COMPLEX TRAJECTORIES

Complex Trajectories:

Longitudinal methodological approaches

Compiled by Janine Jongbloed, IREDU, Université de Bourgogne

Co-funded by the Erasmus+ Programme of the European Union

Author:

Janine Jongbloed

The following document is embedded in a whole MOOC course on Longitudinal analysis techniques developed in the framework of the same Complex Trajectories project. Those interested in doing the MOOC should consider the following procedure:

1. Access the AULAbERTA space through the link: https://aulaberta.uab.pt/

2. Register in the platform a. Select one of the available languages (Portuguese or English) b. Follow the instructions given in the platform

3. After creating the account, they should access the MOOC Longitudinal Analysis through the link: https://aulaberta.uab.pt/course/view.php?id=94

Co-funded by the Erasmus+ Programme of the European Union

Content

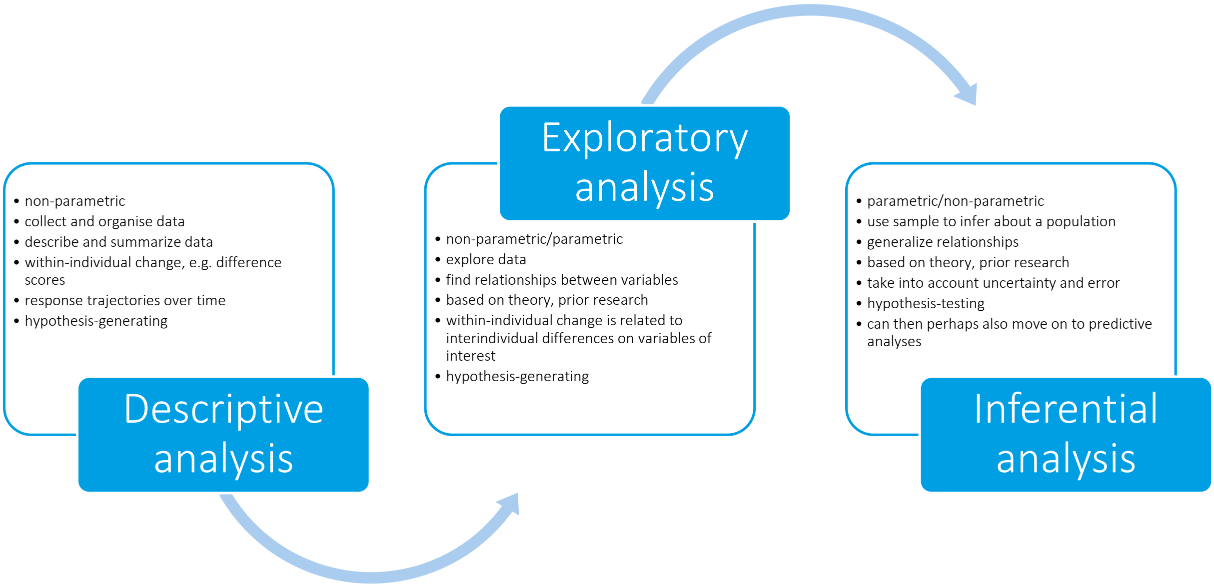
Co-funded by the Erasmus+ Programme of the European Union

1. Introduction

Longitudinal data (or panel data) => follow the same individuals at different points in time, often including repeated observations of the same variables over time for the same individuals (Herle et al., 2020), which allows us to analyse the relationships between earlier events and characteristics and subsequent outcomes (Andres, 2013)

Person-centred analyses => capture the relationships among individuals in order to classify individuals into distinct groups or categories based on their response patterns so that individuals in a particular group are similar to each other and different from individuals in other groups (Muthén and Muthén, 2000)

Variable-centred analyses => identify significant predictors of outcomes, and describe how dependent and independent variables are related (e.g., regression, factor analysis, and structural equation modeling) (Jung and Wickrama, 2008).



Co-funded by the Erasmus+ Programme of the European Union

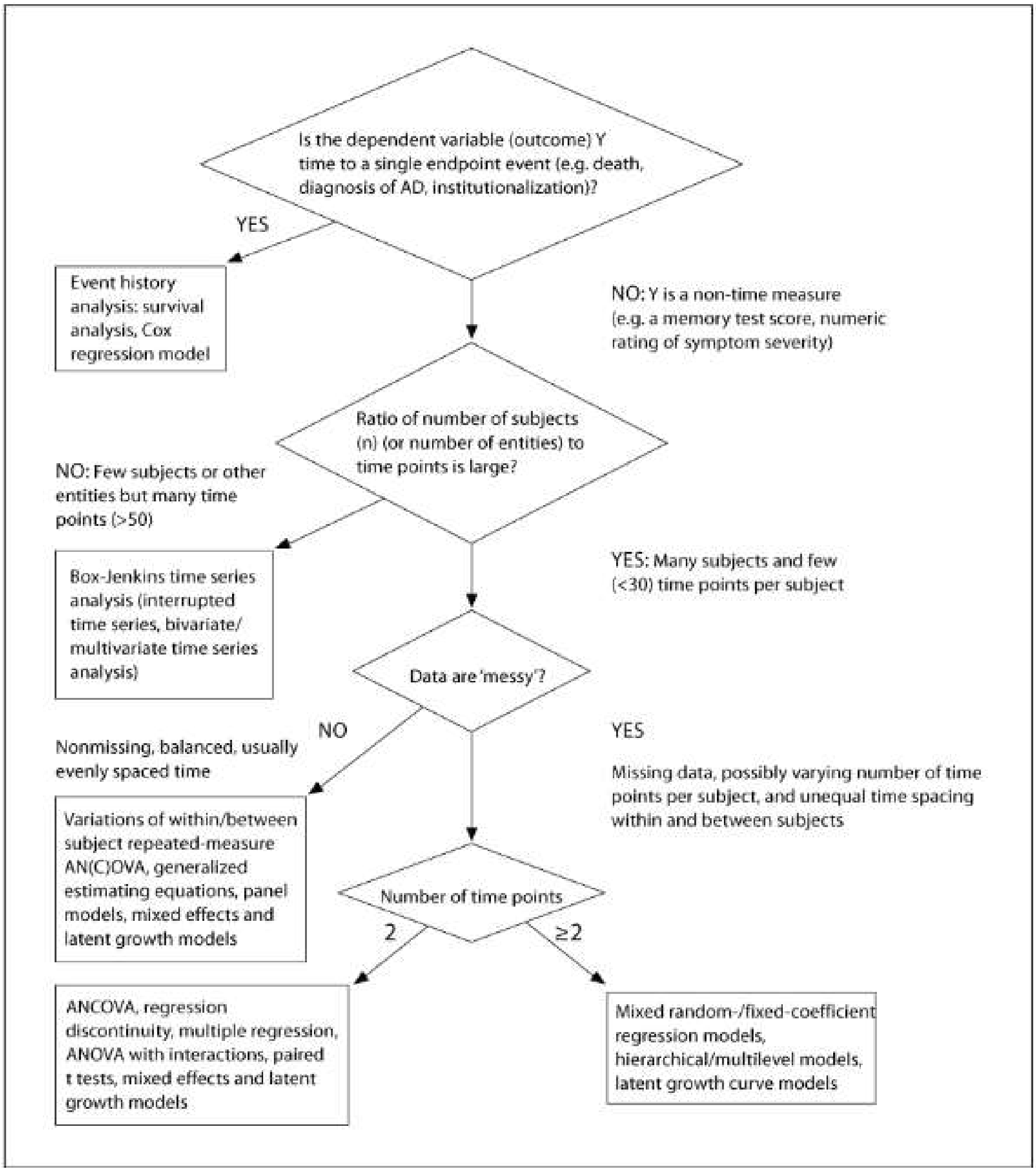
1.1. Challenges:

Because longitudinal data measure the same variables for the same individuals over time, responses are necessarily correlated with one another (a response at time 1 provides information about the likely value at time 2 or time 3). Thus, they are not independent from one another. Furthermore, the variance or variability in the values of responses changes over time. Therefore, many common techniques cannot be used because they rely on an assumption of homogeneity of variance (Fitzmaurice et al., 2008; Fitzmaurice and Ravichandran, 2008).

1.2. Research questions and data:

The choice of longitudinal methods will necessarily depend first and foremost on the research question that one aims to answer, as well as on the data that one has at one's disposal. This may involve pragmatic and iterative choices to make the best of what one can do (or has). Some data availability constraints are illustrated in the data analysis flowchart below (from Locascio and Atri 2011). Key analysis options for our purposes include sequence analysis, growth curve analysis, event history methods, and multilevel mixture models (described later in this document).

Co-funded by the
Erasmus+ Programme
of the European Union



Source: Locascio and Atri (2011, p.339)

2. Types of analyses

2.1. Person-centred analyses

2.1.1. Non-parametric procedures

Non-parametric approaches are statistical procedures that do not make assumptions about the shape or form of the sample's probability distribution. These can be compared to parametric statistical procedures, which do make assumptions about the shape of the distribution (e.g., normal distribution) and parameters (e.g., means and standard deviations) (Hoskin, 2010).

2.1.1.1 K-means, hierarchical cluster, & cluster analysis

Cluster analytic techniques aim at data simplification and prediction through classification (Distefano and Mindrila, 2013; Nadif and Govaert, 2010). This involves taking a large, heterogeneous group and dividing it into smaller, homogeneous groups. Once divided, these groups are similar to within-group members on responses to a set of variables and have different responses from cases in other groups. There are many different clustering algorithms that could be used for classification. One's choice depends on one's purpose in clustering, the type of input variables (more below), etc. Importantly, different methods may produce different results.

Hierarchical agglomerative methods are most popular. They use a proximity matrix to join the two most similar cases, then the next two cases/clusters with the smallest distance and so on for N − 1 iterations. This matrix could be determined in different ways: single linkage (or nearest neighbor) method adds members based on similarities to any other member in a cluster, complete linkage (or furthest neighbor) method to all members of a cluster, etc. The most popular approach is Ward's method, which reaches the minimum variance, or error sum of squares (ESS), within clusters.

The final stage involves examining the dendrogram and other measures related to the cluster solutions to find the solution showing the best fit in terms of statistical differentiation and theoretical interpretability of the group characteristics. Although clustering techniques are not specific to longitudinal analysis, they can be used on longitudinal data measures. Indeed, cluster techniques can be used on various input data, including distance matrices from correspondence analysis (CA) or optimal matching analysis of sequence data (OMA).

We can also cluster qualitative time varying variables using Self-Organizing Maps (SOM) (Rousset and Giret, 2007; Rousset, Giret and Grelet, 2012). A description of SOM techniques can be found here: https://eric.univ-lyon2.fr/~ricco/cours/slides/en/kohonen_som.pdf

