

# Rates Theory

Ivan Hanigan

August 17, 2012



---

# Introduction to rates, standardisation and adjustment

The aim of this project is to explore the methodological issues of rates, standardisation and adjustment in regression models. This topic is relevant in epidemiology and demography (and most human sciences).

The demonstration will utilize the R package *ProjectTemplate*.

# Theoretical background

Different age or sex structures of study populations can be the reason for important differences between the health outcomes of the populations that need to be accounted for when assessing other putative causal relationships such as socio-economic status or the physical environment.

There are two common methods for age standardisation: direct and indirect. There is also an alternative method using Poisson models that adjust for age as a covariate.

# Compared to what?

Question: Hi Epidemiologist, how are you?

Epidemiologist: Compared to what?

# Standardisation strengths and weaknesses

The direct and indirect methods have a long history of use to control for age (and indeed other differences between populations such as sex and other stratifying variables). However, the idea that indirect standardisation is based on the non-study population as the common standard is a misconception which results in a common methodological error: comparing indirectly age standardised incidence ratios (these can only really compare the study population to the standard population, not between study populations).

While direct standardization does provide comparable measures it has other weaknesses, such as greater susceptibility than the indirect method to error with small numbers. Indirect incidence ratios can be compared if you make the assumption that the ratio of rates between the study and standard populations is constant; this is similar to the assumption of proportional hazards in Cox regression.

# Adjustment in regression

The aim of standardisation is to control for a compositional variable. This, of course, is also one of the main aims of regression analysis. One could analyse the data using Poisson regression, a method appropriate for small numbers. An advantage of the regression approach is that one can easily control for multiple confounders. Also, we can test for the presence of an interaction, which would question the validity of the additive model underlying direct standardization.

init



# Additions

# Data

- ▶ get the princeton tutorial

The website by German Rodriguez from Princeton is good  
<http://data.princeton.edu/eco572/std.html> Has some data and methods, comparing with book Demography: measuring and modeling population processes? Samuel H. Preston, Patrick Heuveline, Michel Guillot - 2001. get data from <http://data.princeton.edu/eco572/datasets/preston21long.dat> on 13-4-12

- ▶ get the stata tutorial

We will use data borrowed from Kahn and Sempos (1989, 95-105) that are available on the Stata website, and in the datasets and do-files subdirectory. The problem is (Stata 9, Ref A-J, p. 310), We want to compare 1970 mortality rates in California and Maine, adjusting for age. Although we have age-specific population counts for the two states, we lack age-specific death rates. In this situation, direct standardization is not feasible. We can use the US population census data for the same year to produce indirectly

# Analysis

# Direct standardisation

# Indirect standardisation

# Adjustment using regression

# Control for secular trend

# Uses in spatial epidemiology



# Indirect standardisation controlling for spatial correlation

We'll use the example of the Conditional Autoregressive (CAR) model of Lip cancer in Scotland.

# Regression approach to spatial rates

Mantel and Stark (1968), with reference to an alternative approach to indirect age standardisation. This is useful when the data are being internally standardised (using the data themselves as the standard) and where there is potential confounding. The general approach is to use a regression model with the variable to be standardised (eg age) and with the stratification variable which is potentially confounded (eg area). The standardised rates by the stratification variable can then be found from the regression predictions scaled to the observed total. Note that this approach requires non-zero cells for each stratum (eg at least one event per area).

# Weight by inverse of variance

# References