value decomposition is used to find the vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$, and the associated covariances ω_m, maximum covariance analysis sometimes unfortunately is known as SVD analysis. As illustrated earlier in this chapter and elsewhere in this book, the singular value decomposition has a rather broader range of uses (e.g., Golub and van Loan, 1996). In recognition of the parallels with CCA, the technique is also sometimes called *canonical covariance analysis* in which case the ω_m are called the canonical covariances. The method is also known as Co-inertia Analysis in the biology literature.

There are two main distinctions between CCA and MCA. The first is that CCA maximizes correlation, whereas MCA maximizes covariance. The leading CCA modes may capture relatively little of the corresponding variances (and thus yield small covariances even if the canonical correlations are high). On the other hand, MCA will find linear combinations with large covariances, which may result more from large variances than a large correlation. The second difference is that the vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ in maximum covariance analysis are orthogonal, and the projections v_m and w_m of the data onto them are in general correlated, whereas the canonical variates in CCA are uncorrelated but the corresponding canonical vectors are not generally orthogonal. Bretherton et al. (1992), Cherry (1996), Tippett et al. (2008), and Van den Dool (2007) compare the two methods in greater detail.

It is not unusual to find similar results for CCA and MCA applied to the same data sets. For example, Figure 14.5 shows a pair of MCA-derived homogeneous correlation patterns for winter northern Pacific SSTs (a) and corresponding 500 mb heights (b), which are both very similar to their counterparts in Figure 14.1 that were based on CCA.

Example 14.4 Maximum Covariance Analysis of the January 1987 Temperature Data

Singular value decomposition of (the transpose of) the cross-covariance submatrix $[S_{x,y}]$ in Equation 14.36c yields

$$\begin{bmatrix} 58.07 & 51.70 \\ 81.63 & 110.8 \end{bmatrix} = \begin{bmatrix} .4876 & .8731 \\ .8731 & -.4876 \end{bmatrix} \begin{bmatrix} 157.4 & 0 \\ 0 & 14.06 \end{bmatrix} \begin{bmatrix} .6325 & .7745 \\ .7745 & -.6325 \end{bmatrix}. \tag{14.49}$$

The results are qualitatively similar to the CCA of the same data in Example 14.1. The first left and right vectors, $\boldsymbol{a}_1 = [0.4876, 0.8731]^T$ and $\boldsymbol{b}_1 = [0.6325, 0.7745]^T$, respectively, resemble the first pair of canonical vectors $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$ in Example 14.1 in that both put positive weights on both variables in both data sets. Here the weights are closer in magnitude and emphasize the minimum temperatures rather than the maximum temperatures. The covariance between the linear combinations defined by these vectors is 157.4, which is larger than the covariance between any other pair of linear combinations for these data, subject to $||\ \boldsymbol{a}_1\ || = ||\ \boldsymbol{b}_1\ || = 1$. The corresponding correlation is

$$\begin{aligned}\text{Corr}(v_1, w_1) &= \frac{\omega_1}{(\text{Var}(v_1)\ \text{Var}(w_1))^{1/2}} = \frac{\omega_1}{\left(\boldsymbol{a}_1^T[S_{x,x}]\ \boldsymbol{a}_1\right)^{1/2}\left(\boldsymbol{b}_1^T[S_{y,y}]\ \boldsymbol{b}_1\right)^{1/2}} \\ &= \frac{157.44}{(219.8)^{1/2}(126.3)^{1/2}} = 0.945\end{aligned}, \tag{14.50}$$

which is large, but necessarily smaller than $r_{C_1} = 0.969$ for the CCA of the same data.

The second pair of vectors, $\boldsymbol{a}_2 = [0.8731, -0.4876]^T$ and $\boldsymbol{b}_2 = [0.7745. -0.6325]^T$, are also similar to the second pair of canonical vectors for the CCA in Example 14.1, in that they also describe a contrast between the maximum and minimum temperatures that can be interpreted as being related to the diurnal temperature ranges. The covariance of the second pair of linear combinations is ω_2, corresponding to a correlation of 0.772. This correlation is slightly larger than the second canonical correlation in Example 14.1, but has not been limited by the CCA constraint that the correlations between v_1 and v_2, and w_1 and w_2 must be zero.

