

RADICAL STATISTICS ESSAY COMPETITION 2014

SECOND PLACE: OPEN COMPETITION

Path Analysis for People who Hate Statistics

Dr. Pauline McGovern

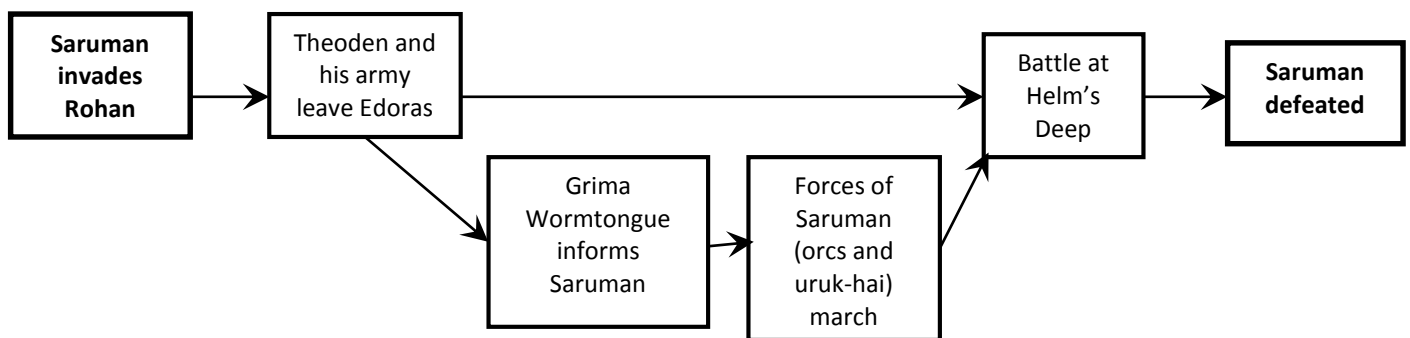
What is path analysis?

Path analysis is an advanced statistical method that enables sociologists to account for what happens in society by tracing the pathways through which causal processes lead to outcomes. It sounds complicated but is actually how we describe what happens a lot of the time! Path analysis is a great tool to use to analyse a lot of information at one time - for example, from surveys.

In a path model, between initial cause and final outcome, there may be a number of intervening factors that are ordered on the basis of time and these have the potential to affect or even negate the outcome. Here's an example of a path model. In the Lord of the Rings, Saruman and his army battled for dominion over Rohan but were defeated by the Rohirrim and their allies at Helm's Deep – cause and effect. However between the initial cause and the final outcome, there were intervening events: Theoden (the king of Rohan) took his army from the castle of Edoras to the fortress of Helm's Deep (more defensible); Grima

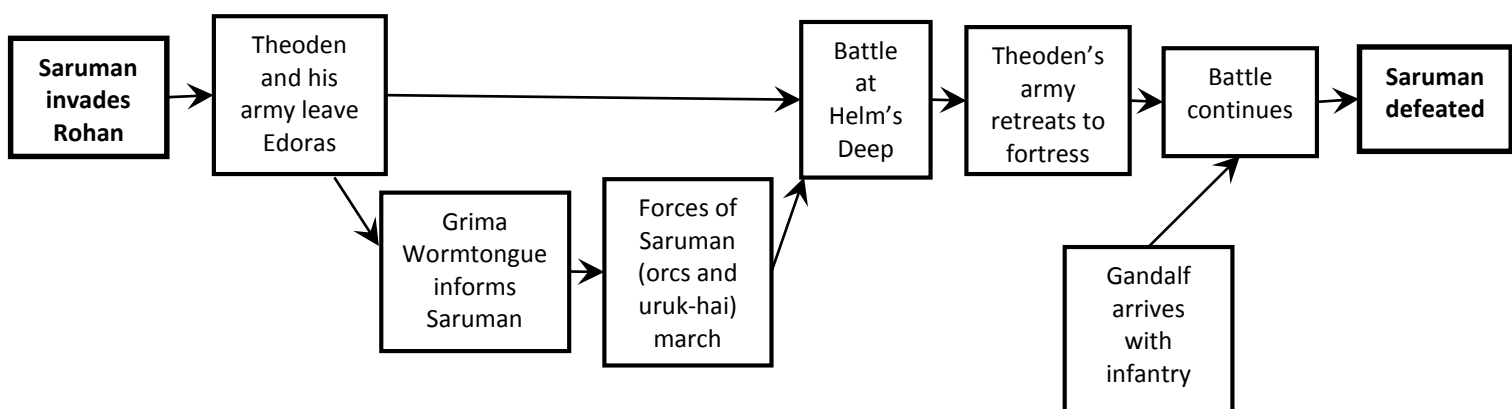
Wormtongue betrayed them; Saruman’s army of orcs and uruk-hai marched on Helm’s Deep; a battle commenced and Saruman’s army was defeated (see Figure 1).

