Probability, Statistics and Modelling II

Tutorial 2 – Linear regression

We will continue with the 'Policing the pandemic' dataset from last week.

Constructs	*Variable*	*Item wording*	*Response options*
Duration it took for someone to finish the survey	duration	N/A	seconds
Confidence in the handling of the COVID-19 crisis	pm	Prime Minister	No confidence at all – A lot of confidence
	nhs	National Health Service	
	pol	Police	
	js	Justice System	
	gov	Government	
Coronavirus status	cov	Have you had Covid-19 (coronavirus)?	Yes, diagnosed and recovered ... Prefer not to say
Coronavirus attitudes	covknow	How would you rate your knowledge level on Covid-19?	Bad-Excellent
	covconc	How concerned are you about getting Covid-19?	Not concerned at all – Very concerned
	covidexp	How long do you expect it will be until the coronavirus outbreak is over and things are back to normal in the UK?	Less than 1 month - Never
Gender	gender	What is your gender?	Male/Female/Non-binary
	male	Binary variables created from gender	Male/Not
	female		Female/Not
	nonbin		Non-binary/Not
Age	age	Which of these age bands do you fall into?	16-24 – 65+
	age1	Binary variables created from age	16-24/Not
	age2		25-44/Not
	age3		45-64/Not
	age4		65+/Not
Area	area	Which city's metropolitan area do you live in?	Birmingham ... None of these
Ethnicity	ethnic	Please select the option which best describes your ethnic group:	Recoded to: Asian ... White
	asian	Binary variables created from ethnic	Asian/Not
	black		Black/Not
	mixed		Mixed/Not
	ethnico		Other ethnicity/Not
	white		White/Not
Key worker	keywork	Are you currently fulfilling any of the government's identified 'key worker' roles (listed below)?	Recoded to: Key worker/not

Table 1 Variables in the dataset

Please carry out the tasks and answer the questions below.

1.Carry out all the preliminary steps: set your working directory, load in the packages and the data, and attach the data so you would not need to identify the dataset for the subsequent commands. If this is your first seminar, please install the packages that we used last week (they are all mentioned in the script).

2.Today, we will have two outcome variables: concern about catching Covid-19 and expectations regarding the length of the pandemic. Look at the descriptive statistics. What are your thoughts looking back at the results three and a half years after the start of the pandemic?

3.Creating crosstabs and running correlation analysis, consider the association between concern about catching Covid-19 and expectations regarding the length of the pandemic. What is your takeaway?

4.Let's consider the association between the two variables, now with linear regression analysis, using them as explanatory and outcome variables one at a time. Please answer each of the questions below:

a) What is the association between these two variables?
b) To what extent are these models different/similar?
c) What is the relationship between the R^2 statistics and the correlation coefficient?
d) How would you interpret the results?

5.Finally, plot the two simple linear regressions discussed today. How do the figures correspond to the outputs?