Co-funded by the Erasmus+ Programme of the European Union

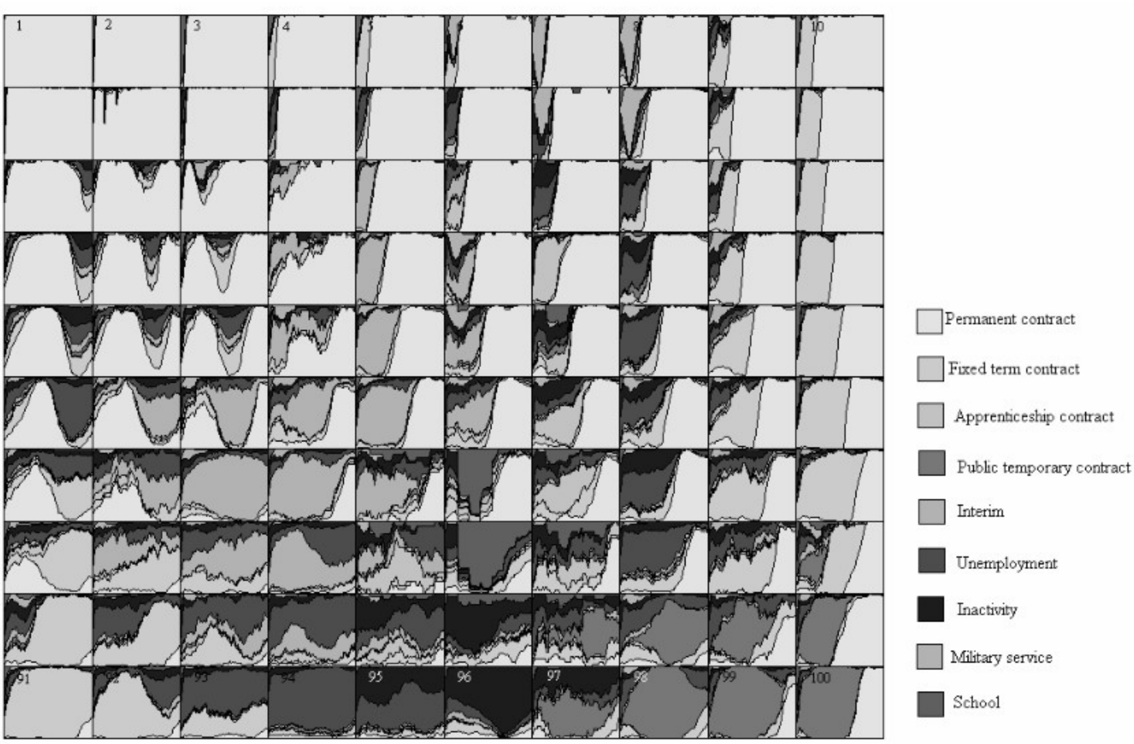


Fig. 1. Typology of career paths with SOM: any unit of the map represents with Chronograms the evolution in time of the proportional contribution (in percentage) of any working position

Source: Rousset and Giret (2007, p.761)

2.1.1.2 Correspondence analysis (CA) & multiple correspondence analysis (MCA)

Correspondence analysis (CA) and multiple correspondence analysis (MCA) involve analysing the pattern of relationships of several categorical dependent variables (Abdi, 2007; Greenacre, 2010, 2007). It is an extension of principal component analysis (PCA) to categorical variables. MCA is a special case of correspondence analysis applied to a matrix of variables coded as a binary variable. It uses nominal data that comprise multiple levels, each of which is coded as 0 or 1. Thus, it is typically used with data that might be described as “qualitative” characteristics. Similarities between individuals on this these variables are examined using distance information on graphs showing which points which tend to group together, and which are further apart. These distance scores can be used to group individuals in further classification analyses, such as cluster techniques.

These analyses are types of multidimensional data analysis (MDA) methods in which objects are represented in a multi-dimensional Euclidean space where distances between points are interpreted as similarities between individuals or categories (Greenacre, 1976, 2009, 2010, 2007).

Co-funded by the Erasmus+ Programme of the European Union

This approach is sometimes referred to as the “French School” and was notably used by Pierre Bourdieu: https://www.politika.io/en/notice/multiple-correspondence-analysis.

Although these techniques are not longitudinal as such, they can be adapted for use with longitudinal data. Correspondence analysis and hierarchical cluster techniques are often combined in the “French School” of sociological data analysis and is referred to as Qualitative Harmonic Analysis (QHA) when used to describe complex trajectories (Robette and Thibault, 2008). To conduct QHA, one must decide on a period of time, divide it into intervals, and then measure the proportion of time that each individual spends in each of the states in each interval. One then does a CA on the resulting matrix in order to summarize this information on the sequence of states, the time when events occur, and the duration in the various states. Clustering hierarchical classification allows one to construct a typology based on this information. These analyses can notably be done quite easily in R using the package FactoMineR: http://factominer.free.fr/factomethods/hierarchical-clustering-on-principal-components.html

2.1.1.3 Sequence analysis & optimal matching

Sequence analysis (SA) is a longitudinal analysis technique borrowed from biology (from the analysis of DNA) that involves ordering a list of elements across time into sequences (in sociology these elements are often a certain status, such as an employment or marital status), where positions of the elements are fixed and ordered by elapsed time (Abbott and Hrycak, 1990; Abbott and Tsay, 2000). It involves analysing a holistic “trajectory” (rather than a discrete “transition”) that then serves as a representation of an individual’s trajectory or “career” over time (Blanchard, Buhlmann and Gauthier, 2014). These can be illustrated using a variety of plots (for example, see the figure below).

A subsequent Optimal Matching Algorithm (OMA) stage allows a refinement of the analysis by comparing each individual sequence with all the others to measure the dissimilarities among them (Halpin, 2009; Lesnard, 2014). Indeed, all sequence pairs are “matched” with one another to create a distance matrix. The distance or dissimilarity scores are determined through the creation of substitution costs as well as setting insertion and deletion costs. They are used to compute the “cost” of transforming one sequence into another, which is done for all sequence pairs. *A potential drawback for our purposes is that we typically require at least 7 or 8 time points (preferably even more) in order to gain meaningful insights from the OM and clustering stages.*