The proportions of variability in the original variables captured by the MCA variables are

$$R_1^2(\boldsymbol{x}) = \frac{[.4876\ .8731]\begin{bmatrix}59.516 & 75.733\\ 75.433 & 185.467\end{bmatrix}\begin{bmatrix}.4876\\ .8731\end{bmatrix}}{59.516+185.467} = 0.897 \tag{14.51a}$$

$$R_2^2(\boldsymbol{x}) = \frac{[.8731-.4876]\begin{bmatrix}59.516 & 75.733\\ 75.433 & 185.467\end{bmatrix}\begin{bmatrix}.8731\\ -.4876\end{bmatrix}}{59.516+185.467} = 0.103 \tag{14.52b}$$

$$R_1^2(\boldsymbol{y}) = \frac{[.6325\ .7745]\begin{bmatrix}61.847 & 56.119\\ 56.119 & 77.581\end{bmatrix}\begin{bmatrix}.6325\\ .7745\end{bmatrix}}{61.847+77.581} = 0.906 \tag{14.53c}$$

and

$$R_2^2(\boldsymbol{y}) = \frac{[.7745-.6325]\begin{bmatrix}61.847 & 56.119\\ 56.119 & 77.581\end{bmatrix}\begin{bmatrix}.7745\\ -.6325\end{bmatrix}}{61.847+77.581} = 0.094 \tag{14.54d}$$

◇

Figure 14.5. Homogeneous correlation maps of (a) average winter sea-surface temperatures in the northern Pacific Ocean, and (b) hemispheric winter 500 mb heights, derived from MCA. These are very similar to the corresponding CCA result in Figure 14.1.

The papers of Bretherton et al. (1992) and Wallace et al. (1992) have been influential advocates for the use of maximum covariance analysis. One advantage over CCA that sometimes is cited is that no matrix inversions are required, so that a maximum covariance analysis can be computed even if $n < \max(I, J)$. However, both techniques are subject to similar sampling problems in limited-data situations, so it is not clear that this advantage is of practical importance, and in any case dimension reduction through use of the leading principal components is often employed. Some cautions regarding maximum covariance analysis have been offered by Cherry (1997) and Hu (1997). Newman and Sardeshmukh (1995) emphasize that the $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ vectors may not represent physical modes of their respective fields, just as the eigenvectors in PCA do not necessarily represent physically meaningful modes.

Poisson ANCOVA

Marc Kéry, in Introduction to WinBUGS for Ecologists, 2010

Publisher Summary

A Poisson analysis of covariance (ANCOVA) can be called a Poisson regression with both discrete and continuous covariates. In most practical applications, Poisson models will have several covariates and of both types. To stress the similarity with the normal linear case, we only slightly alter the inferential setting. This chapter has generalized the general linear model from the normal to the Poisson case to model the effects on grouped counts of a continuous covariate. The changes involved in doing so in WinBUGS were minor, and the inclusion of further covariates is straightforward. The Poisson ANCOVA is an important intermediate step for one"s understanding of the Poisson generalized linear mixed model.

URL: https://www.sciencedirect.com/science/article/pii/B9780123786050000156

Binomial Analysis of Covariance

Marc Kéry, in Introduction to WinBUGS for Ecologists, 2010

Publisher Summary

A binomial analysis of covariance (ANCOVA) can be specified by adding discrete and continuous covariates to the linear predictor of a binomial generalized linear model (GLM). It has been hypothesized that the black color confers a thermal advantage, and therefore, the proportion of black individuals should be greater in cooler or wetter habitats. The value of the linear predictor is again obtained by matrix multiplication of the design matrix (Xmat) and the parameter vector (beta.vec). Moving from the normal and the Poisson to a binomial ANCOVA involves only minor changes in the code of WinBUGS (and also R). Similarly, the concepts of residuals and posterior predictive distributions carry over to this class of models.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780123786050000181

