Inference on Mixed Models

David M. Allen
University of Kentucky

April 5, 2012
This talk presents issues related to inference on $L\beta$ where $\beta$ is the vector of fixed effects in the mixed model and $L$ is a matrix of coefficients specified by the data analyst. In SAS parlance, we are giving the background behind the estimate, contrast, and lsmeans statements.
Estimable functions

Assume $L$ has a one row. We say $L\beta$ is *estimable* if there is a linear combination of the elements of the response vector, $y$, having expected value $L\beta$. Equivalently, $L\beta$ is *estimable* if $L$ can be expressed as a linear combination of the rows of the $X$ matrix. When $L$ has multiple rows, we check each row individually.
Minimum variance unbiased estimation

With REML estimation, the fixed effects are layed aside while we estimate the variance parameters, $\theta$. We calculate $\hat{V} = V(\hat{\theta})$ and pretend it is the actual variance matrix.

If $L\beta$ is estimable, we apply the principle of minimum variance unbiased estimation to find the estimate

$$L(X^t\hat{V}^{-1}X) - X^t\hat{V}^{-1}y$$

and its standard error is

$$\sqrt{L(X^t\hat{V}^{-1}X) - L^t}.$$  

These are not computational formulas.
2 Test statistics

There are a variety of testing methodologies that might be used. Here we consider the Likelihood ratio test and the Wald test.
Likelihood ratio test statistic

Denote the likelihood function for the model by $M(\theta, \beta; y)$. The formula for the likelihood function is the same as the formula for the density function of $y$ but is viewed as a function of $\theta$ and $\beta$ for a fixed $y$. 
Let

- $\Omega$ be the space for all possible values of $\theta$ and $\beta$, and
- $\omega$ be the space for all $\theta$ and all $\beta$ satisfying $L\beta = 0$.

The likelihood ratio statistic for the null hypothesis $L\beta = 0$ against the general alternative hypothesis is

$$\lambda = \frac{\max_{\omega} M(\theta, \beta; y)}{\max_{\Omega} M(\theta, \beta; y)}.$$ 

The test statistic $-2 \log(\lambda)$ is asymptotically distributed with a Chi-square distribution. A major difficulty with likelihood ratio test is that a numerical maximization is required for each different $L$. 

Back
The Wald statistic

The Wald statistic is based on a quadratic approximation of the log likelihood function. The result is a formula that can be evaluated for each different $L$. Specifically, the Wald statistic for the null hypothesis $L\beta = 0$ is

$$W = (L\hat{\beta})^t(L(X^t\hat{V}^{-1}X)L^t)^{-1}L\hat{\beta}$$

The Wald statistic is asymptotically distributed with a Chi-square distribution.
A Comparison

Testing a hypothesis with the likelihood ratio statistic is comparing a value of negative log likelihood to its minimum. Testing a hypothesis with the Wald statistic is comparing a value of quadratic approximation of the negative log likelihood to its minimum.
In the next section we give a generalized Satterthwaite procedure for approximating the distribution of Wald statistics. In this section we review some mathematical techniques that will be needed.
Derivative of the inverse of a matrix

Let $V = V(\theta)$ be a square matrix whose elements are functions of $\theta$. Suppose we require $\frac{d}{d\theta} V^{-1}$. Start with the relationship

$$V V^{-1} = I.$$ 

Take the derivative of both sides to obtain

$$V \left( \frac{d}{d\theta} V^{-1} \right) + \left( \frac{d}{d\theta} V \right) V^{-1} = 0.$$ 

Solving gives

$$\left( \frac{d}{d\theta} V^{-1} \right) = -V^{-1} \left( \frac{d}{d\theta} V \right) V^{-1}$$

Harville [3] is a good reference on matrix algebra for statisticians.
Suppose $\hat{\theta}$ is an estimator of a vector valued parameter $\theta$. Our objective is to estimate the variance of $f(\hat{\theta})$ where $f(\theta)$ is a nonlinear function of $\theta$. We have that

$$\text{Var}(f(\hat{\theta})) \approx d^t \Sigma d$$

where $\Sigma$ is the variance matrix of $\hat{\theta}$, $d_i = \frac{d}{d\theta_i} f(\theta) \bigg|_{\theta=\hat{\theta}}$ and $d = (d_1, \cdots, d_p)^t$. This is called the linearization method.
4 The small sample distribution

A testing procedure based on the chi-square asymptotic distribution of the Wald statistic is usually too liberal. This is because actual distribution of Wald statistic almost certainly has heavier tails than does the chi-square distribution.

