Assignment for the third week

Riccardo Omenti

2023-06-30

Affiliation: Department of Statistical Sciences, University of Bologna

E-mail: riccardo.omenti2@unibo.it

Exercise 1

Consider the following table with the world population growth rates in the following years 1986, 1990, 1994, 1998, 2002 and 2011. These rates originate from the book "Essential demographic methods" by Wachter (2014).

Year	Growth Rate
1986	0.0173050
1990	0.0157940
1994	0.0142830
1998	0.0128710
2002	0.0115567
2011	0.0111520

(a) Type the data in R using the data.frame() function and generate an appropriate scatter plot with Year on the x-axis and Growth Rate on the y-axis. Please assign to the names of the axes appropriate labels.

(b) Run a simple linear regression model considering growth rate as response and calendar year as independent variable. Show the model results with an appropriate R function.

(c) Insert the fitted regression line on the plot that you have generated in point (a).

(d) Determine the differences between the fitted growth rates and the observed growth rates? (Hint: apply the function fitted() to your model)

Exercise 2

Suppose we are interested in describing the population composition of three distinct countries: Japan, Niger and India.

You are provided with a data set containing the following variables:

- *Year*
- *Country*
- *old*: Population aged 65+
- *young*: Population aged 0-14
- *working*: Population aged 15-64

These data are retrieved from the United Nations.

The data are in format *.RData*. Please use the function *load()* to load the data set in your RStudio.

The name of the file to be laoded is "Data_DepRatios.RData".

(a) For each country, calculate the dependency ratio, the young dependency ratio and the aged dependency ratio.

(b) Generate a suitable plot to compare the dependency ratios of the three countries. Please include a suitable legend as well as appropriate labels for the axes.

(c) For each country, generate a suitable plot to compare the aged dependency and young dependency ratios. Please include a suitable legend as well as appropriate labels for the axes.

Exercise 3 (Optional)

In 2023, a group of historical demographers at the University of Bologna managed to construct a rich historical data set by integrating multiple parish records as well as data coming from the rich archive of the Municipality of Bologna.

The data contain the fertility histories of a sample of 200 women that were born in the year 1870 either in Bologna or in the Province of Bologna. All the women in the sample had at least two children, survived at least until the age of 49 and did not experience any pregnancy before age 15. Half of this women lived their entire lives in the Province of Bologna; hence in a rural setting, while the other half lived in the city of Bologna. The data also contain the date of birth of each child.

In particular, the researchers decided to investigate whether women from a rural background had a higher probability of having a third child compared to women who lived in the city.

The data set to be analyzed for this exercise contains the following information:

- *initial_date*: date at baseline (January 1st 1885 for all women)
- *date*: date at the birth of the third child
- *third_birth*: binary variable indicating whether the woman had a third child or not during the historical period 1885-1919
- *background* : background ("Rural" vs. "Urban")

Please, note that if a woman did not give birth to a third child, the date at the birth of the third child is set to December 31st 1919, which coincides with the end of the reproductive period for the cohort of women. This implies that the woman did not give birth to a third child by the end of the study.

The data are in format *.RData*. Please use the function *load()* to load the data set in your RStudio.

The name of the file to be loaded is "Births_Bologna.RData".

The data set is a work of fiction and was simulated.

(a) For each woman experiencing a third birth, calculate the number of days, weeks and years from the beginning of the study to the birth of the third child? (Hint: use the functions in the *lubridate* package)

(b) For each woman having at least a third child, extract the year in which the third birth took place. Also provide the total number of third births by background.

(c) Run a simple logistic regression model with having a third child as response and background as independent variable. Is the propensity to have a third child lower for a mother living in the city than for a mother living in a rural village in the province of Bologna?