Introduction to the Generalized Linear Model

Marc Kéry, in Introduction to WinBUGS for Ecologists, 2010

Publisher Summary

The unification of a large number of statistical methods such as regression, analysis of variance (ANOVA), and analysis of covariance (ANCOVA) under the umbrella of the general linear model was a big advancement for applied statistics. However, even more significant was the unification of an even wider range of statistical methods within the class of the generalized linear model. They showed that a large number of techniques previously thought of as representing quite separate types of analyses, including logistic regression, multinomial regression, Chi-square tests, log-linear models, as well as the general linear model, could all be represented as special cases of a generalized version of a linear model. The two main ideas of the GLM are that, first, a transformation of the expectation of the response E(y) is expressed as a linear combination of covariate effects rather than the expected (mean) response itself. And second, for the random part of the model, distributions other than the normal can be chosen, e.g., Poisson or binomial. Binomial, Poisson, and normal are probably the three most widely used statistical distributions in a GLM. The former two are distributions for non-negative, discrete responses and therefore suitable to describe counts. The normal is the most widely used distribution for continuous responses such as measurements. The GLM is another key concept that appears over and over again in modern applied statistics in empirical sciences such as ecology.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780123786050000132

Fitting Models Using the Bayesian Modeling Software BUGS and JAGS

Marc Kéry, J. Andrew Royle, in Applied Hierarchical Modeling in Ecology, 2016

5.6 Linear Model with Normal Response (Normal GLM): Analysis of Covariance (ANCOVA)

Returning to the illustration of common linear models, we next use BUGS to fit a linear model that underlies a technique called analysis of covariance (ANCOVA). Specifically, within a GLM with normal response, we fit to the mean tit counts the linear model underlying a fixed-effects ANCOVA with interaction effects. For this, we somewhat artificially first construct a factor that classifies the continuous covariate `forest` cover into four levels or groups, with level 1 for values between –1 and –0.5, level 2 corresponding to –0.49 and 0, etc.; see Figure 5.10(a) for the raw relationship between mean tit count and levels of the forest factor (`facFor`). Factors in BUGS must be labeled with integer numbers and not, for instance, with letters or words, and the numbering must start at 1 and end at the number of levels—i.e., have no jumps (e.g., 1, 2, 4, and 5 would cause a crash). We fit the following model in the effects and the means parameterization (see Chapter 3), where *j* indexes the four levels of the forest factor:

Figure 5.10. (a) Relationship between the mean count of great tits and the levels of the forest cover factor (`facFor`). Raw data are shown for each level of `facFor`. (b) Raw data and predicted relationship with elevation under the ANCOVA model with a least-squares fit. Colors denote the four levels of `facFor`.

$$Cmean_i \sim Normal\left(\mu_i, \sigma^2\right)$$

$$\mu_i = \alpha_{0,j} + \alpha_{1,j} * elev_i$$

```
# Generate factor and plot raw data in boxplot as function of
 factor A
facFor <- as.numeric(forest < -0.5)         # Factor level 1
facFor[forest < 0 & forest > -0.5] <- 2     # Factor level 2
facFor[forest < 0.5 & forest > 0] <- 3       # Factor level 3
facFor[forest > 0.5] <- 4                            # Facto
r level 4
```

```
table(facFor)
 # every site assigned a level OK


par(mfrow = c(1, 2), mar = c(5,5,3,2), cex.lab = 1.5, cex.axi
s = 1.5)
plot(Cmean ~ factor(facFor), col = c("red", "blue", "green",
"grey"), xlab = "Forest cover class", ylab = "Mean count of g
reat tits", frame.plot = F, ylim = c(0,20))


# Bundle data
win.data <- list(Cmean = Cmean, M = length(Cmean), elev = ele
v, facFor = facFor)
```

