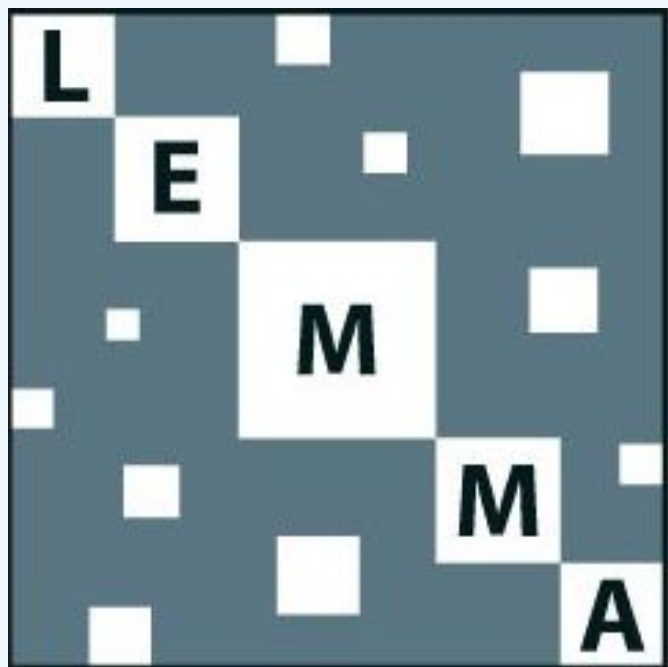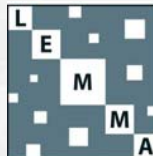# Introductory Multilevel Modelling Workshop for Educational Policy Makers and Researchers

University of Bristol

5-6 Dec 2005

Presenters:

Sally Thomas & Harvey Goldstein

Learning

Environment for

Multilevel

Methodology and

Applications

NCRM node based at University of Bristol

Staff

Kelvyn Jones   Harvey Goldstein   Edmond NG
Fiona Steele   Sally Thomas   Jon Rasbash

Background

- Considerable government investment in data collection (eg PLASC)

- Major developments in models with realistic complexity (eg multilevel modelling, structural equation modelling)

- Growing interest in evidence-based research (eg Campbell collaboration, DFES funded EPPI Centre to provide quality research summaries)
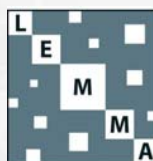
But

- Hefce report on *Strategically important and vulnerable subjects* (June 28th 2005) includes quantitative social science
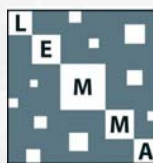
# what are the barriers to using more complex quantitative methods?

• Many researchers in the UK have complex questions and limited statistical modelling expertise.

• This means they either use simple but inappropriate techniques that they can understand or they use complex appropriate techniques that they do not understand.

• The use of simple techniques leads to the criticism that these approaches are too reductionist to ever capture social reality in any meaningful way.

• The attempt to use complex techniques without the necessary background is demoralising for practitioners and leads to bad research.

But…Lack of suitable training and training materials…Lemma (node of the ESRC funded national centre for research methods)

Lemma Has Three Inter-related Aims:

• **Important and much-needed methodological developments in the specification and estimation of multilevel models** which represent a further step-change in the capacity of these models to handle realistic complexity.

• **A variety of integrated flagship projects using this methodology to research important social science questions** Eg Geography of school effects project

• **Extensive capacity building and research training in the analysis of data with complex structure**. Most importantly a multilevel modelling virtual learning environment will be established to support geographically-dispersed researchers who have a variety of needs. See http://lemma.ggy.bris.ac.uk
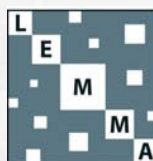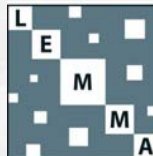
Why do we need to model data?

Current research indicates that unless we can formulate models that approach the underlying structure of the real world and the data gathering process, we will be severely hindered in our understanding of that world and in developing causal explanations.

