

Textbook Notes

""

9/19/2021

Longitudinal Textbook

#Chapter 1 Longitudinal and Clustered Data

The goal of longitudinal analysis is to observe change over time and analyze the factors that contribute to change. By taking repeated measures of the same person at different times, we can observe this change. Cross sectional analysis and other methods are not able to accomplish this. Since each measure is not independent of one another and correlation exists, specific unique methods must be implemented. Not only can we observe change over time in individuals, but we can look at higher-level grouping, such as change in schools, counties, and organizations.

To do longitudinal analysis we implement general regression model: a regression model shows exactly how the mean of the response variable changes depending on the covariates.

Longitudinal analysis can be applied widely to different fields. Most commonly is used when analyzing health data, or measuring the effectiveness of some sort of treatment.

#Chapter 2 Longitudinal Data: Basic Concepts

The key component of longitudinal analysis is that there needs to be 2+ observations of the response variable taken at different times for each individual, but they do not need be spaced out evenly. We refer the measure of individual change as a “response trajectory”, with one type of response trajectory being difference scores. LA and the hypotheses constructed can be done in multiple ways; for example, we could look at just post-treatment results between two groups, or compare the rate of decline in symptoms. Since we are taking multiple observations of the same person, we eliminate some within individual variability such as gender/sex/socioeconomic status that could interfere with results. In LA, missing data is a very common feature. One important component of LA is that there is non-zero correlation among repeated measures, which violates regular regression assumptions. We construct correlation and covariance matrices between all the repeated measures for a given individual, and we see that there is unequal variance between measurements over time, but overall all correlations are positive. Repeated measures that were obtained closer to each other have higher covariance.

There are three sources of variation: (1): between-individual, which stems from the fact that repeated measures of one person are more similar than to other individuals. (2): Within individual biological variation: heart rate, cholesterol (3): Measurement error. Measurements that are unreliable will shrink correlation between data points.

Key takeaway: Correlation is super important! Ignoring correlation leads to overestimating the variation of measure of change.

#Chapter 3 Overview of Linear Models for Longitudinal Data

One of the assumptions of longitudinal analysis is that the responses have a multivariate normal distribution, but it is not required.

Structure of covariates: If there are p covariates, there exists a $p \times n$ matrix that captures all covariates in the analysis for each distinct measurement. There are FIXED covariates (like gender) that do not change

and those that change over time. The linear regression model follows a more complex version of $y = \beta x + e$, where β 's are unknown regression coefficients that relate the mean response to the covariate.

We assume that the errors have a distribution with a mean centered at 0. Conditional distribution of the response variable has a mean centered at $X_i\beta$, and can be multivariate normally distributed. Univariate normality in separate components DOES NOT equal multivariate normality. Assumptions of multivariate normality are not needed when data are complete.

We can plot mean response over time, and use smoothing techniques like LOWESS to estimate missing responses. Imagine a window centered at time t . A straight line is fitted inside the window, but observations closer to the center are given more weight. Missing values are predicted by using the fitted line.

The mean is the parameter of interest we want to model but modeling covariances is important as well. There are several ways to model them, one being estimating every possible covariance pair, but this can be very time consuming. We can also use random effects to model the covariance, for example random effects: using a specific single individual random effect variable to account for correlation among repeated measures. This is also known as a randomly varying intercept.

Chapter 4

Maximum likelihood: We use maximum likelihood as a way to estimate what the values of β and σ^2 are. We choose the value of the estimates that are most likely to occur based on the data observed. In other words, we are maximizing the joint probability of the RV occurring based on the observed data. When we assume that observations are independent of one another, we can use the “ordinary least squares” estimate for β (this is only for beta, not variance). However, if we do not assume that repeated measures are independent, and we assume that the covariance is known, we can use the GLS estimator for beta, and it also will be unbiased. If we don't know the covariance the estimator isn't unbiased, but it is still consistent.