We can define the model in the effects or the means parameterization, and we show both. In either case, we define vector-valued parameters using the handy *nested indexing* in the BUGS language. We will fit the model in WinBUGS, JAGS, and compare with the MLEs obtained by using the least-squares method by way of the function `lm` in R.

```
# Specify model in BUGS language in effects parameterization
cat(file = "ANCOVA1.txt","
model {


# Priors
alpha ~ dnorm(0, 1.0E-
06)             # Prior for intercept = effect of level 1 of
forest factor
beta2 ~ dnorm(0, 1.0E-
06)                  # Prior for slope = effect of elevation
for level 1 of forest factor
beta1[1] <- 0                                    # Set to zero
 effect of first level of facFor
beta3[1] <- 0                                    # Set to zero
 effect of first level of facFor of elevation
for(k in 2:4){
  beta1[k] ~ dnorm(0, 1.0E-
06)     # Prior for effects of factor facFor
  beta3[k] ~ dnorm(0, 1.0E-
06)     # Prior for effects of factor facFor
}
tau <- pow(sd, -2)
sd ~ dunif(0, 1000)                          # Prior for dispe
rsion on sd scale


# Likelihood
for (i in 1:M){
  Cmean[i] ~ dnorm(mu[i], tau)     # precision tau = 1 / vari
ance
```

```
  mu[i] <- alpha + beta1[facFor[i]] + beta2 * elev[i] + beta3
[facFor[i]] * elev[i]
}
}
")
```

We must not give any initial values for fixed quantities (here, `beta1[1]` and `beta3[1]`); note that in place of the initial for the first element of the parameter vectors `beta1` and `beta3`, we have an "NA."

```
# Initial values
inits <- function() list(alpha = rnorm(1,,10), beta1 = c(NA,
rnorm(3,,10)), beta2 = rnorm(1,,10), beta3 = c(NA, rnorm(3,,1
0)))

# Parameters monitored
params <- c("alpha", "beta1", "beta2", "beta3", "sd")

# MCMC settings
ni <- 6000   ;   nt <- 1   ;   nb <- 1000   ;  nc <- 3

# Call WinBUGS or JAGS from R (ART <1 min)
out3 <- bugs(win.data, inits, params, "ANCOVA1.txt", n.chains
 = nc, n.thin = nt, n.iter = ni, n.burnin = nb, debug = TRUE,
 bugs.directory = bugs.dir, working.directory = getwd())

out3J <- jags(win.data, inits, params, "ANCOVA1.txt", n.chain
s = nc, n.thin = nt, n.iter = ni, n.burnin = nb)
# traceplot(out3J)

# Fit model using least-
squares (yields equivalent estimates as MLE)
(fm <- summary(lm(Cmean ~ as.factor(facFor)*elev)))

Coefficients:
                                          Estimate  Std.
Error   t value  Pr(>|t|)
(Intercept)                          0.3353         0.2301
     1.457   0.14633
as.factor(facFor)2             0.4244         0.3231      1.
313   0.19028
as.factor(facFor)3             1.2690         0.3083      4.
115   5.2e-05 ***
as.factor(facFor)4             3.7205         0.3162    11.7
66   < 2e-16  ***
elev                                    -0.6013         0.
4203    -1.431   0.15377
```

```
as.factor(facFor)2:elev  -0.6866         0.5999   -1.145   0.
25345
as.factor(facFor)3:elev  -1.2116         0.5427   -2.232   0.
02644 *
as.factor(facFor)4:elev  -1.6164         0.5708   -2.832   0.
00499 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.783 on 259 degrees of freedom
Multiple R-squared:  0.4941,    Adjusted R-squared:  0.4804
F-statistic: 36.13 on 7 and 259 DF,  p-value: < 2.2e-16
```

```
# Summarize posteriors
print(out3, 3)
                          mean       sd           2.5%
    25%             50%            75%         97.5%      Rha
t  n.eff
alpha              0.337  0.231      -0.114        0.183
    0.337        0.491         0.799  1.001  15000
beta1[2]         0.422  0.324      -0.211        0.205
 0.422        0.641         1.051  1.001  15000
beta1[3]         1.268  0.310        0.664        1.062
  1.267        1.473         1.887  1.001  15000
beta1[4]         3.721  0.318        3.093        3.509
  3.721        3.931         4.350  1.001  15000
beta2             -0.602  0.421      -1.442      -0.885      -
0.600     -0.319         0.222    1.001  15000
beta3[2]     -0.687  0.605      -1.859      -1.101      -0.69
2     -0.277         0.503   1.001  15000
beta3[3]     -1.215  0.544      -2.290      -1.581      -1.21
8     -0.847       -0.158    1.001  15000
beta3[4]     -1.611  0.578      -2.744      -1.999      -1.61
0     -1.223       -0.456    1.001    6100
sd                   1.791  0.078        1.648        1.737
       1.788        1.842         1.953    1.001  15000
deviance  1067.483  4.339   1061.000  1064.000   1067.000  10
70.000   1078.000    1.001  15000

DIC info (using the rule, pD = Dbar-Dhat)
pD = 9.0 and DIC = 1076.5
DIC is an estimate of expected predictive error (lower devian
ce is better).
```