Figure 1: Path Model of the Battle for Rohan



A path model is based on ideas about how things fit together (a theory) and may miss out some essential element and, as a result, be a poor explanation of what happened. For example, I haven’t included the fact that Gandalf saved the day. When the walls of Helm’s Deep were breached by Saruman’s army and Theoden’s army had to retreat, he arrived with 1,000 infantry from Rohan and the battle was won. Here is the amended path model:

Figure 2: Modified Path Model - the Factors that Affected the Outcome of the Battle for Rohan



Real life example: does former occupation affect the health of retired people?

Path analysis is fantastic for exploring a range of possible causes (predictors) and assessing both their total effect on a particular outcome and the effect of each predictor on the outcome. I have been using path analysis to explore whether former occupation has an effect on the health of retired people. Research suggests that there is a link between occupation and health for employed people. A solicitor or architect has more likelihood of being healthy than a street cleaner or shop assistant. Of course this does not mean that *all* people of high social class are healthier in every way than all people of lower class - these are probabilities not certainties - but it does suggest that health inequalities are class-related.

The relationship between occupation and health gets more complicated for retired people, because they are more distant from their working conditions. So, how might the former occupation of retired people affect their health status? Path analysis can test possible answers that are suggested by theory. Retired people generally have space for leisure pursuits and, having a specific former occupation and the income that went with it may affect their present lifestyle choices - their tastes, habits and interests. A path model can be constructed that gives these variables a time order in relation to health.

My data are from five waves of the *English Longitudinal Survey of Ageing* (2002/03 to 2010/12) and respondents are 50+ years old. Using computer software, the hard work of path analysis is done for you. I used Amos (part of a well-known software package called SPSS).

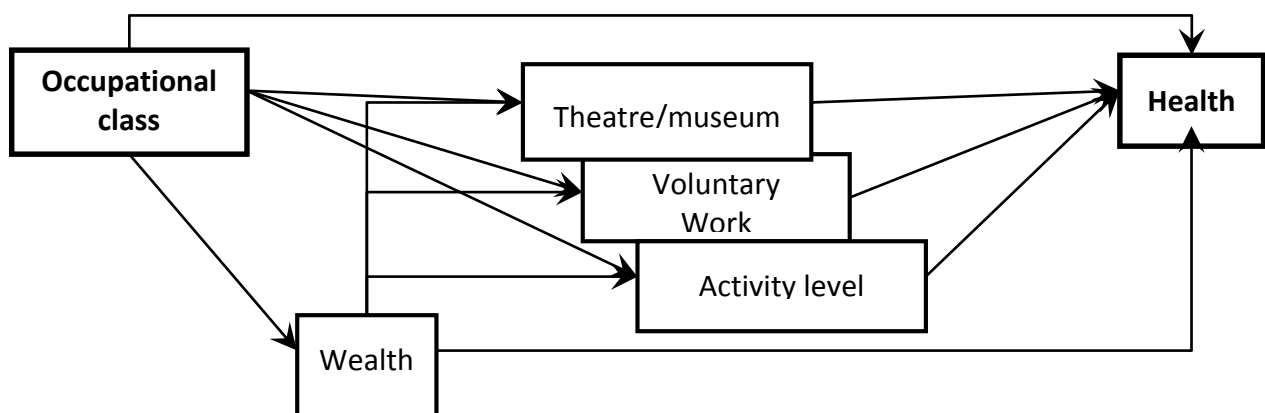
The path model of former occupation and health

With path analysis, we start by drawing a diagram of what we think are the relationships between variables that predict an outcome (we *hypothesise* the relationships based on our theory). Lines with arrows show the direction of cause and effect and the predictors are ordered from left to right on the basis of time. In this example, the predictors are lagged at an earlier time point than the health outcomes so the

hypothesised direction of causality is not open to debate. The predictors are taken from the first *ELSA* wave (2002/03) and I have taken the worst health score that respondents have in the next four waves as the outcome variable.

