Supporting, delivering and facilitating high quality opportunities and training by engaging with students, staff and partners to positively impact the social science community in Scotland

Modelling Trajectories

2025-06-25

Dr Diarmuid McDonnell

Scottish Graduate School of Social Science
Sgoil Cheumnaichean Saidheans Sòisealta na h-Alba

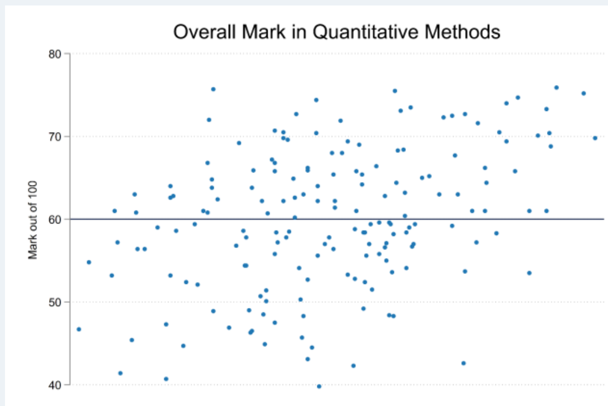
UKRI Economic and Social Research Council

Scottish Funding Council
Promoting further and higher education

Outline

1. Recap of statistical modelling
2. Multilevel statistical modelling
3. Growth-curve models
4. Examples
5. Extensions

Recap of statistical modelling

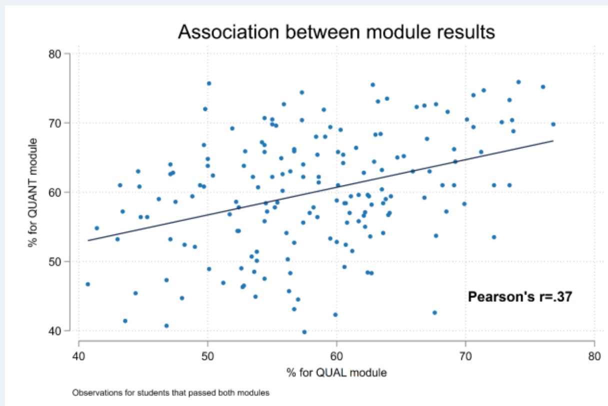


Let's start with the simplest statistical model: the mean.

We have an outcome y which we want to predict for a sample of individuals at a single point in time.

In the absence of further information about these individuals, the prediction that minimises the variance is the mean. We can represent our prediction using a straight line, also known as the intercept.

Recap of statistical modelling



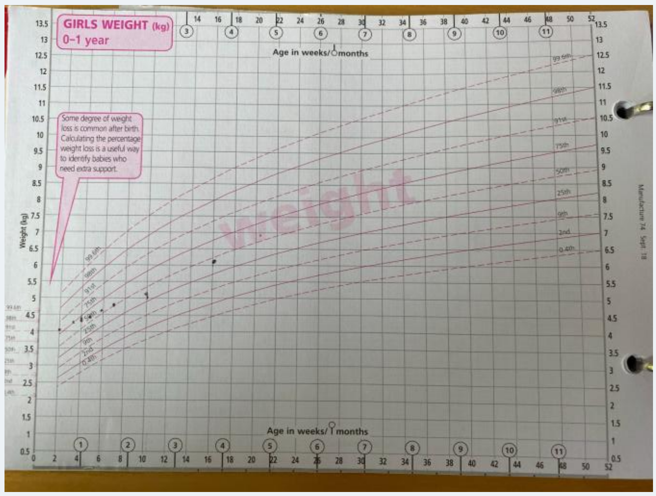
Now let's improve our predictive power (i.e., reduce the variance further) by using information on a student's mark in the qualitative methods module in the previous term.

We still have an intercept but now there is a slope also which better represents the correlation between the modules' marks.

The combination of an intercept and slope gives us our regression line (or our rate of change).

Usually we are only interested in this line for the purposes of making predictions about y and/or examining residuals (i.e., the distance between the observed and predicted values). But sometimes the line itself is the substantive focus of a piece of research.

Typical example of trajectories



An example from public health is the measurement of infant weight and height (and head circumference).

What we observe is the weight of a child at discrete time points (think of surveying the same person across multiple waves of the UK HLS).

What we infer is the trajectory that produces these observed values.

Therefore we want a method that allows us to take observed values for the outcome and time, and estimate the shape of the trajectories.

