Part 1: Infant Death Rates

Arden Miller

Executive Summary

There is a wide range in infant death rates for the countries included in this dataset. Different regions of the world have very different infant death rates with the lowest rates being in Western Europe, North America, Australia, and Japan and the highest rates in Africa. The relationship between the infant death rate of a country and its death rate is quite interesting. There is a strong positive relationship between these rates for countries from Asia, Africa, the Middle East, and South America but they are essentially unrelated for countries from Eastern Europe, Western Europe, North America, Australia, and Japan. This difference may be related to differences in wealth. If countries are grouped by gross national product (GNP), then as GNP increases the relationship between infant death rate and death rate gets much weaker. It would appear that for rich countries death rates are being affected by different factors than infant death rates. A similar pattern is observed for the relationship between infant death rates and birth rates. For poor countries (low GNP) there is a strong positive relationship but as GNP increases this relationship becomes much weaker.

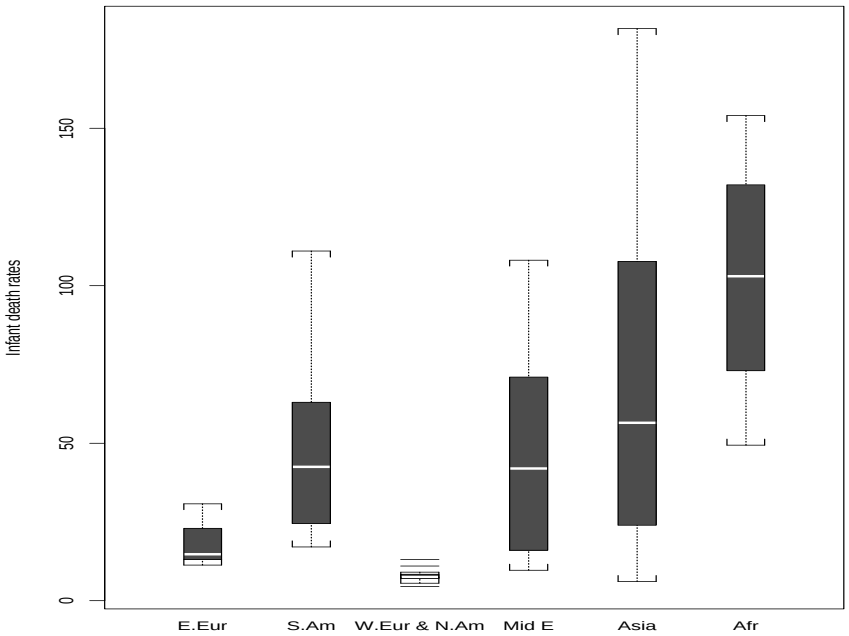

Figure 1: Boxplot of infant death rates for different regions of the world

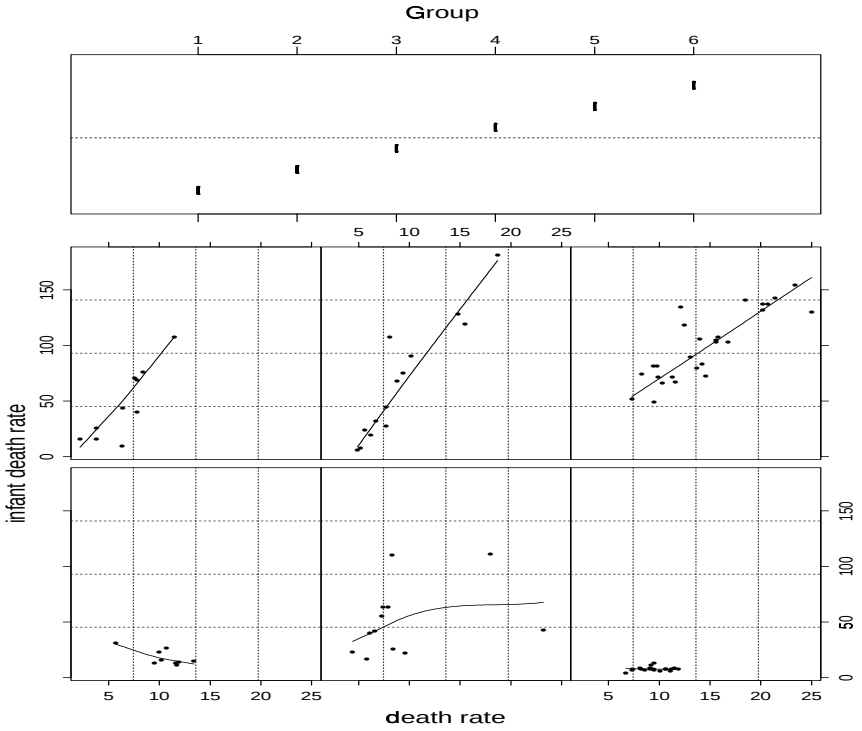

Figure 2: Plots of infant death rate versus death rate for different regions.

