TESTING THE PRODUCT OF SLOPES IN RELATED REGRESSIONS

Christopher H. Morrell\textsuperscript{a,b}, Veena Shetty\textsuperscript{c}, Terry Phillips\textsuperscript{d}, Thiruma V. Arumugam\textsuperscript{e}, Mark P. Mattson\textsuperscript{f} and Ruiqian Wan\textsuperscript{f}

\textsuperscript{a}Mathematics and Statistics Department
Loyola University Maryland
4501 North Charles St.
Baltimore, MD 21210-2699, U. S. A.
e-mail: chm@loyola.edu

\textsuperscript{b}Laboratory of Cardiovascular Sciences
National Institute on Aging
Biomedical Research Center
251 Bayview Boulevard
Baltimore, MD 21224, U. S. A.

\textsuperscript{c}MedStar Research Institute
Hyattsville, MD, U. S. A.

\textsuperscript{d}Laboratory of Bioengineering and Physical Science
National Institute of Biomedical Imaging and Bioengineering
9000 Rockville Pike, Bethesda, MD 20892, U. S. A.

\textsuperscript{e}School of Biomedical Sciences
The University of Queensland
Australia

Received: June 24, 2013; Accepted: August 5, 2013
2010 Mathematics Subject Classification: 62-XX, 92-XX.
Keywords and phrases: multivariate multiple regression, linear mixed-effects model, non-linear mixed-effects model, repeated measures, delta method.
Abstract

A study was conducted of the relationships among neuroprotective factors and cytokines in brain tissue of mice at different ages that were examined on the effect of dietary restriction on protection after experimentally induced brain stroke. It was of interest to assess whether the cross-product of the slopes of pairs of variables vs. age was positive or negative. To accomplish this, the product of the slopes was estimated and tested to determine if it is significantly different from zero. Since the measurements are taken on the same animals, the models used must account for the non-independence of the measurements within animals. A number of approaches are illustrated. First a multivariate multiple regression model is employed. Since we are interested in a non-linear function of the parameters (the product) the delta method is used to obtain the standard error of the estimate of the product. Second, a linear mixed-effects model is fit that allows for the specification of an appropriate correlation structure among repeated measurements. The delta method is again used to obtain the standard error. Finally, a non-linear mixed-effects approach is taken to fit the linear mixed-effects model and conduct the test. A simulation study investigates the properties of the procedure.

1. Introduction

Age and excessive energy intake/obesity are risk factors for cerebrovascular disease, but it is not known if and how these factors affect the extent of brain damage and outcome in ischemic stroke (stroke caused by decrease in the blood supply). Stroke, a major cause of disability and mortality in the elderly, occurs when a cerebral blood vessel is occluded and/or ruptured resulting in ischemic damage and death of neurons [2, 6]. Studies using cell culture and animal models have identified several different
proteins and signaling pathways that can protect neurons against ischemic injury including: neurotrophic factors, protein chaperones, and antioxidant enzymes.

A study was conducted that investigated the interactions of age and energy intake on the outcome of ischemic brain injury [1]. A panel of neurotrophic factors, cytokines and cellular stress resistance proteins in brain tissue samples were measured in young, middle age, and old mice that had been maintained on control or energy restricted diets prior to middle cerebral artery occlusion and restoration of blood flow. Intermittent fasting (IF) increased levels of protective proteins and decreased inflammatory cytokines in young, but not in old mice. Reduction in dietary energy intake differentially modulates neurotrophic and inflammatory pathways to protect neurons against ischemic injury, and these beneficial effects of IF are compromised during aging resulting in increased brain damage and poorer functional outcome.

The goal of the study was to evaluate some possible associations among protective factors and inflammatory products with age by dietary energy restriction. The product of the slopes for each pair of chemicals with age was of primary interest. A significant negative product indicated that an age-, diet-, and ischemia-dependent elevation in one particular factor was associated with a reduction in another factor.

Section 2 describes a number of statistical approaches to estimating and conducting tests of the product of the slopes of the regression lines. Section 3 illustrates these approaches by applying them to a pair of variables from the study. Section 4 reports on the results of a simulation study that investigates the properties of the procedures and, finally, Section 5 provides some conclusions of the study.

