Statistical Literacy, Globalisation, and the Internet

Jim Ridgway, James Nicholson and Sean McCusker
SMART Centre, School of Education
University of Durham

Statistical Literacy – big picture

Kathrine Wallman (1993) Presidential Address to ASA

- … the ability to understand and critically evaluate statistical results that permeate our daily lives –
- ... to appreciate the contribution that statistical thinking can make in public and private, professional and personal decisions

Globalisation and the Internet

- Acceleration
- Emergence
  - Unpredictable aspects of the impact of technology
    - *Need to be nimble... AND retail core values*

Savoir Liberateur

Nicolas de Condorcet

jim.ridgway@durham.ac.uk
www.dur.ac.uk/smart.centre

Some Global Trends

- Unprecedented information flow
- Need for ‘super- government’ initiatives
  - Concerns about the need to ‘think INSIDE the box’ [e.g. FOE – in 2001 we were running at about 20% over global biocapacity; UK if scaled up globally consumes at a rate of 3.1 planets]
  - Growing consensus on the need to go ‘beyond GDP’
  - Milenium Development Goals

Some Global Trends

- Mistrust of government and government agencies
- Alienation
- Misinformation

Internet - Information Explosion

- Government and NGO websites: *data.gov; data.gov.uk; UNdata; OECD Statsportal, Google's public data explorer; tv channel data services*
- 'Semantic web' (web 3.0) – linking data sets
  - *Search by featured data sets, keywords, etc*

- Standard application programming interfaces (APIs)
    - *ProgrammableWeb documents 3500+ API*

- Apps *-Govt; private e.g.datamasher*
  - Apps for mobile phones

Internet - Information Explosion

- Mash-ups
  - Automatic updates
  - *ProgrammableWeb links to 6000 mashups*

- Communication
  - *Easy to embed* mashups *in facebook, YOuTube etc*

- Site analysis tools - ratings and commentaries

Internet – Information Explosion

- Media: New York Times, Washington Post -interactive documents
- Discussion forums VERY VARIABLE usage…
- **Information and misinformation explosion**
  - **Astroturfing, *wikiscanner***

Hey! Problem Solved…

- “It is now as easy to analyse data as it is to buy a pair of shoes or a TV on-line” (August 2011)

http://www.socrata.com/datagov/new-data-gov-platform-video-overview/

Examples:Obesity

- *Datamasher*
  - “Kids who are fat because their parents are fat”

Examples:Obesity

- Datamasher
  - “Kids who are fat because their parents are fat”
    - % obese children TIMES adult obesity rate

Examples:Obesity

- Datamasher
  - Fast food restaurants PLUS adult obesity rate PLUS % overweight children
  - % obese children DIVIDED by % obese adults
  - Number of suicide deaths PLUS % obese children

Problem Solved…??

- “It is now as easy to analyse data as it is to buy a pair of shoes or a TV on-line”
  - Small gap between Web 3.0 rhetoric and reality?

Internet - Opportunities

- NSOs educational resources: NZ, Finland, Portugal, *Canada...*
- OECD eXplorer story telling – Mikael Jern
- Gapminder and viral video – Hans Rosling
- Data driven, live updates of displays e.g. INTERSOS
- Animations, interactivity, linking

- BUT... *people find it hard to read box-plots...*

Savoir Liberateur and Stat Lit

Nicolas de Condorcet

jim.ridgway@durham.ac.uk
www.dur.ac.uk/smart.centre

Statistical Literacy is...

- Societistics….

- As well as all the familiar stuff

Challenge…

Every interesting problem in health, crime, poverty, environment, education, personal well being…

Challenge…

Every interesting problem in health, crime, poverty, environment, education, personal well being…

is multivariate

Challenge…

Every interesting problem in health, crime, poverty, environment, education, personal well being…

is multivariate

has non-linear relationships

Challenge…

Every interesting problem in health, crime, poverty, environment, education, personal well being…

is multivariate

has non-linear relationships

has confounding variables

Challenge…

Every interesting problem in health, crime, poverty, environment, education, personal well being…

is multivariate

has non-linear relationships

has confounding variables

So we might have some problems developing societistics

Statistics is ?