A Graphical Analysis of the U.N.E.S.C.O. Data

The birth rates, death rates, and infant death rates for 91 countries were obtained from the *U.N.E.S.C.O. 1990 Demographic Year Book* and *The Annual Registrar 1992*. The birth rates and death rates are recorded as number per 1000 of population and the infant death rate is number per 1000 of population less than one year old. Infant death rates vary from 4.5 (Japan) to 181.6 (Afghanistan) with a median of 43. Figure 1 gives comparative boxplots of infant death rates for different regions of the world. Clearly, infant death rates are lowest in western Europe and North America (median = 7.5) and highest in Africa (median = 103).

It would be expected that the infant death rate for a country would be strongly correlated to the overall death rate. Interestingly, this is true for some regions of the world but not others. Figure 2 contains plots of infant death rate versus birth rate for the following six regions (the scale at the top indicates the group number):

1. Eastern Europe
2. South America and Mexico
3. Western Europe, North America, Japan, and Australia
4. Middle East
5. Asia
6. Africa

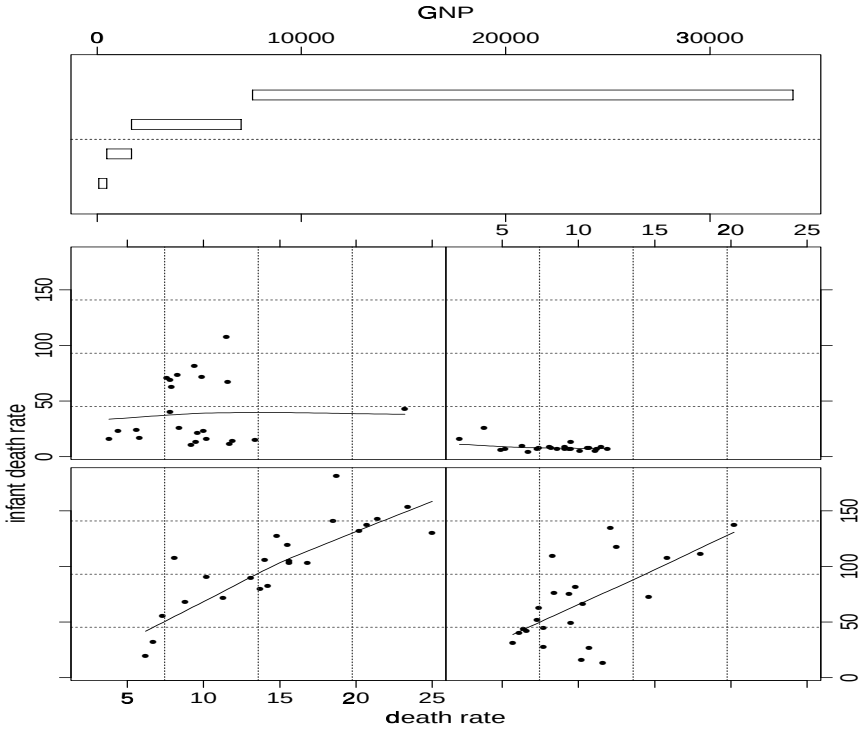

Figure 3: Plots of infant death rate versus death rate for different levels of GNP.

There is a strong positive relationship for South America, the Middle East, Asia, and Africa. There is evidence of a weak positive relationship for Eastern European countries and no evidence of any relationship for Western European and North American countries. The differences between regions can be explained by the wealth of the countries. For the plots in Figure 3 the gross national product of the countries has been used to group the data. The scale across the top of the graph indicates the range of GNP included in each plot. These plots indicate that as GNP increases the relationship between infant death rate and birth rate becomes weaker. Thus for poor regions of the world there is a strong relationship between infant death rate and the overall death rate whereas for rich regions the relationship is much weaker. Thus suggests that in poor countries much the same factors determine both infant death rate and death rate but rich countries these rates are determined by different factors.

