Modeling the Mean (II)

Biostatistics 653

Applied Statistics III: Longitudinal Analysis

Strengths of Analyzing Response Profiles

- Straightforward when design is balanced, timing of measures is common across individuals, and when all covariates are discrete
- Allows arbitrary patterns in mean response over time
- Allows arbitrary patterns in covariance of responses
- Thus has some robustness from risk of bias due to misspecification of mean and covariance models (makes minimal assumptions)

Weaknesses of Analyzing Response Profiles

- Not well-suited to handle measurements at different times across individuals
- Ignores time ordering of repeated measurements
- May have low power to detect group differences in specific trends over time (for example, testing linear trend in mean response over time)
- In saturated model, number of estimated mean and covariance parameters grows rapidly with number of measurement occasions (e.g., 12 parameters for 2 groups measured at 3 occasions; 75 parameters for 2 groups measured at 10 occasions!)

Parametric Curves

Fitting parametric curves in time has a number of advantages:

- Easily handles different timing of measurements across individuals
- Handles large numbers of repeated measurements
- Accounts for time ordering of repeated measurements
- Leads to more parsimonious model
- Typically greater power for testing when reasonable shapes of time trend are known (narrows the range of alternatives)

Parametric Curves

We will discuss three simple approaches for fitting parametric curves in time:

- linear trends over time
- quadratic trends over time, and
- linear splines

Higher-order polynomials or splines may also prove useful.

Linear Trend

A straight line is the simplest possible curve for describing changes in the mean response over time. In a two-group study, we can fit the linear trend model

$$E(Y_{ij}) = \beta_0 + \beta_1 Time_{ij} + \beta_2 Group_{ij} + \beta_3 Time_{ij} Group_{ij}$$

which reduces to the model

$$E(Y_{ij}) = \beta_0 + \beta_1 Time_{ij}$$

for subjects in the control group and

$$E(Y_{ij}) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) Time_{ij}$$

for subjects in the treatment group.

Linear Trend

Parameter Interpretation:

- β_0: average response for the control group at time = 0.
- β_1: time slope for the control group.
- β_2: group difference in average response at time = 0.
- β_1: group difference in time slope.

Linear Trend

In this model

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$$

and

$$X_{ij} = \left(1, Time_{ij}, Group_{ij}, Time_{ij}Group_{ij}\right)^T$$

To test whether the two groups differ in their changes in mean response over time, we test the hypothesis $H_{04}: \beta_3 = 0$.

Higher-order Polynomials

- If we model change over time using higher-order polynomials, changes in mean response are no longer constant (as in the linear trend model) over time.
- Instead, the rate of change over time may depend on whether we look early or late in study time.

Quadratic in Time

A quadratic model for a two-group study is given by

$$E(Y_{ij}) = \beta_0 + \beta_1 Time_{ij} + \beta_2 Time_{ij}^2 + \beta_3 Group_{ij} + \beta_4 Time_{ij} Group_{ij} + \beta_5 Time_{ij}^2 Group_{ij}$$

which reduces to the model

$$E(Y_{ij}) = \beta_0 + \beta_1 Time_{ij} + \beta_2 Time_{ij}^2$$

for subjects in the control group and

$$E(Y_{ij}) = (\beta_0+\beta_3) + (\beta_1 + \beta_4) Time_{ij} + (\beta_2+\beta_5) Time_{ij}^2$$

for subjects in the treatment group.

Quadratic in Time

In this model

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T$$

and

$$X_{ij} = \left(1, Time_{ij}, Time_{ij}^2, Group_{ij}, Time_{ij}Group_{ij}, Time_{ij}^2 Group_{ij}\right)^T$$

In this model, the rate of change in the control group is given by $\beta_1 + 2\beta_2 Time_{ij}$ and in the treatment group is given by $(\beta_1+\beta_4) + 2(\beta_2+\beta_5)Time_{ij}$.

To test whether the two groups differ in their changes in mean response over time, we test the hypothesis $H_{05}: \beta_4 = \beta_5 = 0$.

Centering

- We note that when using polynomials, centering time before creating polynomials can alleviate computational problems associated with collinearity.
- One would not have to subtract the mean but any fixed constant (for all study subjects).
- Scaling the data might also help computational problems; for example, we could have used gestational age in months instead of gestational age in weeks.

