Durham University
Research Methods Centre

MATH43515: Multilevel Modelling

Lecture 3: Hierarchical Data Structures

Module Convenor/Tutor:
Andy Golightly

Outline

- Recap of F-test, diagnostics for record sales case study
- Hierarchical data structures (motivation, terminology)
- Caveats from historical approaches
- Example: student popularity data

Goodness of Fit

Compare two models through the F-test.

Reduced model (p1 predictors) versus full model (p2 predictors)

Calculate $MSR=SSE(R) - SSE(F) / (p2-p1)$, $MSE=SSE(F)/(n-p2)$

The F statistic is $F^*=MSR / MSE$

What it does:

- Tests the **null hypothesis that the additional predictors in the full model have zero coefficients** versus the alternative that at least one coefficient is non-zero
- If SSE(F) is close to SSE(R), it makes sense to use the simpler model.
- If SSE(F) is substantially smaller than SSE(R), it makes sense to use the full model.

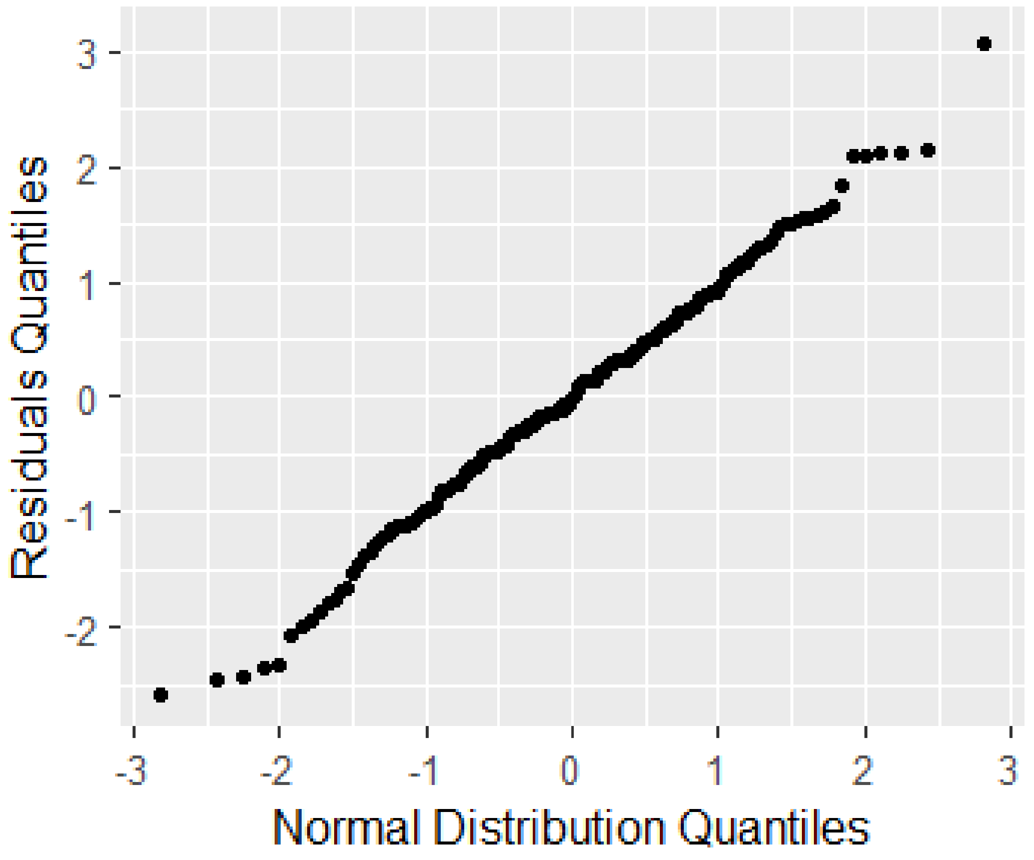
Case Study 2

Recall multiple linear regression with **Airplay** and **Longevity** added (the response variable is **record sales**)
Airplay: number of times the song was played on the radio in the week before the album was released.
Longevity of the band: years since band was formed.

	Estimate	Std error	t-value	p-value
(Intercept)	-26.61	17.35	-1.53	0.13
Adverts	0.08	0.01	12.26	<0.001
Airplay	3.37	0.28	12.12	<0.001
Longevity	11.09	2.44	4.55	<0.001
R^2=0.66				