The relationship between infant death rate and birth rate follows a similar pattern. If countries are divided into groups according to GNP, see Figure 4, then there is a strong positive relationship for countries with low GNP but as GNP increases the relationship becomes weaker. Figure 5 contains plots of infant death rate versus birth rate for the six different regions. For poor regions there is a strong relationship between infant death rate and birth rate and little or no relationship for rich regions.

Statistical Appendix

I decided to concentrate on the infant death rates and how they were related to the other variables. My report includes the graphs that I found most useful.

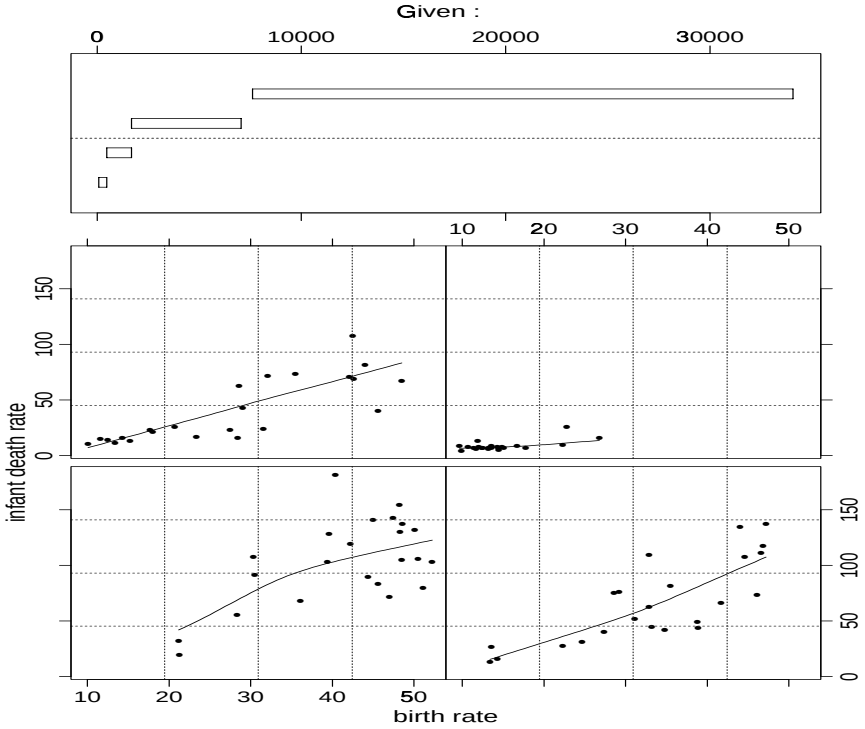

Figure 4: Plots of infant death rate versus birth rate for different levels of GNP.