We see the usual close numerical agreement between the Bayesian estimates and the MLEs obtained with function `lm` in R. Next, we fit the model using the means

parameterization, where we fit directly the effect of each level of factor `facFor` (note the changed parameter naming in the output). We do not need any change in the data bundle. In addition, we also illustrate how we can estimate custom contrasts as derived quantities—i.e., differences or other functions of parameters. We estimate all pair-wise differences between the group means `beta[1:4]`. Of course, we could also easily compute these derived quantities in R using posterior samples of the vector `beta` produced by BUGS.

```
# Specify model in BUGS language
cat(file = "ANCOVA2.txt","
model {

# Priors
for(k in 1:4){
   alpha[k] ~ dnorm(0, 1.0E-06)     # Priors for intercepts
   beta[k] ~ dnorm(0, 1.0E-06)       # Priors for slopes
}
tau <- pow(sd, -2)
sd ~ dunif(0, 1000)                          #
Prior for dispersion on sd scale

# Likelihood
for (i in 1:M){
  Cmean[i] ~ dnorm(mu[i], tau)      #
precision tau = 1 / variance
  mu[i] <- alpha[facFor[i]] + beta[facFor[i]] * elev[i]
}

# Derived quantities: comparison of slopes (now you can forge
t the delta rule !)
for(k in 1:4){
   diff.vs1[k] <- beta[k] - beta[1]    #
Differences relative to beta[1]
   diff.vs2[k] <- beta[k] - beta[2]    # ... relative to beta
[2]
   diff.vs3[k] <- beta[k] - beta[3]    # ... relative to beta
[3]
   diff.vs4[k] <- beta[k] - beta[4]    # ... relative to beta
[4]
}
}
")

# Initial values
inits <- function() list(alpha = rnorm(4,,10), beta = rnorm(4
,,10))
```

```
# Parameters monitored
params <- c("alpha", "beta", "sd", "diff.vs1", "diff.vs2", "d
iff.vs3", "diff.vs4")

# MCMC settings
ni <- 6000   ;   nt <- 1   ;   nb <- 1000   ;  nc <- 3

# Call WinBUGS or JAGS from R (ART <1 min) and summarize post
eriors
out4 <- bugs(win.data, inits, params, "ANCOVA2.txt", n.chains
 = nc, n.thin = nt, n.iter = ni, n.burnin = nb, debug = TRUE,
 bugs.directory = bugs.dir, working.directory = getwd())

system.time(out4J <- jags(win.data, inits, params, "ANCOVA2.t
xt", n.chains = nc, n.thin = nt, n.iter = ni, n.burnin = nb))
traceplot(out4J)



print(out4, 2)
Inference for Bugs model at "ANCOVA2.txt", fit using WinBUGS,
Current: 3 chains, each with 6000 iterations (first 1000 disc
arded)
Cumulative: n.sims = 15000 iterations saved
                          mean        sd      2.5%        25
%        50%        75%   97.5%   Rhat    n.eff
alpha[1]           0.33    0.23   -0.13      0.18     0.33
   0.49     0.79         1    15000
alpha[2]           0.76    0.23      0.31      0.61     0.76
     0.91     1.21         1    15000
alpha[3]           1.60    0.21      1.19      1.46      1.60
      1.74     2.00         1    15000
alpha[4]           4.06    0.22      3.64      3.91     4.05
     4.20     4.49         1    15000
beta[1]           -0.60    0.42   -1.41    -0.88    -0.60   -