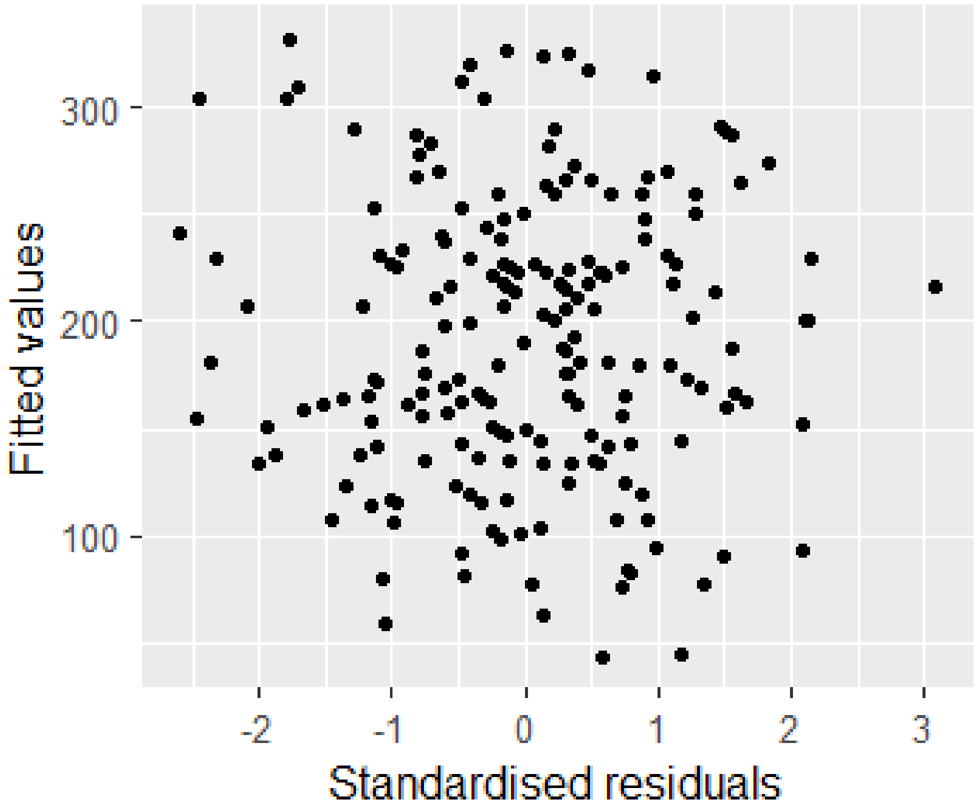
Durham University

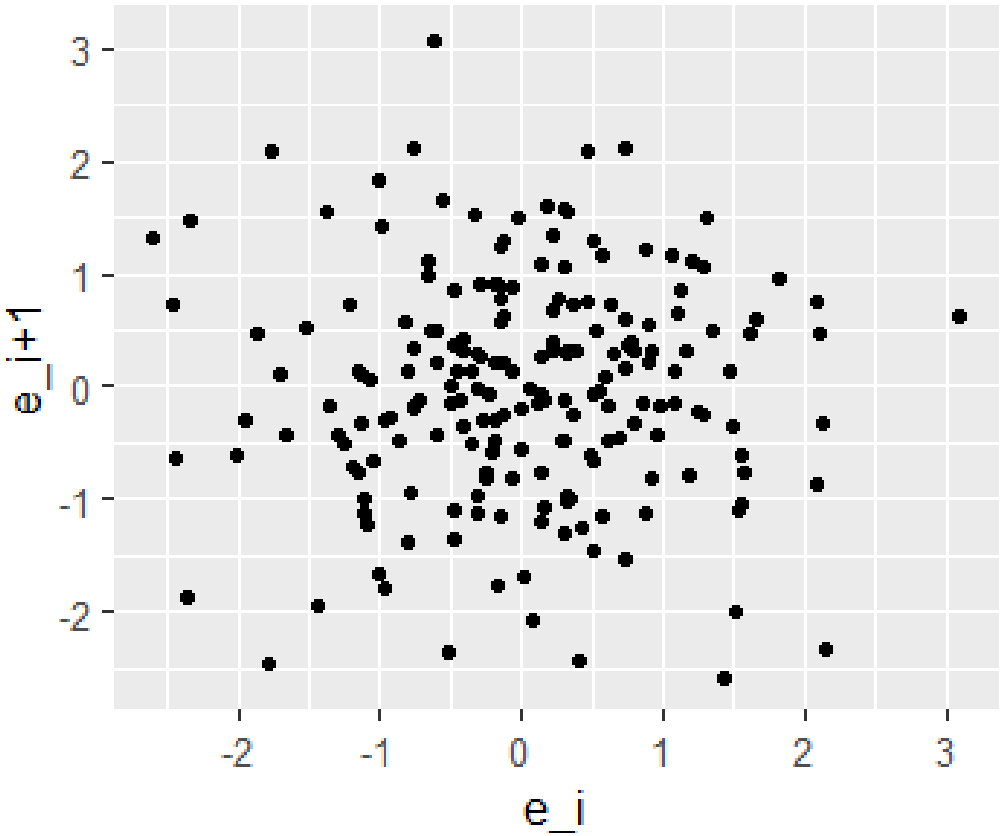
Case Study 2



Durham University

Case Study 2





Durham University

Case Study 2

VIF:

adverts	airplay	longevity
1.014593	1.042504	1.038455

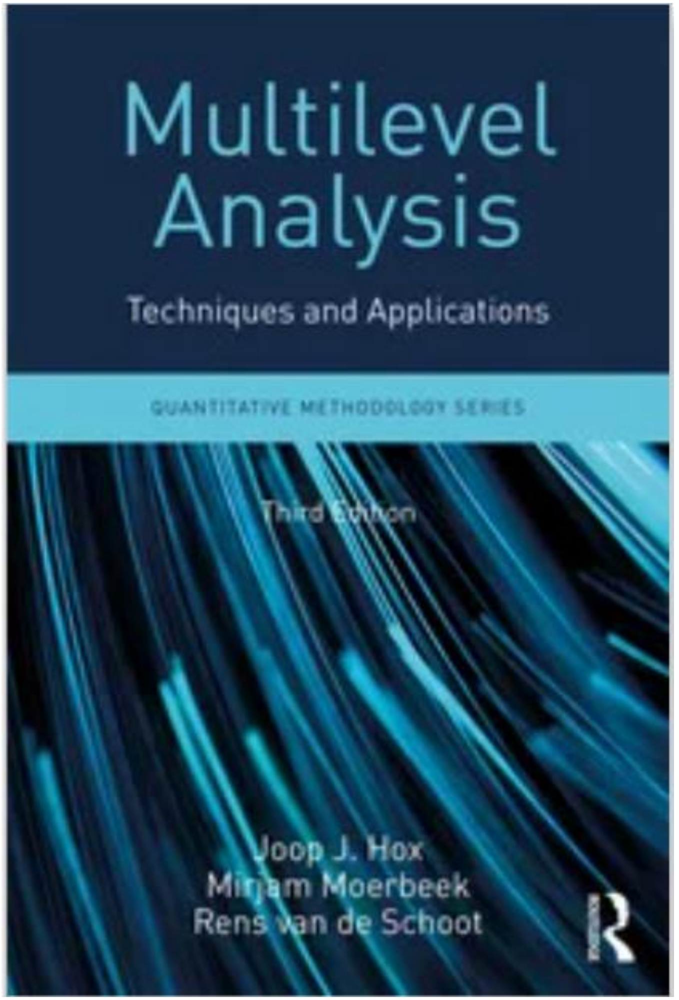
What do we conclude?

Model	predictors	SSE	F*
Simple regression with only adverts	1	866645	
Multiple regression	3	442113	94.6

Which model is better and why?

Assumptions are met, large F* (much larger than the upper 5% critical value of 3.9)

Literature



Lectures 3 to 6 follow roughly Section 2 to 5 of this book.

This includes the use of some data examples.

See

https://library.dur.ac.uk/record=b2733258~S1

for electronic access via the library.

Hierarchical data structures (1)

Children in one class tend to show more similar behaviour/performance (whatever is measured) than children in different classes.

Classes within one school tend to perform more similarly than classes from different schools.

…..

Hierarchical Data Structures (2)

Individuals living within the same household tend to behave more similarly than individuals in other households (e.g, in terms of nutrition).

Households within the same neighbourhood tend to behave more similarly than households in other neighbourhoods.

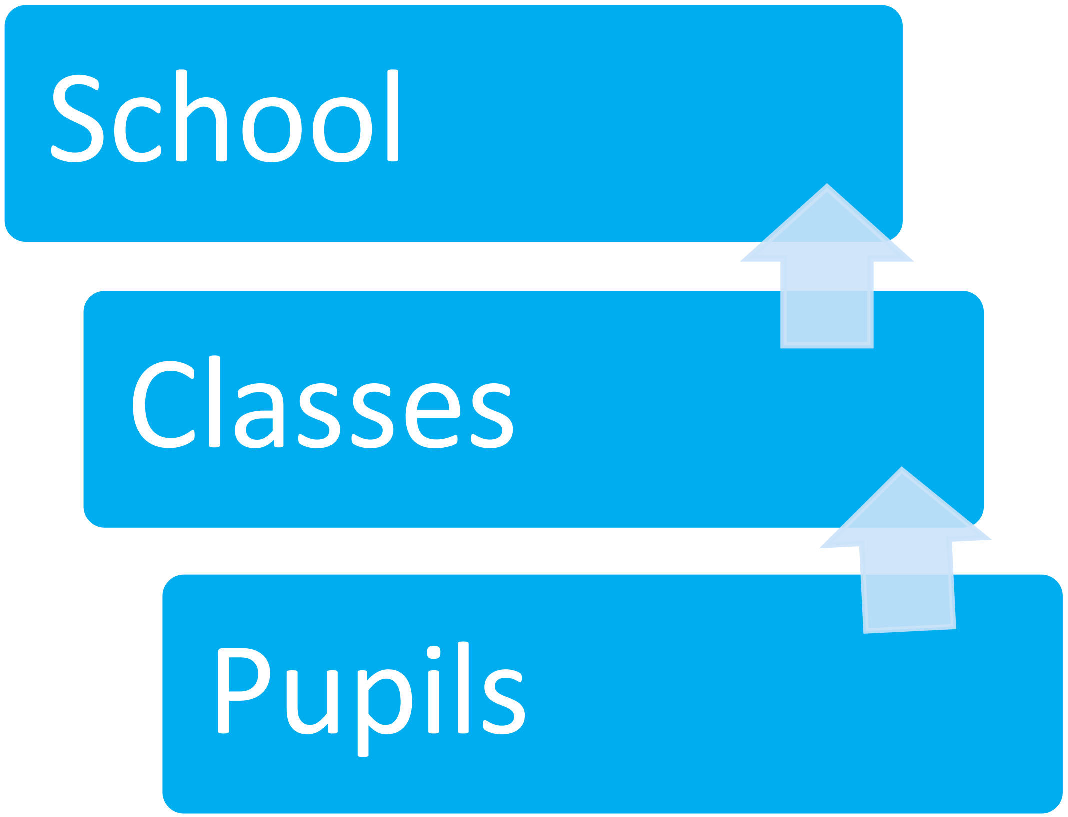
Neighbourhoods in the same county tend to behave more similarly than neighbourhoods in other counties…

