Covariance Properties of Hierarchical Models

by

Frongillo, E.A., Casella, G. and McCulloch C.E.

BU-1215-M July 1993

Covariance Properties of Hierarchical Models

Edward A. Frongillo, Jr., Ph.D., George Casella, Ph.D.,

and Charles E. McCulloch, Ph.D.

Edward A. Frongillo, Jr., Division of Nutritional Sciences, B36 Savage Hall, Cornell University, Ithaca, New York 14853-6301

George Casella, Biometrics Unit, 320 Warren Hall, Cornell University, Ithaca, New York 14853.

Charles E. McCulloch, Biometrics Unit, 338 Warren Hall, Cornell University, Ithaca, New York 14853

Keywords: covariance, hierarchical model, randomized block model

Submitted to Biometrics 7/30/93

SUMMARY

Previously applied hierarchical models have assumed conditional independence, based on three independence assumptions, with a consequence that units are uncorrelated and subunits within units are positively correlated. This paper demonstrates that these independence assumptions can be relaxed to yield hierarchical models with more general covariance properties. A theorem demonstrates that, using minimal assumptions on the distributions and moments, the marginal covariance is a simple function of the first- and second-stage correlation and variance parameters. This approach introduces greater flexibility into the hierarchical structure, and allows for randomized block models that can have negative covariances.

1. Introduction

There are many statistical problems for which information is available from multiple units that are similar, and for which the combining of this information from the multiple units is advantageous. Examples of such problems are meta-analysis, blocking designs, and models based on a mixture or compound distribution. For these problems, variability among units as well as within units is expected. Two-stage hierarchical models are a useful way to approach these problems because these models allow the specification of parameters both within and among units.

Hierarchical models that have been applied in the literature have been assumed to be conditionally independent. This conditional independence is based on three independence assumptions. First, subunits within each unit are conditionally independent given the unit effects. Second, subunits in different units are conditionally independent. Third, units are independent. A consequence of these assumptions is that units are uncorrelated and that subunits within units are positively correlated (Kass and Steffey, 1989).

Although these conditionally independent hierarchical models are reasonable for many problems, there are other problems for which the specification of more general covariance properties is desirable. One example of such problems is an animal experiment with litters as blocks for which competition is expected within the litters. Another example is a blocking experiment for which the blocks are expected to be correlated because of clustering of the blocks due to some factor not modeled.

This paper demonstrates a class of hierarchical models with more general covariance properties. A theorem demonstrates that, using minimal assumptions on the distributions and moments, the marginal covariance is a simple function of the first- and second-stage correlation and variance parameters. Covariance properties, a possible elaboration, and implications of these models are presented. Two randomized block models are developed that allow for negative covariances; these models are alternatives to the one given by Hocking (1985). Results from the literature are brought together to show how covariances behave under non-linear transformations. Two applications are then presented.

2. Covariance in a Hierarchical Model

2.1 Theorem

We begin with a theorem specifying a general covariance structure in a two-stage hierarchical model with a two-parameter density at each stage. Let f be the first-stage density of a two-stage hierarchical model. Let Y_1 and Y_2 be two observations from f with possibly different mean and variance parameters. The parameters θ_i, σ_i^2, and ρ are, respectively, the means, variances, and correlation in the first stage. The first stage can be written as:

1) $Y_1|\theta_1, \sigma_1^2, \rho \sim f(y_1|\theta_1, \sigma_1^2, \rho)$

$Y_2|\theta_2, \sigma_2^2, \rho \sim f(y_2|\theta_2, \sigma_2^2, \rho)$

with $E[Y_1|\theta_1, \sigma_1^2, \rho] = \theta_1$

$E[Y_2|\theta_2, \sigma_2^2, \rho] = \theta_2$

$\text{var}[Y_1|\theta_1, \sigma_1^2, \rho] = \sigma_1^2$

$\text{var}[Y_2|\theta_2, \sigma_2^2, \rho] = \sigma_2^2$

$\text{corr}[Y_1, Y_2|\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho] = \rho$