One rationale for this view is that in the social sciences 'cause and effect' relationships are not necessarily clear-cut in the sense that:

• the outcome occurs only if the causal factor has operated

• the action of a factor always produces the outcome

Why do we need to model data?

Also …to make 'cause and effect' inferences it is necessary to assume external influences are excluded from the results

- This is often undertaken via an experimental design involving 'randomised controlled trials' (RCTs)

- However in real world situations - such as schools - RCTs are sometimes ethically difficult, expensive to set up and the assumption of no extraneous influences can be problematic.

So social scientists are often in the position where data can only be collected in the form of measures or observations of the real world – for example large-scale surveys (eg PLASC data) – these measures capture the existing natural variation in the population being studied.

Why do we need to model data?

Therefore we need techniques to model the realistic complexity of substantive research problems with the aim of uncovering underlying patterns and producing credible evidence of a relationship.

The results are non-deterministic, so we need:

• Quantitative evidence of the size of effect;

•Quantitative evidence of the size of the effect taken account of other possible causal factors

•A quantitative assessment of uncertainty about size of effect

ESRC National Centre for
Research Methods

Learning
Environment for
Multilevel
Methodology and
Applications

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

What does the Data look like?

- DATASET = VARIABLES + STRUCTURES

- VARIABLES = Attributes/characteristics measured on a variety of quantitative scales

- STRUCTURES = Simple or Complex

  Simple: non hierarchical, list of cases
  Complex: hierarchical, group membership, time series, spatial series, who /what did the measuring

- In order to model data a distinction must be made between types of variable:

  ➢ response/ outcome/ independent variable(s) and
  ➢ predictor/ explanatory/ dependent variable(s)

  (ie causal relationship is usually 'suggested' by your research question and this is 'imposed' via the selection of the response and predictor variables)
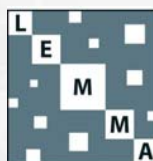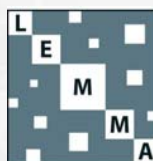
Modelling Data

Linear and multiple regression techniques are a basic form of data modelling - used to describe or 'to suggest/predict causal relationships' between a dependent variable (eg pupil performance) and a variety of explanatory variables (eg prior attainment, gender, entitlement to free school meals)

However, these basic techniques are limited:

☐• First because they do not take account of the structure or hierarchical nature of data in the social sciences.

☐• Second because they do not allow the model to take into account various forms of missing data/survey designs.

Multilevel modelling is a statistical technique developed in the last 15 years that addresses these limitations. Essentially these methods are a generalised form of multiple regression.

ESRC National Centre for
Research **M**ethods

Learning Environment for Multilevel Methodology and Applications

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

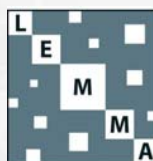# What does the Model look like?

All statistical regression models have a common form

Response= Fixed part + Random part

Fixed part = average relationship between the response and predictors

Random part = the remaining variation in the response after taking account of all included predictors

