MULTIVARIATE, MULTIVARIABLEMODELS FOR DEFENSIBLE INFERENCE ABOUT SHAPE

Keith E. Muller
Associate Professor, Department of Biostatistics
University of North Carolina at Chapel Hill

OVERVIEW

1 STATEMENT OF THE PROBLEM

1.1 Motivation

Interpretation of my assignment:
Describe methods which provide defensible statistical inference about shape analysis tools.

Achieve success by demonstrating:

1 *Validity* (outcomes reflect “truth”)

2 *Reliability* (same shape if repeat study)

3 *Differences* between methods
 a) computer method A better than method B
 b) computer method better than human
 c) computer method as good as human

4 *Generalizability* to population of interest

1.2 Assumptions and Implications

1) Shape analysis aimed at natural objects.

Implies need to draw inference about the *population* of objects by considering a *sample* of objects.

Common types of samples:
Population based, random survey (Gallup poll)
Sample of convenience
Multiple samples of convenience
(multicenter clinical trial)

Social and epidemiologic sciences have many variations of observational study designs, based on systematic sampling schemes

2) Nature of the sampling scheme determines the nature of the inference.

Many statisticians reserve the term *experiment* for studies that have random assignment to treatment.
Example: randomly assign images to method A or B

Random assignment allows ascribing any systematic differences between such groups as reflecting only the treatment (and/or chance).

"Causal" inference, if you are so inclined.

Greatly reduces the need for and value of controlling for nuisance variables (covariates),
such as gender, age, other demographics.

Observational studies require great care in controlling nuisance variable to maximize defensibility and precision of inference.

3) Limits of sample (implied by sampling scheme) determine the limits of inference.

Do 20 healthy kidneys represent billions?
Obviously not;
population conclusions require population samples.

4) Most measures of interest are
continuous variables (distance, angle)

5) Independent images often expensive to acquire

6) Repeated measures on each image typical;

7) Patients that contribute images vary greatly

2 A MODEL FOR TOOL DEVELOPMENT

2.1 Developing Ethical Pharmaceuticals to Achieve US FDA Approval

Phase I: *in vitro*; test tube, state promising model.

Phase II: *in vivo*; animal studies,
limited human studies of safety and efficacy.
Seek reliable and valid differences.

Phase III: multicenter clinical trials demonstrating
efficacy in diverse patients and clinics
Seek generalizability.

Achieve **approval to market** the product.

Phase IV: post-marketing adverse event reports
Seek full generalizability.

Progressively more expensive in time, $, and risk.
Proceed to next phase only after success in previous.

2.2 Developing Shape Tools

For example, automatic segmentation method.

Phase I: state promising model and describe analytic properties.

Phase II: **a)** artificial images, simple phantoms.
b) demonstrations with real cases.
c) true experiments with samples of convenience.
Seek reliable and valid differences.

Phase III: multicenter derived images demonstrating efficacy in expected clinical range of diversity of patients and clinical practice.
Seek generalizability.

Phase IV: post-marketing adverse event reports
Seek full generalizability.

Progressively more expensive in time, $, and risk.

Proceed to next phase only after success in previous.

3 IMPLEMENTATION STRATEGIES

3.1 Aim 1: Design

Implement the principles of design in
Muller, Barton, and Benignus (1984)

Kleinbaum, Kupper, Muller and Nizam (1998; ch. 1, 2, 9, 11, 12, 15, 16; skip 21)

Muller and Fetterman (2002; ch. 7, 10, 11, 16, 17)

1.1 Choose a well-focused and practical design.

1.2 Balance type I (false positive) and type II (false negative) errors rates to maximize efficiency.
Use best, appropriate data and power analysis.

1.3 Implementing multiple-study strategies to support simultaneous exploratory and confirmatory goals typical of most research.

1.4 Use principles, not pixels, to choose responses.

3.2 Aim 2: Statistical Analysis

2.1 Use best available standard methods if appropriate.

2.2 Use more general, nonstandard, "state-of-the-art" methods, when necessary.

2.3 Adapt recently developed statistical methods for medical imaging applications, when necessary.

3.3 Aim 3: Research Data Management

3.1 Archive data and programs from all statistical analyses to meet ethical needs of documentation. Creates a digital laboratory notebook.

3.2 Provide economies across studies by reusing well documented designs, analyses, and code developed for earlier studies.

3.3 Insure ease of retrieving data from old studies for planning new study with
a) statistical power calculations and
b) exploratory analysis to refine choice of analysis.

Earlier studies provide extremely valuable information (such as variance estimates).

3.4 Aim 4: Adapt New Methods

An *internal pilot study design* uses early data to estimate the variance and update sample size.