Statistics is… modelling

- A collection of big ideas
- A set of modelling tools…
  - With a **history**
- Question
  - What were the tools FOR?

The Weight of 1920s History…

- Data gathering was expensive
  - Aim: to generalise from samples to populations
- There was no computational power
  - Aim: find models that can work using hand calculations
- SO make assumptions about distributions (e.g. normality)
  - assume linearity
  - assume homogeneity

The Weight of History…

- Social Sciences have embraced 1920s statistics…
  - models designed in the distant past for different purposes
- Decompose, ask binary questions, sew everything back together at the end
  - hypothesis testing
- Modelling the dominant model…
  - school physics not school biology

Societistics vs 1920s Statistics

- 1920s models are often invalid because of
  - Assumptions about
    - Linearity e.g. Regression, SEM, factor analysis
    - Homogeneneity
  - Emphasis on hypothesis testing
    - Statistical significance, not effect sizes

Societistics vs 1920s Statistics

What Statistical literacies do we need for *Societistics?*

- In Societistics, we generalise from populations to subsets…

How Fair is Britain?

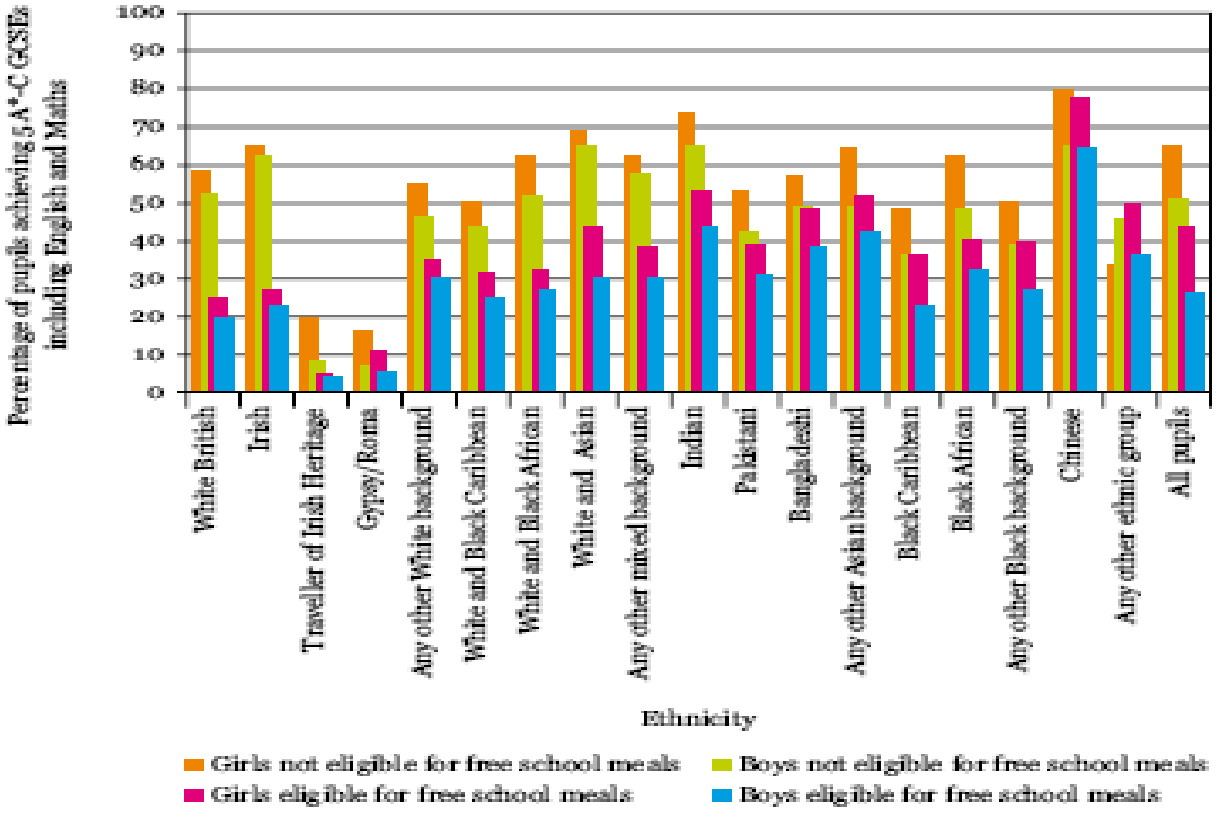
Equality and Human Rights Commission: triennial review:

How Fair is Britain? Equality, Human Rights and Good Relations in 2010

It is 748 pages with lots of multivariate data, but can it be undestood?

http://www.equalityhumanrights.com/key-projects/triennial-review/full-report-and-evidence-downloads/

Figure 10.4.1 Percentage of pupils achieving 5 A*-C GCSEs including English and Maths 2009 by gender, FSM status and ethnic group in England, 2008-09[78]



Source: Key Stage 4 attainment by Pupil Characteristics, in England 2008/09.

Education WP – foreword by the Prime Minister and Deputy Prime Minister

- *Children on free school meals do significantly worse than their peers at every stage of their education*
- *They are just half as likely to get good GCSEs as the average*
- *Chinese girls on free school meals for example … significantly outperform the national average*

A Problem – Policy Cycle

- Problem definition
- Problem exploration
  - Data
    - What is relevant?
    - What is available?
    - Is it reliable?
  - Data exploration
  - Modelling
  - Theorising
    - Informal
    - Formal
- Policy formulation, implementation, revision

Things we need to make progress

- ‘Biological’ thinking – symbiosis, mutation, evolution
- More Players – politicians, journalists, NSIs, citizens…

Implications…

You have to be able to DESCRIBE the phenomena before you begin

DESCRIPTION brings you face to face with big statistical ideas – quality of data, study design, measurement error, interaction, effect size…

Two Strong Claims

Semi-qualitative descriptions often give more ‘bangs per buck’ than quantitative analyses of complex situations

You get to big statistical ideas if you work

- Top-Down (look at data on teenage drinking) – not Bottom-Up (t-test to anova...)

Statistical Literacy – big picture

Kathrine Wallman (1993) Presidential Address to ASA

- … the ability to understand and critically evaluate statistical results that permeate our daily lives
- ... to appreciate the contribution that statistical thinking can make in public and private, professional and personal decisions

Statistical Literacy 2011 is…

- Good ‘that’s a lie’ antennae
  - Awareness of rhetorical devices

- Knowing a lot of FACTS about the domain of interest!!!

Statistical Literacy 2011 is…

- Awareness of the ‘politics of data’
  - the choice of measures reflects values and philosophies
  - Aesthetics of measurement

- Sympathy for the role of evidence in shaping beliefs and policies (not determining them) – so some Bayesian thinking
  - Understanding meta analysis

- Understanding risk, and utility

Statistical Literacy - heuristics

- Check the effect size is a lot bigger than the likely error of measurement [think about confidence intervals]
- Focus on effect size not significance level
- Identify variables that have the strongest effects
- Look at absolute levels – are they big enough to be worth worrying about?

Statistical Literacy - heuristics

- Look for non-linear relationships
  - Explore the effects over different values of each variable
- Look for changes over time
- Look for interactions, and think about ‘data surfaces’
- Think about possible confounding variables

- Disaggregate data, are the patterns the same?