ESRC National Centre for

esearch
ethods

Learning
Environment *for*
Multilevel
Methodology *and*
Applications

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
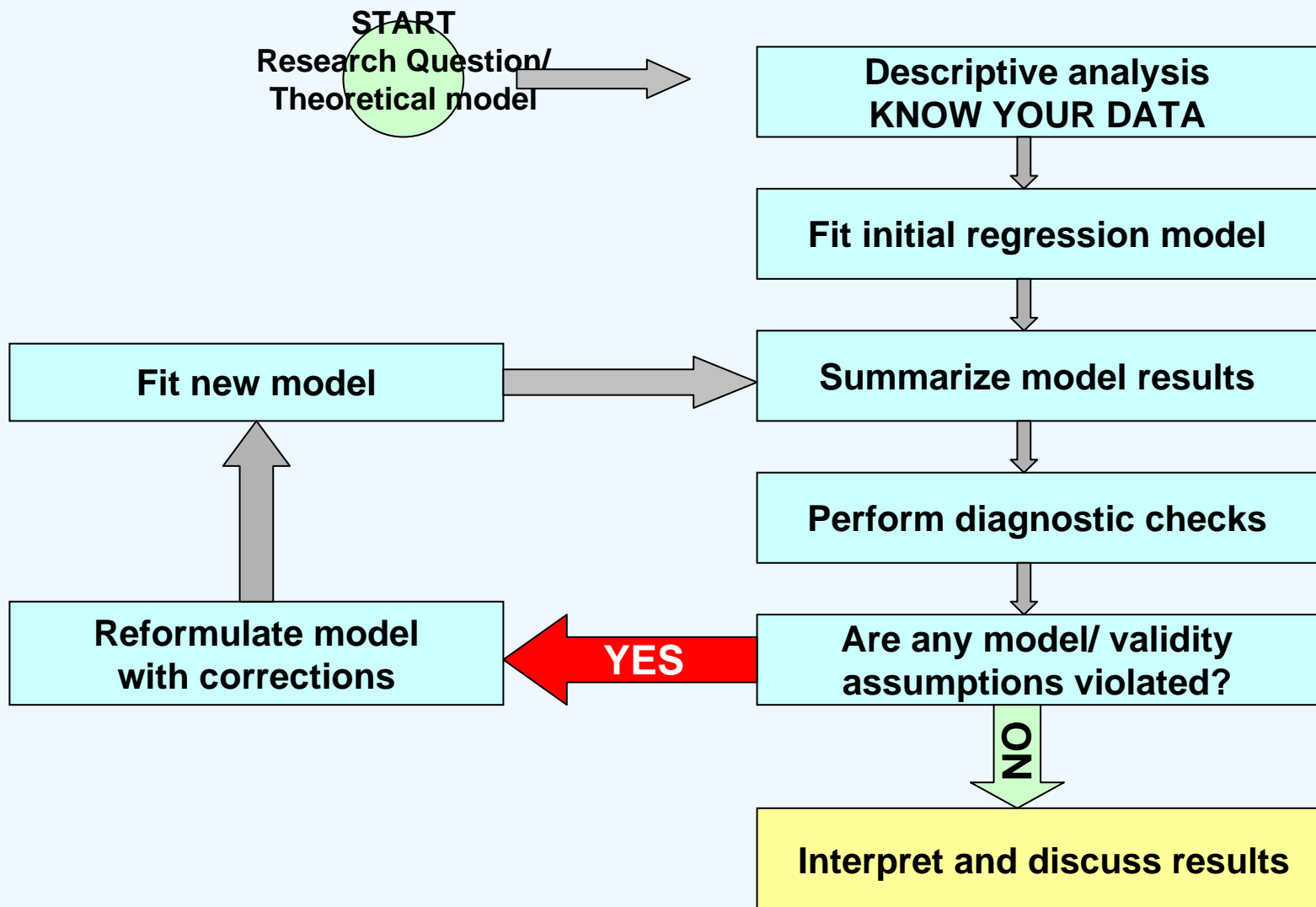COUNCIL

Graphs for relationships

Research Questions questions that can be tackled by modelling data

(assuming a random sample of pupils from a random samples of schools)

- Do Boys make greater progress than Girls (F)

- Are boys more or less variable in their progress than girls?(R)

- What is the between-school variation in progress? (R)

- Is School X different from other schools in the sample in its effect? (F)

- Are schools more variable in their progress for pupils with low prior attainment? (R)

- Does the gender gap vary across schools? (R)

ESRC National Centre for
Research Methods

Learning Environment for Multilevel Methodology and Applications

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

An ideal model building process

START
Research Question/
Theoretical model

→

Descriptive analysis
KNOW YOUR DATA

↓

Fit initial regression model

↓

Fit new model → **Summarize model results**

↓

Perform diagnostic checks

↓

Reformulate model with corrections ← **YES** ← **Are any model/ validity assumptions violated?**

NO ↓

Interpret and discuss results

ESRC National Centre for
Research Methods

Learning
Environment for
Multilevel
Methodology and
Applications

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Validity Issues

Including for example:

- Are assumptions of model violated?