2. Statistical Approaches

A number of approaches to fitting linear models to pairs of variables with age, obtaining estimates of the product of the slopes and the standard error of the product, and conducting the test that the product is zero are described.
First a multivariate multiple regression model is employed. Since we are interested in a non-linear function of the parameters (the product), the delta method [3] is used to obtain the standard error of the estimate of the product. Second, a linear mixed-effects model is fit that allows for the specification of an appropriate correlation structure among the repeated measurements. The delta method is again used to obtain the standard error of the product. Finally, a non-linear mixed-effects approach is taken to fit the linear mixed-effects model and conduct the test of the product of the slopes.

Data is available on a number of animals of different ages. On each animal a number of measurements of neurotrophic factors, protein chaperones, an antioxidant enzyme, and inflammatory cytokines were obtained. Here we focus on a particular pair of these response variables. The goal is to fit a simple linear regression model for each response variable with age and to determine if the slopes of the two regressions have the same or different signs. To achieve this we estimate the product of the slopes and conduct inference on this product.

2.1. Multivariate multiple regression

The first approach to fitting a model to the bivariate data is the standard multivariate multiple regression model [4]. The model we wish to estimate is

\[ y_{ij} = \beta_{0j} + \beta_{1j} x_i + \epsilon_{ij} \quad \text{for} \quad j = 1, 2, \]

where

\[
\begin{pmatrix}
\sigma_{11}^2 & \sigma_{12}^2 \\
\sigma_{12}^2 & \sigma_{22}^2
\end{pmatrix} = R_1
\]

The quantity of interest is the product of the slopes, \( \beta_{11} \times \beta_{12} \). An estimate of this product is obtained by inserting the individual estimates of the \( \beta_{1j} \) from the fitted model. To conduct a test to determine whether this product is significantly different from 0, the standard error of the estimated product is required. Since the product is not a linear function of the parameters, the delta method is employed. This produces:

\[
Var(\hat{\beta}_{11} \times \hat{\beta}_{12}) = \hat{\beta}_{11}^2 Var(\hat{\beta}_{11}) + 2 \times \hat{\beta}_{11} \times \hat{\beta}_{12} Cov(\hat{\beta}_{11}, \hat{\beta}_{12}) + \hat{\beta}_{12}^2 Var(\hat{\beta}_{12}).
\]
Then a $t$-statistic can be computed as $t = \frac{\hat{\beta}_{11} \times \hat{\beta}_{12} - 0}{\sqrt{\text{Var}(\hat{\beta}_{11} \times \hat{\beta}_{12})}}$ with degrees of freedom given by the error degrees of freedom from the regression models. However, to compute this quantity, the covariance matrix of the parameter estimates is needed. Strangely, this quantity is not produced directly by SAS. However, SAS does provide $(X^T X)^{-1}$ and the error sum of squares and cross products matrix ($E$). The covariance matrix of the complete estimated parameter vector is then $\text{Cov}(\hat{\beta}) = (X^T X)^{-1} \otimes E \frac{1}{df_{\text{error}}}$, where $\otimes$ denotes the Kronecker product. The appropriate elements of this matrix can be extracted to compute the standard error of the estimate of the product of the parameters in (2).

2.2. Linear mixed-effects (LME) model

An alternative approach to fitting the model is to apply a linear mixed-effects model [7] to the repeated measurements (the two response variables) with an appropriate error structure. The model we wish to estimate is $y_i = X_i \beta + \varepsilon_i$, where $y_i$ is a $2 \times 1$ vector of observations on animal $i$ and $\text{Cov}(y_i) = \text{Cov}(\varepsilon_i) = R_i = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ as in (1). This can be accomplished using the repeated statement in proc mixed. The parameters of the model may be estimated using either maximum likelihood (ML) or restricted maximum likelihood (REML). The covariance matrix of the parameter vector is readily available from proc mixed and is used, as above, along with the delta method to obtain the standard error (2) and test statistic for the product of the slope estimates.