Let π_1 be a second-stage density for θ_i, with parameters μ, τ^2, ϕ that represent, respectively, the mean, variance, and correlation. Let π_2 be a second-stage density for σ_1^2, with parameters δ^2 and υ. The second-stage can be written as:

2) $\theta_1,\ \theta_2 | \mu,\ \tau^2,\ \phi \sim \pi_1\ (\theta | \mu,\ \tau^2,\ \phi)$

$\sigma_1,\ \sigma_2 | \delta^2, \upsilon \sim \pi_2\ (\sigma | \upsilon, \delta^2 - \upsilon^2)$

with $E\ [\theta_1 | \mu,\ \tau^2,\ \phi] = E\ [\theta_2 | \mu,\ \tau^2,\ \phi] = \mu$

$\mathrm{var}\ [\theta_1 | \mu,\ \tau^2,\ \phi] = \mathrm{var}\ [\theta_2 | \mu,\ \tau^2,\ \phi] = \tau^2$

$\mathrm{corr}\ [\theta_1,\ \theta_2 | \mu,\ \tau^2,\ \phi] = \phi$

$E\ [\sigma_1^2 | \delta^2, \upsilon] = E\ [\sigma_2^2 | \delta^2, \upsilon] = \delta^2$

$E\ [\sigma_1 | \delta^2, \upsilon] = E\ [\sigma_2 | \delta^2, \upsilon] = \upsilon$

$E\ [\sigma_1 \sigma_2 | \delta^2, \upsilon] = E\ [\sigma_1 | \delta^2, \upsilon]\ E\ [\sigma_2 | \delta^2, \upsilon] = \upsilon^2$

With these assumptions, the theorem then concludes that the marginal density has the following mean, variance, and covariance:

$E\ [Y_1] = E\ [Y_2] = \mu$

$\mathrm{var}\ [Y_1] = \mathrm{var}\ [Y_2] = \delta^2 + \tau^2$

$\mathrm{cov}\ [Y_1,\ Y_2] = \rho \upsilon^2 + \phi \tau^2.$

The proof follows directly from repeated use of iterated expectations. First,

$E\ [Y_1] = E\ [E\ [Y_1 | \theta_1,\ \sigma_1^2,\ \rho]] = E\ [\theta_1] = \mu$

$E\ [Y_2] = \mu$

Next, using the formula for marginal variances,

$$\begin{aligned}\mathrm{var}\ [Y_1] &= E\ [\mathrm{var}\ [Y_1 | \theta_1,\ \sigma_1^2,\ \rho]] + \mathrm{var}\ [E\ [Y_1 | \theta_1,\ \sigma_1^2,\ \rho]] \\ &= E\ [\sigma_1^2] \qquad\qquad + \mathrm{var}\ [\theta_1] \\ &= \delta^2 + \tau^2\end{aligned}$$

Similarly,

$$\text{var}\ [Y_2] = \delta^2 + \tau^2$$

Calculation of the covariance of Y_1 and Y_2 also uses iterated expectations:

$$\begin{aligned}\text{cov}\ [Y_1,Y_2] &= E\ [\ \text{cov}\ [Y_1,Y_2|\theta_1,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2,\ \rho]\ ]\ + \\ &\qquad \text{cov}\ [\ E\ [Y_1|\theta_1,\ \sigma_1^2,\ \rho],\ E\ [Y_2|\theta_2,\ \sigma_2^2,\ \rho]\ ] \\ &= E\ [\rho\sigma_1\sigma_2] + \text{cov}\ [\theta_1,\theta_2] \\ &= \rho\upsilon^2 + \phi\tau^2\end{aligned}$$

Alternately, the covariance of Y_1 and Y_2 can be calculated as the difference of the variances of Y_1 and Y_2 from the variance of $Y_1 + Y_2$ (Frongillo, 1991).