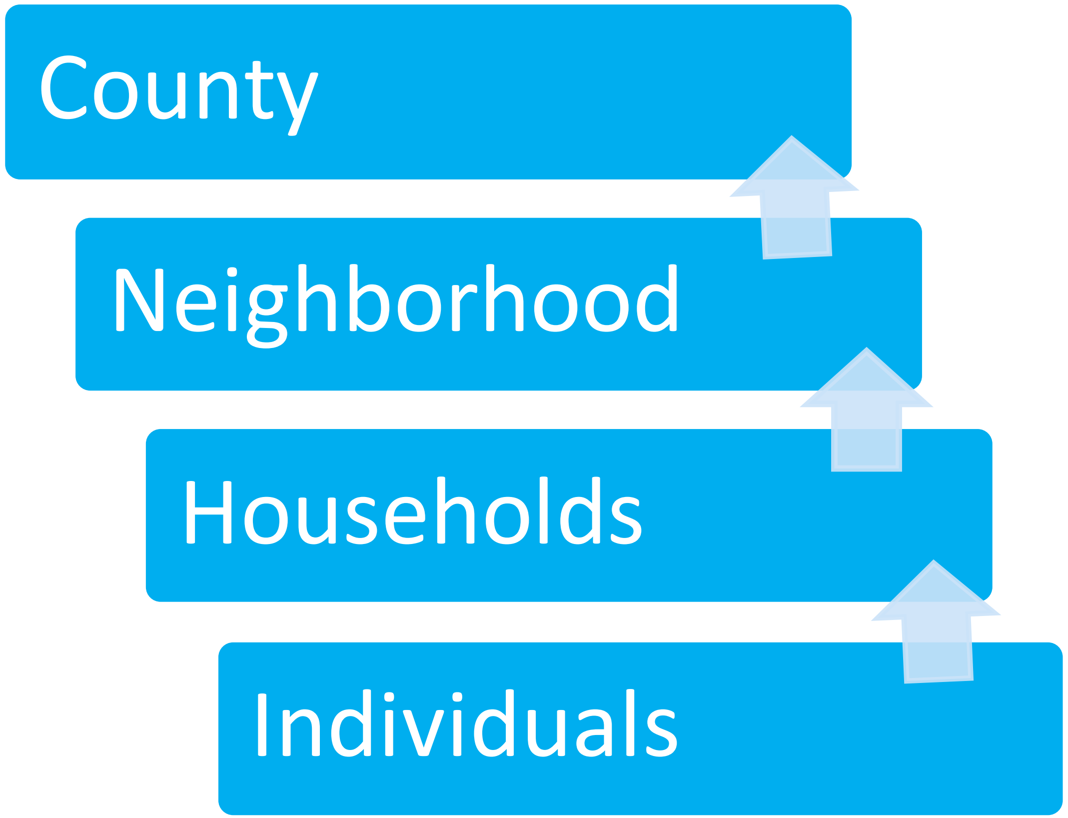
Levels

In both preceding examples, a hierarchical data structure is induced due to the presence of different levels. In the first example, there are three levels, in the second there are four levels.

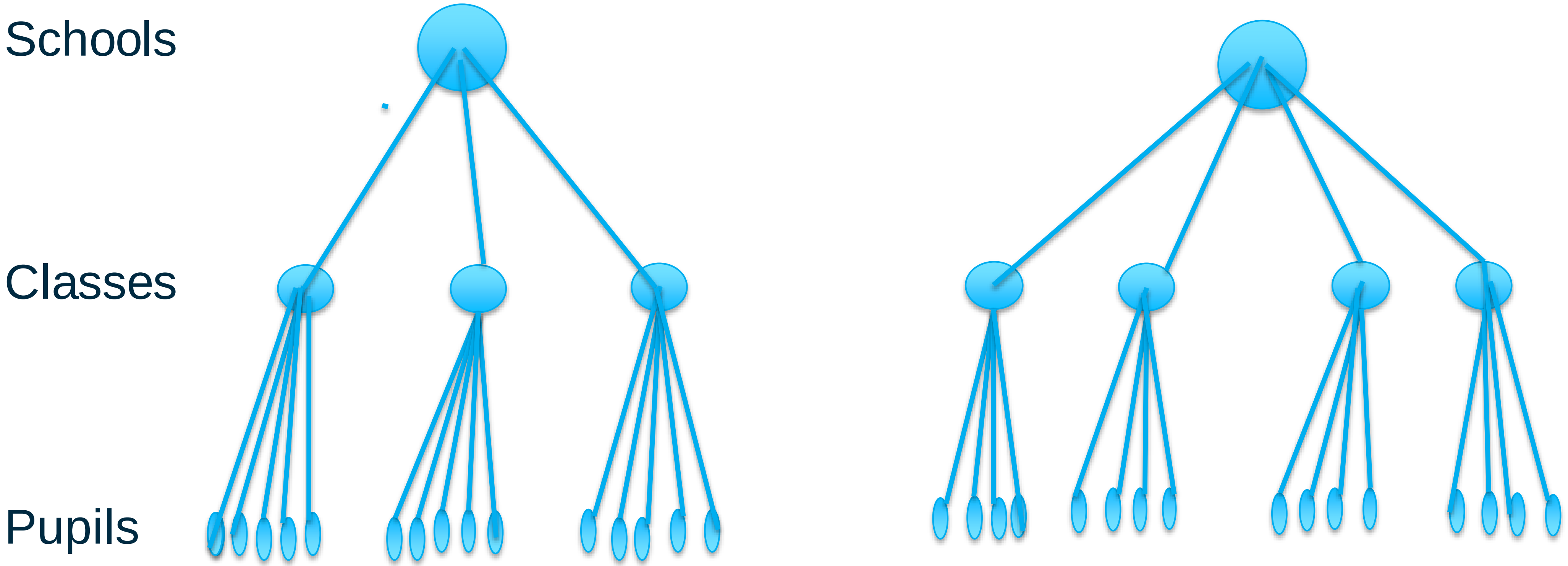
We have that lower level units are nested in higher-level-units. For instance, pupils are nested in classes, which are nested in schools…

Classification diagrams





Unit diagrams



Durham University

Levels and variables

Variables are attributes of levels. Usually, a variable can be clearly allocated to a certain level; i.e. the level at which it is measured.

For instance, gender or age would be measured at pupil level.

Class size would be measured at class level.

School type (faith school, free school, academy,…) would be measured at school level.

The older literature refers to variables which are used at their proper level as global variables (a superfluous term from a modern multilevel modelling point of view).

Why multilevel models?

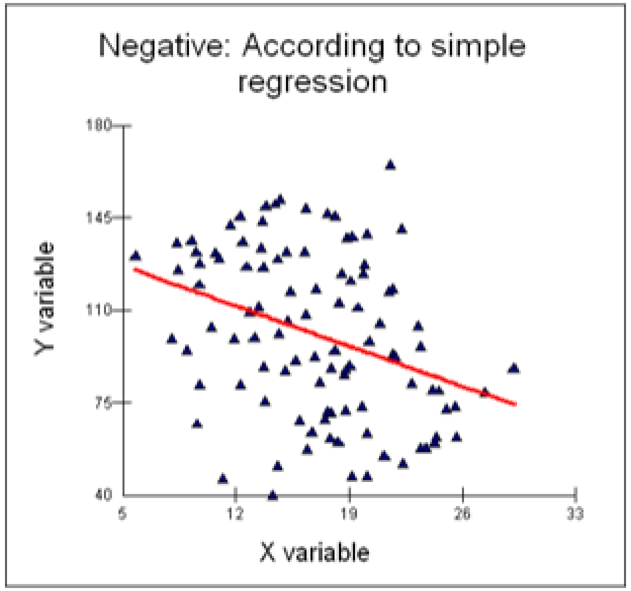
The similarities of lower-level-units belonging to the same upper-level-items causes correlation between the units.

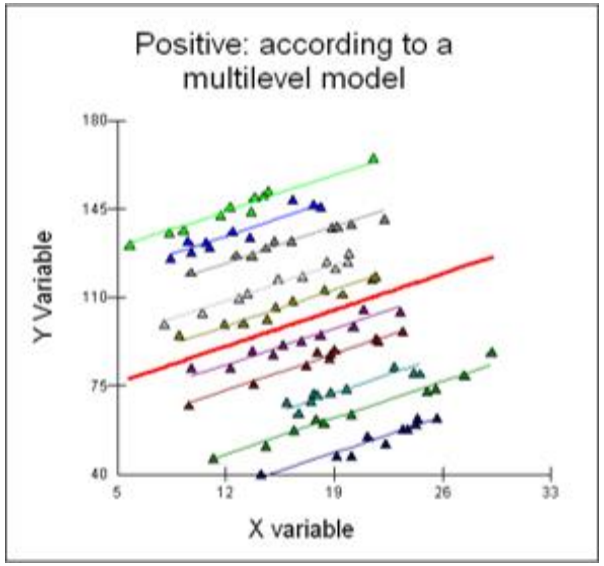
Ignoring these correlations in the analysis can lead to incorrect conclusions.