Co-funded by the
Erasmus+ Programme
of the European Union

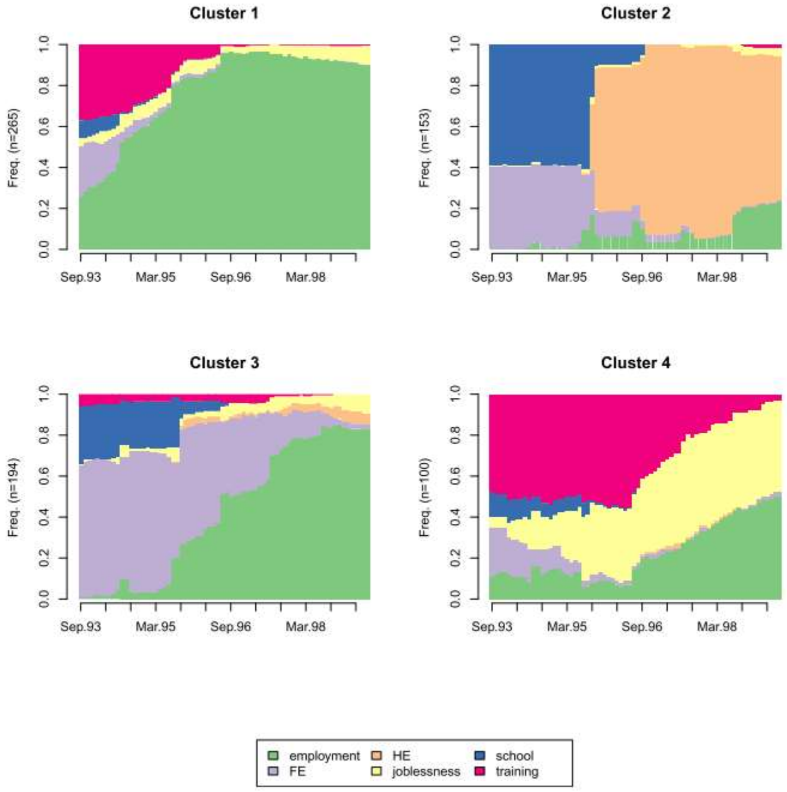


Figure 1: State distribution plots by cluster.

Source: Gabadinho, Ritschard, Müller, & Studer (2011, p.7).

Then, classification techniques such as clustering can be employed on this dissimilarity matrix to create groups of similar sequences where we have maximum similarity within groups and minimal similarity between groups. Clustering is a critical step in SA, as it allows complexity reduction with the intent of determining whether meaningful groupings exist in the data. More complex methods can also be used (Helske and Helske, 2019). The resulting groups can be used in regression analyses, for example, as a dependent variable in a multinomial logistic regression examining predictors of following each trajectory (Boylan, 2020).

This overall approach can be broken down into five steps: (1) describing key sequences via aggregated measures; (2) visualizing key sequences via sequence index plots; (3) comparing sequences via optimal-matching (OM) or alternative techniques; (4) grouping sequences into clusters via cluster or multidimensional scaling methods; and (5) associating patterns with other variables within regression models (Mills, 2014b).

Co-funded by the
Erasmus+ Programme
of the European Union

2.1.2. Parametric procedures

2.1.2.1 Latent class analysis (LCA), LPA, LTA, & RM-LCA…

Latent class analysis (LCA) uses categorical observed variables to measure a categorical latent variable. (When indicators are continuous it is then called latent profile analysis (LPA)). Levels of this latent variable represent subgroups of individuals that are called latent classes. LCA allows us to find groups of individuals who are similar on a categorical latent variable (Muthén and Muthén, 2000). Thus, this technique relies on the assumption that individuals can be divided into subgroups based on an unobservable construct of interest. We cannot know the “true” latent class membership of any particular individual because we do not measure the latent variable itself, but rather multiple indicator variables that all have measurement error.

LCA is not in itself a longitudinal method, but latent transition or trajectory analysis (LTA) is an extension of LCA used with longitudinal data. In this longitudinal adaptation of the technique, individuals transition between different latent classes over time. Class membership thus becomes dynamic and follows a development path through different stages. Thus, LTA uses multiple indicators at each time point to define a latent class variable for each time point in order to evaluate the probability of a transition from one class at a particular time point to another class at the next time point (Muthén and Muthén, 2000). The results of these analyses are comparable to those of sequence analysis (Barban and Billari, 2012). More specific techniques are described below. More information can be found at: https://www.methodology.psu.edu/ra/lca/model-faq/

In another variation of LCA, repeated-measures latent class analysis (RM-LCA) takes a person-centred approach and uses repeated measures to identify latent patterns of responding to categorical items over time: More information can be found here: https://ctri.wisc.edu/researchers/behavior-change-analysis/rmlca

2.1.2.2 Growth modelling

In practice, growth analysis uses a multilevel regression model based on two levels of analysis, where level one examines how an outcome variable increases, decreases, or remains constant over time, and level two describes how change varies between individuals. These level two variables allow us to consider what time-invariant covariates explain variance between the individual slopes and intercepts fit within the level-1 model. These approaches are useful when for theoretical reasons we think that are likely distinct subpopulations within our population (Jung and Wickrama, 2008). For some practical examples, see this helpful presentation: https://www.bgsu.edu/content/dam/BGSU/college-of-arts-and-sciences/center-for-family-and-demographic-research/documents/Workshops/2018-Growth-Curve-Model-in-Stata-final.pdf

Co-funded by the
Erasmus+ Programme
of the European Union

Types of Generalized Linear Growth Models

Name	Acronym	Alternative Name (acronym)	Link	Distributional Assumptions	
				Level 1	Level 2 +
Hierarchical Linear Model	HLM	Random Effects Model Random Coefficients Model Mixed Model Multilevel Model (MLM)	Identity	Normal	Normal
Hierarchical Generalized Linear Model	HGLM	Generalized Linear Mixed Model (GLMM)	May be identity, log, logistic or logit, probit, complementary log-log, etc.[a]	Member of Exponential Family[b]	Normal
Growth Mixture Model	GMM	Growth Mixture Analysis Finite Mixture Model	May be identity, log, logistic or logit, probit, complementary log-log, etc.	Exponential Family	Multinomial[c] & Normal
Latent Class Growth	LCGA	Semi-parametric Group-Based Model (SGBM)		Exponential Family	Multinomial
Latent Profile Analysis			Identity	Normal	Multinomial
Longitudinal Latent	LLCA	Latent Class Analysis (LCA)	May be identity, logistic or logit, probit, complementary log-log, etc.	Multinomial	

[a]Depends on the distribution of the observed data.

[b]Exponential family of distributions includes normal, exponential, gamma, multinomial, binary or Bernoulli, Poisson, zero-inflated Poisson, negative binomial, etc.