We can learn a lot from estimating the trajectories – in this example:

- Not a single intercept or slope for all infants
- Variation in the weight of infant girls increases over time (fanning out)
- Growth seems steepest for the largest infants at birth
- Growth is initially rapid for all children but then the gradient eases for all infants

Therefore a reasonably simple modelling approach can yield nuanced, rich substantive insights. How do we estimate these models?

Modelling framework

Multilevel model (MLM) for change.

MLM respects and studies dependency in the data.

Special interest lies in disentangling social processes operating at different levels of analysis.

Data structures motivate the need for MLM.

MLMs allow us to model the dependency, contextuality and heterogeneity in the data. (Leckie & Browne, 2023)

Other names include growth-curve models, latent growth models…

Dependency = observations are not independent of each other i.e., they are correlated, in their values or their residuals. Therefore you need to respect this from a statistical perspective, and potentially study it from a substantive perspective.

Are the differences between levels the same as the differences within levels? Which explanatory variables operate at which level?

There is no need to estimate a MLM when there is no dependency in the data. The research design / data structure determines whether it is applicable.

MLMs allow us to model the *dependency* (e.g., more conservative standard errors), *contextuality* (e.g., coefficients for higher-level units) and *heterogeneity* (e.g., varying coefficients).

Modelling framework

Longitudinal data are also multilevel:

- Repeated measures within individuals
- Waves within households
- Years within countries

Growth-curve models are concerned with differences in initial levels (i.e., intercepts) and growth rates (i.e., slopes). We want to use these models to describe growth and explain patterns in relation to relevant covariates.

By longitudinal we mean repeated contacts = the same units are observed over time. This is different from repeated cross-sectional = different units from the same population are observed over time.

Growth-curve models

"Growth-curve models are a special case of random-coefficient models where it is the coefficient of time that varies randomly between subjects." (Rabe-Hesketh & Skrondal, 2021)

"…the contemporary use of the term growth curve model typically refers to statistical methods that allow for the estimation of inter-individual variability in intra-individual patterns of change over time" (Curran et al., 2010)

Though note that the intercept can vary randomly between subjects also e.g., infants can have different initial birth weights.

Growth-curve models

LEVEL 1 $Y_{it} = \beta_{0i} + \beta_{1i}X_{1t} + \varepsilon_{it}$ $\varepsilon_{it} \sim N(0, \sigma_{\varepsilon}^2)$

LEVEL 2a $\beta_{0i} = \gamma_0 + \zeta_{0i}$ $\zeta_{0i} \sim N(0, \sigma_{\zeta 0}^2)$

LEVEL 2b $\beta_{1i} = \gamma_1 + \zeta_{1i}$ $\zeta_{1i} \sim N(0, \sigma_{\zeta 1}^2)$

SINGLE EQ. $Y_{it} = \{\beta_0 + \beta_1 X_{1t}\} + \{\mu_{0i} + \mu_{1i} + \varepsilon_{it}\}$

A lot going on here so let's unpack.

Growth-curve models

LEVEL 1 $Y_{it} = \beta_{0i} + \beta_{1i}X_{1t} + \varepsilon_{it}$

LEVEL 2a $\beta_{0i} = \gamma_0 + Z_i + \zeta_{0i}$

LEVEL 2b $\beta_{1i} = \gamma_1 + Z_i + \zeta_{1i}$

SINGLE EQ. $Y_{it} = \{\beta_0 + \beta_1 X_{1t} + \beta_2 X_{2i}\} + \{\mu_{0i} + \mu_{1i} + \varepsilon_{it}\}$

Now let’s add some covariates to our multilevel model.

Zi = Binary variable e.g., child is a boy or girl at birth.

Growth-curve models

Growth-curve models can estimate:

- Average level at beginning of period (baseline)
- Average change over time
- Correlation between level and change
- Variation around baseline
- Variation around change
- Predictors of baseline and change

Extensions and considerations

Unbalanced data = units are not observed the same number of times / at the same times.

Functional form = is change a linear / curvilinear process?

Autocorrelation = differences from expected values are probably correlated over time.

Variance functions = changes in variance over time.

Binary / count outcomes = estimating logistic or count regression models.

Questions and comments

Conclusion:

https://github.com/SGSSSonline/modelling-trajectories-summer-school-2025

Contact me: diarmuid.mcdonnell@uws.ac.uk