0.31     0.23          1    15000
beta[2]           -1.29    0.43   -2.14    -1.58    -1.29   -
0.99    -0.43         1    15000
beta[3]           -1.82    0.34   -2.49    -2.05    -1.82   -
1.59    -1.13         1      6700
beta[4]           -2.22    0.39   -2.97    -2.48    -2.22   -
1.96    -1.45         1    15000
sd                       1.79    0.08     1.65       1.74
 1.79      1.84     1.96         1      9900
```

```
diff.vs1[1]     0.00    0.00     0.00      0.00     0.00
 0.00     0.00         1            1
diff.vs1[2]   -0.69    0.60   -1.88    -1.10   -0.69    -0.28
     0.48         1    15000
diff.vs1[3]   -1.22    0.55   -2.30    -1.58   -1.22    -0.86
    -0.15         1    14000
diff.vs1[4]   -1.62    0.57   -2.73    -2.01   -1.63    -1.23
    -0.49         1    15000
diff.vs2[1]     0.69    0.60   -0.48      0.28      0.69
 1.10     1.88         1    15000
diff.vs2[2]     0.00    0.00     0.00      0.00      0.00
  0.00     0.00         1            1
diff.vs2[3]  -0.53    0.56   -1.63    -0.90    -0.53   -0.16
    0.57         1    15000
diff.vs2[4]  -0.93    0.58   -2.07    -1.32    -0.93   -0.54
    0.21         1    15000
diff.vs3[1]     1.22    0.55     0.15      0.86      1.22
 1.58     2.30         1    14000
diff.vs3[2]     0.53    0.56   -0.57      0.16      0.53
 0.90     1.63         1    15000
diff.vs3[3]     0.00    0.00     0.00      0.00      0.00
  0.00     0.00         1            1
diff.vs3[4]   -0.40    0.52   -1.42    -0.75    -0.40   -0.05
     0.62        1      9700
diff.vs4[1]     1.62    0.57     0.49      1.23      1.63
  2.01     2.73         1    15000
diff.vs4[2]     0.93    0.58   -0.21      0.54      0.93
 1.32      2.07         1    15000
diff.vs4[3]     0.40    0.52   -0.62      0.05      0.40
 0.75     1.42        1      9700
diff.vs4[4]     0.00    0.00     0.00      0.00      0.00
  0.00     0.00         1            1
```

```
# Fit model using maximum likelihood
(fm <- summary(lm(Cmean ~ as.factor(facFor)*elev-1-elev)))

Coefficients:
                                          Estimate   Std.
Error  t value     Pr(>|t|)
as.factor(facFor)1             0.3353         0.2301      1.
457  0.146328
as.factor(facFor)2             0.7596         0.2269      3.
348  0.000935 ***
as.factor(facFor)3             1.6042         0.2052      7.
816  1.37e-13 ***
```

```
as.factor(facFor)4              4.0558         0.2169    18.7
00     < 2e-16 ***
as.factor(facFor)1:elev  -0.6013         0.4203   -1.431  0.1
53772
as.factor(facFor)2:elev  -1.2880         0.4280   -3.009  0.0
02880 **
as.factor(facFor)3:elev  -1.8129         0.3433   -5.281  2.7
3e-07 ***
as.factor(facFor)4:elev  -2.2177         0.3862   -5.743  2.6
0e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.783 on 259 degrees of freedom
Multiple R-squared:  0.6689,    Adjusted R-squared:  0.6587
F-statistic: 65.42 on 8 and 259 DF,  p-value: < 2.2e-16
```