[c]Multinomial includes both ordered and unordered categorical

[e]Model lacks latent growth factors

Source: Feldman, Masyn and Conger (2009, p.677)

With these approaches, we are interested in questions about individual differences in change over time and what explains or predicts these differences. At least three waves of data are necessary. Importantly, these techniques can be used on different measures of student persistence from transcript data, such as credit hours attempted, credit hours completed, credit completion ratio, and enrolment (Nathan Marti, 2008).*

2.1.2.3 Latent trajectory, latent class growth & growth mixture modelling

Latent trajectory analysis is a longitudinal adaptation of LCA (as described above) and thus another person-centred analytical technique that focuses on a single outcome variable measured at multiple points in time. Unlike regression and ANOVA, which rely on overall sample means and variance around the mean, this approach assumes that there are latent groups with independent measures of central tendency within the population. These (unobserved) subpopulations are modelled using a mixture distribution (Van De Schoot, 2015). This approach evaluates latent trajectories, which are unobserved, but estimated by a regression equation. Individuals are classified to the latent group whose trajectory they most resemble. Latent Growth Mixture Modelling (LGMM) or Latent Class Growth Modelling (LCGA) can be used (discussed below).

Co-funded by the Erasmus+ Programme of the European Union

These approaches both assume that an unmeasured latent variable, composed of several classes, explains the way the variables evolve over time (Herle et al., 2020).

Latent class growth analysis (LCGA) is used to define a latent class model of a single repeated-measure variable where the latent classes correspond to different growth curve shapes on the outcome variable (Muthén and Muthén, 2000). The different growth curve shapes are described as well as the probability of belonging to each class. The results are the proportional odds computed using a cumulative logit link function. Using this approach, the influence of background variables on class membership can also be measured. This captures the latent change or growth trajectory classes of individuals with similar trajectories (Marcoulides and Heck, 2013).

Growth mixture modelling (GMM) combines features of conventional growth modelling and LCGA. In traditional growth modelling, all individuals in the sample are assumed to come from a single population and individual variation around the mean is measured, while in LCGA separate mean growth curves are estimated for each class without any allowance of *individual* variation. GMM combines these to estimate both the mean growth curves for each class and individual variation around them (Muthén and Muthén, 2000). In essence, we assume that there are multiple mixed effects models for each subgroup (or “class”) of trajectories that are defined by their common mean and shape (Herle et al., 2020).

Co-funded by the
Erasmus+ Programme
of the European Union

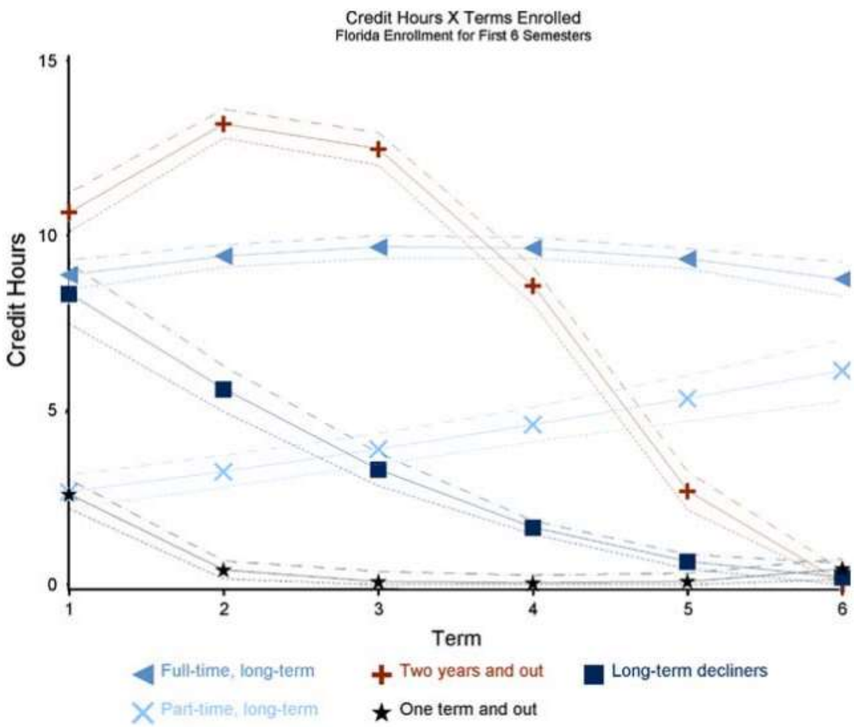


Fig. 1 Latent postsecondary persistence trajectories with 95% confidence intervals for Florida community college system student database

Source: Marti (2008, p.326)

An example of a research question for this type of analysis is: “To what extent do students who drop out of high school pattern into different subgroups based on their trajectory in teacher assigned grades?” (Bowers and Sprott, 2012).*

More examples and explanations can be found in Fiona Steele’s informative presentations, for example: http://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/longitudinal.pdf

Co-funded by the
Erasmus+ Programme
of the European Union

2.2. Variable-centred analyses

2.2.1. Traditional approaches

2.2.1.1 *Descriptive statistics*

The goal of these types of analyses is to identify important features of data and to prepare for subsequent model-based analyses (Singer and Willett, 2003). Typically, within-individual change is summarized as the changes in the repeated measurements on each individual during the period of observation (Fitzmaurice and Ravichandran, 2008). The simplest measure is the change score or difference score, which describes the difference between two points in time as a function of another variable (Tang et al., 2013). However, this does not take into account measurement error. For time-to-event data, life table and Kaplan-Meier estimates can be used (Mills, 2014a).