2.2 Covariance Properties

Fixing the values of the correlation parameters illustrates the covariance properties of these models. Consider a setting where the θ's correspond to units, and the Y's correspond to sub-units.

1) $\rho = 0$

If Y_1 and Y_2 are conditionally uncorrelated, then cov $[Y_1, Y_2] = \phi\tau^2$. This is the marginal covariance of two random variables when their subunits are conditionally uncorrelated but their units are correlated.

2) $\phi = 0$

If θ_1 and θ_2 are uncorrelated, then cov $[Y_1, Y_2] = \rho\upsilon^2$. This is the marginal covariance of two random variables when their units are conditionally uncorrelated but their subunits are correlated.

3) $\phi = 1$

If, for example, Y_1 and Y_2 have the same conditional expectation, then cov $[Y_1, Y_2] = \rho\upsilon^2 + \tau^2$. This is the marginal covariance of two random variables when their subunits are conditionally correlated and their units are perfectly positively correlated.

4) $\phi = 0$ and $\rho = 0$

If θ_1 and θ_2 are uncorrelated and Y_1 and Y_2 are conditionally uncorrelated, then cov $[Y_1, Y_2] = 0$. This is the marginal covariance of two random variables when their subunits are conditionally uncorrelated and units are uncorrelated. An example of this situation is the classical randomized block design, where Y_1 and Y_2 are from different blocks.

5) $\phi = 1$ and $\rho = 0$

If, for example, Y_1 and Y_2 have the same conditional expectation and Y_1 and Y_2 are conditionally uncorrelated, then cov $[Y_1, Y_2] = \tau^2$. This is the marginal covariance of two random variables when their subunits are conditionally uncorrelated and units are perfectly positively correlated. An example of this situation is the classical randomized block design, where Y_1 and Y_2 are from the same block.

We state some other properties of these models. Since cov $[Y_1, Y_2] = \rho\upsilon^2 + \phi\tau^2$ and $-1 \leq \rho, \phi \leq +1$, cov $[Y_1, Y_2]$ can be positive or negative. Since var $[Y_1 + Y_2] = 2(\delta^2 + \rho\upsilon^2) + 2\tau^2 (1 + \phi)$ and $-1 \leq \rho, \phi \leq +1$, this variance can take on values between $2(\delta^2 - \upsilon^2)$ (when both ρ and ϕ are -1) and $2(\delta^2 + \upsilon^2) + 4\tau^2$ (when both ρ and ϕ are +1). For a single observation, Y_1, the unconditional or marginal variance $(\delta^2 + \tau^2)$ is always greater than or equal to the conditional variance (δ^2). Likewise, for the sum, $Y_1 + Y_2$, the unconditional variance is always greater than or equal to the conditional variance of the sum. However, the unconditional variance of the sum, $Y_1 + Y_2$, can be smaller than the conditional variance of Y_1 if $\rho < 0$.

The preceding theorem and analyses assumed that there was no second-stage distribution on ρ, i.e., ρ = E [ρ] with probability 1. An elaboration would assume: $\rho|\gamma,\Gamma \sim \pi_3 (\gamma,\Gamma)$ with $E[\rho|\gamma,\Gamma] = \gamma$, $\text{var}[\rho|\gamma,\Gamma] = \Gamma$. Then, each Y could be conditional on a different ρ, with the ρ's tied together with this prior.