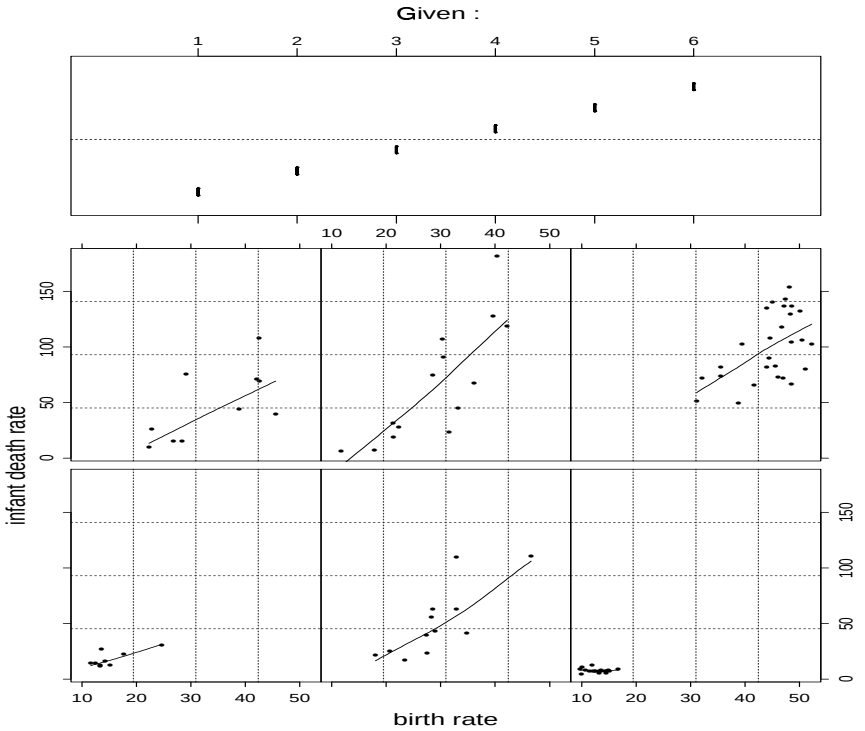

Figure 5: Plots of infant death rate versus birth rate for different regions.

Part 2: Life Expectancies

Executive Summary

There is a strong linear relationship between the life expectancies of women and men. In general, the life expectancy of women is several years longer than for men in most countries. The size of this difference increases as the life expectancy of men increases. Thus for countries that have male life expectancies of 40 years, the average life expectancy of women will be from 1 to 3 years longer whereas countries that have male life expectancies of 70 years will have an average female life expectancy that is between 5.1 and 6.3 years longer.

A Regression Analysis of Life Expectancies

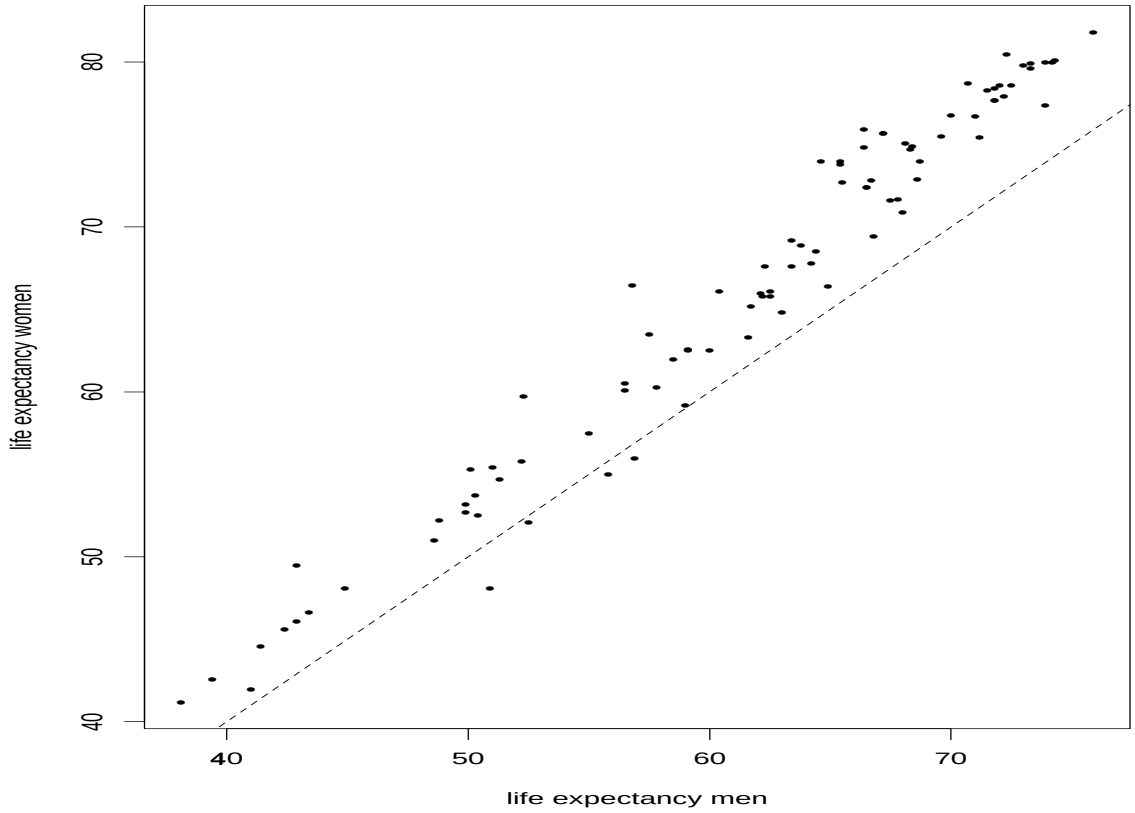

Figure 6: life expectancy women versus life expectancy men

Figure 6 contains a scatterplot of the life expectancies for women versus those for men for the 91 countries in the dataset. There is a strong linear relationship which suggests that a simple regression model would be appropriate. The dashed line on the plot indicates points where the life expectancy of woman and men are the same. Thus we can see for most countries the life expectancy for women is higher than that for men.