2.3. Non-linear mixed-effects (NLME) model

The linear mixed-effects model described in 2.2 can also be fit using non-linear mixed-effects model [5] software (for example proc nlmixed in SAS). This approach is useful as proc nlmixed allows for the estimation of non-linear functions of the parameters (using the estimate statement).
and standard errors are computed using the delta method without requiring the user to derive and then code the formula for the standard error of the product. One drawback is that proc nlmixed only uses maximum likelihood to estimate the model parameters. In addition, proc nlmixed has the ability to handle random effects but cannot directly accommodate the variance structure needed for the error term. Consequently, a model with error variance and random effects structures is used that replicates the needed error covariance structure (see Dale McLerran, The University of Georgia SAS-L Archives-April 2004, week 1 (#225), http://www_listserv.uga.edu/cgi-bin/wa?A2=ind0404A&L=sas-l&P=25548).

In general, the model we wish to estimate is \( y_i = X_i \beta + \varepsilon_i \), where \( y_i \) is a 2 \times 1 \ vector and \( \text{Cov}(y_i) = \text{Cov}(\varepsilon_i) = R_1 = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \). As mentioned, this can be accomplished directly using the repeated statement in proc mixed as described in Subsection 2.2.

If random effects are included the model becomes \( y_i = X_i \beta + Z_i d + \varepsilon_i \). Let the two columns of \( Z_i \) correspond to intercept \((1, 1)^T\) and an indicator variable for the second variable \((0, 1)^T\). Let us assume independent errors with homogeneous error variances, \( \text{Cov}(\varepsilon_i) = R_2 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \) and let the covariance matrix of the random effects vector, \( d \), be \( \text{Cov}(d) = D = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} \). The marginal covariance matrix is \( \text{Cov}(y_i) = R_2 + ZDZ^T = \begin{pmatrix} \sigma^2 + d_{11} & d_{11} + d_{12} \\ d_{11} + d_{12} & \sigma^2 + d_{11} + 2d_{12} + d_{22} \end{pmatrix} \). If \( \sigma^2 \) is set to zero and \( \text{Cov}(y_i) \) is equated to \( R_1(1) \), then we obtain \( d_{11} = \sigma_1^2 \), \( d_{12} = \sigma_{12} - \sigma_1^2 \), and \( d_{22} = \sigma_2^2 - 2\sigma_{12} + \sigma_1^2 \). Consequently, if the error variance is constrained to
be very small, then the use of the given random effects along with independent and homogenous errors will produce a marginal covariance matrix of the response vector that will replicate the error variance/covariance matrix when using the repeated statement in proc mixed. This approach can be used in proc mixed and in proc nlmixed to obtain the desired marginal covariance structure.

3. Results

For illustration of the approaches described above, we investigate the association of glucose regulated protein 78 (GRP78) and tumor necrosis factor-alpha (TNFα) with age in mice on an intermittent feeding regimen and with induced ischemic stroke. Figure 1 displays the data for the two variables. The plot shows that GRP78 decreases with age (least squares slope = −5.452) while TNFα increases with age (least squares slope = 17.944) so that the product of the estimated slopes is negative (−97.824). While the association of TNFα does not appear linear, when added, the quadratic terms was not statistically significant ($p = 0.113$).

Table 1 provides the estimates of the slopes of the two linear regression models along with the estimate of the product of the slopes and the standard errors and $t$-values from the various approaches described in Section 2. The slope estimates (and their product) are the same for all methods. The multivariate multiple regression approach and the linear mixed-effects model using REML provide identical standard errors of the product and consequently identical $t$-statistics. Similarly, the three approaches using maximum likelihood (the linear mixed-effects model using the repeated statement or with random effects, and the non-linear mixed-effects model) lead to identical standard errors of the product and $t$-statistics.
Figure 1. Plots of GRP78 and TNFα vs. Age.

Table 1. Results of various modeling approaches to estimating and testing the product of slopes

<table>
<thead>
<tr>
<th></th>
<th>GRP78: $\hat{\beta}_{11}$</th>
<th>TNFα: $\hat{\beta}_{12}$</th>
<th>$\hat{\beta}<em>{11} \times \hat{\beta}</em>{12}$</th>
<th>$se(\hat{\beta}<em>{11} \times \hat{\beta}</em>{12})$</th>
<th>$t$-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>REML</td>
<td>-5.452</td>
<td>17.944</td>
<td>-97.824</td>
<td>28.364</td>
<td>-3.45</td>
</tr>
<tr>
<td>ML</td>
<td>-5.452</td>
<td>17.944</td>
<td>-97.824</td>
<td>27.293</td>
<td>-3.58</td>
</tr>
</tbody>
</table>