These problems include incorrect standard errors of estimated parameters (which may lead to effects incorrectly labelled as “significant”), but the consequences can be much more severe.

Durham University

Why multilevel modelling (cont.)





Durham University

Source: Centre for Multilevel Modelling, Univ. of Bristol

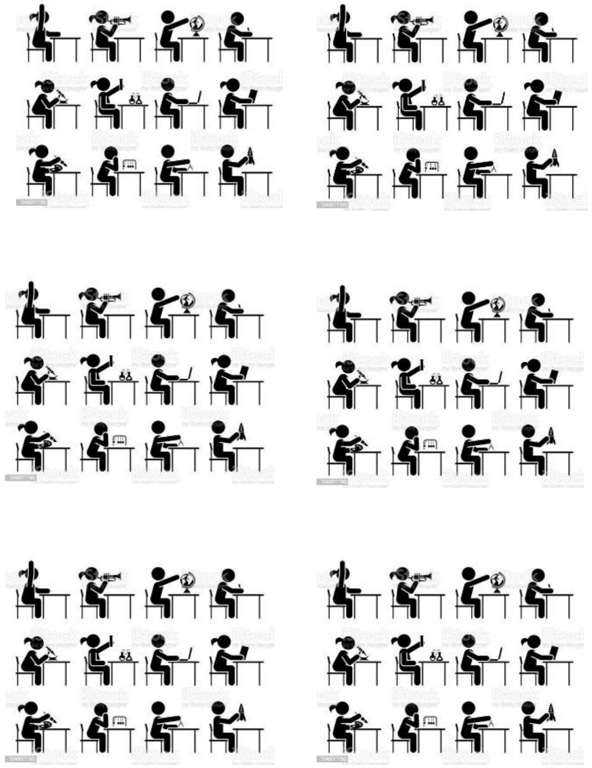
Historical approaches: Aggregation and disaggregation

Historically, multilevel structures (assuming they have been recognized!) have been dealt with in one of two ways (incorrect from a modern point of view!):

- Disaggregation: All units at the lower level receive the value of a variable at the higher level at which they were measured, resulting in contextual variables (For instance, each pupil gets associated to the size of their class).
- Aggregation: Operating at a higher level, one takes averages or other aggregated measures of lower-level units to use as predictor at that level, resulting in structural variables. For instance, a class may be given an attribute “mean intelligence”.

Durham University

Fallacy of disaggregated data



Assume there are J=6 classes with n=12 children each.
As well as individual level covariates, suppose we have an average reading score for each class.

Disaggregation yields a data set of size n’=72. Analysis of this data set in a (multiple) linear regression model would assume that these 72 pieces of information are independent.

They are not!

Disaggregation “blows up” data, leading to artificially large samples sizes, incorrect standard errors, and spurious significances!

Fallacy of aggregated data

"Ecological fallacy", "Robinson effect"

Robinson (1950) highlighted that the (at the time) commonplace practice of

1. aggregating individual-level data to group- or class-means,
2. estimating correlations, effects, etc on the upper (group) level,
3. drawing from these conclusions for individuals on the lower level

…. is fundamentally flawed (since group-level variables may act as moderators).

It is almost 60 years since WS Robinson showed, in 1950, that differing results could be obtained when the same data set is analysed at individual and aggregate levels.[1] Analysing the 1930 US census, Robinson found a correlation of 0.77 between percent black and percent illiteracy at the state-level, while the correlation between illiteracy and race (black vs rest) at the individual-level was 0.20.[1] He showed an even more striking discrepancy between ecological and individual correlations between illiteracy and nativity (foreign-born vs rest), such that the state-level correlation was negative ($r = -0.53$), while the individual correlation was positive ($r = 0.12$).[1] The incongruous empirical correlations at the aggregate- and individual-level led Robinson to conclude that: 'the purpose of this paper will have been accomplished if it prevents the future computation of meaningless correlations and stimulates the study of similar problems with use of meaningful correlation between the properties of individuals' (p. 357).[1] Use of ecological analysis since Robinson has been charged with the methodological crime of 'ecological fallacy', a term coined in 1958 by Selvin,[2] referring to 'the invalid transfer of aggregate results to individuals'.

Revisiting Robinson: the perils of individualistic and ecologic fallacy, IJE, 2009

2-level example [Hox et al.]
Student popularity data

Lower level (level 1): pupils

Upper level (level 2): classes

Data from $J = 100$ classes, with n_j pupils in each class, where $\sum_{j=1}^{J} n_j = 2000$.

Outcome variable (pupil level):

Popularity measured on a continuous scale ranging from 0 (very unpopular) to 10 (very popular) by a sociometric procedure.

Student popularity data

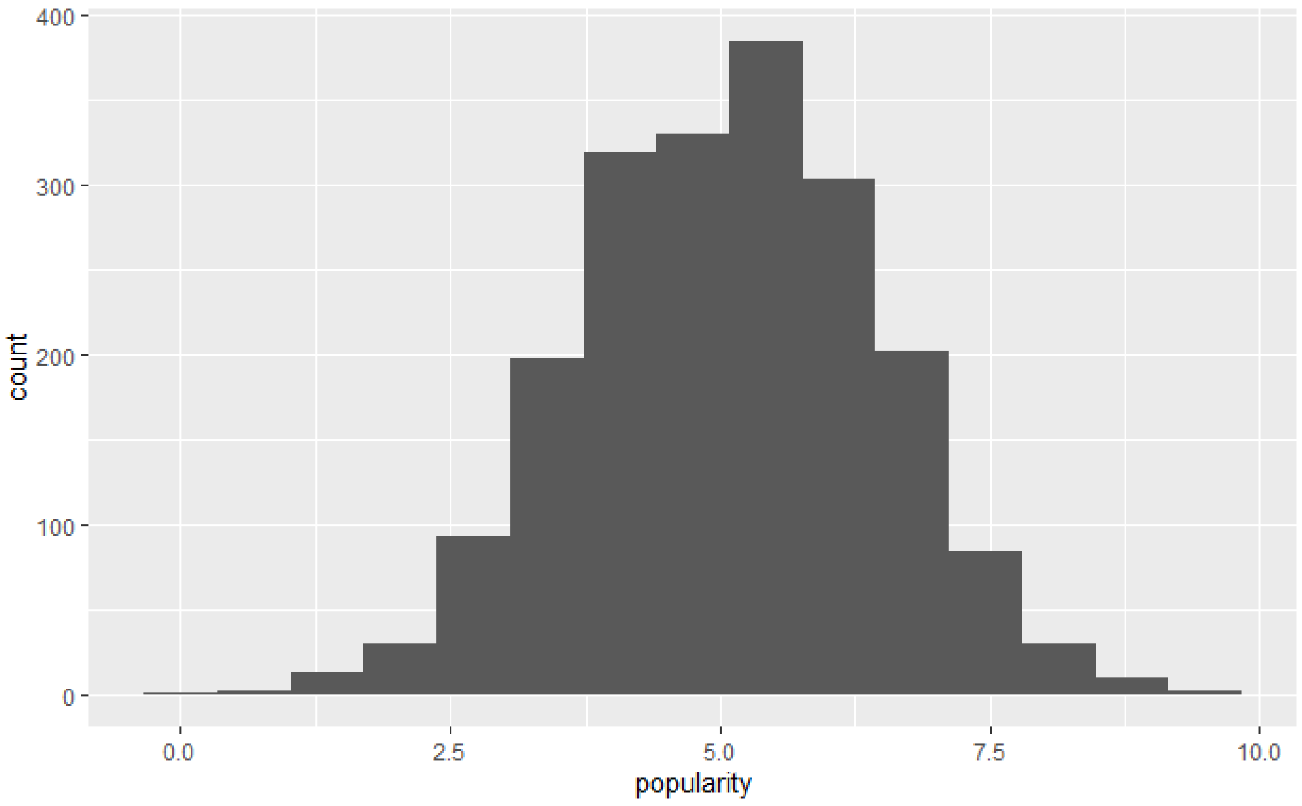
Explanatory variables on pupil level:

- Pupil extraversion (x_1), measured on a self-rating 10-point scale from 1 (min) to 10 (max)
- Pupil gender (x_2), with values $x_2 = 0$ (boy) and $x_2 = 1$ (girl)

Explanatory variable on class level:

- Teacher experience (z), in years, ranging from 2-25, integer-valued

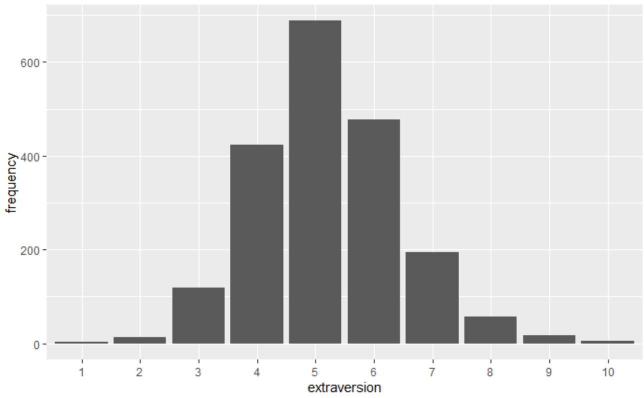
Some exploratory analysis (outcome variable)

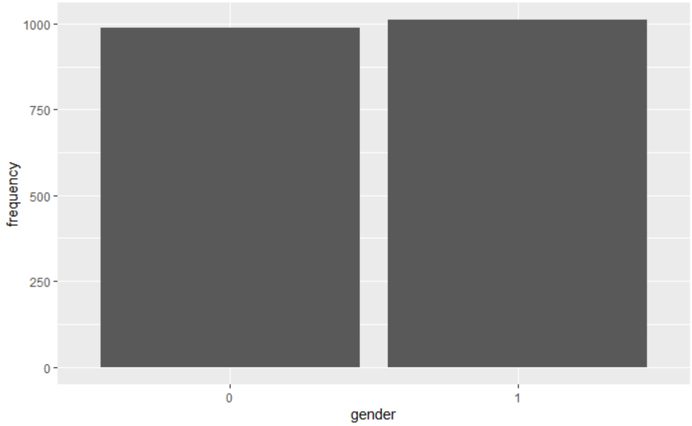


Popularity is reasonably symmetric, could be deemed normally distributed. This is “good”, and is actually a model assumption, to ensure correct inferences.

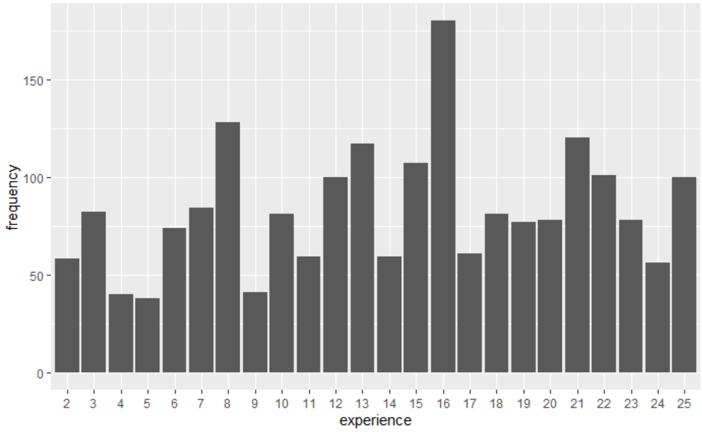
Some exploratory analysis (predictor variables)

pupil level



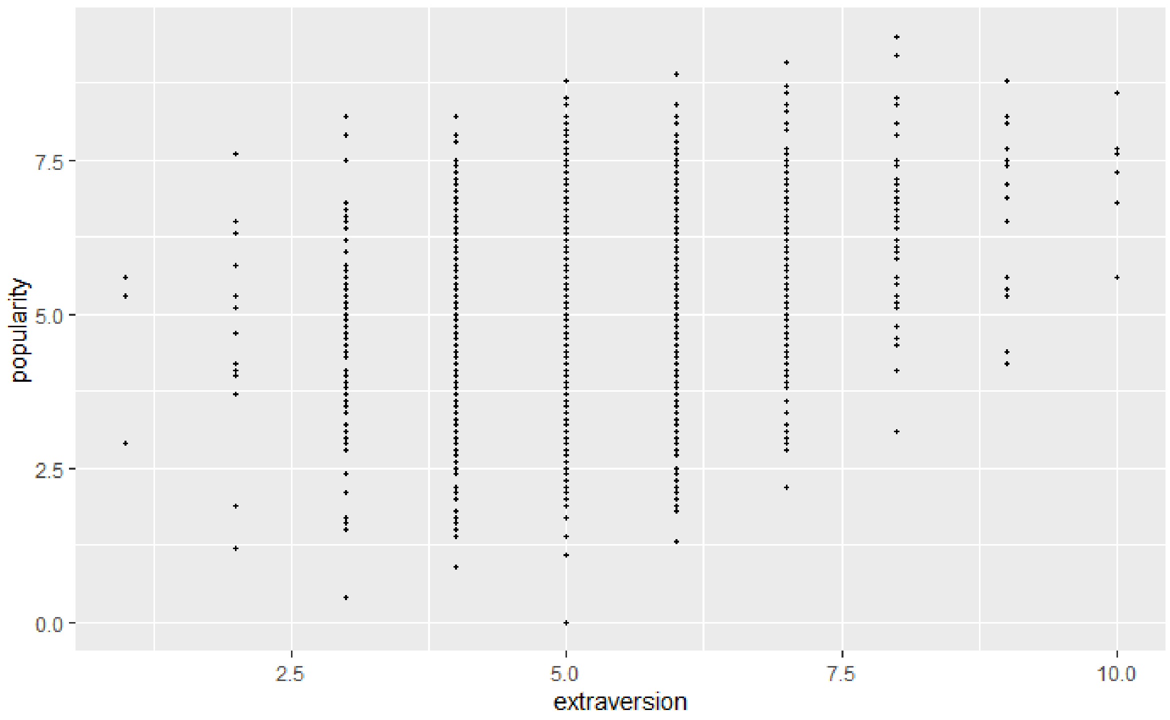


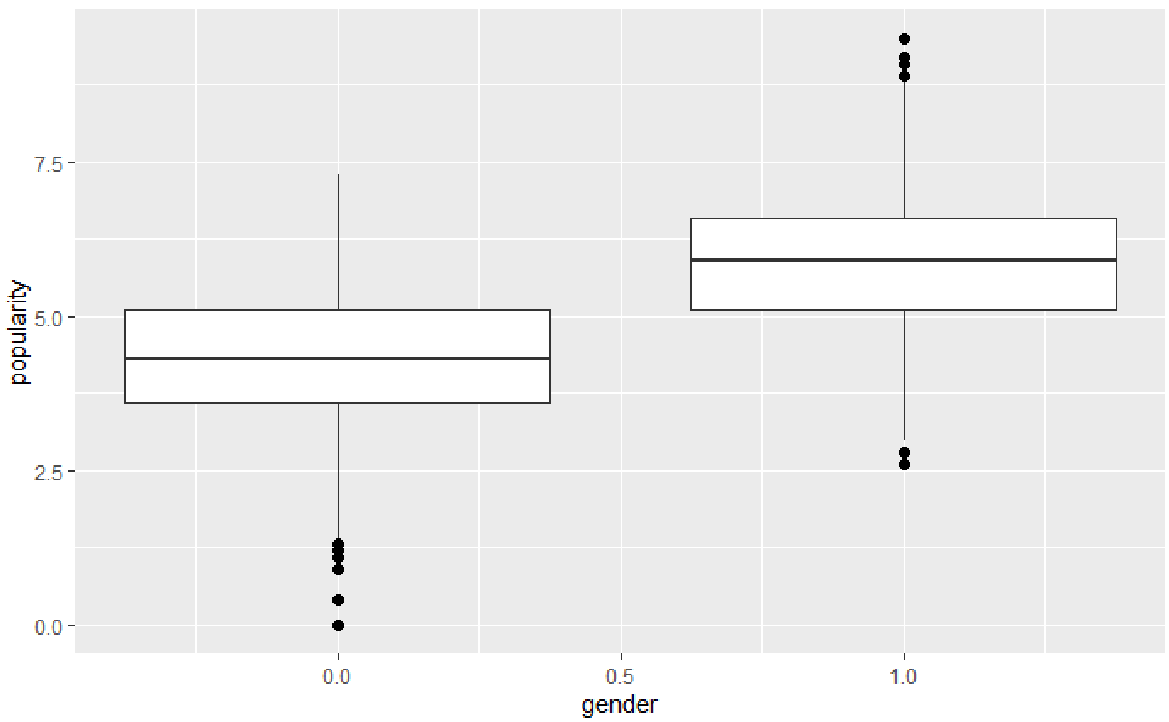
class level



It is not important for regression analysis in general (or multilevel analysis in particular) that these are normally distributed or follow any other `convenient' distribution.