3. Randomized Block Models

The usual randomized block model arises from a hierarchical model and is given by

$$Y_{ij} = \mu_i + B_j + e_{ij} \qquad i = 1, \ldots, t; \quad j = 1, \ldots, b \tag{3.1}$$

where μ_i refers to the ith of t treatment means, B_j refers to the jth of b block means, and e_{ij} refers to the random error for the ith treatment and jth block combination. Combined with the usual error assumptions

$$Y_{ij}|\mu_i,\ B_j,\ \sigma^2 \sim N\ (\mu_i + B_j,\ \sigma^2) \quad \text{uncorrelated}$$

$$B_j|\tau^2 \sim N\ (0,\ \tau^2) \quad \text{uncorrelated}$$

this specification results in the marginal distribution

$$Y_{ij}|\mu_i,\ \sigma^2,\ \tau^2 \sim N\ (\mu_i,\ \sigma^2 + \tau^2)$$

with covariances

$$\text{cov}\ [Y_{ij},\ Y_{kj}|\mu_i] = \tau^2 \quad \text{same block}$$

$$\text{cov}\ [Y_{ij},\ Y_{ik}|\mu_i] = 0 \quad \text{different block}$$

In this model, since τ^2 is a variance, the marginal covariance must be positive or zero.

Hocking (1985, p. 325) gives an alternate specification that allows the covariance, defined as α, to be negative. This specification does not give a model as in (3.1), but rather gives the following expectations and covariances:

$$E\ [Y_{ij} | \mu_i] = \mu_i$$

$$\text{var}\ [Y_{ij} | \mu_i] = \sigma^2 + \alpha$$

$$\text{cov}\ [Y_{ij},\ Y_{kj} | \mu_i] = \alpha \qquad \text{same block}$$

$$\text{cov}\ [Y_{ij},\ Y_{ik} | \mu_i] = 0 \qquad \text{different block}$$

This specification has two disadvantages. First, it cannot arise from a hierarchical model because, if α is negative, then the marginal variance is less than σ^2. Second, this specification has negative marginal variance if α is negative and $|\alpha| > \sigma^2$.

A hierarchical model that allows for negative covariance is one in which the correlation within blocks is non-zero. This model does not have the disadvantages of Hocking's alternate specification. This model is the same as (3.1), but with different distributional assumptions:

$$Y_{ij} = \mu_i + B_j + e_{ij} \qquad i = 1,\ \ldots,\ t; \quad j = 1,\ \ldots,\ b \tag{3.2}$$

$$Y_{ij} | \mu_i,\ B_j,\ \sigma^2,\ \rho \sim N\ (\mu_i + B_j,\ \sigma^2)$$

$$\text{with corr}\ (Y_{ij},\ Y_{kj} | \mu_i,\ B_i,\ \sigma^2,\ \rho) = \rho$$

$$\text{and}\ \ \text{corr}\ (Y_{ij},\ Y_{km} | \mu_i,\ B_i,\ \sigma^2,\ \rho) = 0$$

$$B_j | \tau^2 \sim N\ (0,\ \tau^2)\ \text{uncorrelated}$$

This specification results in the marginal distribution

$$Y_{ij} | \mu_i, \sigma^2,\ \tau^2,\ \rho \sim N\ (\mu_i,\ \sigma^2 + \tau^2)$$

and covariances

$$\text{cov}\ (Y_{ij},\ Y_{kj} | \mu_i,\ \sigma^2,\ \tau^2,\ \rho) = \rho\sigma^2 + \tau^2 \qquad \text{same block}$$

$$\text{cov}\ (Y_{ij},\ Y_{km} | \mu_i,\ \sigma^2,\ \tau^2,\ \rho) = 0 \qquad \text{different block}$$

The same-block covariance is less than 0 when

$$0 > \rho\ \sigma^2 + \tau^2$$

$$-\rho\ \sigma^2 > \tau^2$$

$$\rho < \frac{-\tau^2}{\sigma^2}$$

Note that ρ cannot be negative if $\tau^2 > \sigma^2$; this restriction of the range of ρ is a consequence of the hierarchical structure. An application of this model follows in section 5.