Missing data: There are two types of missing data: those that are missing completely at random (MCAR), and those that are missing at random (MAR). MCAR is when the probability of observations being missing does not depend on the obtained observations nor the underlying population. However, for MAR, the probability of observations being missing depends on the observed responses, but not the underlying population that should have been sampled. For example, if observations were stratified by their similarity, their probabilities of missingness would be different. The difference between the two types of missing data effects the biasedness and accuracy of the estimators. When there is no assumption of multivariate normal distribution and there is MAR data, then the GLS estimator is no longer valid. However, estimates of beta are valid with MAR when the data is assumed to be from a multivariate normal distribution.

Statistical Inference: It is possible to construct confidence intervals using similar methods for proportions and means. We can also use the likelihood ratio tests to determine whether slopes are suitable for LA. The use of standard normal and chi-squared distributions is valid when the covariance is known or the number of observations is large enough.

Restricted maximum likelihood estimation: in terms of estimating the variance, the ML estimate produced a biased estimate. But if we use REML estimation we will get an unbiased estimator for the variance, because it separates the data used for estimating beta from when it is used to estimate the covariance. In REML we eliminate β from the likelihood function so it is only defined in terms of the covariance. It is best to use REML for estimation of covariance and use the GLS estimator for estimating beta.

Chapter 5: Modelling the mean: analyzing response profiles

There are three effects of interest when analyzing response profiles in LA: 1. group x time interaction effect (are the mean response profiles similar in groups over time?) 2. time effect (assuming mean response profiles are flat, are the means constant over time?) 3. Group effect (do the mean response profiles coincide?)

1. is the primary interest of analysis. How we analyze response profiles depends on whether the study is randomized or observational. When the study is randomized and baseline measurement is taken before treatment assignment, the mean response at occasion 1 is independent of the group. For an observational study, there is no assumption that the groups have the same mean response at the beginning. To test for significance of the group x time effect, we have a null hypothesis that the difference in means between the n groups is constant over time.

General linear model formation: To express the model for LA for G groups and n occasions of measurement, we have G x n parameter for the G mean response profiles. For example, for 2 groups measured at 3 occasions, we have 6 slope parameters. if $\beta_1 - \beta_3$ represent slope parameters for mean responses in group 1 and $\beta_4 - \beta_6$ represent slope parameters for mean responses in group 2, our null hypotheses would be that $(\beta_1 - \beta_4) = (\beta_2 - \beta_5) = (\beta_3 - \beta_6)$. We can also write these equations using $L\beta$, a linear transformation? If we have a response missing at a specific occasion, say the last one.. we can remove the last row from full data design matrix for subjects from the first group. Once the covariance of Y_i is specified, then mlm estimation and tests of group x time interaction are possible.

We can run wald's test to identify whether there are differing patterns of change from baseline across the two groups, but only REML estimate of β will give us actual slope interpretations.

One DF tests for group x time interaction For studies with large number of measurements there might be too many degrees of freedom causing the test to become less sensitive to interaction effects. We can calculate a 'contrast', for example computing the mean response from occasion 2:n (excluding baseline measurement) and subtracting the mean baseline for each group. We can also calculate the area under the curve by subtracting the baseline mean from each mean and calculating the area under the trapezoid.

Adjusting for baseline response: - ANCOVA method: the baseline measurement gets taken out of the response vector and becomes a covariate. - Contrast method: subtracting the mean baseline measure from the mean of the other measurements. If it is an observational study, don't use ANCOVA approach because the baseline variable might have a relationship with the other covariates in the model, which could create a confound. However this is not a problem for randomized studies. ANCOVA is a more powerful test and can yield smaller errors, but is not suitable for every type of study. The interpretation for ANCOVA is also conditional: whether an individual belonging to one group is expected to change more/less than someone else, given that they both have the same baseline. ANCOVA does impose a structure on the covariance matrix, so it is also okay to just retain the baseline measurement as part of the response vector and assume the group means are equal at baseline (when it is appropriate to do so based on the context of the situation)

Strength and weaknesses of analyzing response model Strength: can easily be extended to handle cases where individuals can be grouped by more than 1 factor. Weakness: not good at dealing with mistimed measurements, and because response profiles allow for arbitrary patterns in mean response, the results of the analysis provide only a general/broad statement about the group differences.