4. Simulation Study

A simulation study is conducted to assess the properties of the estimators of the product, their standard errors, and the power of the test. Since it has been illustrated above that the three approaches to REML provide identical results and the three approaches to ML provide identical results, the simulation uses the linear mixed-effects model to obtain both ML and REML estimates. The simulation study investigates 1152 combinations of: the two slopes, the two error variances, the correlation between the errors from the two regression models, and a number of sample sizes. One thousand samples are drawn for each of these combinations and the product of the slopes and its standard error are estimated using ML and REML from each sample. The $t$-value and $p$-value are computed for each sample. The proportion of times the test rejects at a number of significance levels is determined. It is of interest to look at:

(i) the estimates of the products of the slopes to determine if the procedures provide unbiased estimates of the product;
(ii) the standard error of this estimated product as estimated by the delta method compared to the standard deviation of the estimated products among the simulated estimates which estimates the true variability of the estimated product; and

(iii) the values of the tests statistic and its \( p \)-value as a way of assessing the type I error rate and the power of the procedure.

All combinations of the following set of parameters are used:

\[
\begin{align*}
\beta_{11} & : -1, 0, 1 \\
\sigma_1 & : 1 \\
\beta_{12} & : 0, 0.5, 1, 2, 5, 10 \\
\sigma_2 & : 1, 2, 5, 10 \\
\rho & : 0.0, 0.25, 0.5, 0.75 \\
n & : 15, 30, 90, 150
\end{align*}
\]

To assess the bias in the estimates, the true product is subtracted from the mean of the 1000 estimates for each combination of the parameters and these differences are plotted against the sample size for each of the other four parameter. Figure 2a shows that the estimates tend to slightly overestimate the true value and that as the sample size increases the bias decreases. There appears to be little difference in bias for the different values of the slope of the second variable. Figure 2b indicates when the slope of the first variable was \(-1\), the bias tended to be larger than when the slope was 0 or 1. In these latter two cases the bias again decreased with sample size. Figure 2c examines the bias for different values of the error correlation. The bias increases with the correlation but decreases with sample size. Finally, Figure 2d investigates the effect of the error variance on the bias. As the variance increases the bias increases.

To investigate how well the standard errors computed using the delta method measures the true variability in the estimated products, the standard deviation of the products is subtracted from the mean of the standard errors and plotted for both ML and REML (See Figure 3). All plots show that as the sample size increases the estimated standard errors converge towards the standard deviation of the products. Particularly in small sample sizes, REML
tends to provide standard errors that are closer to the standard deviations than ML. There seems to be little effect of the slope of the second line or the error correlation on this difference. The difference tends to be larger as the error variance of the second variable increases, particularly using ML.

Finally, to determine how the power of the test is related to these variables, the proportion of times the test rejects at the 5% significance level is plotted for both ML and REML (see Figure 4). The figures show that, except when the true product = 0 (when either $\beta_{12} = 0$ or $\beta_{11} = 0$, i.e., the null hypothesis is true), the proportion of times the test rejects increases with the sample size. Interestingly, Figures 4a(i) and (ii) and 4b(i) and (ii) show that when either $\beta_{12} = 0$ or $\beta_{11} = 0$ the proportion rejecting at the 5% level tends to be lower than the nominal level. The error correlation has little effect on the rejection proportion while, as expected, the power of the test decreases as the error variance of the second variable increases.

**Figure 2.** Plots of difference between the mean of the estimates and the true product vs. sample size for various values of $\beta_{12}$, $\beta_{11}$, $\rho$, and $\sigma_2$. 
Figure 3. Plots of difference between the mean of the standard errors of the estimates and the standard deviation of the estimates of the product vs. sample size for ML and REML and for various values of $\beta_{12}$, $\beta_{11}$, $\rho$ and $\sigma^2$. 
Figure 4. Plots of the proportion of tests of $H_0 : \beta_{11} \times \beta_{12} = 0$ vs. $H_a : \beta_{11} \times \beta_{12} \neq 0$ rejected at the 5% significance level vs. sample size for ML and REML and for various values of $\beta_{12}, \beta_{11}, \rho$ and $\sigma_2$. 