Centering

$\mathbf{X}$	$\mathbf{X^2}$	$\mathbf{Z = X - 20}$	$\mathbf{Z^2}$
10	100	-10	100
10	100	-10	100
10	100	-10	100
20	400	0	0
20	400	0	0
20	400	0	0
20	400	0	0
20	400	0	0
30	900	10	100
30	900	10	100
30	900	10	100

$$Corr(X, X^2) = 0.986, \quad Corr(Z, Z^2) = 0.$$

Splines

- Non-linear trends in mean response sometimes cannot be approximated by simple polynomials.
- Example: mean response increases (decreases) rapidly at first and more slowly afterwards.
- Spline models can be useful in such situations.
- Spline method is a technique to decompose a continuous predictor in piecewise polynomials.
- The pieces are defined on intervals based on the domain of the covariate.
- Endpoints of the intervals are called the “knots”.

Linear Splines

- divide the time axis into segments
- consider a model comprised of piecewise linear trends
- segment lines may have different slopes but are joined together at fixed times (knots)
- The most simple model has just one knot at time s_1. Generally, a spline model can have T knots at times $s_1, \cdots, s_T$.

Linear Splines

For the two-group design, a linear spline model with one knot is given by

$$E(Y_{ij}) = \beta_0 + \beta_1 Time_{ij} + \beta_2 (Time_{ij} - s_1)_+ + \beta_3 Group_{ij} + \beta_4 Time_{ij} Group_{ij} + \beta_5 (Time_{ij} - s_1)_+ Group_{ij}$$

where

$$(x)_+ = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Linear Splines

Thus

$$E(Y_{ij}) = \beta_0 + \beta_1 Time_{ij} + \beta_2 (Time_{ij} - s_1)_+$$

for subjects in the control group and

$$E(Y_{ij}) = (\beta_0 + \beta_3) + (\beta_1 + \beta_3) Time_{ij} + (\beta_2 + \beta_5)(Time_{ij} - s_1)_+$$

for subjects in the treatment group.

Linear Splines

In this model

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T$$

and

$$X_{ij} = \left(1, Time_{ij}, \left(Time_{ij} - s_1\right)_+, Group_{ij}, Time_{ij}Group_{ij}, \left(Time_{ij} - s_1\right)_+ Group_{ij}\right)^T$$

- For example, in the dental data if we have a knot at s1 = 10, then the Xi matrix for a male subject is given by

$$\mathbf{X}_i \begin{pmatrix} 1 & 8 & 0 & 1 & 8 & 0 \\ 1 & 10 & 0 & 1 & 10 & 0 \\ 1 & 12 & 2 & 1 & 12 & 2 \\ 1 & 14 & 4 & 1 & 14 & 4 \end{pmatrix}$$

Linear Splines

- In this model, the rate of change in the control group is given by $\beta_1 + \beta_2 I(Time_{ij} > s_1)$ and in the treatment group is given by $(\beta_1 + \beta_4) + (\beta_2 + \beta_5) I(Time_{ij} > s_1)$.
- To test whether two groups differ in their changes in mean response over time, we test the hypothesis $H_{05}: \beta_4 = \beta_5 = 0$.

Quadratic Splines

- We could also consider a quadratic spline as

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 (X - s_1)_+ + \beta_5 (X - s_1)_+^2$$

- Making $\beta_4 = 0$ is equivalent to imposing the restriction that the X-slope of the two-pieces are the same at s_1
- Necessity of two pieces can be verified by testing $\beta_4 = \beta_5 = 0$.

Polynomial Curves

- When using polynomial curves in time, one is not restricted to using unstructured covariance matrices but may explore more parsimonious models for covariance. We will explore covariance model selection shortly.

Mean Model Selection

Given one specified covariance structure, the following methods for mean model selection are often used:

- Likelihood ratio tests
- Wald test
- AIC or BIC

LRT

- The likelihood ratio test for two nested models (the larger called the full model, and the smaller called the reduced model) is constructed by comparing their maximized log-likelihoods, $\hat{l}_{full}$ and $\hat{l}_{reduced}$.
- The larger the difference between the maximized loglikelihoods, the more evidence there is that the reduced model is not adequate. The statistic $T_{LRT} = 2(\hat{l}_{full} - \hat{l}_{reduced})$ can be compared to a chisquared distribution with degrees of freedom equal to the difference between the number of parameters in the full and reduced models.
- We must base the likelihood ratio test for mean parameters on the likelihood of the data rather than the residual likelihood (that is, we must use ML, not REML, for testing mean parameters).