Chapter 6: Modeling the mean: parametric curves.

Intro: Analyzing the mean response profile has some flaws, such as that it ignores the time-ordering of measurements, and that the null hypothesis of no time x group interaction is too broad and lacks power. Instead, we can fit parametric curves. Since most measurements in LA increase/decrease in a smooth pattern, fitting curves is suitable, and the alternative hypotheses are more specific compared to response profiles.

Polynomial trends in time: In an example comparing a treatment group to a control group, we can fit a linear curve using the following equation: $E(Y_{ij}) = \beta_1 + \beta_2 Time_{ij} + \beta_3 Group_i + \beta_4 Time_{ij} \times Group_i$ If $\beta_4 = 0$, then the two groups do not differ in terms of changes in the mean response over time.

For quadratic trends, the changes in mean are no longer constant, and the rate of change depends on the time. $E(Y_{ij}) = \beta_1 + \beta_2 Time_{ij} + \beta_3 Time_{ij}^2 + \beta_4 Group_i + \beta_5 Time_{ij} \times Group_i + \beta_6 Time_{ij}^2 \times Group_i$

To test for significance, test for higher power first. If it isn't significant, remove and test next order. In quadratic modeling, time and time² are heavily correlated. To remove this, we can center time in relation to the mean of all time values.

Linear spines Sometimes when the mean response increases rapidly for some duration and then more slowly thereafter, we use linear spines, which fits different intercepts across time measurements. The points at which the time is divided up are called knots. The location of knots depends on the context of the data and previous research.

An example of that model is here, where t^* represents the time of a knot. $+$ means that the expression is only evaluated if it is positive, otherwise it is 0.

$$E(Y_{ij}) = \beta_1 + \beta_2 Time_{ij} + \beta_3 (Time_{ij} - t^*)_+ + \beta_4 Group_i + \beta_5 Time_{ij} \times Group_i + \beta_6 (Time_{ij} - t^*)_+ \times Group_i$$

Chapter 7: Modeling the covariance

Intro: Fitting the covariance model correctly is important for accurately estimating the regression parameters, especially when data are missing. If there is correlation between repeated measures, the variability of within-individual differences is always smaller than the variability of between-individual differences. In addition, the larger the correlation, the smaller the variability of between-individual differences. When the correlation is not accounted for, estimates of β are inaccurate and standard errors are too large.

Longitudinal data need to be modeled in two ways: the conditional mean response over time and the conditional covariance. These two models are interdependent, and the covariance model chosen ultimately depends on how the conditional mean response is modeled. There are several covariance models:

Unstructured Covariance Model

Suitable model when the number of measurements is small and individuals are measured at the same times. We can allow all covariance values to be arbitrary. The advantages of this model are that it resembles reality where variances are rarely constant over time. However, it can be time consuming to estimate.

There are n measurements and $\frac{n \times (n-1)}{2}$ parameters.

Compound Symmetry

Variance is constant across occasions, and $Corr(Y_{ij}, Y_{ik}) = \rho$

$$\begin{pmatrix} 1 & p & \dots & p \\ p & 1 & \dots & p \\ \dots & p & 1 & \dots \end{pmatrix}$$

This model only has two parameters to estimate but makes very strong assumptions that don't align with LA, where the correlation between measurements are expected to decrease with separation in time.

Toeplitz

Any pair of responses that are equally separated in time have the same correlation

$$Cov(Y_i) = \begin{pmatrix} 1 & p_1 & p_2 & \dots & p_{n-1} \\ p_1 & 1 & p_1 & \dots & p_{n-2} \\ p_2 & p_1 & 1 & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

Measurements need to be made at the same intervals. Has n parameters.

Autoregressive constant (special case of Toeplitz model)

$$\begin{pmatrix} 1 & p & p^2 & \dots & p^{n-1} \\ p & 1 & p & p^2 & \\ \dots & p & 1 & \dots & \end{pmatrix}$$

Correlations decline over time as separation increases between measurements, but they normally don't decline this quickly.