- Has the structure of the data been correctly specified?

- Is there a large/representative enough sample size to generalise?

- Is measurement error a problem?

- Have all relevant variables/statistical controls been included?

- Exploratory data analysis – are the results meaningful or substantively important?
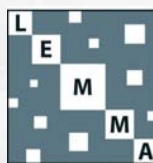
Problem: the Texas sharp shooter
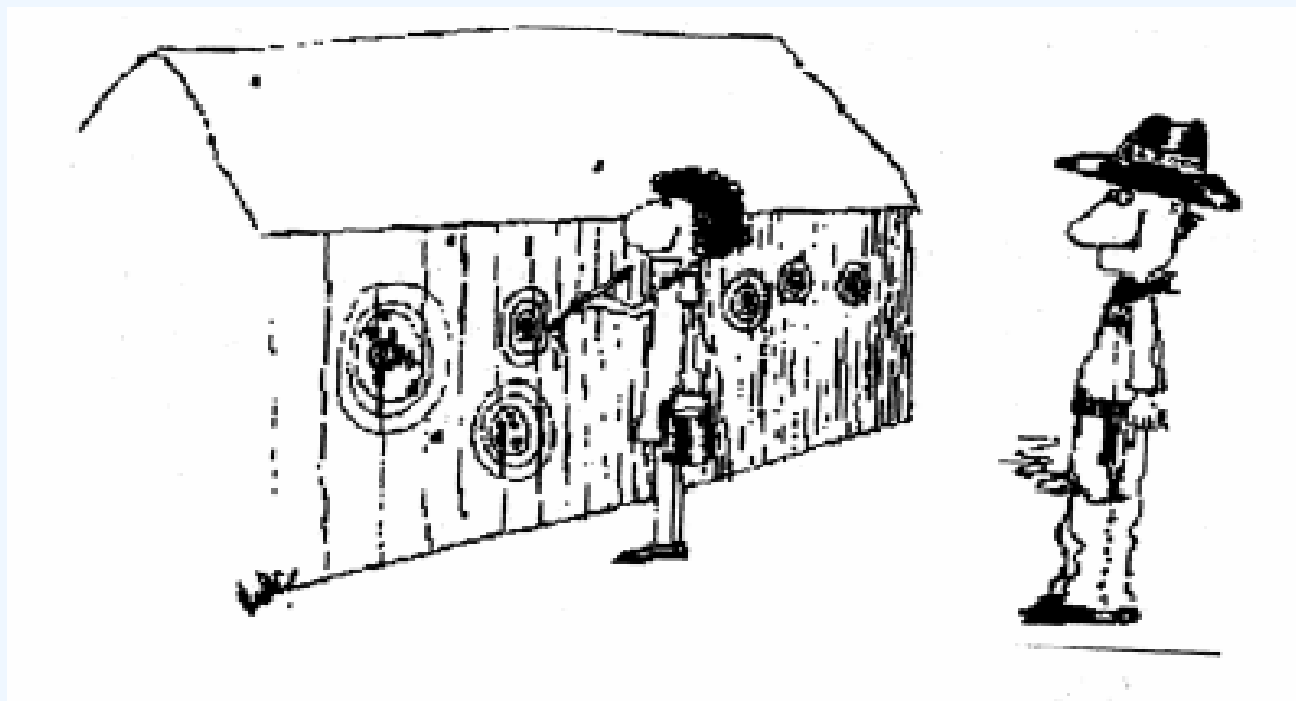
Group Task

- Think of examples of educational research questions that can be addressed by modelling data (10 mins in pairs/threes)

- Discussion

- Example of LEA Research Project which employs multilevel modelling: Lancashire LEA Value Added project