Christopher H. Morrell et al.
5. Conclusions

This paper has described how to test if slopes of two regression lines have the same or opposite slopes when the data for the two lines are from a single set of subjects. To achieve this, the product of the slopes of the two regressions must be estimated and tested to determine whether this product is zero. This can be achieved in a number of ways: using multivariate multiple regression, using a linear mixed-effects model, and using a non-linear mixed-effects model. When using the first two approaches the researcher will need to derive and code the standard error using the delta method. The advantage of the non-linear mixed-effects approach is that the `estimate` statement of `proc nlmixed` allows for the estimation of non-linear functions of the parameters and automatically computes the standard errors using the delta method without requiring the user to determine the formula and code. The disadvantages of the non-linear mixed-effects approach is that only maximum likelihood is available and the user must estimate the marginal covariance matrix through the use of random effects rather than through the direct specification of the error structure.

We have illustrated the approaches using data from an experiment on variables from brain tissue samples from laboratory mice at different ages. A simulation study investigates the properties of the estimator. When sufficient data is available the procedure appears to provide unbiased estimates of the product and reasonable estimates of the variability of the product.

Acknowledgement

This research was supported by the Intramural Research Program of the NIH, National Institute on Aging. A portion of that support was through a R&D contract with MedStar Research Institute.

References


Appendix A: SAS Code for fitting the models and conducting the tests

A.1. Multivariate multiple regression

PROC GLM data=reg_data;
    MODEL y1 y2=age/i solution;
    manova h=age/printe;
    ods output invxpx = xpxi errorsscp = esscp OverallANOVA = anova parameter estimates = pests;
RUN;

data xpxi;
    set xpxi;
    if Parameter in ('Intercept', 'age');
    keep Intercept age;
run;

data esscp;
    set esscp;
    keep y1 y2;
run;
* The following two data steps extract needed information and store the results in macro variables to be used in IML;
data anova;
    set anova;
    if _n_ = 5 then CALL SYMPUT('df', DF);
run;
data pests;
    set pests;
if _n_ = 2 then CALL SYMPUT('beta1hat', Estimate);
if _n_ = 4 then CALL SYMPUT('beta2hat', Estimate);
run;

PROC IML;
use xpxi;
read all into xpxi;
close xpxi;
use esscp;
read all into esscp;
close esscp;
esscp2=esscp *(1/&dfe);
covbm=xpxi@esscp2;
std=sqrt(vecdiag(covbm));
prodslopes = (&beta1hat)*(&beta2hat);
c33=covbm[3,3]; c34=covbm[3,4]; c44=covbm[4,4];
varprod=(&beta2hat)**2*c33+2*(&beta1hat)*(&beta2hat)*c34+(&beta1hat)**2*c44;
seprod=sqrt(varprod);
tprod=prodslopes/seprod;
print prodslopes varprod seprod tprod;
quit;

A.2. Linear mixed-effects model

proc mixed data=mixed_data noclprint method = reml;
class var id;
model y= int1 int2 x1 x2 /noint ddfm=satterth s covb corrb;
repeated var / type=un subject=id R;
ods output R=cov solutionF = betas Covb = Covb;
run;
data covb;
set covb;
drop row effect;
run;
data betas;
set betas;
betahat=estimate;
keep betahat;
run;
PROC IML;
use betas;
read all into betahat;
close betas;
use Covb;
read all into Covb;
close Covb;
prodslopes = betahat[3]*betahat[4];
c33=Covb[3,3]; c34=Covb[3,4]; c44=Covb[4,4];
varprod = ((betahat[4])**2)*c33+2*(betahat[3])*(betahat[4])*c34+((betahat[3])**2)*c44;
seprod=sqrt(varprod);
tprod=prodslopes/seprod;
print prodslopes varprod seprod tprod;
quit;
A.3. Non-linear mixed-effects model

The linear mixed-effects model with random effects may be first used to obtain starting estimates for the non-linear mixed-effects model.

```
proc nlmixed data = nonlin_data maxiter = 1;
parms b1 = 110 b2 = 770 b3 = -5 b4 = 18 s1 = 54 s12 = -215 s2 = 17830;
fit = b1*i1 + b2*i2 + b3*x1 + b4*x2 + u1 + u2*i2;
model y ~ normal(fit, 1e-8);
random u1 u2 ~ normal([0, 0], [s1, s12, s2]) subject = id;
estimate 'ProdSlopes' b3*b4;
run;
```