The path model of the relationship between occupation and health for retired people is in Figure 3. The order of variables and direction of arrows is based on our theory, which is influenced by Pierre Bourdieu's seminal work, *Distinction* (1979). Bourdieu argued that practice (in this case, health) derives from class. He drew a distinction between the primary occupational characteristics by which people are classified and secondary properties of class that reflect lifestyle. I use the National Statistics socio-economic classification (NS-SEC) as the primary component because it classifies occupations as Bourdieu himself did, in terms of working relations and conditions. For the secondary lifestyle properties, I use measures of respondents' total wealth, theatre/museum-going, volunteering and level of exercise. In this example, health status is measured using people's own perceptions of their health (they are asked how good their health is and the scale runs from 'poor' to 'excellent').

Figure 3: Path Model of the Hypothesised Relationships Between Occupation and Health



Path analysis makes it possible to assess the direct, indirect and total effects of the predictors on health. In paths drawn in this path model, there is a *direct* path from former occupation to lifestyle and wealth and *indirect* paths to health via lifestyle and wealth. Notice that there is also a direct path from occupation to health as well, because occupation may have an effect on health that does not go through wealth and lifestyle.

Finally, if this isn't too confusing, as well as direct and indirect effects, we can calculate the *total* effects of each predictor on health (the combination of direct and indirect effects).

Other aspects of path analysis

Apart from the predictors in our path model, there are other known causes of ill-health in older people that have to be taken into account. We know that the older a person is, the more probability there is that he or she may be ill. A 90 year old is unlikely to be as healthy as someone of 60. Males and females may also suffer different kinds of ill-health and women tend to live longer than men (we have to have some advantage). Because these are well-known causes, I controlled for the possible effects of age and sex by adding them to the path model as predictors of everything! Sometimes people call this a parsimonious analysis because it accounts for the effects of known predictors thoroughly. Also the effect of age is not constant – the rate at which people get ill speeds up as they get older. Squaring the value of a person's age (age^2) allows for this and so this extra variable was also added as a control.

When a path analysis is run using computer software, there are three key statistics that are used to assess the relative effects of predictors on the outcome. Path analysis is based on linear regression. Linear regression describes relationships between predictors and outcomes in terms of how closely the change in one relates to a change in the other. So, if Alice grew taller in relation to the amount of the growing potion that she drank, then the relationship between the potion and height would be linear. For each path between two variables in our model, path analysis provides a statistic that measures the extent to which these two variables are related. This is called a *path coefficient* and I use the *standardised* coefficient to account for their different measurement scales when comparing their relative effects on health.

Path analysis also provides measures of the overall fit between all the predictors and the outcome variable. This is generally the most important statistic to look at first because if the model does not fit the data, then it is a non-starter. If you buy clothes for an elephant at Asda, they are not going to fit and it is no use saying that the arm length is okay. There are several *goodness-of-fit* measures that are used for different sample sizes. For this analysis, with a sample of 6,000 retired people, I chose a goodness-of-fit indicator that is used for large samples. This is the Root Mean Square Error of Approximation (RMSEA). RMSEA is based on chi-square. If RMSEA is over 0.1, this generally means that the path model is badly fitting.

There is another useful statistic produced in path analysis called R^2 . This is an assessment of the amount of variability in the outcome (in this case, health) that is caused by the predictors in the path model. It

can be converted to a percentage to make it easier to understand. This shows us how much is unexplained by the model and, if most of the change in health is not due to the predictors, this might suggest we are missing some important causes of good health in older people (such as not smoking, eating lots of ice cream or boogie-ing the night away).

The results

Now that the hard bit is done, we can look at the results. It was good news to find that the model was a reasonably good fit (RMSEA=0.082). However, the predictors in the model accounted for only a very small amount of variability in health status ($R^2 = 0.075$ or, put another way, 7.5%). It is not surprising to find that a big chunk of what explains health status is missing in this model. Any quantitative analysis reduces the real world to a few elements and I am not suggesting that this is a complete explanation of the health status of older people. The point of this analysis is not to completely explain health but to understand some aspects of the relationship between health and class.