Empirical growth plots can also be used to find trends. In econometrics, a multiple time series plot is often used to plot a response versus time, connecting observations over a common subject. This allows us to (1) detect patterns in the response, by subject and over time, (2) identify unusual observations and/or subjects and (3) visualize the heterogeneity (Frees, 2004). More information can be found in this (freely available) book: https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Longitudinal%20and%20Panel%20Data/Book/Chapters/FreesFinal.pdf

See below a link to Chapter 2 of *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* by Judith D. Singer and John B. Willett along with practical examples of data analysis in Stata:

www.investigadores.cide.edu/crow/Clases/MetCuant12/Lectura/CH 2 Exploring Longitudinal Data on Change.pdf

Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence by Judith D. Singer and John B. Willett Chapter 2: Exploring Longitudinal Data on Change | Stata Textbook Examples (ucla.edu)

Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence by Judith D. Singer and John B. Willett Chapter 10: Describing Discrete-time Event Occurrence Data | Stata Textbook Examples (ucla.edu)

2.2.1.2 *RM-ANOVA, MANOVA*

ANOVA stands for analysis of variance, and can be used to assess differences in an outcome measured at different time points in the same subjects (Schober and Vetter, 2018). Repeated-measures ANOVA (RM-ANOVA) examines the relationship between the study outcome variable

Co-funded by the Erasmus+ Programme of the European Union

and a set of covariates, comparing the mean outcome at multiple time points or between groups (Ma, Mazumdar and Memtsoudis, 2012). There exist univariate repeated-measures ANOVA or a multivariate ANOVA (MANOVA). Both can model group, time, and group-by-time interaction effects (Fitzmaurice et al., 2008; Fitzmaurice and Ravichandran, 2008). However, these are parametric approaches that make quite strong assumptions about the underlying data. Specifically, (1) the outcome variable must be continuous and normally distributed, (2) the covariates must be discrete (categorical) variables, (3) the outcome must have constant variance across time and a constant correlation between any 2 time points (this is called the "assumption of sphericity"), and (4) all subjects must have the same number of repeated measurements (i.e., we must drop those subjects who have missing observations), which may result in a biased sample. A schema is shown below.

Figure 1. Schematic representation of a one-way repeated-measures ANOVA. An outcome is repeatedly measured or observed for each subject (here: subjects A–D) at each time point or under each condition, allowing to assess how the outcome value changes within each subject. Therefore, a factor in which each subject's outcome variable is repeatedly measured at each factor level (here: time point or condition), is referred to as a within-subject factor. One-way repeated-measures ANOVA compares the mean values of the outcome variable between the factor levels. ANOVA indicates analysis of variance.

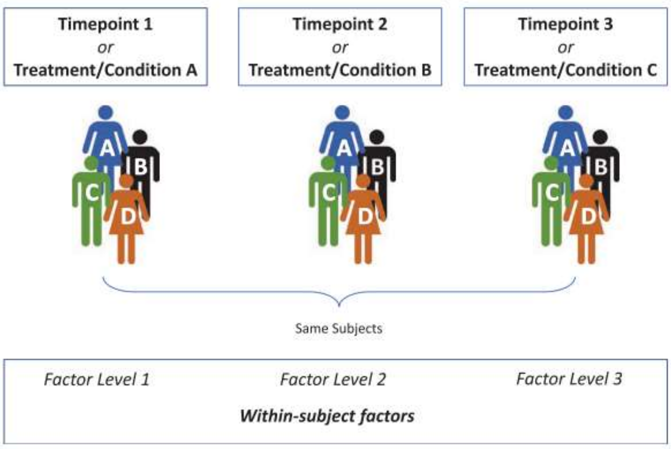
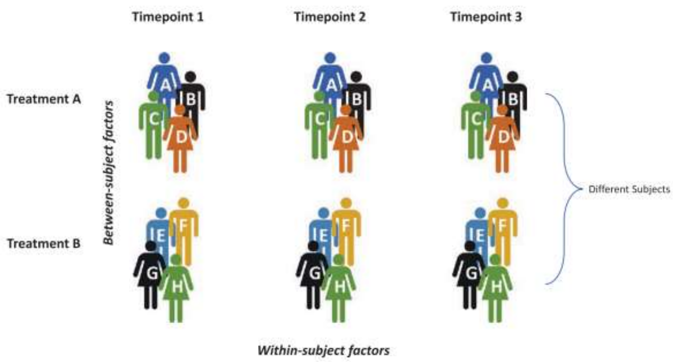


Figure 2. Schematic representation of a two-way (two-factor) mixed ANOVA. The model includes a within-subject factor (Figure 1), but also a between-subject factor. Factors that vary between subjects are between-subject factors (here: a part of the subjects are assigned to treatment A, but other subjects are assigned to treatment B). A two-way mixed ANOVA tests for differences in the mean values of the outcome variable between the factor levels of the within-subject factor, between the factor levels of the between-subject factor, as well as the interaction of the 2. ANOVA indicates analysis of variance.



Source: Schober & Vetter (2018, p.571)

Co-funded by the Erasmus+ Programme of the European Union

2.2.1.3 Analysis of response profiles

This is an extension of ANOVA to a longitudinal setting. Analysis of response profiles compares the sequence of mean responses over time among different groups (i.e., comparing their mean response profiles) (Fitzmaurice et al., 2008; Fitzmaurice and Ravichandran, 2008). If the pattern of change in the mean response is the same in the treatment groups, the response profiles should have the same shape. A variety of complex methods can be used to analyse profile behaviour including Profile “Level” and Profile “Shape,” for example, in engineering (Snee, 2020).*

2.2.1.4 Event History Analysis (EHA)/survival analysis

This approach deals with questions about timing and duration until the occurrence of an event (time-to-event) or “when?” (Mills, 2014c; Willett and Singer, 1991). The dependent variable is the “hazard rate” or the conditional probability that an event occurs at a particular time interval—or in a discrete-time model, the odds (Mills, 2014c). Survival time is the duration or time that it takes before an event occurs (Mills, 2014c). (The hazard rate measures failing (i.e., experiencing the event) while the survivor function measures surviving (i.e., not experiencing the event).) Time can be measured on a continuous or discrete scale. (In our case, since we have yearly data, we will only know that the event occurred within a particular year, which is a discrete time scale. The analyses will therefore be discrete time logistic regression models.)