Another hierarchical model that allows for negative covariance is one in which, ϕ, the block correlation in the second stage is negative rather than zero:

$$Y_{ij} = \mu_i + B_j + e_{ij} \qquad i = 1, \ldots, t; \quad j = 1, \ldots, b \tag{3.3}$$

$$Y_{ij} | \mu_i, B_j, \sigma^2 \sim N(\mu_i + B_j, \sigma^2) \quad \text{uncorrelated}$$

$$B_j | \tau^2, \phi \sim N(0, \tau^2) \qquad \text{with corr}(B_j, B_k | \tau^2, \phi) = \phi$$

Then, the marginal distribution is

$$Y_{ij} | \mu_i, \sigma^2, \tau^2, \phi \sim N(\mu_i, \sigma^2 + \tau^2)$$

and the covariances are

$$\text{cov}[Y_{ij}, Y_{kj} | \mu_i] = \tau^2 \qquad \text{same block}$$

$$\text{cov}[Y_{ij}, Y_{km} | \mu_i] = \phi\tau^2 \qquad \text{different block}$$

This model differs from that of Hocking in that the marginal variance is always positive and is greater than the first-stage variance in this model. The correlation ϕ can not be less than -1/(b-1) under the assumption of compound symmetry (Searle, 1982). An application of this model follows in section 6.

The results of Section 2 have been used to write two randomized block models that allow for negative covariances. Both of these models differ from the usual model (3.1) only in the distributional assumptions. Model (3.2) allows for negative covariances within blocks. This model might be useful in analyses in which competition within blocks (e.g., animal litters) is expected. Model (3.3) allows for negative covariances among blocks. This model might be useful in analyses in which a factor that relates several blocks to each other is not modeled, thereby inducing correlation among blocks. Although both of these models allow for negative covariances, the hierarchical structure imposes restrictions on the magnitude of negative covariances for both models.

4. Covariance Properties Under Non-linear Transformations

Restrictions in the range of a correlation parameter have been reported by Kupper and Haseman (1978), Paul (1985, 1987), and Prentice (1986) for binary data. These authors have developed extensions of the betabinomial model for binary data that allow for non-zero correlations within units; each of these extended models resulted in a correlation parameter with restricted range. A useful aim of further research would be to understand whether such restrictions of range in the extended betabinomial models are due to an underlying hierarchical model, to the fact that the variance of the binomial distribution is a function of its mean, or to both.

For binary data, a reasonable alternative to the extended betabinomial models that allows for non-zero correlations within units is the logit-normal hierarchical model. The first stage of this model uses the logistic transformation (logit link function), with the transformed variable assumed to be normally distributed. The second stage assumes that the first-stage location parameter is normally distributed. Results of Lehmann (1966) and Esary et al. (1967) can be used to show that the sign of the correlation parameter is preserved under the inverse-logistic transformation (see Appendix). That is, if we assume that Z_1 and Z_2 are distributed bivariate normal and that $Y_i = k(Z_i) = \exp(Z_i)/(1+\exp(Z_i))$ for i=1,2, then $\mathrm{corr}(Z_1,Z_2) \geq 0$, if and only if $\mathrm{corr}(Y_1,Y_2) \geq 0$.

For random variables that are distributed bivariate normal on the logit scale, the magnitude of the correlation on the binomial scale will be less than that on the logit scale (Lancaster, 1957). Lancaster's theorem on the magnitude of the correlation applies to all transformations that result in finite variance. Lancaster's result is consistent with other results that correlations among non-normal random variables do not achieve the theoretical bounds of -1, 1 commonly understood for correlations (Cox and Wermuth, 1992; Shih and Huang, 1992).

5. Application of Model (3.2) to Randomized Block Data

We apply model (3.2) to data that Hocking (1985, p. 326) modified and used to illustrate his alternate specification. This example has 25 data points from a design with five treatments and five blocks. The estimates of σ^2 and α in Hocking's specification were, respectively, 8.81 and -0.07. For model (3.2), 8.81 is the estimate for σ^2 and -0.07 is the estimate for $\rho\sigma^2 + \tau^2$. Because there is only one replicate for each treatment and block combination, ρ and τ^2 can not be separately estimated. The possible values for τ^2 and ρ are: $0 \leq \tau^2 \leq 8.74$, and $-1 \leq \rho \leq -0.0079$, indicating a negative correlation within blocks in these data.