Exploratory analysis: pupil-level variables





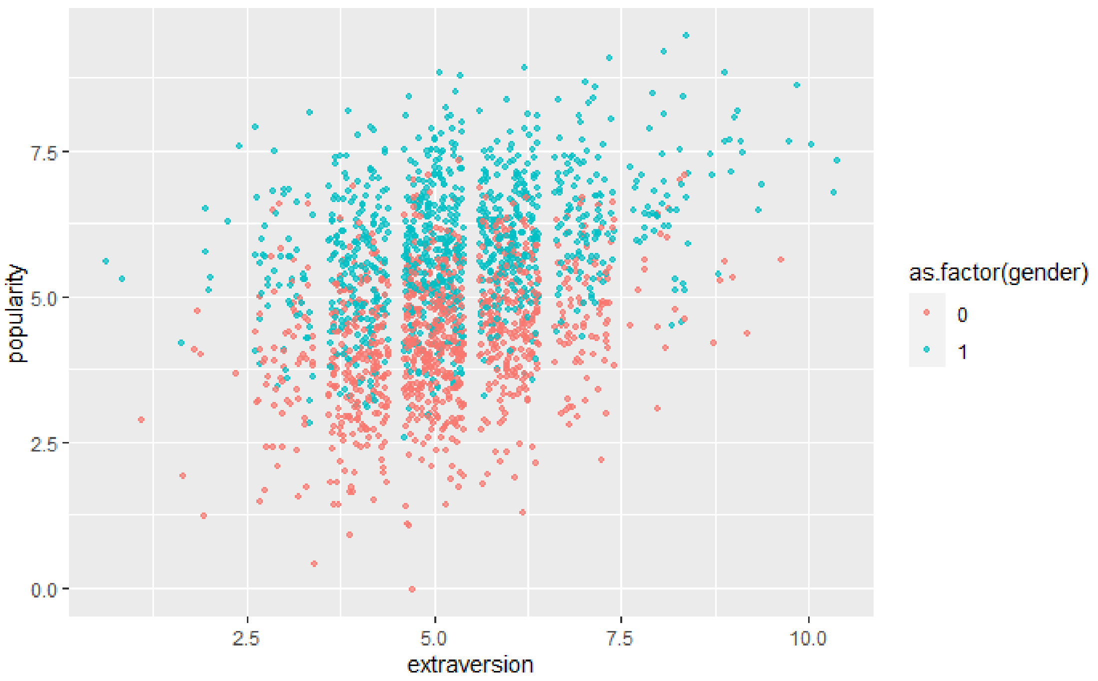
weak linear association
(r=0.32)

stronger association

Durham University

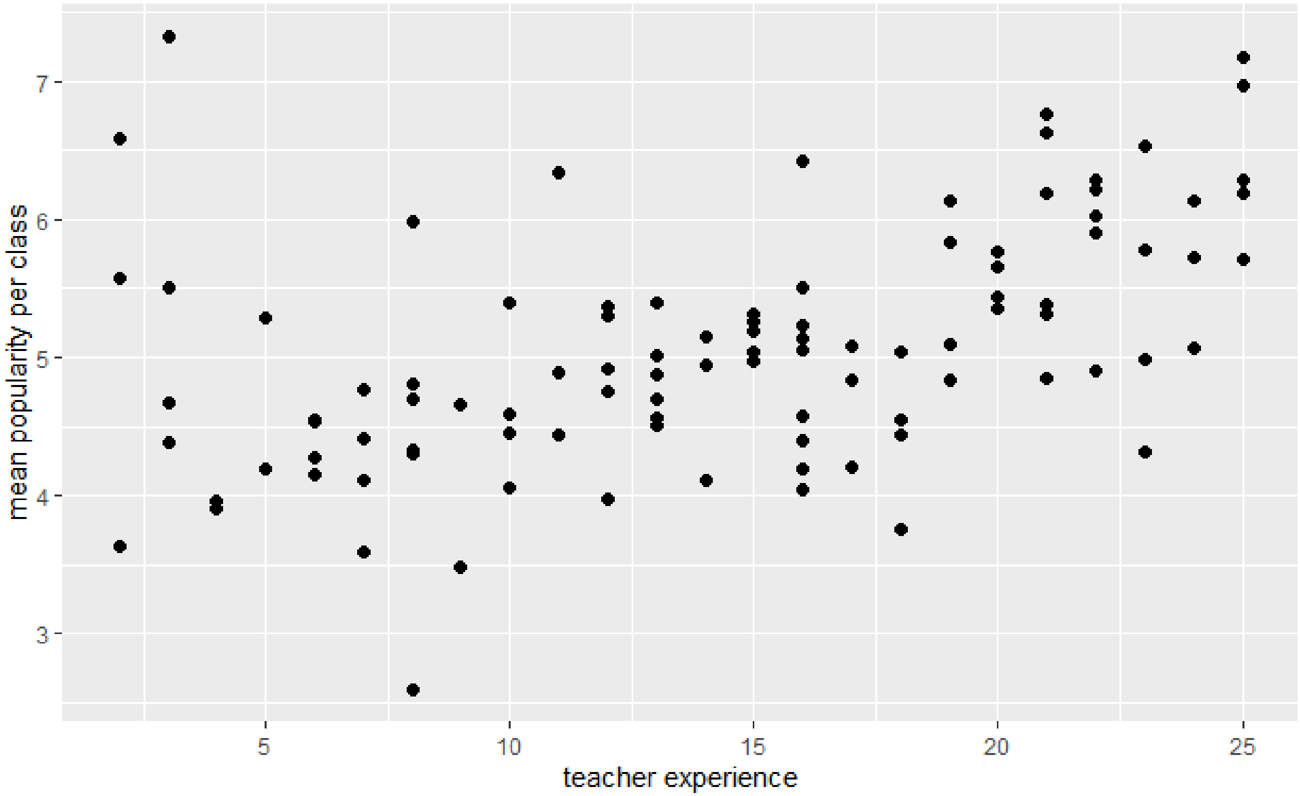
Exploratory analysis: pupil-level variables (cont'd)

Same information as on previous slide, but presented within one single plot, and jittered for better visibility:



Durham University

Exploratory analysis: class-level variable



moderate association (r=0.47)

Question: Why do we need to be careful when interpreting this plot?

