Introduction

Repeated measures:

- Defined broadly as data in which the response of each experimental unit is observed on multiple occasions or under multiple conditions
  - the response could be univariate or multivariate
  - "multiple" usually means "more than two"

Longitudinal data:

- Generally refers to data in which the repeated measurement factor is time

Using these definitions, longitudinal data are a subset of repeated measures data

Introduction

An alternative distinction is sometimes made:

Longitudinal:

- data collected over an extended period of time, often under uncontrolled conditions

Repeated measures:

- data collected over a relatively short time period, frequently under experimental conditions

Using these definitions, repeated measures data can be regarded as a subset of longitudinal data

As a result, we shall use "repeated measures" and "longitudinal" somewhat interchangeably

Strengths of Longitudinal Studies

- Economizes on subjects
- Subjects serve as their own controls
- Between-subjects sources of variability are excluded from the experimental error
- Provide more efficient estimators of some parameters than cross-sectional designs with the same number and pattern of measurements
- Information is more reliably quantified than in a cross-sectional study
- Provide information concerning individual patterns of change

Difficulties of Longitudinal Studies

- Other than for multivariate normal data, analysis methods are not as well-developed
- Lack of software
- Existing methods are often computationally intensive
- Dependence of multiple measurements
- Unbalanced designs due to missing data and/or attrition
- Time-dependent covariates
- Carry-over effects
  (primarily of concern when the repeated measurement factor is condition or treatment, rather than time)

General Notation

- n subjects (indexed by $i = 1, \ldots, n$)
- t_i measurement times for the ith subject (indexed by $j = 1, \ldots, t_i$)
- p covariates (indexed by $k = 1, \ldots, p$)

  between-subject covariates (time-independent)

  within-subject covariates (time-dependent)
- y_{ij} is the response from subject i at time j, for $i = 1, \ldots, n$, $j = 1, \ldots, t_i$
- x_{ij} is the corresponding $p \times 1$ vector of covariates
- Since the data may be incomplete:

$$\delta_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ and } x_{ij} \text{ are observed} \\ 0 & \text{otherwise} \end{cases}$$

General Notation

Subject	Time Point	Missing Indicator	Response	Covariates		
1	1	δ_{11}	y_{11}	x_{111}	$\cdots$	x_{11p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	j	δ_{1j}	y_{1j}	x_{1j1}	$\cdots$	x_{1jp}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	t_1	δ_{1t_1}	y_{1t_1}	x_{1t_11}	$\cdots$	x_{1t_1p}
i	1	δ_{i1}	y_{i1}	x_{i11}	$\cdots$	x_{i1p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	j	δ_{ij}	y_{ij}	x_{ij1}	$\cdots$	x_{ijp}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	t_i	δ_{it_i}	y_{it_i}	x_{it_i1}	$\cdots$	x_{it_ip}
n	1	δ_{n1}	y_{n1}	x_{n11}	$\cdots$	x_{n1p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	j	δ_{nj}	y_{nj}	x_{nj1}	$\cdots$	x_{njp}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	t_n	δ_{nt_n}	y_{nt_n}	x_{nt_n1}	$\cdots$	x_{nt_np}

Factors Involved in the Analysis of Repeated Measurements

- Nature of the response variable:
  - continuous, normally-distributed
  - categorical (binary, polytomous, ordinal, count)
  - continuous, nonnormal
- Number of subjects (n)
- Number of time points per subject (t_i)
  - $t_i = t$, no missing data
  - varying number of observations per subject
- Number and type of covariates:
  - one sample ($p = 0$)
  - multiple samples (one categorical covariate)
  - multiple samples (p categorical covariates)
  - regression (at least one continuous covariate)
  - time-dependent covariates

Notation for Special Cases

- A common set of t measurement times

Subject	Time Point	Missing Indicator	Response	Covariates		
1	1	δ_{11}	y_{11}	x_{111}	$\cdots$	x_{11p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	j	δ_{1j}	y_{1j}	x_{1j1}	$\cdots$	x_{1jp}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	t	δ_{1t}	y_{1t}	x_{1t1}	$\cdots$	x_{1tp}
i	1	δ_{i1}	y_{i1}	x_{i11}	$\cdots$	x_{i1p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	j	δ_{ij}	y_{ij}	x_{ij1}	$\cdots$	x_{ijp}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	t	δ_{it}	y_{it}	x_{it1}	$\cdots$	x_{itp}
n	1	δ_{n1}	y_{n1}	x_{n11}	$\cdots$	x_{n1p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	j	δ_{nj}	y_{nj}	x_{nj1}	$\cdots$	x_{njp}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	t	δ_{nt}	y_{nt}	x_{nt1}	$\cdots$	x_{ntp}

Notation for Special Cases

Multiple Samples:

- In some applications, repeated observations are obtained from s subpopulations (groups)
  - group h contains n_h subjects ($\sum_{h=1}^{s} n_h = n$)
- The s groups may be defined by the:
  - s levels of a single covariate
  - cross-classification of several variables
- In terms of the general notation:
  - the s groups can be described in terms of $p = s - 1$ categorical and time-independent covariates
  - y_{hij} now denotes the response at time j from subject i in group h

Notation for the Special Case of Multiple Samples

		Time Point				
Group	Subject	1	...	j	...	t
1	1	y_{111}	...	y_{11j}	...	y_{11t}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	i	y_{1i1}	...	y_{1ij}	...	y_{1it}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	n_1	y_{1n_11}	...	y_{1n_1j}	...	y_{1n_1t}
h	1	y_{h11}	...	y_{h1j}	...	y_{h1t}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	i	y_{hi1}	...	y_{hij}	...	y_{hit}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	n_h	y_{hn_h1}	...	y_{hn_hj}	...	y_{hn_ht}
s	1	y_{s11}	...	y_{s1j}	...	y_{s1t}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	i	y_{si1}	...	y_{sij}	...	y_{sit}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	n_s	y_{sn_s1}	...	y_{sn_sj}	...	y_{sn_st}

Notation for Special Cases

One Sample:

- In this case, the data can be displayed in an $n \times t$ matrix, as follows:

	Time Point				
Subject	1	...	j	...	t
1	y_{11}	...	y_{1j}	...	y_{1t}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
i	y_{i1}	...	y_{ij}	...	y_{it}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
n	y_{n1}	...	y_{nj}	...	y_{nt}

- The corresponding missing value indicators are defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$$

Outline of Topics

Types of response variables:

- continuous, normally-distributed
- categorical (binary, polytomous, ordinal, count)
- continuous, nonnormal

Types of statistical methodology:

- normal-theory
- weighted least squares (GSK)
- randomization model (CMH)
- generalized linear model extensions
- nonparametric

Types of data:

- one sample ($p = 0$)
- multiple samples (one categorical covariate)
- multiple samples (p categorical covariates)
- regression (quantitative covariates)