Wald Tests

- For nested models, we can also construct Wald tests of parameters of interest.
- In order to test the hypothesis $H_0: \boldsymbol{L}_{r\times p}\boldsymbol{\beta}_{p\times 1} = \boldsymbol{0}_{r\times 1}$, we can compute

$$W^2 = N\widehat{\boldsymbol{\beta}}^T_{\widehat{\boldsymbol{\Sigma}}^{-1}}\boldsymbol{L}^T\left(\boldsymbol{L}\widehat{\boldsymbol{C}}_{\boldsymbol{\Sigma}^{*-1}}\boldsymbol{L}^T\right)^{-1}\boldsymbol{L}\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Sigma}}^{-1}}$$

and compare this to a χ^2 distribution with degrees of freedom equal to the number of rows of L. This test is often called a multivariate Wald test.

Wald Tests

- We note that we have discussed the use of the chi-squared rather than the F distribution for testing hypotheses (SAS PROC MIXED provides F tests by default).
- The chi-squared tests are based on the large sample properties of the sampling distribution of the MLEs of β.
- One could argue that the use of these distributions in smaller samples leads to a procedure that is too "liberal" (confidence intervals too narrow) because there is an implicit assumption of infinite denominator degrees of freedom. (Note however that once we reach N=100 in a univariate sample, the normal and t distributions are essentially identical.)

Wald Tests

- One problem with using t and F distributions with longitudinal data is that the denominator degrees of freedom associated with tests is not easy to determine except in special cases. A number of methods are available in SAS for computing the denominator degrees of freedom; however, their small sample properties in longitudinal data analysis have not been extensively studied.

AIC and BIC

- Often, investigators wish to choose among models that are not nested. For example, in the dental data, we may wish to compare a model with a linear term in age to a model with a linear term in the natural logarithm of age. In this case, likelihood ratio or Wald tests do not apply.
- Many information criteria (IC) follow the form

$$IC = -2\hat{l} + qp$$

where $\hat{l}$ is the maximized log-likelihood (using ML for mean model selection) of the model under consideration, q is a penalty for model complexity, and p is the number of parameters.

- When we are doing mean model selection, we use maximum likelihood estimation, and p is the number of rows in β.

AIC and BIC

- One simple IC is proportional to the log-likelihood itself (q=0). The model with the biggest log-likelihood (and thus the smallest value of an IC with q=0) is in some sense the “best” model.
- However, it is known that adding parameters to a model does not decrease the log-likelihood, and we would like to have a model that is parsimonious as well as a good fit to the data.
- In order to do this, we make the IC larger by a penalty q that is scaled by the number of parameters in the mean model, p.

AIC and BIC

The penalty q determines which IC is being used. Some popular IC for mean model selection are described below:

- AIC: q = 2
- BIC (Schwartz Bayesian Criterion or SBC): q = log(N)
- AICc: finite-sample corrected version of AIC: AICc=AIC+$\frac{2p(p+1)}{N-p}$

AIC and BIC

- The AIC and BIC are the most popular information criteria in use.
- The BIC places a much larger penalty on model complexity; however, it can be derived as a Laplace approximation to the posterior model probability and will rank models in the same order as posterior model probabilities under certain conditions.
- One drawback to the IC is that they are not calibrated; that is, we are not in general able to assess whether one model is "significantly" better than the other if its IC takes a smaller value.