Actions

- Rethink ‘statistical literacy’ *continuously*
- Capitalise on new visualisations *continuously*
- Treat statistics tutorials like metadata
  - Ubiquitous; tutorial support; self test and diagnosis
- Distribute ‘themed’ high data applets with commentaries
- Make societistics a major focus of the school curriculum

Conclusions

- Urgent need for ‘evidence informed’ policy – NSOs need to do this to survive...
- Continue to capitalise on new tools
- Rebalance ideas on ‘the big ideas in statistics’
- Rethink ‘hard to understand’ – embed tutorial support everywhere
- We are going to be redefining ‘statistical literacy’ for rather a long time…

Statistical Literacy, Globalisation, and the Internet

Jim Ridgway, James Nicholson and Sean McCusker
SMART Centre, School of Education
University of Durham

And…

- Statistics is ‘hard’ *[is it?]*
- There is a need for a long ‘statistical apprenticeship’ learning to master difficult technique *[is this true?]*

HOWEVER people like to make sense of things, and *do* use evidence…

Progress Indicators

We are *FAILING* if…

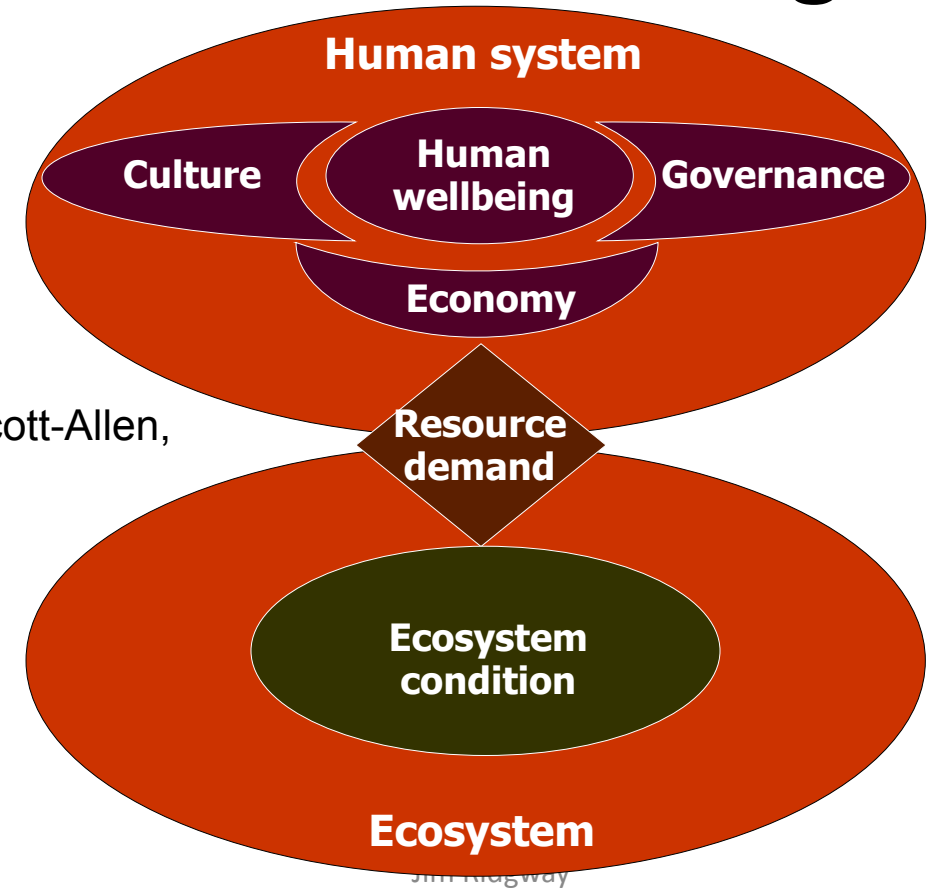
- It is OK to say
  - ‘my feeling about this is…’ G. Bush (senior)
  - ‘I can’t do mathematics’ celebrity interviews

Progress Indicators

We are *SUCCEEDING* if…

- MV data are used routinely…
- In the media
- In government policy documents
  - E.g. release of beta versions with interactive data displays
- In one's private life