6. Application of Model (3.3) to Data on Soup Kitchens in New York State

We apply model (3.3) using data that Rauschenbach et al. (1990) reported from a survey of guests receiving meals at soup kitchens in New York State. The purpose of the survey was to determine social, economic, and demographic characteristics of the guests. Data were collected from 501 guests, sampled from 28 of the 34 soup kitchens operating in five cities:

City	Soup Kitchens Sampled	Total Meals Sampled
Albany	6	110
Buffalo	8	155
Westchester	7	112
Rochester	3	44
Syracuse	4	80

Meals were sampled in each soup kitchen using random sampling within early-month and late-month strata proportionate to the population of meals previously established by census; a questionnaire was administered to the guest served a sampled meal. Soup kitchens from the first three cities were sampled in April 1987 and from the last two cities in October 1987.

Rauschenbach et al. (1990) found substantial variation among soup kitchens using a betabinomial model, and, for some variables measured, substantial variation as well among cities as estimated by analysis of variance. If soup kitchens are regarded as the units and meals as the subunits, then one might expect to find a compound symmetric structure because of the correlation induced by the fact that soup kitchens are clustered within cities. For some variables, because guests with different characteristics may choose one soup kitchen over another within a city, a negative correlation among soup kitchens might be found.

The model for analysis of these data can be written as

1) $Y_{ij} | \theta_{ij}, \sigma_{ij}^2 \sim \text{normal}(\theta_{ij}, \sigma_{ij}^2)$

2) $\theta_{ij} | \mu, \tau^2, \phi \sim \text{normal}(\mu, \tau^2)$

$\text{corr}(\theta_{ij}, \theta_{ik} | \mu, \tau^2, \phi) = \phi$ for $j \neq k$

$\text{corr}(\theta_{ij}, \theta_{hk} | \mu, \tau^2, \phi) = 0$ for $i \neq h$.

The index i refers to city and ranges from 1 to 5. The index j refers to a soup kitchen within a city. Y_{ij} is the observed logodds for a binary variable representing the response for soup kitchen i,j. Y_{ij} was calculated using the empirical logistic transformation: $Y_i = \ln[(n_i\hat{p}_i+0.5)/(n_i-n_i\hat{p}_i+0.5)]$, which has the property $E[Y_i] = \theta_i + O(1/n_i^2)$; any other choice of constant besides 0.5 has bias $O(1/n_i)$ (McCullagh and Nelder, 1989; Cox and Snell, 1989). θ_{ij} is the true response for soup kitchen i,j, and σ_{ij}^2 is the true variance of that response. In this model, the σ_{ij}^2 are assumed to be known. σ_{ij}^2 is the bias-adjusted variance of a binary variable calculated according to the suggested formula of McCullagh and Nelder (1989): $\sigma_i^2 = (n_i\hat{p}_i+0.5)^{-1} + (n_i-n_i\hat{p}_i+0.5)^{-1}$.

Maximum likelihood estimation was implemented using the Maxlik applications module of Gauss386, version 2.01-5.00 (Schoenberg, 1990). This module provides various iterative routines for doing maximum likelihood estimation by maximizing a log-likelihood calculated in a user-written procedure. We used the steepest descent and Newton-Raphson routines without analytic derivatives. The estimates of the parameters representing the overall mean (μ), the variance among soup kitchens (τ^2), and the correlation among soup kitchens (ϕ) are presented in Table 1 for ten variables.

Estimates of the μ ranged from -1.762 to 0.065 for nine of the variables; these correspond to proportions from 0.147 to 0.484. The other variable, below poverty, had 1.616 as an estimate for μ, corresponding to a proportion of 0.834. Some of these results were of particular interest because they counter prevailing views that soup kitchen guests are predominantly single, homeless males. About 1/2 of guests received Food Stamps and about 1/3 received other public assistance. About 1/5 had a child in the household and 39% were female. Most (83%) were below the poverty line, but only 18% were homeless.