Most general model treatment in
Coffey and Muller (1999, 2000a, 2000b, 2001)

4.1 Choose study designs, power and data analyses that allow valid application of internal pilot designs

4.2 Create prototype software to implement the methods for the research at hand (part of MIP).

4.3 Develop new methods to extend internal pilots to repeated measures analysis
(basis of NCI grant, funding expected Jan. 2003)

4 LIST OF ANALYSIS METHODS

4.1 Box-Cox Transformation

(Box and Cox, 1964; Muller & Fetterman, ch. 7, 10)

$$Y_{i*} = \begin{cases} Y_i^{\pi} & \pi \neq 0 \\ \ln(Y_i) & \pi = 0 \end{cases}$$

Family of transformations indexed by π

Use regression assumption diagnostics
(on residuals, error estimates, not original data):

- skewness (symmetry)
- kurtosis (tail heaviness)
- test of Gaussian distribution

Ratio scale data (radius, distance)
often have heavy right tail.

- Suggests $\pi < 1$

Moving up ladder thickens left tail and thins right; moving down thins the left tail and thickens right.

Half-Steps on Ladder of Power Transformation

π	Transform		Description
	$\vdots$		
-2	y^{-2}		
$-3/2$	$y^{-3/2}$		
-1	y^{-1}		Reciprocal
$-1/2$	$y^{-1/2}$	$= 1/\sqrt{y}$	
“0”	$\lim_{\pi \to 0} y^{\pi}$	$= \ln y$	
$1/2$	$y^{1/2}$	$= \sqrt{y}$	Square root
1	y^{1}	original data;	Identity
$3/2$	$y^{3/2}$		
2	y^{2}	$=$	Square
	$\vdots$		

4.2 MULTIVARIABLE (MANY X's) MODELS: GLUM

One continuous response variable, transformable to provide Gaussian residuals, independent values.

Many predictors, any combination of continuous and categorical, or interactions.

Use the **General Linear Univariate Model** for "multiple regression" (one to many continuous X's)

(fixed predictor) Analysis of Variance (ANOVA)

independent groups t-test

and many useful combinations as special cases

Texts:
Kleinbaum, Kupper, Muller and Nizam (1998)
Muller and Fetterman (2002)

Matrix statement of GLUM:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

$\boldsymbol{y}$ is $N \times 1$ and contains response values.
Rows are subjects/images.

$\boldsymbol{X}$ is $N \times q$, and contains predictor values.
Rows are subjects/images and
columns are predictors: age, gender, etc.

$\boldsymbol{\beta}$ is $q \times 1$, contains unknown parameters:
regression model slopes, means, *etc*.

$e_i \sim \mathcal{N}\left(0, \sigma^2\right)$ independent of all others.

Equivalently, $\boldsymbol{e} \sim \mathcal{N}_N\left(\mathbf{0}, \sigma^2 \boldsymbol{I}_N\right)$

4.3 MULTIVARIATE (MANY Y's) MODELS: GLMM

N independent sampling units,
with p observations for each gives $N \cdot p$ total obs

Repeated measures: responses commensurate
"Multivariate": many responses in different metrics

Handled with a single theory for the
General Linear Multivariate Model

Includes many special cases:
MANOVA, (categorical X, many Y)
“multivariate” approach to repeated measures
“univariate” approach to repeated measures
Discriminant analysis
Canonical correlation
All multivariate versions of all multivariable models

Matrix statement of GLMM:

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}$$

$\boldsymbol{y}$ is $N \times p$ and contains response values.
Rows are subjects/images (independent).
Cols are repeated measures or distinct responses

$\boldsymbol{X}$ is $N \times q$, and contains predictor values.
Rows are subjects/images and
columns are predictors: age, gender, etc.
$\boldsymbol{X}$ is identical as for univariate model!

$\boldsymbol{\beta}$ is $q \times p$, contains unknown parameters:
regression model slopes, means, etc.
one column per response variable (col of $\boldsymbol{Y}$)

$[\text{row}_i(\boldsymbol{E})]' \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ indep if $i \neq i'$

$\Leftrightarrow \text{vec}(\boldsymbol{E}') \sim \mathcal{N}_{N \cdot p}(\mathbf{0}, \boldsymbol{I}_N \otimes \boldsymbol{\Sigma})$

4.4 “MIXED” MODEL

Last resort, nearly always except with large ratio of #indep sampling units /#obsrvatns per samplng unit

Desire $N/p \gg 1$

Due to very biased (optimistically small) variance estimates, and hence optimistic p values otherwise.

1) allows stating explicit model of covariance matrix among repeated measures.
For example, much work in “spatial statistics”

2) Allows *repeated covariates*, i. e., predictors that change within image/subject/repeated measure

2) allows missing data (not a big deal for us)

3) allows mistimed data (not a big deal for us)

Why no matrix formulation?
Two pages of explanation just to define notation
Also, more than one basic class of models.

4.5 USE APPROPRIATE SOFTWARE

To the mechanic with only a hammer,
every problem is a nail.

Excel®, Matlab®, and Mathematica® have few built in multivariate statistical analysis tools.