We will often see the linear model that underlies an ANOVA (analysis of variance) or an ANCOVA using nested indexing in the BUGS language. Let us plot the predicted response as a function of the explanatory variables `facFor` and `elev` (Figure 5.10(b)). We use the parameter estimates from the least-squares fit (= MLEs), but clearly could also use the Bayesian posterior means.

```
plot(elev[facFor==1], Cmean[facFor==1], col = "red", ylim = c
(0, 20), xlab = "Elevation", ylab = "",       frame.plot = F)
points(elev[facFor==2],   Cmean[facFor==2], col = "blue")
points(elev[facFor==3],   Cmean[facFor==3], col = "green")
points(elev[facFor==4],   Cmean[facFor==4], col = "black")
abline(fm$coef[1,1],   fm$coef[5,1], col = "red")
abline(fm$coef[2,1],   fm$coef[6,1], col = "blue")
abline(fm$coef[3,1],   fm$coef[7,1], col = "green")
abline(fm$coef[4,1],   fm$coef[8,1], col = "black")
```

To further illustrate how simple it is to test custom hypotheses in an MCMC-based analysis, let us compute the probability that the difference in the slopes between level 3 of `facFor` and the other levels of that factor is greater than 1. We plot the histograms of these contrasts (Figure 5.11) and then compute the proportion of the area under the curve that lies to the right of 1.

Figure 5.11. Posterior distributions of the difference between the slope of the regression of the mean tit counts on elevation in `facFor` level 3 versus levels 1 (a), 2 (b), and 4 (c). The probability that this difference is greater than 1 is represented by the area under the curve to the right of the red line for each posterior distribution.

```
attach.bugs(out4)     # Allows to directly address the sims.list

str(diff.vs3)

par(mfrow = c(1, 3), mar = c(5,5,3,2), cex.lab = 1.5, cex.axis = 1.5)

hist(diff.vs3[,1], col = "grey", breaks = 100, main = "", freq=F, ylim = c(0, 0.8))

abline(v = 1, lwd = 3, col = "red")

hist(diff.vs3[,2], col = "grey", breaks = 100, main = "", freq=F, ylim = c(0, 0.8))

abline(v = 1, lwd = 3, col = "red")

hist(diff.vs3[,4], col = "grey", breaks = 100, main = "", freq=F, ylim = c(0, 0.8))

abline(v = 1, lwd = 3, col = "red")


# Prob. difference greater than 1

mean(diff.vs3[,1] > 1)

mean(diff.vs3[,2] > 1)

mean(diff.vs3[,4] > 1)

 [1] 0.6554667

 [1] 0.1981333

 [1] 0.003733333
```

Hence, there is a 66% probability that the difference between the slopes in groups 1 and 3 of `facFor` is greater than 1, and we find corresponding probabilities of 20% and of essentially 0% for the analogous slope differences between group 3 and groups 2 and 4, respectively.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780128013786000059

Space–Time Metrics

George Christakos, in Spatiotemporal Random Fields (Second Edition), 2017

Example 2.1

Some worth-noticing special cases of Eqs. (2.1a) and (2.1b), to be used later in the context of the "geometrical metric–physical covariance" analysis, are presented in a matrix–vector format by

$$\frac{\partial}{\partial \boldsymbol{h}} c_X(\Delta \boldsymbol{p}) = \frac{\partial \Delta p}{\partial \boldsymbol{h}} \frac{\partial c_X}{\partial \Delta \boldsymbol{p}}, \tag{2.2a–b}$$