Dimensions of Progress

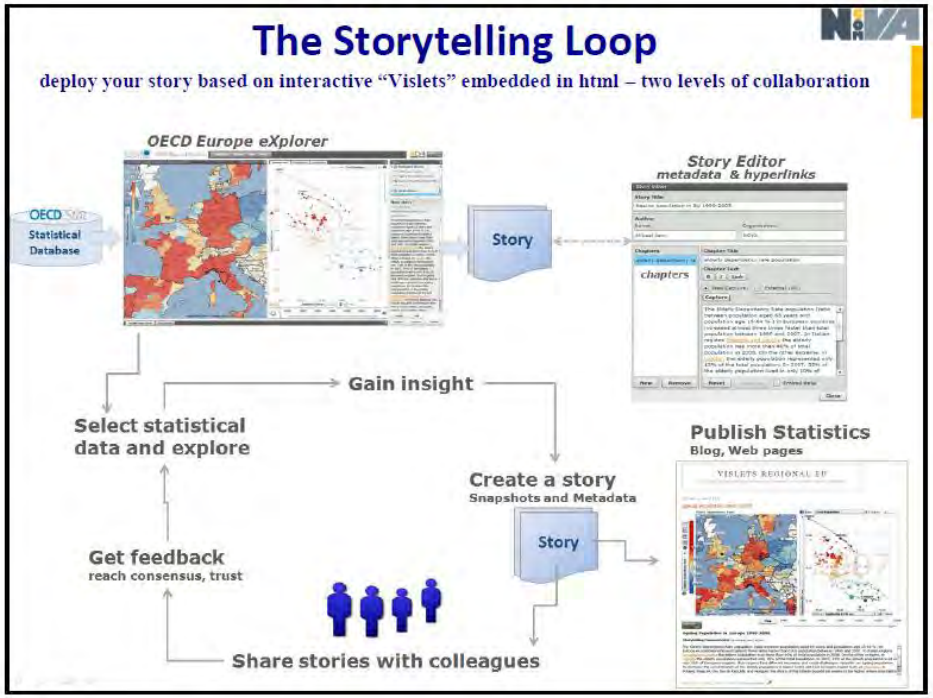


Source: Robert Prescott-Allen, 2008

Examples:Obesity

- National obesity Comparison tool
  - http://public.tableausoftware.com/views/contributorstoobesity/Eatyourvegtables?:embed=yes&:toolbar=yes

Mikael Jern NCVA



From ‘Statistics’ to ‘Societistics’?

‘Measuring Progress’ won’t work if:

- The measures are not shared
- They are not communicated to the whole society
- Citizens do not understand them

SMART Centre Research

Understanding Sense Making Based on Evidence

- Rethinking 'statistics' – *what?* And *where?*
- Interface design NZ wages, calculating axes, water fleas
- Defining and describing 'new literacies'
  - What skills are critical for dealing with mis/information?
  - Are there hierarchies of knowledge?
  - What heuristics are useful?
- Misconceptions are?
  - diagnostic actions should be?
- Engagement with a variety of communities

Conceptual and Research Issues

The development of both a semi-quantitative and semi-qualitative approach to understanding evidence

- For any situation
  - What are the vices and virtues of applying quantitative methods?
  - What are the vices and virtues of applying qualitative methods?
  - How do we maximise understanding by using both approaches?

Understanding ‘New Literacies’

- Tracking a moving target [PISA, PIAAC]
- Understanding user understanding
- Revise conceptions of ‘Statistics’
  - Media
  - NSOs
  - School and university
  - Social science

Distribute provocative applets

- Alcohol consumption
- Sexually transmitted diseases

Press Release Horror

- ***Longer Life Expectancy for Men and Women***
- *The report* ***Measuring Ireland's Progress, 2007****, …..*

Activity

Life Expectancy

- What are the salient features of the data?
- What issues does the data raise?
- What would YOU write in a press release?
- What ‘statistical thinking’ did you use?

- *Life expectancy at birth was provisionally estimated at 81.5 years for Irish women and 76.7 years for Irish men in the period 2004-2006. In comparison with 2001-2003, men's life expectancy increased by 1.6 years and women's by 1.2 years, reducing the gap between men and women to 4.8 years in 2004-2006, the lowest it has been since the 1970-1972 period (Table 6.3).*