Banded

Correlation is 0 beyond a pre-specified interval

For example, a band size of 3 indicates that correlations between measurements taken 3+ instances apart will be zero.

$$\begin{pmatrix} 1 & p_1 & 0 & \dots \\ p_1 & 1 & p_1 & \dots \\ 0 & p_1 & 1 & \dots \end{pmatrix}$$

Again, similar to the autoregressive constant, it makes a strong assumption about how quickly correlation decreases.

Exponential

When measurements are not equally spaced we can use t_i, \dots, t_{in} for the i th person. The correlation between measurements decreases exponentially with separations between them. Again, like banded and autoregressive constant, it assumes correlations rapidly creases to 0, which is rarely observed.

$$Corr(Y_{ij}, Y_{ik}) = \rho^{|t_{ij} - t_{ik}|}$$

Hybrid Models

We can combine compound symmetry with exponential model to explain the covariation. In this way, the correlation between replicate measurements is less than 1 (to account for measurement error), and as separation increases, correlation has a minimum value of $\frac{p_1\theta_1^2}{\theta_1^2 + \theta_2^2}$ rather than 0.

Choice among covariance models: The model for the covariance is trying to account for the residuals of the model for the mean. So the covariance model should be based on the "maximal" model for the mean that minimizes any errors. Choosing between covariance models depends on comparing maximized likelihoods of each of the models. If a pair of models are nested, we can do a Likelihood Ratio Test, where reduced model is the "special case" model. For example, the compound symmetry model is nested within the Toeplitz model. We compare the maximized REML log-likelihood of the two models.

If models are not nested, we can use AIC, and select the model that minimizes it. The AIC attempts to capture how complex a model is and its adequacy of fit, but penalizes models for every additional parameter.

Chapter 8 Linear Mixed Effects Models

In linear mixed effects models, there is a distinction between fixed effects, which are shared by all individuals in a population, and random effects that are unique to an individual. The model accounts for both of these characteristics. Some regression coefficients are allowed to vary randomly from one individual to another.

Introducing random effects changes the covariance model, in that they can be expressed as functions of time and with few parameters. Linear mixed effects models do not require balanced data, or same number of observations, or consistent timing of measurements.

Random Intercept model

$$Y_{ij} = X'_{ij}\beta + b_i + \epsilon_{ij}$$

Where b_i is the random subject effect and ϵ is the measurement error. Both effects are random, with mean 0 and $Var(b_i) = \sigma_b^2, Var(\epsilon_{ij}) = \sigma^2$

The conditional mean is $E(Y_{ij}|b_i) = X'_{ij}\beta + b_i$ Marginal mean is $E(Y_{ij}) = X'_{ij}\beta$ Where β represents the patterns of change in the pop of interest, and b_i describes how an individual deviates from the population average.

More on LME

LME can expressed as $Y_i = X_i\beta + Z_i b_i + \epsilon_i$

β is a $p \times 1$ vector of fixed effects b_i is a $q \times 1$ vector of random effects X_i is a $n \times p$ matrix of covariates Z_i is a $n \times q$ matrix of covariates

The subset of regression parameters that vary randomly are found in $Z_i b_i$ comes from a multivariate normal distribution with mean 0 and covariance matrix G . ϵ_i are independent of b_i , come from multivariate normal distribution with mean 0 and covariance matrix R_i . sampling measurement errors are assumed to be uncorrelated with each other.

The covariance matrix $Cov(Y_i)$ can be expressed as a function of time. The covariance matrix G for random effects should be left unstructured. We can also center the times of measurements in order to avoid problems of collinearity. We can center the times of measurement at some fixed age within the age range of the study.

Random effects covariance structure

Conditional Covariance: $Cov(Y_i|b_i) = Cov(\epsilon_i) = R_i$ Marginal Covariance: $Cov(Z_i b_i) + Cov(\epsilon_i) = Z_i G Z_i' + R_i$
The introduction of random effects b_i creates correlation among components of Y_i .