The models are actually just regression models, but with a rate as the dependent variable (we can also think of this as “risk”). We can include fixed or time-varying independent variables to see how they influence the rate. One of the unique features of EHA is that it takes censoring into account. (Right) censoring is when the event of interest does not take place during the time period examined : (Mills, 2014c, p.5). Another is the fact that we can include time-varying independent variables. However, this approach often necessitates data restructuring where we organise data into discrete units in a subject-period file.

An example applied to the sociology of education is to study the event of leaving a higher education institution, where one can focus on (at least) two different types of event: transferring to a different institution and leaving higher education (dropout) (Hovdhaugen, 2011).* One can analyse these events separately using binary logistic regression because the hazard functions are quite different.

Co-funded by the
Erasmus+ Programme
of the European Union

2.2.2. Models conditioned to the subject:

2.2.2.1 Multilevel/hierarchical linear models (HLM)/ Mixed effects models

A conventional growth model is a multilevel, random-effects model (Bryk and Raudenbush, 1992; Jung and Wickrama, 2008). When we use multilevel models for longitudinal data, time is at level 1 and the individual is at level 2. In other words, time is nested within individuals. We can examine what kinds of time varying characteristics (level 1) and "person" characteristics (level 2) affect the initial level of a given variable (intercept) and change over time (slope). Time to event occurrence for recurring events using event history analysis also uses a multilevel (logistic) modelling framework (Steele, 2008).

As with earlier examples from *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* by Judith D. Singer and John B. Willett, you can see worked examples on the UCLA website: https://stats.idre.ucla.edu/other/mult-pkg/seminars/mlm-longitudinal/

Mixed effects models estimate an average population trajectory, parametrised in terms of fixed effects, and the variation of the individual trajectories around this average as subject-specific random effects (Herle et al., 2020). Individuals are assumed to have their own unique "curve" that describes longitudinal change in the response (Fitzmaurice et al., 2008; Fitzmaurice and Ravichandran, 2008). This approach relates outcomes collected on the same individual to their observation times, allowing for the shape of this relationship to vary across individuals. In these types of models, we typically use maximum likelihood (ML) estimation and can also use structural equation modelling (SEM) techniques.

2.2.2.2 Fixed effects models

An example of a research question is: "Does a change in an exposure cause a change in the outcome?" This approach can also be a useful exploratory tool for longitudinal data because it controls for all time-invariant confounding by using only the changes in exposure occurring within individuals to estimate the outcome (Gunasekara et al., 2014). However, this approach is equivalent to including a fixed intercept for each individual, so this does not work well for large samples (many dummy variables to include in model). Time series models are another approach used in economics, but these require many time points (typically 30 or more) (Frees, 2004). More information can be found in this (freely available) book: https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Longitudinal%20and%20Panel%20Data/Book/Chapters/FreesFinal.pdf

Co-funded by the Erasmus+ Programme of the European Union

2.2.2.3 Non-conditional, marginals, & population-averaged models

In a conditional or mixed model, the coefficients have cluster-specific interpretations. In a marginal model, the coefficients have population-averaged interpretations. Thus, marginal models are population-average models whereas conditional models are subject-specific. In practice, marginal models are an alternative form of the multilevel model in which “the marginal mean, rather than the conditional mean given random effects, is regressed on covariates” (Heagerty and Zeger, 2000, p.1). Some authors prefer random-effects over marginal models (Lee and Nelder, 2004). However, estimates may vary considerably (Carrière and Bouyer, 2002). Differences between these models become statistically complex.

Co-funded by the
Erasmus+ Programme
of the European Union

3. Bibliography and other resources

Abbott, A. and Hrycak, A. (1990) Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers. *American Journal of Sociology*. 96 (1), pp.144–185.

Abbott, A. and Tsay, A. (2000) Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect. *Sociological Methods & Research*. [online]. 29 (1), pp.3–33.

Abdi, H. (2007) Multiple correspondence analysis. *Encyclopedia of measurement and statistics*. [online]. 95 (2), pp.116–28.

Andres, L. (2013) *Longitudinal Studies of Education Sociology of Education : An A-to-Z Guide*. Sage Publications, Inc., 440–441.

Barban, N. and Billari, F.C. (2012) Classifying life course trajectories: A comparison of latent class and sequence analysis. *Journal of the Royal Statistical Society. Series C: Applied Statistics*. [online]. 61 (5), pp.765–784.

Blanchard, P., Buhlmann, F. and Gauthier, J.A. (2014) *Advances in Sequence Analysis: Theory, Method, Applications* Blanchard, P., Gauthier, J.-A. and Buhlmann, F. (eds.) Heidelberg, Germany, Springer International Publishing.

Bowers, A.J. and Sprott, R. (2012) Examining the multiple trajectories associated with dropping out of high school: A growth mixture model analysis. *Journal of Educational Research*. [online]. 105 (3), pp.176–195.

Boylan, R.L. (2020) Predicting Postsecondary Pathways: The Effect of Social Background and Academic Factors on Routes through School. *Socius: Sociological Research for a Dynamic World*. [online]. 6, p.237802311989517.

Bryk, A.S. and Raudenbush, S.W. (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods* Newbury Park, CA, Sage Publications, Inc.

Carrière, I. and Bouyer, J. (2002) Choosing marginal or random-effects models for longitudinal binaryresponses: Application to self-reported disability among older persons. *BMC Medical Research Methodology*. [online]. 2, pp.1–10.

Distefano, C. and Mindrila, D. (2013) Cluster analysisIn: *Handbook of Quantitative Methods for Educational Research*. New York, NY, Sense Publishers, 103–122.

Feldman, B.J., Masyn, K.E. and Conger, R.D. (2009) New approaches to Studying Problem Behaviors: A Comparison of Methods for Modeling Longitudinal, Categorical Adolescent Drinking Data. *Developmental Psychology*. 45 (3), pp.652–676.

Fitzmaurice, G. et al. (2008) *Longitudinal data analysis Longitudinal Data Analysis*. [online].

Fitzmaurice, G.M. and Ravichandran, C. (2008) A primer in longitudinal data analysis. *Circulation*. [online]. 118 (19), pp.2005–2010.