$$\frac{\partial^2}{\partial \boldsymbol{h}^T \partial \boldsymbol{h}} c_X(\Delta \boldsymbol{p}) = \frac{\partial^2 \Delta p}{\partial \boldsymbol{h}^T \partial \boldsymbol{h}} \frac{\partial c_X}{\partial \Delta \boldsymbol{p}} + \left(\frac{\partial \Delta p}{\partial \boldsymbol{h}}\right)^T \left(\frac{\partial \Delta p}{\partial \boldsymbol{h}}\right) \frac{\partial^2 c_X}{\partial \Delta \boldsymbol{p}^2},$$

where, $\Delta p = \left(\Delta \boldsymbol{p}^2\right)^{\frac{1}{2}}$ as usual. At this point, I would like to reiterate the significant point that an apparent feature of the above CDF, which is useful for physical metric determination purposes (discussed in Section 3), is that they decompose the ordinary covariance derivatives with respect to space lags ($h_1 \ldots h_n$) and time lag ($h_0 = \tau$), rather commonly encountered in physical covariance laws, in terms of the space–time physical metric and the corresponding covariance derivatives with respect to this metric.

By combining MDF with CDF, some interesting expressions are found that explicitly contain the space- and time-independent metric coefficients. This is shown in the following proposition.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780128030127000039

Statistical Prediction

M.E. Borsuk, in Encyclopedia of Ecology, 2008

ANOVA and ANCOVA models

When the predictors, $\boldsymbol{x}$, of a linear model consist entirely of categorical variables, and yet the response, $\boldsymbol{y}$, remains continuous, the model is referred to as an 'Analysis of Variance' (ANOVA) model. If an additional, continuous predictor is also included, and it is assumed that the effect of that covariate on the response is equal for all categories, the model is referred to as an 'Analysis of Covariance' (ANCOVA) model. Traditionally, the use of ANOVA and ANCOVA-type models has focused on hypothesis testing, but the models can also be used for prediction.

There are three basic classes of ANOVA models:

1. 'Fixed effects models' in which the data are assumed to come from normallydistributed populations which differ only in their means. In these models, the parameter β is simply regarded as a vector of constants.
2. 'Random effects models' in which the data are assumed to come from a hierarchy of populations, and differences are constrained by the hierarchy. In these models, β is treated as a vector of random variables.

3. ‘Mixed effects models’ in which both fixed and random effects are present.

Because the random components in ANOVA and ANCOVA models are assumed to be normally distributed, the methods used to fit and apply simple and multiple linear regression models can continue to be used.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780080454054002342

SEWALL WRIGHT

James F. Crow, in Philosophy of Biology, 2007

Publisher Summary

Sewall Wright was one of the great trio—R. A. Fisher, J. B. S. Haldane, and himself—who, starting about the time of World War I, founded the mathematical theory of microevolution and population genetics. Wright had an interest in statistics from his earliest scientific studies, in his early study of guinea pig weights in 1917; he was able to subdivide product moments into between- and within-group components. In effect, he had discovered what was later called the analysis of covariance, a subject invented independently and carried farther by R. A. Fisher. His most important contribution to statistical methodology is his method of path analysis. In 1925, Wright published a monumental analysis of the production and prices of corn and hogs in the period from the Civil War to World War I. In the early years after Wright's discovery of the method, path analysis was very popular among animal breeders. Wright's approach was always quantitative. Today, his inbreeding coefficient, discussed earlier, and its extension to hierarchical population structure are among his greatest accomplishments, now widely used in the study of population structure, including humans. Although most of Wright's work in population genetics was theoretical, he entered several collaborations with experimentalists, especially Th. Dobhansky. He gave the shifting balancing theory in 1968-78. He also had a long interest in the philosophy of organism and the mind-body problem. For a man of such accomplishment, Wright was excessively modest with a self-deprecating wit, although he did not hesitate to argue a point on which he had an opinion—like politics or the shifting balance theory.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B978044451543850006X