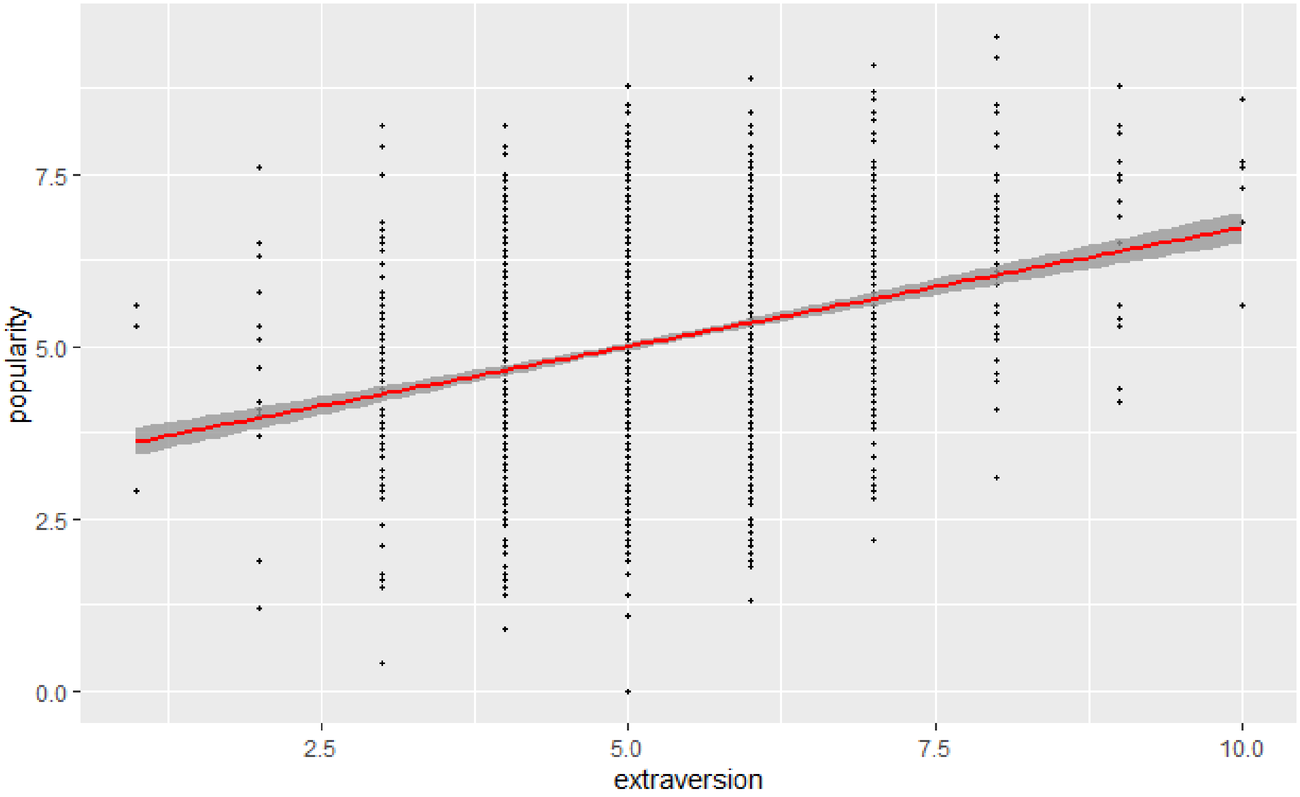
Naïve analysis on pupil-level

Ignore 2-level structure, gender, and teacher experience for now;
i.e. consider the single covariate x_1 =extraversion, and denote this by x for simplicity

```
> simple.linear.model<- lm(popularity~extraversion, data=pop.data)
> summary(simple.linear.model)$coef

              Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)   3.2728356  0.12473523 26.23826  1.237371e-130
extraversion  0.3458513  0.02324748 14.87694  1.481801e-47
```

Naïve analysis on pupil-level (cont'd)



Simple linear regression

$$y = \hat{a} + \hat{b}x$$

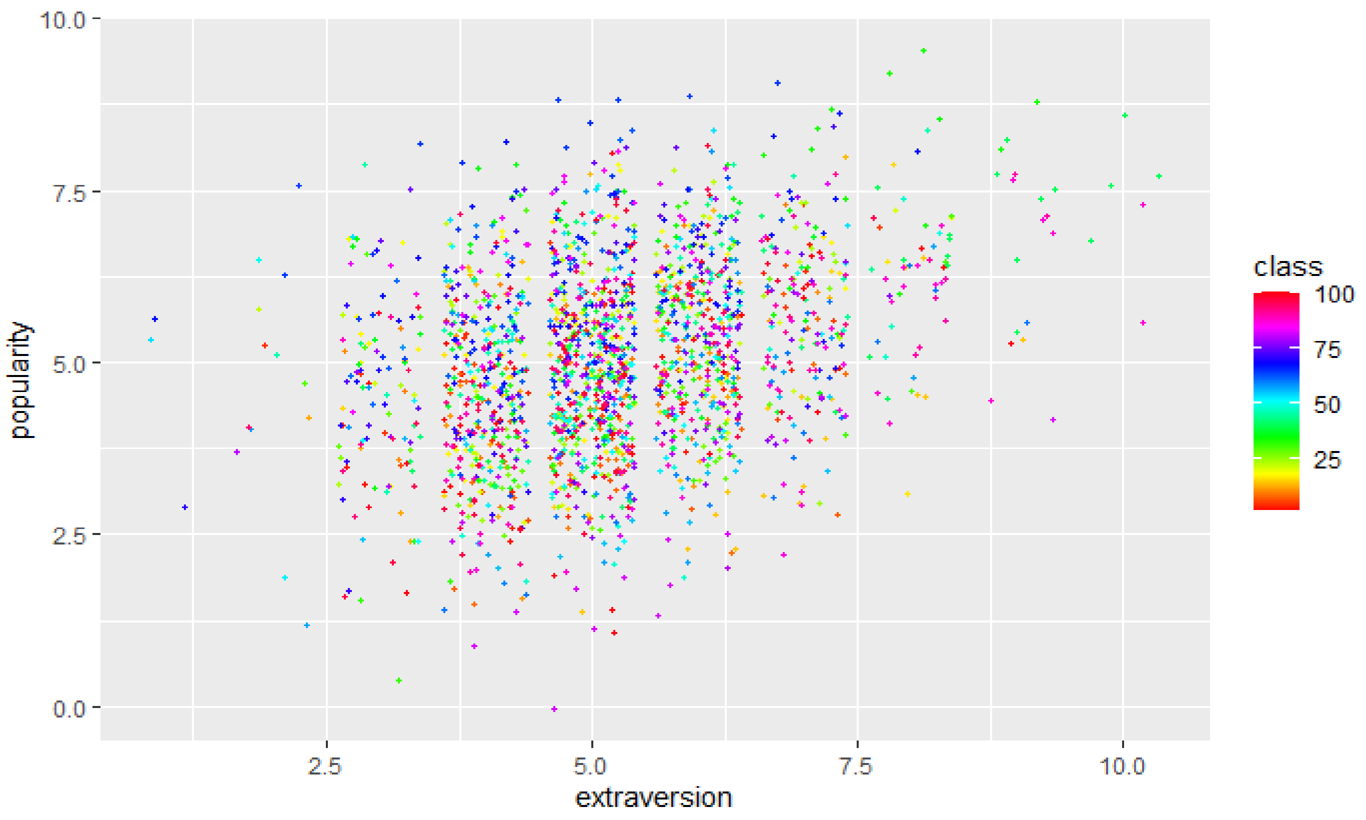
Where

$$\hat{a} = 3.27,$$
$$\hat{b} = 0.35$$

(with standard error band).