Frees, E.W. (2004) *Longitudinal and panel data: Analysis and applications for the social sciences*

Gabadinho, A. et al. (2011) Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*. [online]. 40 (4), pp.1–37.

Greenacre, M.J. (2007) Constructing maps of data using correspondence analysis. (6), pp.16–17.

Greenacre, M.J. (2010) Correspondence analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*. [online]. 2 (5), pp.613–619.

Co-funded by the Erasmus+ Programme of the European Union

Greenacre, M.J. (2009) *The Use of Correspondence Analysis in the Exploration of Health Survey Data*

Greenacre, M.J. (1976) *Tying up the loose ends in simple, multiple and joint correspondence analysis* Barcelona, Spain.

Gunasekara, F.I. et al. (2014) Fixed effects analysis of repeated measures data. *International Journal of Epidemiology*. [online]. 43 (1), pp.264–269.

Halpin, B. (2009) Optimal matching analysis and life-course data: The importance of duration. *Sociological Methods and Research*. [online]. 38 (3), pp.365–388.

Heagerty, P.J. and Zeger, S.L. (2000) Marginalized multilevel models and likelihood inference. *Statistical Science*. [online]. 15 (1), pp.1–19.

Helske, S. and Helske, J. (2019) Mixture hidden Markov models for sequence data: The seqhmm package in R. *Journal of Statistical Software*. [online]. 88 (1).

Heo, M. et al. (2003) Hierarchical linear models for the development of growth curves: An example with body mass index in overweight/obese adults. *Statistics in Medicine*. [online]. 22 (11), pp.1911–1942.

Herle, M. et al. (2020) Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. *European Journal of Epidemiology*. [online]. 35 (3), Springer Netherlands, pp.205–222.

Hoskin, T. (2010) Parametric and nonparametric: demystifying the termsIn: *More Good Reasons to Look at the Data*. Mayo Clinic, 1–5.

Hovdhaugen, E. (2011) Do structured study programmes lead to lower rates of dropout and student transfer from university? *Irish Educational Studies*. [online]. 30 (2), pp.237–251.

Jung, T. and Wickrama, K.A.S. (2008) An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass*. [online]. 2 (1), pp.302–317.

Lee, Y. and Nelder, J.A. (2004) Conditional and marginal models: Another view. *Statistical Science*. [online]. 19 (2), pp.219–238.

Lesnard, L. (2014) Using Optimal Matching Analysis in Sociology: Cost Setting and Sociology of Time. [online]. pp.39–50.

Locascio, J.J. and Atri, A. (2011) An Overview of Longitudinal Data Analysis Methods for Neurological Research. *Dementia and Geriatric Cognitive Disorders Extra*. [online]. 1 (1), pp.330–357.

Ma, Y., Mazumdar, M. and Memtsoudis, S.G. (2012) Beyond repeated-measures analysis of variance: Advanced statistical methods for the analysis of longitudinal data in anesthesia research. *Regional Anesthesia and Pain Medicine*. [online]. 37 (1), pp.99–105.

Marcoulides, G.A. and Heck, R.H. (2013) Mixture models in education. *Handbook of Quantitative Methods for Educational Research*. [online]. pp.347–366.

Mills, M. (2014a) An Introduction to R and Data Exploration via Descriptive Statistics and Graphics. *Introducing Survival and Event History Analysis*. [online]. pp.18–46.

Mills, M. (2014b) Sequence analysisIn: *Introducing Survival and Event History Analysis*. [online]. Sage Publications, Inc., 213–226.

Mills, M. (2014c) The Fundamentals of Survival and Event History Analysis. *Introducing Survival and Event History Analysis*. [online]. pp.1–17.

Co-funded by the
Erasmus+ Programme
of the European Union

Muthén, B.O. and Muthén, L.K. (2000) Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*. [online]. 24 (6), pp.882–891.

Nadif, M. and Govaert, G. (2010) *Cluster Analysis Data Analysis*. [online].

Nathan Marti, C. (2008) Latent postsecondary persistence pathways: Educational pathways in American two-year colleges. *Research in Higher Education*. [online]. 49 (4), pp.317–336.

Robette, N. and Thibault, N. (2008) Comparing Qualitative Harmonic Analysis and Optimal Matching: An Exploratory Study of Occupational Trajectories. *Population-E*. [online]. 63 (04), pp.533–556.

Rousset, P. and Giret, J.-F. (2007) Classifying Qualitative Time Series with SOM : The Typology of Career Paths in FranceIn: Sandoval, F. (ed.) *IWANN*. Berlin, Springer-Verlag, 757–764.

Rousset, P., Giret, J.-F.F. and Grelet, Y. (2012) Typologies de parcours et dynamique longitudinale. *BMS Bulletin of Sociological Methodology/ Bulletin de Methodologie Sociologique*. [online]. 114 (1), pp.5–34.

Schober, P. and Vetter, T.R. (2018) Repeated measures designs and analysis of longitudinal data: If at first you do not succeed-try, try again. *Anesthesia and Analgesia*. [online]. 127 (2), pp.569–575.

Van De Schoot, R. (2015) Latent trajectory studies: The basics, how to interpret the results, and what to report. *European Journal of Psychotraumatology*. [online]. 6 (July).

Singer, J.D. and Willett, J.B. (2003) Exploring Longitudinal Data on ChangeIn: *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. [online]. 15 (1), 583–605.

Snee, R.D. (2020) An effective strategy for the analysis of response profiles. *Quality Engineering*. [online]. 33 (1), Taylor & Francis, pp.85–99.

Steele, F. (2008) Multilevel Models for Longitudinal Data. *Journal of the Royal Statistical Society*. 171 (1), pp.5–19.

Tang, W. et al. (2013) *Longitudinal data analysis Modern Clinical Trial Analysis*. [online].

Willett, J.B. and Singer, J.D. (1991) From Whether to When : New Methods for Studying Student Dropout and Teacher Attrition. *Review of Educational Research*. 61 (4), pp.407–450.

Co-funded by the Erasmus+ Programme of the European Union