Having found the model to be a good fit, the next stage is to look at the size of the direct, indirect and total effects of the predictors on health using the standardised coefficients. A main finding is that former occupation had the biggest direct effect of all the predictors on health. Half of its total effect was direct. To this extent, retired people are still marked by the occupations of their working years, just as employed people are. However, the beauty of path analysis is that we can see that half of the effect of occupation on health is indirect – it goes through the wealth that people have and their lifestyles.

The findings also show that having lots of money is likely to give people better health! 85% of the total effect of personal wealth on health was direct rather than through its influence on a person's lifestyle choices. It is likely that the direct effect of wealth presupposes indirect effects via a range of material factors that promote good health that are missing from our path model. There are lots of things that money buys in addition to going to the theatre and keeping active. These include the extent to which people have choice in their houses and location, in their diet and in whether they take holidays. This finding might lead to a revision of the original path model to include other predictors that are missing.

Surprisingly, going to the theatre or to museums also had an effect on health that was comparable to that of wealth and nearly as much as that of former occupation. This is a big effect! Another unexpected

finding was that volunteering did not appear to have much effect on the health of people in this sample.

I should sound a note of caution about these results. In path analysis, the model used is based on theoretical assumptions – and using a different theory with other predictors and outcome variables, ordered in a different way might lead to different conclusions!

So why does this matter?

This path analysis suggests that class-related health inequalities do persist for older people. It shows that former occupation has an enduring effect on the health of people who are no longer in the workplace. It also shows that the amount of wealth that older people have is linked to their health status, that taking part in cultural activities is good for you, and that both of these factors are related to and explain some of the occupational class effect. And, yes, these results could have been found using other quantitative research techniques.

What path analysis adds is about *how* the effect of occupation on health is filtered through these other characteristics. It suggests that a large amount of the effect of former occupation on the health of older people is not direct but indirect – through their personal wealth and lifestyle. This potentially provides evidence for government policy. There are ten million people in the UK over 65 years old (that is one in six people) and this number is growing. In 2007/08, the average care bill for NHS services to retired households was over £5,000 per household per year (www.parliament.uk). If the government want to reduce costs to the NHS from ill-health in the elderly, it may be a good thing to subsidise museums and theatre groups and sports activities in order to keep the minds and bodies of older people as active as possible for as long as possible.

A final thought on the future of Middle Earth...

So what does path analysis tell us about how Middle Earth was saved? In the battle for Rohan, the direct effect of Theoden and his army at Helm's Deep was weaker than the total effect of Saruman's forces. Otherwise, Theoden would not have had to retreat to the fortress and, in so doing, put the future of Middle Earth at risk. Luckily for Middle Earth, the direct effect of Gandalf galloping to the rescue with the Rohirrim was to tip the balance of total effect in favour of Theoden. The forces for good became stronger and the battle was won!

References and Further Reading

To explore the relationship between social class and health further, the Joseph Rowntree Foundation website, publications page (www.jrf.org/publications) is a good place to start:

- Exworthy M., Stuart M., Blane D., Marmot M. (March, 2003) Tackling Health Inequalities Since the Acheson Inquiry *Joseph Rowntree Foundation*
<http://www.jrf.org.uk/publications/tackling-health-inequalities-acheson-inquiry> - look at the 4 page summary.
- Rowlingson K. (September, 2011) Does Income Inequality Cause Health and Social Problems? *Joseph Rowntree Foundation*
<http://www.jrf.org.uk/publications/income-inequality-health-social-problems> - look at the 4 page summary.

For an example of a health-related path analysis:

- Honjo K., Tsutsumi A., Kawachi I., Kawakami N. (2006) What Accounts for the Relationship Between Social Class and Smoking Cessation? Results of a Path Analysis *Social Science and Medicine* 62:317-328 – this is not easy to read but look at the path model on p. 318 and try reading the Discussion that starts on p. 325.

There are no simple books on path analysis but if you are interested to explore this technique further, look at:

- Mertler C.A., Vannatta R.A. (2005) *Advanced and Multivariate Statistical Methods: Practical Application and Interpretation* California:Pyrczak Publishing – or for a free download, try: <http://www.gobookee.net/mertler-and-vannatta/>

Dr. Pauline McGovern, Centre for Positive Ageing, University of Greenwich, p.mcgovern@greenwich.ac.uk