For a Wald statistic equal four, the $p$-value assuming a chi-square distribution with one degrees of freedom and the $p$-value assuming a F distribution with one and five degrees of freedom are shown in the following graph.
Error in the $p$-value

- $F$ with 1 and 5 df
- $\chi^2$ with 1 df

$w = 4.0$
One numerator degree of freedom

Giesbrecht and Burns [2] suggest a procedure to determine denominator degrees of freedom when there is one numerator degree of freedom. Their methodology applies to variance component models with unbalanced data. We present an extension of their technique to the general mixed model.
A one degree of freedom test is synonymous with $L$ having a single row. When $L$ has a single row the Wald Statistic can be re-expressed as

$$W = \frac{(L\hat{\beta})^2}{L(X^t\hat{V}^{-1}X)^{-1}L^t}$$

$$= \frac{U_1}{U_2}$$

where

$$U_1 = \frac{(L\hat{\beta})^2}{L(X^t\hat{V}^{-1}X)^{-1}L^t}$$

and

$$U_2 = \frac{L(X^t\hat{V}^{-1}X)^{-1}L^t}{L(X^t\hat{V}^{-1}X)^{-1}L^t}.$$
The distribution of the $U_1$ should be well approximated by $\chi^2(1)$.

Our objective is to find a $\nu$ such that the $U_2$ is approximately distributed $\chi^2(\nu)/\nu$. We gloss over the requirement the numerator and denominator must be independent if the ratio is to have an $F$ distribution.
The equation for $\nu$

Let us assume (correctly or not) that

$$\left(\frac{L(X^t \hat{V}^{-1} X)^{-1} L^t}{L(X^t V^{-1} X)^{-1} L^t}\right) \sim \chi^2(\nu)/\nu.$$  

Then

$$\text{Var} \left(\frac{L(X^t \hat{V}^{-1} X)^{-1} L^t}{L(X^t V^{-1} X)^{-1} L^t}\right) = \frac{2}{\nu}.$$  

Our approach is to approximate the left hand side of the above expression and solve for $\nu$. 

---

Back
Now

$$\text{Var} \left( \frac{L(X^t \hat{V}^{-1} X)^{-1} L^t}{L(X^t V^{-1} X)^{-1} L^t} \right) = \frac{\text{Var}(L(X^t \hat{V}^{-1} X)^{-1} L^t)}{(L(X^t V^{-1} X)^{-1} L^t)^2}$$

so

$$\nu = \frac{2(L(X^t V^{-1} X)^{-1} L^t)^2}{\text{Var}(L(X^t \hat{V}^{-1} X)^{-1} L^t)}.$$

The numerator of $\nu$ is obtained by straightforward matrix computations though $\hat{V}$ must be substituted for $V$. The rest of this section is devoted to approximating the denominator of $\nu$. 
The denominator of ν

The denominator of ν is approximated using the linearization method. Newton’s method provides an estimate of variance matrix of \( \hat{\theta} \), and \( d \) is found by applying the derivative of the inverse formula twice. The key part of this operation is

\[
\frac{d}{d \theta_i} (X^t V^{-1} X)^{-1} = -(X^t V^{-1} X)^{-1} \frac{d}{d \theta_i} (X^t V^{-1} X)(X^t V^{-1} X)^{-1}
\]

\[
= (X^t V^{-1} X)^{-1} X^t V^{-1} \left( \frac{d}{d \theta_i} V \right) V^{-1} X (X^t V^{-1} X)^{-1}
\]
For the multiple degrees of freedom case, we generalize the results of Fai and Cornelius [1]. This paper is from Alex Fai’s University of Kentucky Ph.D. dissertation. Suppose $L$ in our expression for the Wald statistic has $q > 1$ rows. Compute an orthogonal matrix $R$ and a diagonal matrix $\Lambda$ such that $L(X^t \hat{V}^{-1} X)^{-1} L^t = R \Lambda R^t$. The Wald statistic can be re-expressed as

$$W = (R^t L \hat{\beta})^t \Lambda^{-1} (R^t L \hat{\beta}) = \sum_{i=1}^{q} \frac{(r_i^t L \hat{\beta})^2}{\lambda_i}$$

where $r_i$ is the $i$-th column of $R$, and $\lambda_i$ is the $i$-th diagonal element of $\Lambda$. 
Each term in $W$ above is similar in form to the expression for a single degree of freedom Wald statistic. For the $i$-th term apply the formula for $\nu$ and denote the result by $\nu_i$. Assuming the $i$-th term is distributed $F(1, \nu_i)$, the expected value of $W$ is $\sum_{i=1}^{q} \frac{\nu_i}{\nu_i - 2}$ which we denote $E_W$.

Assuming $W$ is distributed $F(q, \nu)$ its expected value is $\frac{q\nu}{\nu - 2}$. Equating this to $E_W$ and solving gives

$$\nu = \frac{2E_W}{E_W - q}.$$
References