Durham University

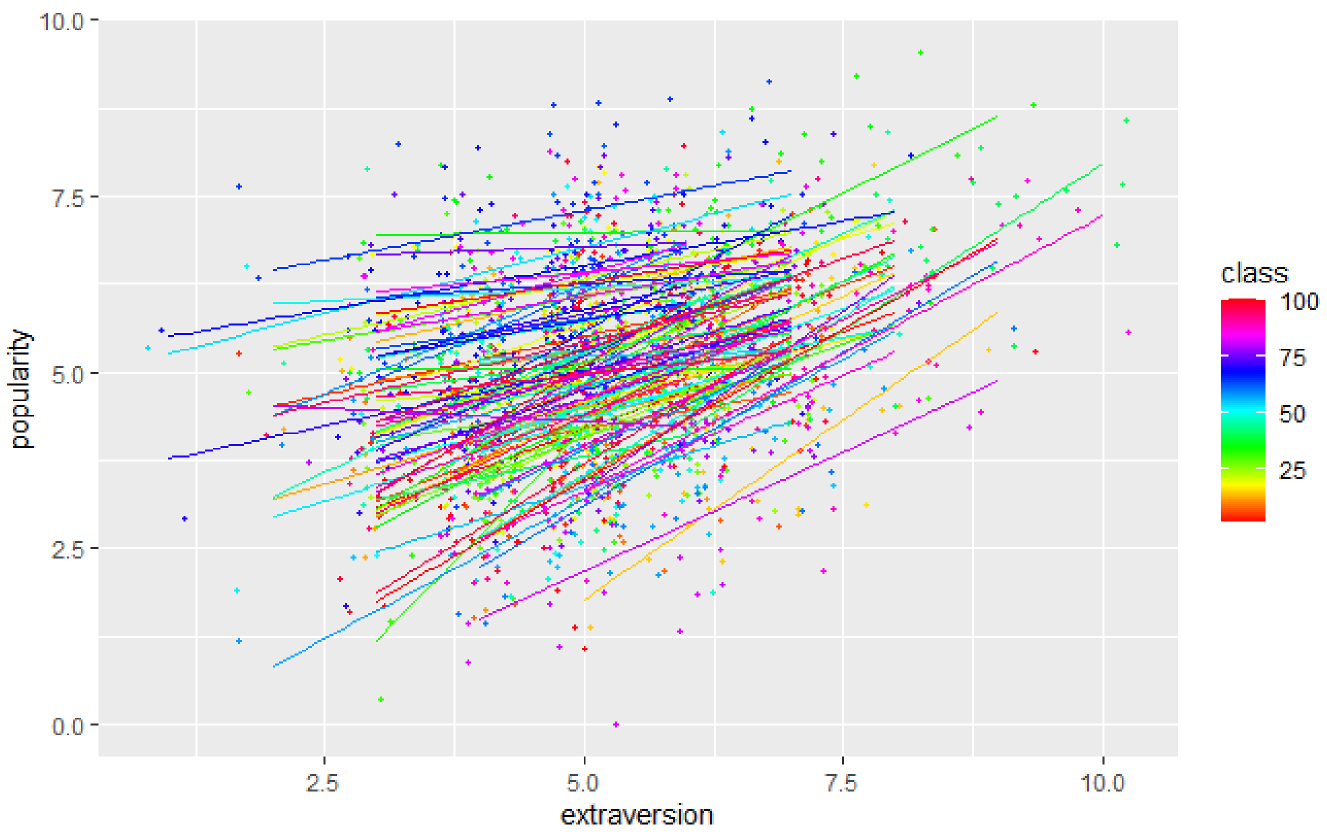
Does one line fit it all?



Same data, but coloured by class (higher level), again jittered for better visibility.

Durham University

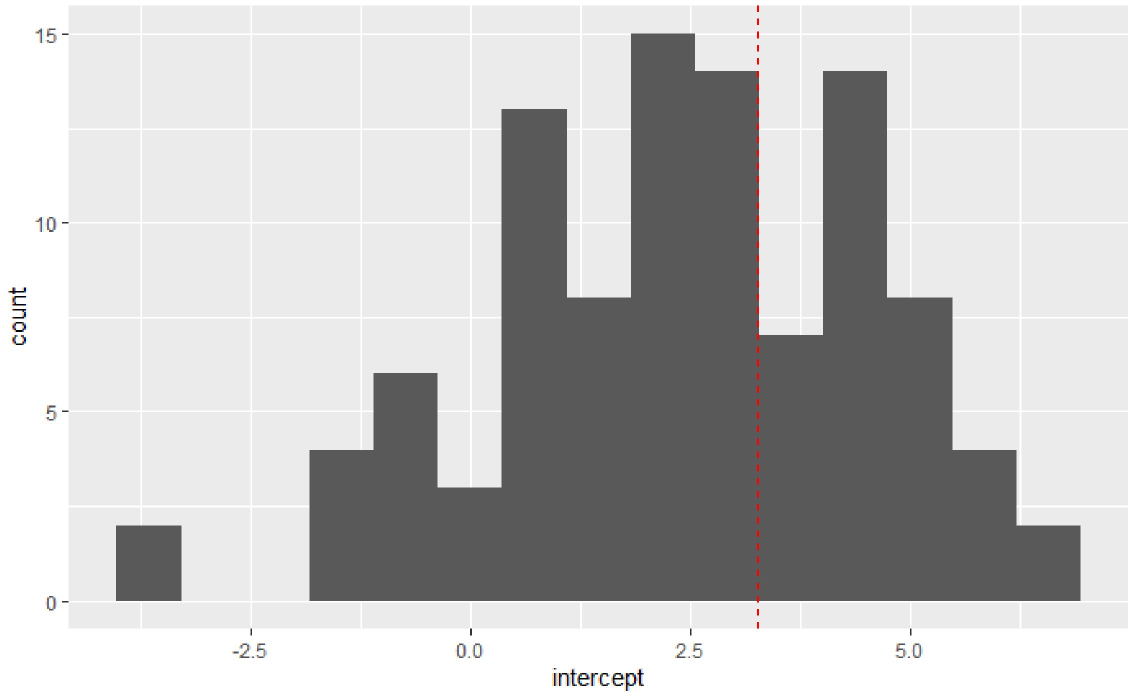
Fit separate regression lines for each class



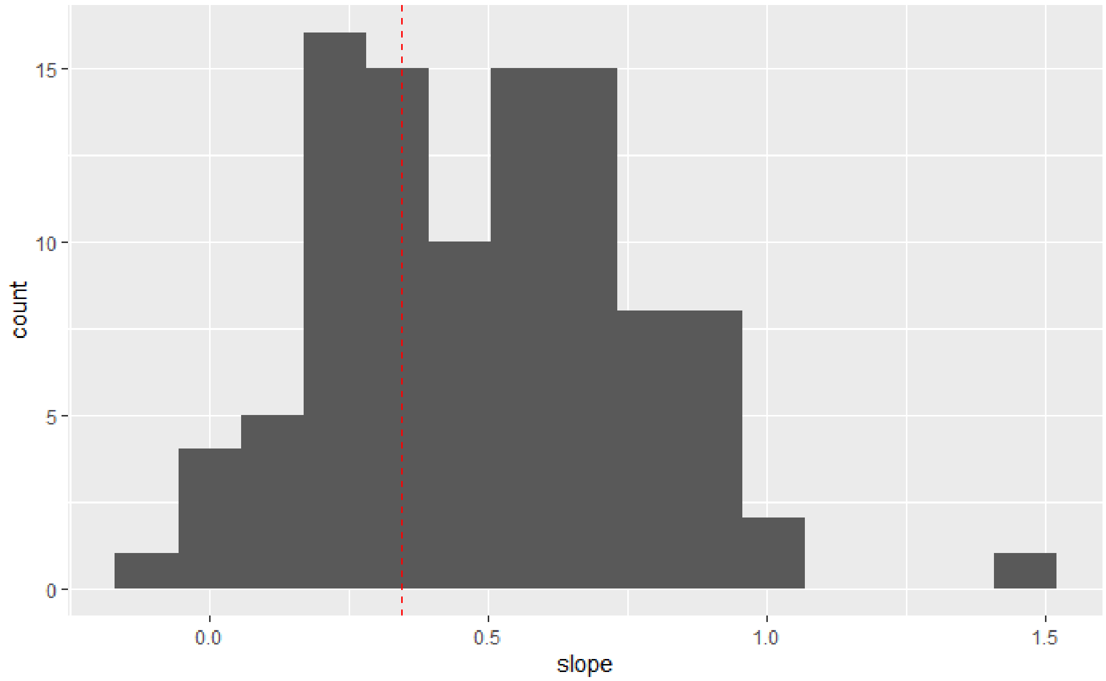
Clearly, the lines have different intercept, possibly also different slopes. However, they are not wildly different; the variation between lines appears to be governed by some sort of distribution.

Distributions of 100 regression lines

Distributions of class-wise intercepts



Distribution of class-wise slopes



Durham University

Question

(to be answered in the next lecture – will also touch on in tomorrow's lab!)

How we can we capture the inter-class (and within-class) variability of the extraversion-popularity relationship without fitting 100 separate regression models?

Durham
University

Thank you!!!!!

Durham University
Research Methods Centre

andrew.golightly@durham.ac.uk