Lesson 10

Data visualisation with R

Alessandro Bramucci

27-06-2023

Contents

1 The most populous countries in 2007

Table 1 reports the three most populous countries in 2007 broken down by continent.

2 Life expectancy and per capita GDP

From Figure 1 we can observe the existence of a positive relationship between per capita GDP growth and life expectancy.

2.1 Regression analysis

This document is written using the package *bookdown* (Xie 2023). Data are obtained from the *Gapminder* dataset (Bryan 2023).

2.1.1 Per capita GDP growth and life expectancy

2.1.2 Inclusion of dummy variables

In econometrics, a dummy variable is a binary variable used to represent categorical or qualitative variables in regression analysis and to capture the effects of different categories on the dependent variable. Dummy variables are created by assigning a value of 1 to one category or group, while all other categories or groups

Table 1: The three most populous countries by continent in 2007. Source: Gapminder.

Country	Population
Africa	
Nigeria	135'031'164
Egypt	80'264'543
Ethiopia	76'511'887
Americas	
United States	301'139'947
Brazil	190'010'647
Mexico	108'700'891
Asia	
China	1'318'683'096
India	1'110'396'331
Indonesia	223'547'000
Europe	
Germany	82'400'996
Turkey	71'158'647
France	61'083'916
Oceania	
Australia	20'434'176
New Zealand	4'115'771

Table 2: Relationship between per capita GDP growth and life expectancy, 1952-2007.

	Dependent variable:
	dlifeExp
pcGDPgr	0.012***
	(0.002)
Constant	1.490***
	(0.053)
Observations	1,562
R^2	0.017
Note:	*p<0.1; **p<0.05; ***p<0.01

are assigned a value of 0. Since we have a variable called "continent" with five categories, i.e. the five continents, we can create a dummy variable called "Europe" and assign a value of 1 to countries in Europe and 0 to countries in other continents.

By including dummy variables in regression models, we can estimate the effect of different categories on the outcome variable.[1] When we want to include dummy variables in a regression, we must avoid the so-called dummy variables trap. What is the dummy variables trap? The dummy variable trap arises when we include dummy variables for all categories of a categorical variable in a regression model, for example all continents in our case. Including dummy variables for all categories creates a linear dependency among the variables, causing perfect multicollinearity.

[1] There are several ways in R to create dummy variables.

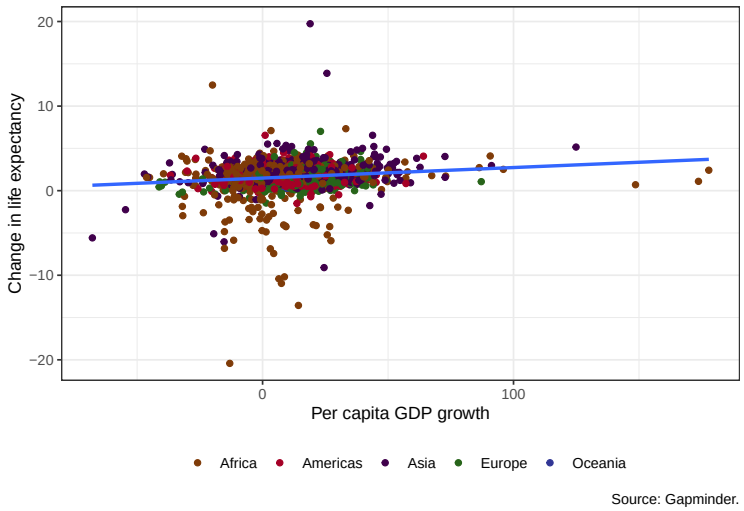

Figure 1: Relationship between per capita GDP growth and change in life expectancy, 1952-2007.

To understand why this happens, consider a categorical variable with k categories like our five continents, for example. If we create *k-1* dummy variables, where one category is excluded as the reference, the information about the excluded category is captured in the constant term of the regression equation. We must then interpret the coefficient of the dummy variable with respect to the constant term, i.e. the benchmark category. In Table 3 the benchmark category is set to "Europe".[2] If different dummy variable sets are included, such as for example for continents but also whether the country is located north or south of the hemisphere, the interpretation of the constant as a reference category will no longer be possible.

[2]Here we have used the function *relevel(continent, ref = "Europe")*.

Table 3: Relationship between per capita GDP growth and life expectancy including continent dummy variables, 1952-2007.

	Dependent variable:	
	dlifeExp	lifeExp
	(1)	(2)
pcGDPgr	0.012^{***}	0.016
	(0.002)	(0.012)
Americas	0.318^{***}	-22.699^{***}
	(0.123)	(0.615)
Asia	0.718^{***}	-6.788^{***}
	(0.144)	(0.720)
Africa	1.015^{***}	-11.272^{***}
	(0.133)	(0.667)
Oceania	-0.097	2.292
	(0.387)	(1.932)
Constant	1.021^{***}	72.329^{***}
	(0.104)	(0.519)
Observations	1,562	1,562
R^2	0.059	0.514
Note:	*p<0.1; **p<0.05; ***p<0.01	

References

Bryan, Jennifer. 2023. *Gapminder: Data from Gapminder*.

Xie, Yihui. 2023. *Bookdown: Authoring Books and Technical Documents with r Markdown*. https://github.com/rstudio/bookdown.