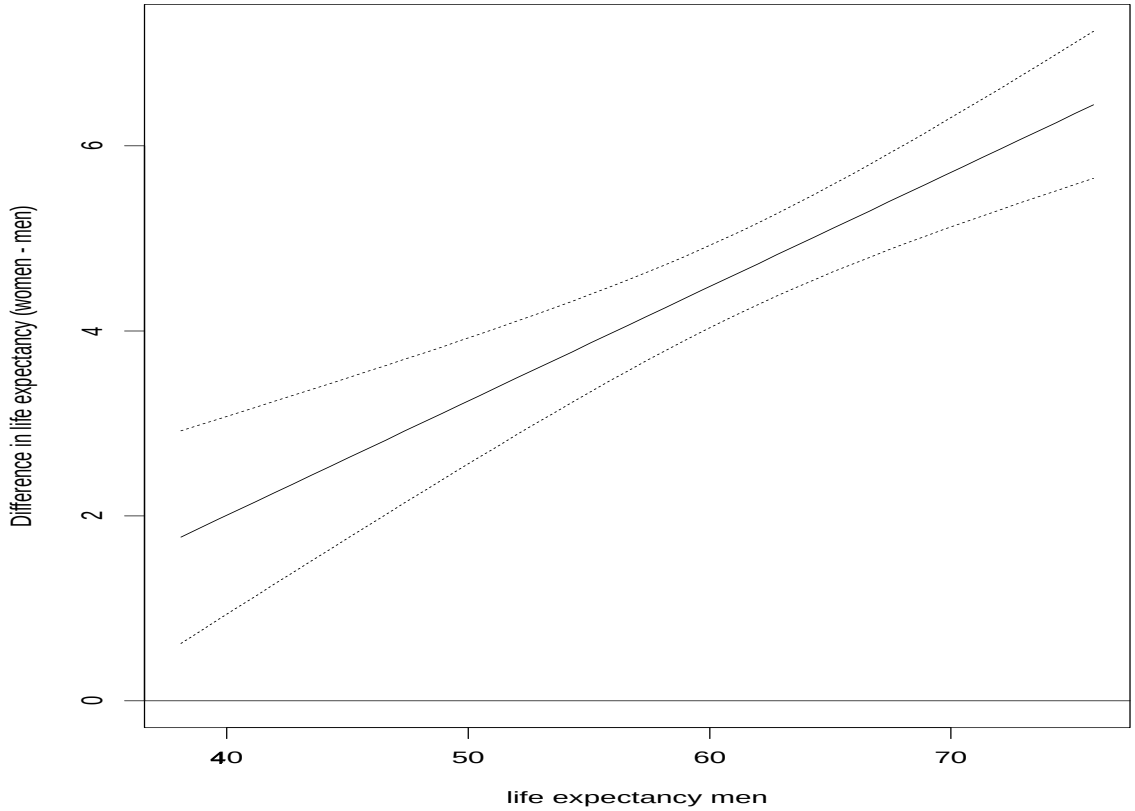

Figure 7: life expectancy women versus life expectancy men

To further investigate the difference in life expectancy for men and women the data was used to estimate the following regression model that relates the life expectancy of women to that for men.

$$\text{life expectancy women} = -2.94 + 1.12 \times (\text{life expectancy men})$$

This model can be used to predict the life expectancy of women for a country given the life expectancy of men for that country. The difference in life expectancies can be estimated by subtracting the life expectancy for men from the predicted life expectancy for women. Figure 7 contains a plot of the estimated difference in life expectancies versus the life expectancy for men. The solid line represents the estimated difference and the dotted lines indicate the range of plausible values for this difference. It is clear that the difference in life expectancy increases as the life expectancy of men increases. For countries that have a life expectancy for men of 40 years then we expect that the life expectancy for women will be, on average, between 1 and 3 years longer whereas for countries that have life expectancies for men of 70 years the average for women will be between 5.1 and 6.3 years longer.