Estimates of τ^2 ranged from 0.109 to 0.499, meaning that the variability of soup kitchens across the ten variables differed by a factor of five. That is, the variability for homeless was five times that for Food Stamps, meaning that soup kitchens tended to have similar proportions of guests receiving Food Stamps, but varying proportions of guests who were homeless.

For seven of the variables, the estimate for ϕ was positive; the seven estimates ranged from 0.209 to 0.991. For the other three variables, the estimate was negative. The estimate of ϕ reported in Table 1 for these three variables was from the iteration just prior to that for which the estimate of ϕ was less than the theoretical boundary of -0.143 (= -1/7), due to the assumption of compound symmetry; this boundary is the reciprocal of one less than the minimum number of soup kitchens per city. The interpretation of large, positive estimates for the variables receiving Food Stamps and receiving other public assistance is that soup kitchens within cities were similar on those characteristics. On the other hand, negative estimates for the variables having a telephone, having a child in the household, and below poverty indicates that soup kitchens within cities were dissimilar on those characteristics. Across the ten variables, the estimates of τ^2 and ϕ were correlated 0.56, indicating that variables with greater variability tended to be similar among soup kitchens within cities.

6. Conclusions

Previously applied hierarchical models have assumed conditional independence with a consequence that units are uncorrelated and subunits within units are positively correlated. This paper demonstrates that the three independence assumptions can be relaxed to yield hierarchical models with more general covariance properties. Using minimal assumptions on the distributions and moments, the marginal covariance is a simple function of the first- and second-stage correlation and variance parameters. This approach introduces greater flexibility into the hierarchical structure, and allows for randomized block models that can have negative covariances without problematic parameters.

In the soup kitchen example, a three-stage conditionally independent hierarchical structure could have been used to describe the study, with the three stages being: 1) individuals, 2) soup kitchens, and 3) cities. This structure restricts covariances among soup kitchens to be positive, whereas the two-stage structure presented in this paper does not have this restriction. Another restriction under the assumption of compound symmetry did impose a boundary of -1/7 on the correlation parameter ϕ.

ACKNOWLEDGEMENTS

Dr. Casella's research was supported by NSF Grant DMS9100839 and NSA Grant 90F-073.

REFERENCES

Cox, D.R. and Wermuth, N. (1992) A comment on the coefficient of determination for binary responses. *American Statistician* **46**, 1-4.

Esary, E.J., Proschan, F., and Walkup, D.W. (1967) Association of random variables with applications. *Annals of Mathematical Statistics* **38**, 1466-1474.

Hocking, R.R. (1985) *The Analysis of Linear Models.* Monterey, CA: Brooks/Cole.

Kass, R.E. and Steffey, D. (1989) Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association* **84**, 717-726.

Kupper, L.L. and Haseman, J.K. (1978) The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics* **34**, 69-76.

Lancaster, H.O. (1957) Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika* 44, 289-292.

Lehmann, E.L. (1966) Some concepts of dependence. *Annals of Mathematical Statistics* **37**, 1137-1153.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models.* London: Chapman and Hall.

Paul, S.R. (1985) A three-parameter generalization of the binomial distribution. *Communications in Statistics--Theory and Methods.* **14(6)**, 1497-1506.

Paul, S.R. (1987) On the beta-correlated binomial (BCB) distribution-- A three parameter generalization of the binomial distribution. *Communications in Statistics--Theory and Methods* **16(5)**, 1473-78.

Prentice, R.L. (1986) Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association* **81**, 321-327.

Rauschenbach, B.S., Frongillo, E.A. Jr., Thompson, F.E., Andersen, E.J.Y. and Spicer, D.A. (1990) Dependency on soup kitchens in urban areas of New York State. *American Journal of Public Health* **80**, 57-60.

