Probability, Statistics and Modelling II

Tutorial 4 –Multiple linear regression, standardised estimates, moderation, model fit

We will continue with the 'Policing the pandemic' dataset.

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.We begin the tutorial by revisiting the models from last week. Compare these models to each other. Which (partial) associations are significant, and which are not? How would you interpret these coefficients? How do these models change from model 1 to model 5? What do these changes indicate? If you were helping the government's behavioural science unit, what would be your advice?

2.Now the behavioural science unit wants to understand which variables have the strongest association with the outcome variable (relative to each other). To answer this, estimate the standardised coefficients first, by revisiting the two different solutions for model 1 (from tutorial 2), and then by deriving the standardised coefficients for model 5. How would you interpret the results? What would you tell the behavioural science unit based on these results?

3.The behavioural science unit believes that women key workers were more likely to expect that the pandemic would last longer. Similarly, they also hypothesised that the association between concern about contracting the virus and the increased expectations regarding the length of the pandemic depended on one's gender identity. Test these two hypotheses using moderation analysis. How would you interpret the results? What do the figures tell you?

4.The behavioural science unit asks you to consider the fit of each of the models. Estimate all of the relevant model fit criteria. Purely based on these, which model is preferable and why? Why do simpler (i.e. more parsimonious) models have better fit?

5.To what extent should you make a decision based on model fit statistics? What can be the advantages and the disadvantages of maximising model fit? From the seven models fitted, which one would you pick and why?

6.Let's consider a few alternative models with the variables used in the first week: confidence in the various institutions' handling of the pandemic. Using the confidence in the prime minister's handling of the pandemic as the outcome variable and the confidence in the NHS, the government, and the police as explanatory variables, how do each of the models fare? Which model has the best fit? Which model would you consider the best? Why?