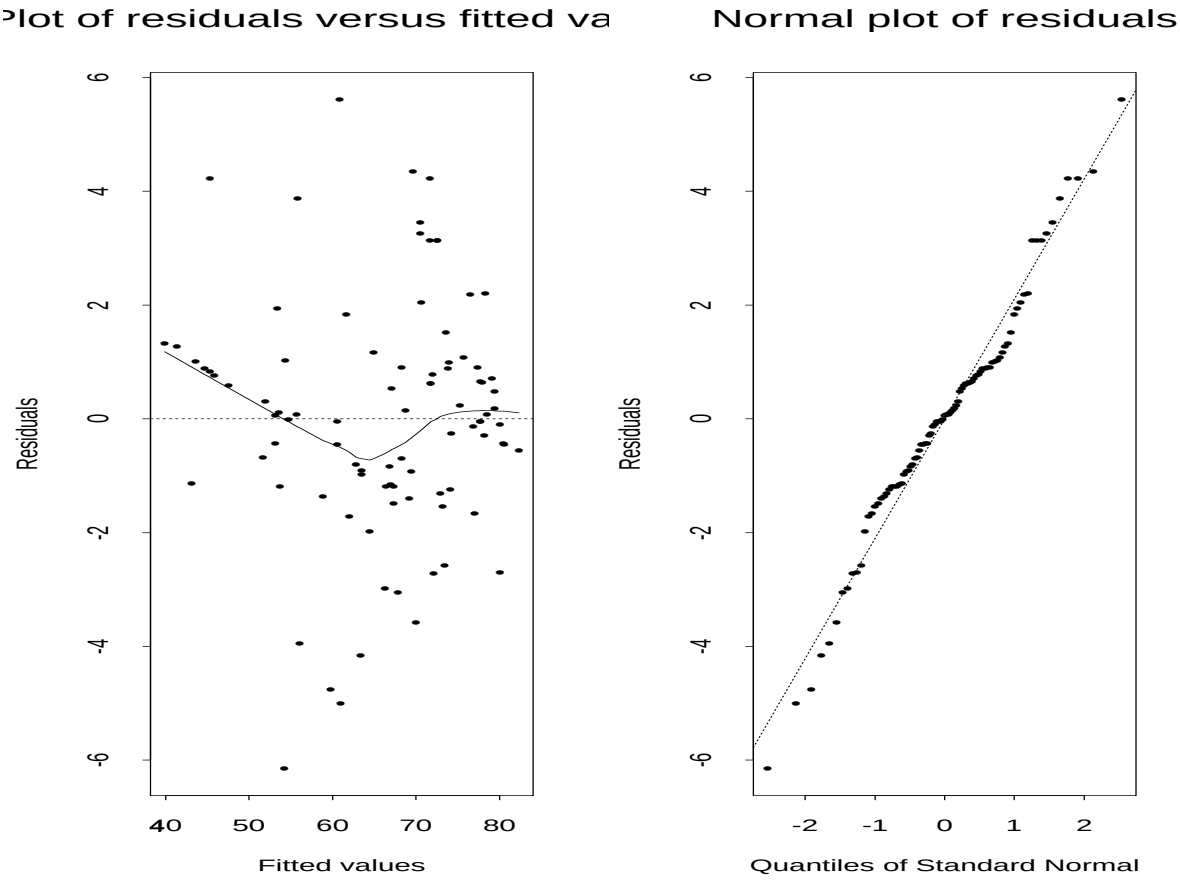

Figure 8: Residual plots for the fitted regression model

Statistical Appendix

To investigated the difference in life expectancies I used my regression model to predict the life expectancy of women for 50 life expectancies of men from 38.1 to 75.9 (these were the range in the data set). A 95% confidence interval was calculated for the mean life expectancy of women at each point and the life expectancy of men subtracted to get an interval for the difference. These were joined to produce the bands in Figure 7.

Figure 8 contains a plot of residuals versus fitted values and a Normal plot of residuals. There is a slight indication of non-linearity in the first plot but it is not enough to cause real concern. There is no indication of non-Normality in the Normal plot. Figures 9 and 10 contain trellis plots of life.f versus life.m conditioned on group (Figure 9) and GNP (Figure 10). These plots indicate that the relationship between life.f and life.m is very similar for different groups and for different ranges of GNP. Thus fitting a single model to the entire dataset is reasonable.