Schoenberg, R. (1990) *Gauss Applications Manual Version 2.01: Maximum Likelihood Estimation*. Kent, WA: Aptech Systems.

Searle, S.R. (1982) *Matrix Algebra Useful for Statistics*. New York: John Wiley and Sons.

Shih, W.J. and Huang, W.M. (1992) Evaluating correlation with proper bounds. *Biometrics* **48**, 1207-1213.

APPENDIX

Lehmann (1966) defined Positive Quadrant Dependence (PQD) as the property where, for random variables Z_1 and Z_2,

$$P(Z_1 \le z_1, Z_2 \le z_2) \ge P(Z_1 \le z_1) P(Z_2 \le z_2) \ \forall \ z_1, z_2.$$

Negative Quadrant Dependence (NQD) is defined similarly:

$$P(Z_1 \le z_1, Z_2 \le z_2) \le P(Z_1 \le z_1) P(Z_2 \le z_2) \ \forall \ z_1, z_2.$$

Lemmas 1 and 3 given in Lehmann (1966) and a lemma in Esary et al. (1967) can be combined to yield the following set of relations for all non-decreasing functions f and g:

$$\begin{array}{ccc} (Z_1, Z_2) \text{ is PQD} & \Rightarrow & [f(z_1), g(z_2)] \text{ is PQD } \forall \ f, g \\ \Downarrow & & \Downarrow \\ \text{cov}(Z_1, Z_2) \ge 0 & \Leftarrow & \text{cov}[f(z_1), g(z_2)] \ge 0 \ \forall \ f, g \end{array}$$

Lehmann, Lemma 1 is the top relation. Lehmann, Lemma 3 is the left and right relations. The lemma of Esary et al. is the bottom relation. Lehmann also showed that, if (Z_1, Z_2) is bivariate normal with $\text{cov}(Z_1, Z_2) \ge 0$, then (Z_1, Z_2) is PQD. This last result means that, if (Z_1, Z_2) is bivariate normal, then each of the sufficient implications in the above relation become necessary and sufficient.

These results are useful for the logistic transformation. Assume that Z_1 and Z_2 are distributed bivariate normal and that $Y_i = k(Z_i) = \exp(Z_i)/(1+\exp(Z_i))$ for $i=1,2$. If $\text{cov}(Z_1, Z_2) \ge 0$, then $\text{cov}[k(Z_1), k(Z_2)] \ge 0$, i.e., $\text{cov}(Y_1, Y_2) \ge 0$, because $k(\cdot)$ is an increasing function. Since (Z_1, Z_2) is bivariate normal, then either (Z_1, Z_2) is PQD or NQD. If $\text{cov}[k(Z_1), k(Z_2)] \ge 0$, i.e., $\text{cov}(Y_1, Y_2) \ge 0$, then (Z_1, Z_2) must be PQD and therefore $\text{cov}(Z_1, Z_2) \ge 0$.

Table 1. Maximum likelihood parameter estimates for selected variables from soup kitchen study.

	Parameter estimates for Among-soup-kitchen		
	Mean	Variance	Correlation
	μ	τ^2	ϕ
Food Stamps	-0.065	0.109	0.690
Other Public Assistance[a]	-0.781	0.121	0.991
Wages	-1.023	0.155	0.588
Subsidized Housing	-1.013	0.115	0.419
Telephone[a]	-0.255	0.300	-0.056
Female	-0.453	0.224	0.118
Disability	-1.762	0.132	0.477
Child in Household[a]	-1.311	0.165	-0.066
Homeless	-1.529	0.499	0.209
Below Poverty	1.616	0.325	-0.088

[a] These models did not converge because the estimate of correlation ϕ became either less than -0.143, the theoretical lower boundary for these data, or greater than 1. The reported estimates were from the iteration previous to the one for which the boundary was encountered.