

Towards a Critical Epistemology of Analytical Statistics

Wendy Olsen and Jamie
Morgan, British Sociological
Association, March 2004
(York)

Please write to the authors for a copy of the paper. It
is under review for a journal. Write to:
wendy.olsen@man.ac.uk

Analytical statistics

- ...is a process of analysis which creates survey data, links it to other forms of data, transforms the data, and analyses the results, finally *interpreting* the data
- ...involves both qualitative and quantitative data
- ...is in danger of presenting as fact that which is only a representation of reality

Induction: Facts from Details?

An epistemological dilemma

The ‘problem of induction’ (Hacking)

Solution 1. Suppose the data are facts. χ

Solution 2. Suppose the data are representations but the world around us is real. ✓

Do the ontological limitations of analytical statistics render it incompatible with other aspects of good research? No.

Analytical statistics have a limited role.

- They might be appropriate in the investigation of aspects of society that tend to exhibit or approximate regularity in their relations in a given milieu at a given time.

...But an open-systems approach to reality is needed.

...methodological closure is not the same as assuming that closure exists in reality.

The argument has 8 stages.

Stage 1 dissociates our notion of methodological closure from closure in reality.

METHODOLOGICAL
CLOSURE

VS.

CLOSURE IN REALITY

- Closure in reality is a much stronger assumption and is not supported by evidence.
- The realist *a priori* is not a simple essentialism but rather a dialectical assumption

Methodological closure

- 1. The variables would be self-contained and of sufficient interest;
- 2. Regularities among the variables and cases highlight differentiable parts of the whole;
- 3. These regularities may have some continuity or duration over space/time.

Closure in reality

- 1. A non-permeable boundary to the system being looked at (not correct)
- 2. Causal mechanisms that are separable (not correct)
- 3. No emergent properties (not correct)

Re-locating data as non-factual

- A variable is not to be equated with a causal mechanism.
- Variables should be seen as ficts.
- Variables are not facts and are not truth claims. They do not constitute knowledge.

An example of a descriptive statistic

- The odds ratio
- Consider Londoners 3.2 to 1 odds of labour-market participation (2000, BHPS data)
- Outside London this figure was 2.4:1
- The ratio is $3.2/2.4 = 1.5$, much greater than a 1:1 odds ratio.

Authors advising the use of descriptive statistics from a realist perspective

- Andrew Sayer (1984, 2nd ed 1992; 1997; 1999)
- Tony Lawson (1989; 1997; 1999; 2003)

Realist authors arguing for more sophisticated statistical techniques

- Walby (2001)
- Olsen (2003a, b), and this paper.
- E.g.:
- Logistic regression offers opportunities for improved knowledge claims, compared with the odds ratio, and helps us interpret the odds ratio as well as qualitative data.

Logistic regression

- Multiple regression framework
- A qualitative outcome
- Labour-market participation
- Gender issues can be un-picked, e.g. mothering vs. wife-hood and the husband's income which affects a woman's likelihood of participating at a given time

Regression Results for 2000

log of the odds of employment =

$$-1.47(\text{LTLI}) + 0.27 * \text{London} + .61 * \text{Degree} - \\ 0.76 * \text{Noqual} + 0.92 * \text{Wife} + .61$$

Each number shows whether the odds of being employed are raised or lowered by the presence of a given characteristic. In this equation, the following definitions are used:

LTLI = Long-term limiting illness

(specifically, the person reports that they are unable to do some forms of work, due to an illness or other disabling condition)

London= Lives in Inner or Outer London or the rest of the Southeast

Degree = Has a degree and/or a higher degree

Noqual= Has no qualifications ie no CSEs or O-Levels or other qualifications

Wife = Is married or cohabiting, and is female

Intellectual work done by interpreter

- The people doing interpretation are socially located, grounded, responsible, and time/place-specific
 - (This has the formal label ‘epistemic relativism’)
- The interpreter is pluralist at the level of theory
- And is methodological pluralist if other data-types and sources of data, or other viewpoints, are being considered (see New and Carter, eds., forthcoming book)
- This approach allows for structure-agency dialectic

Conclusions

- THE WORLD HAS A MULTI-LEVEL STRUCTURE, WHICH IS NON-NESTED, COMPLEX AND INCLUDES STRUCTURE-AGENCY DIALECTICAL RELATIONS.
- ANALYTICAL STATISTICAL TECHNIQUES CAN CONTRIBUTE TO OUR KNOWLEDGE OF OUR WORLD.
- COUNTER-PHENOMENAL DISCOVERIES CAN OCCUR.