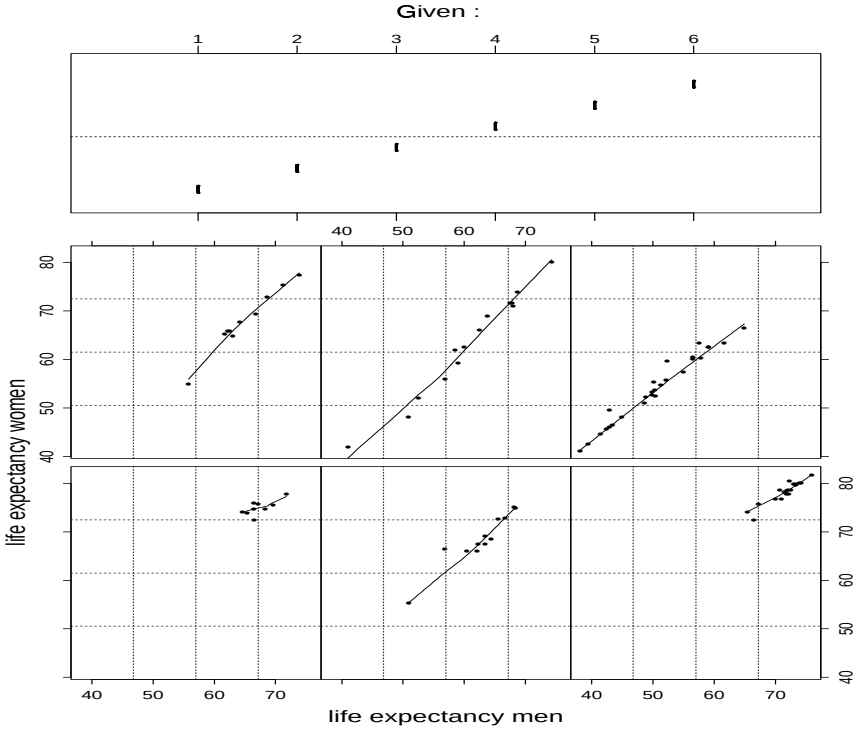

Figure 9: Plots of the life expectancy of women versus men for different regions

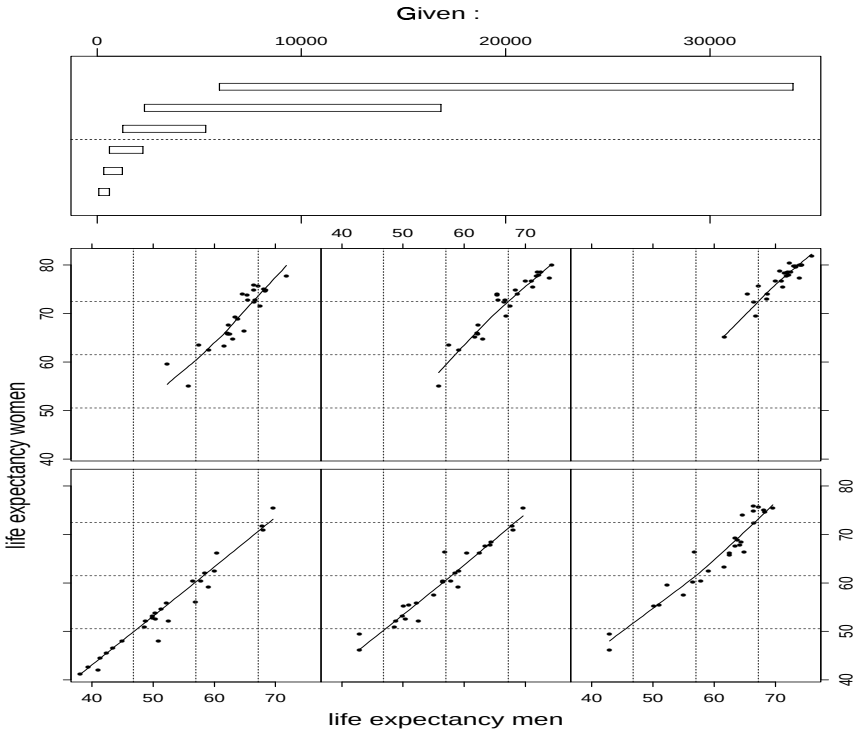

Figure 10: Plots of the life expectancy of women versus men for different levels of GNP