Methodological note

PIN Flash 21 – Reducing Road Deaths for Young People Aged 15-30

Regression estimation of the average annual percentage change in road deaths or mortality rates over the past decade

To estimate the average yearly percentage change in road deaths occurring over a given period, one should make use of the whole time series of count, not just the counts in the first and the last year.

The task is to estimate the average annual change in road deaths in the period 2001-2010 using the yearly recorded values in each country. The procedure is the same for mortality rates

We assume a priori a reduction in deaths or mortality rate over time, so to fix the sign of a change; we will assume reduction, so that a minus sign indicates an increase. Let the average reduction per year as a percentage of the previous year be p. If D_n is the number of deaths in year n, then we wish to fit a model $D_n = K{*}D_0(1 - p/100)^n$, where K is typically close to 1 and takes account of the extent to which D_0 was abnormally high or low. In this case year 0 is 2001 and $n = 9$ in 2010.

This is equivalent to $\ln(D_n/D_0) = \ln K + n. \ln(1 - p/100)$ so if we fit $\ln(D_n/D_0) = a + bn$ by linear regression, then a is the estimate of $\ln K$, b is the estimate of $\ln(1 - p/100)$ and p is estimated by $100(1 - e^b)$.

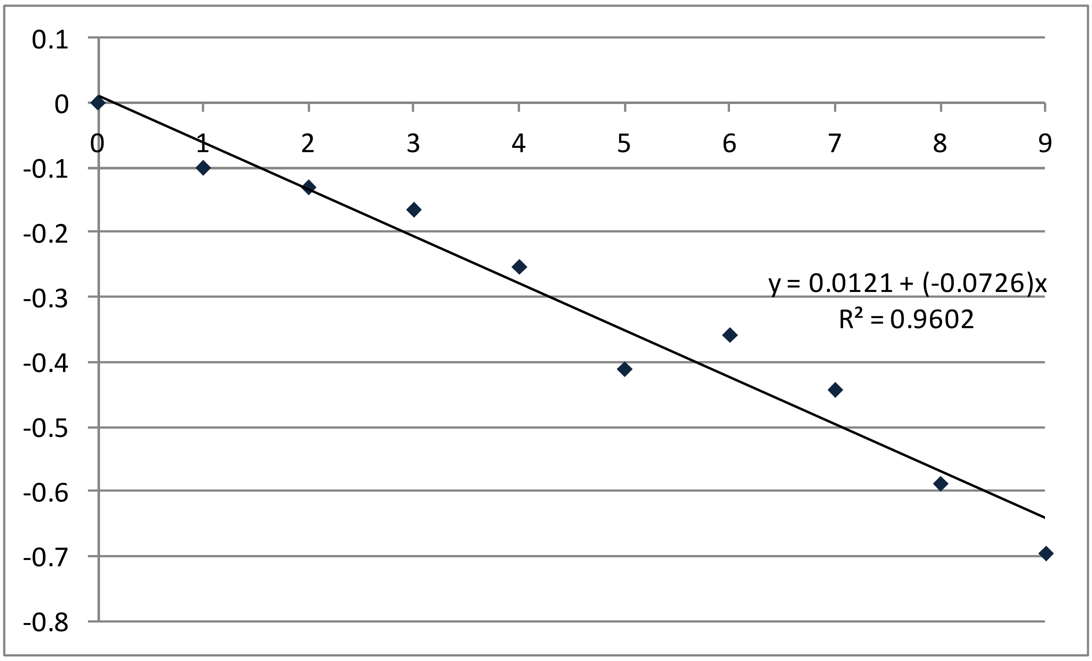

Fig.1: Linear regression function for logarithmically transformed changes in death counts with 2001 as the baseline year.

In this figure illustrating the use of the method and constructed for Austria, the function $\ln(D_n/D_0) = a + bn$ corresponds to the function $y=a+bx$, so the b is equal -0.0726 and K, which is estimated by e^a, is 1.013, close to 1 as expected. The p can now be estimated as $100(1 - e^b) = 100(1 - e^{-0.0726}) = 7.01$. Average yearly reduction in road deaths is thus estimated as 7.0%. One can conclude that the average annual reduction in road deaths over the period 2001-2010 has been about 7%.

ETSC has revised the method of estimating the annual average percentage change from previous estimations which up to now used a linear regression model of the form *y=ax.* Allowing the regression line intercept to be different than 0 allows for a better estimate of the linear model (higher R^2 value). Moreover, in the $y=ax$ model the regression line was forced to go through the origin thus giving too much weight to the starting year in the time series (because the origin corresponds to the baseline year as $\ln(D_n/D_0)=0$ for n=0). By allowing the intercept to be different than 0 in the $y=a+bx$ model the baseline year for the time series is thus given equal weight in calculating the slope b of the regression line and the model is adjusted for the possibility that the baseline year in the time series might have been one with an abnormally high (or low) number of road deaths.