Example: Dental Study

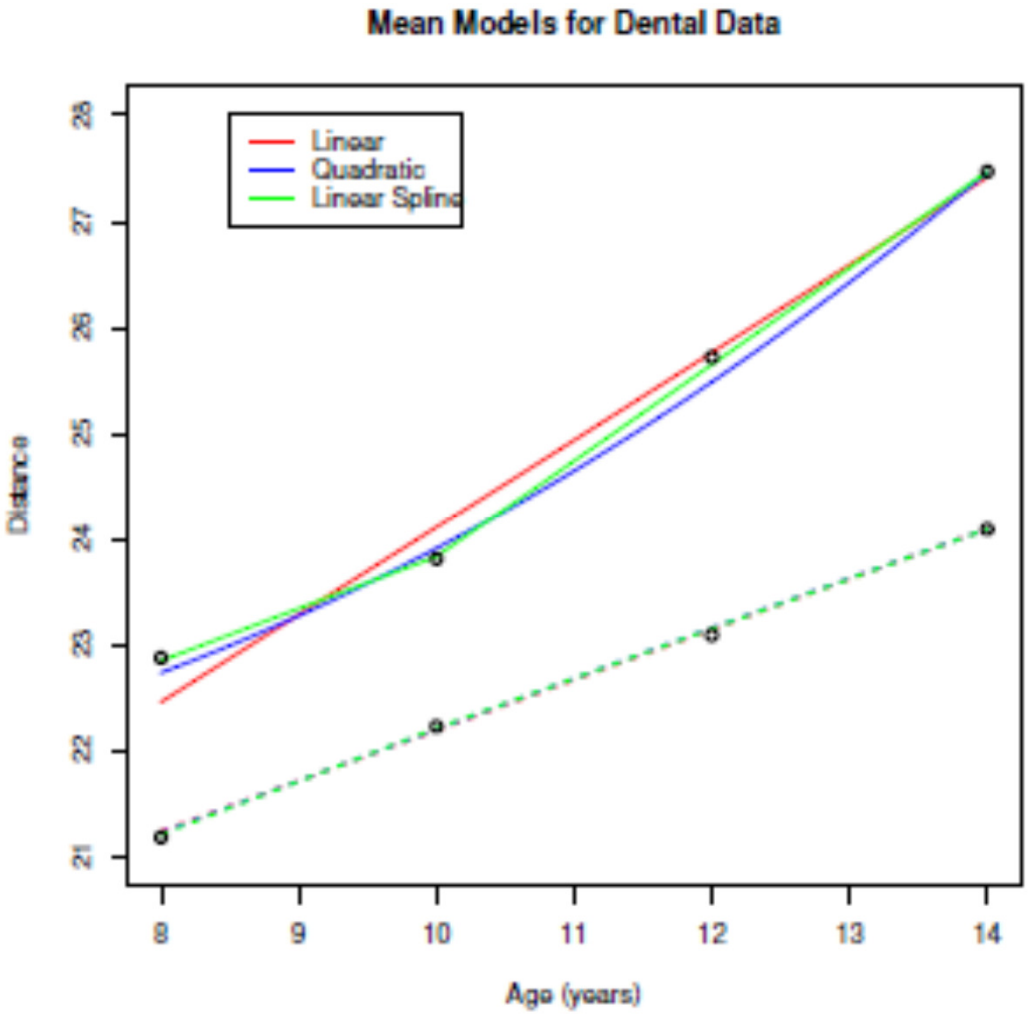
- Using the dental study data, we consider linear and quadratic polynomial models as well as a linear spline with a single knot at age=10.
- For the sake of simplicity, we assume a common unstructured covariance across gender in all cases, so that $V(Y_i) = \Sigma$.

Example: Dental Study

Mean models fitted:

- $E(Y_{ij}) = \mu_{ij}$
- $E(Y_{ij}) = \beta_0 + \beta_1 Time_{ij} + \beta_2 Male_{ij} + \beta_3 Time_{ij} Male_{ij}$
- $E(Y_{ij}) = \beta_0 + \beta_1 Time_{ij} + \beta_2 Time_{ij}^2 + \beta_3 Male_{ij} + \beta_4 Time_{ij} Male_{ij} + \beta_5 Time_{ij}^2 Male_{ij}$
- $E(Y_{ij}) = \beta_0 + \beta_1 Time_{ij} + \beta_2 (Time_{ij} - 10)_+ + \beta_3 Male_{ij} + \beta_4 Time_{ij} Male_{ij} + \beta_5 (Time_{ij} - 10)_+ Male_{ij}$

Example: Dental Study



Example: Dental Study

Time model	p	$-2\widehat{l}$	*AIC*	*BIC*
Linear	4	419.5	447.5	465.6
Quadratic	6	417.0	449.0	469.7
Cubic	8	416.5	452.5	475.8
Spline	6	416.6	448.6	469.3
Saturated model	8	416.5	452.5	475.8