The “atoms” and syntaxes of the languages do not match the needs of statistical data analysis.

Use SAS®, SPSS®, or something competitive.

Use appropriate power analysis software.
Commercial: PASS®, NQuery®

Freeware: POWERLIB (only via matrix form)
http://www.bios.unc.edu/~muller
UNIFYPOW (O'Brien, CCF.org)

4.6 ACQUIRE APPROPRIATE ADVICE

Job security for statisticians!

Many prominent Psychometricians in Europe, as well as Biostatisticians.

Use appropriate texts and articles.
See *bibliography* and
my web site, *http://www.bios.unc.edu/~muller*

****

END OF PLANNED PRESENTATION

****

(OPTIONAL)
5. EXAMPLE GLMM SHAPE ANALYSIS

BIBLIOGRAPHY

Agresti, A. 1991, *Categorical Data Analysis*, New York: Wiley.

Coffey C.S. and Muller K.E. (1999) Exact test size and power of a Gaussian error linear model for an internal pilot study, *Statistics in Medicine*, **18**, 1199-1214.

Coffey C.S. and Muller K.E. (2000a) Properties of doubly-truncated Gamma variables, *Communications in Statistics, Theory and Methods*, **29**, 861-857.

Coffey C.S. and Muller K.E. (2000b) Some distributions and their implicaitons for an internal pilot study with a univariate linear model, *Communications in Statistics, Theory and Methods*, **29**, 2677-2791.

Coffey C.S. and Muller K.E. (2001) Controlling test size while gaining the benefits of an internal pilot design, *Biometrics*, **57**, 625-631.

Cohen, J. 1977, 1988, *Statistical Power Analysis for the Behavioral Sciences*. (1st ed., reprinting) New York: Academic Press.

Crowder, M.J., Hand D.J. 1990, *Analysis of Repeated Measures*. London: Chapman Hall.

Davidson, M.L. 1972, Univariate versus multivariate tests in repeated measures experiments. *Psychological Bulletin*, **77**, 446-452.

Davis, C.S. 2002, *Statistical Methods for The Analysis of Repeated Measurements*, Springer New York.

Gatsonis, C., Sampson, A.R. 1989, Multiple correlation: exact power and sample size calculations. *Psychological Bulletin*, **106**, 516-524.

Jones, B., Kenward, M. G. 1989, *Design and Analysis of Cross-Over Trials*, New York: Chapman and Hall.

Keselman H.J., Keselman J.C. 1984, The analysis of repeated measures designs in medical research. *Statistics in Medicine*, **3**, 185-195.

Kirk, R.E. 1982, *Experimental Design*, Monterey, Ca: Brooks/Cole.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E. 1988, *Applied Regression Analysis and Other Multivariable Methods* second edition, Boston: PWS-Kent.

Koch, G. G., Elashoff, J. D., Amara, I.A. 1987, Repeated measurements studies, design and analysis. In *Encyclopedia of Statistical Sciences*, N.L. Johnson and S. Kotz (eds.), New York: Wiley.

Kraemer, H. C., Thiemann, S. 1987, *How Many Subjects? Statistical Power Analysis in Research.* Newbury Park, Ca: Sage.

Kupper, L. L., Hafner, K.B. 1989, How appropriate are popular sample size formulas? *The American Statistician*, **43**, 101-105.

Louis, T.A. 1988, General methods for analysing repeated measures. *Statistics in Medicine*, **7**, 29-45.

Lipsey, M.W. 1990, *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, Ca.: Sage.

Maxwell, S.E. Delaney, H.D. 1990, *Designing Experiments and Analyzing Data*. Belmont, Ca: Wadsworth.

Monlezun, C.J., Blouin, D.C., Malone, L.C. 1984, Contrasting split plot and repeated measures experiments and analyses (also see Comments in same issue), *American Statistician*, **38**, 21-27.

Muller, K.E., Barton, C.N., Benignus, V.A. 1984, Recommendations for appropriate statistical practice in toxicology. *Neurotoxicology*, **5**, 113-126.

Muller, K. E., Benignus, V.A. 1992, Increasing scientific power with statistical power. *Neurotoxicology and Teratology*, **14**, 211-219.

O'Brien, R.G., Kaiser, M.K. 1985, MANOVA method for analyzing repeated measures designs: an extensive primer. *Psychological Bulletin*, **97**, 316-333.

O'Brien, R.G., Lohr, V.I. 1984, Power analysis for linear models: the time has come. *Proceedings of the Ninth Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc.

O'Brien, R.G., Muller, K.E. 1992, A unified approach to statistical power for t-tests to multivariate models. Chap 8, *Applied Analysis of Variance in Behavioral Sciences*, L.K. Edwards, ed. New York: Marcel Dekker.

Rao, C.R., Kleffe, J. 1988, *Estimation of Variance Components and Applications.* New York: Elsevier.

Timm N.H. 2002, *Applied Multivariate Analysis*, Springer, New York.