Choosing covariance models: Number of covariance parameters is the same regardless of # of measurements, but number of correlated random effects can be tested using the likelihood ratio test. ## Prediction of random effects

To predict individual response profiles, we need to predict b_i , which is a random variable. Predictor aka BLUP: $E(b_i|Y_i) = GZ_i' Cov(Y_i)(Y_i - X_i \hat{\beta})$ when incorporating estimated $Cov(Y_i)$ using REML, we can get the BLUP predictor. the empirical BLUP estimator brings the subject's predicted response profile towards the population-averaged mean response profile, and depends on how large R_i aka within-subject variability

Chapter 9: Fixed Effect Models

Introduction

Developed in order to remove potential confounding effects of time invariant covariates. Useful for when time variant covariates are of interest .

Linear mixed effects model implicitly makes stronger assumptions about the stable, time-invariant, characteristics of individuals. Specifically, mixed effects models treat the α as random, rather than fixed, and

assume that they are uncorrelated with the measured covariates included in the regression model; that is, the α_i are assumed to be uncorrelated with X_i . However, with the fixed effects model the fixed effects (rather than random effects) are assumed to be correlated with the covariates.

The linear fixed effects model: $Y_{ij} = X'_{ij}\beta + W'_i y + \alpha_i + \epsilon_i$

Where y represents the covariates with fixed effects. This model can only estimate coefficients for time-varying covariates and not time-invariant covariates because of their collinearity with α_i . β can be estimated with normal OLS regression.

Fixed effects model removes potential for bias by removing confounding effects of time invariant covariates which are captured in α_i

The positives of fixed effects models is that since it allows for correlation between α_i and X_i , it can reduce the bias of the estimates. However, fixed effects model ignores between subject variation and focuses only on within-subject variation, which is sometimes not our research goal, and can produce larger standard errors.

Combination of fixed effect and mixed effects models

There is a model that is able to recognize both the “cross sectional” and “longitudinal” aspects of longitudinal data: 1) responses can vary between subject and 2) response varies over time

$$Y_{ij} = X_{ij}^* \beta^{(L)} + \bar{X}_i' \beta^{(c)} + W'_i y + Z'_{ij} b_i + \epsilon_i$$

Chapter 10: Residual Analysis and Diagnostics

Plotting residuals

Residuals are a measurement of the difference between the observed response and the predicted response, and can assess the fit of a model. In LA, there is a vector of residuals for each individual $r_i = Y_i - X_i \hat{\beta}$. This vector has a mean of zero. We can plot this vector r_i against the predicted mean response $\hat{\mu}_i = X_i \hat{\beta}$. In the standard interpretation of linear models, we should expect this scatterplot to have no systemic pattern and have a random scatter around mean 0. However, the components in the r_i vector are correlated and do not have constant range, and residuals may also be correlated with covariates, leading to a possibility that we would see patterns in the scatterplot.

Thus, we can “de-correlate” residuals so they resemble those from standard linear regression. Using the “Cholesky decomposition” transformation, we can transform residuals to have constant variation and 0 correlation. The interpretation of the residuals also changes so that all elements after the first represent standardized deviations from mean response *given* the previous observation. Now that we have a transformed set of residuals, usual residual diagnostics can be applied. We can also re-estimate β using transformed values of Y_i^* and X_i^* that undergo the same transformation.

We can also calculate a Mahalanobis distance summary measure that captures the distance between the observed vs fitted responses. Once this value is compared to a chi-sq distribution, we can identify outliers which will have small p-values.

Aggregating Residuals

An alternative to plotting residuals, whose interpretations can be subjective, is to aggregate residuals over both the covariates and fitted values. We compare these aggregated residuals to a reference distribution that has an assumed distribution under the assumption that the correctly specified model for the mean response was chosen. If there is any deviations in the comparison, it suggests that the model chosen to fit the observed values is not adequate.

$W_k(x)$ represents the sum of the residuals over covariate k that have a value less than or equal to x , and $W_f(x)$ represents a similar concept but for all fitted values. $W_f(x)$ is good for assessing linearity between response and covariates. We can also use a “moving” sum rather than cumulate that focuses on residuals within a window of size b .

Assessing covariance model