DSR: Exercise 11 (graded)

This exercise focuses on fitting and interpreting *multiple* linear regression models.

Make sure to read extensively on the topic prior to completing the exercise, especially if you have not worked with linear models for a while.

This is a *graded* exercise, to be completed within your group, and within your group *only* — do not communicate between groups. All other school regulations regarding student ethics and plagiarism obviously apply.

All answers are to be provided directly in the 'answers' script. The code should produce the results reported in the 'Answers' section at the top of the script.

Some further important coding requirements:

- It must run (execute) properly.
- It should be **as simple and legible** as possible.
- Consequently, it should use **as few packages** as possible.
- Concision (coding/writing that is **short** and **clear**) is *very much* valued.
- Comment lines should *not* go over 80 characters per line.
- Include the code to load packages, but *not* the code to install them.
- There will be grades assigned to all requirements above.

Scenario

You are working for a think tank interested in testing the contemporary relevance of the results published in Tim Immerzeel and Frank van Tubergen, "Religion as Reassurance? Testing the Insecurity Theory in 26 European Countries", *European Sociological Review* 29(2): 359–372, 2013.

Instructions

1. Access a dataset

Download a copy of the European Social Survey (ESS), Round 9, Edition 3.2, in SPSS format (`.sav`), and place it in the `data` folder.

Import the dataset in R, writing your code in Section 1 of the answers script.

2. Recode some variables

Find the variables below in the ESS documentation, and recode them as instructed below, writing your code in Section 2.2 of the answers script:

- **Age** should not be modified, and should thus measure age in years.
- **Sex** should be coded as a 'dummy' coding `1` for females.
- **Employment relationship** should not be modified, and should thus code `1` for employees, `2` for self-employed, and `3` for working for own family business.
- **Marital status** should be recoded into three groups, as follows:
  - legally married or in a civil union
  - separated, divorced or widowed
  - none of those (never married or else)
- **Subjective income** ("Feeling about household's income nowadays") should not be modified, and should thus contain 4 ordered response items.

Last, recode **religious attendance** (variable `rlgatnd`) in order for the variable to measure religious attendance from `0` for 'never' to `6` for 'every day.'

Answer Questions 2.1, 2.2 and 2.3 at the end of that section, rounding all numbers at one-digit precision. The answers should go at the top of the script, in the header section, on the lines starting with `[ANSWER 2.1]`, `[ANSWER 2.2]` and `[ANSWER 2.3]`. The same logic applies to all questions below.

3. Write a linear model

Using the variables created in the previous section, predict religious attendance from age, sex, employment relationship, marital status and subjective income, for the entire ESS data sample.

In that model, include an interaction between sex and marital status, in order for your model to include coefficients for those predictors *as well as for their product*.

Answer Question 3 at the end of that section. This question is a simple check of your model formula.

4. Interpret regression results

Spend some time understanding the results of your model.

Answer Questions 4.1, 4.2 and 4.3 at the end of that section.

5. Diagnose a linear regression model

Write up the combination of regression diagnostics that you think will provide the best assessment of whether the model performed well.

Answer Question 5 at the end of that section.

6. Start thinking beyond 'flat' models

Re-estimate the same model as before, but include the country of residence of the respondent.

Answer Questions 6.1 and 6.2 at the end of that section. Note that Question 6.2 is a hard question, which you should skip if you find it too difficult.

Submission instructions

Please email me your group's answers as a single R script called `exercise-11-Group-01.r`, where `01` is your group number, as stated in the groups spreadsheet on Google Drive.

Please send me that email **along with your answer to Exercise 12, before our last class together**.

You do *not* need to also attach the data and/or the rest of the `exercise-11` folder – the answers will be enough.

Use the email subject **"DSR: Answers to Exercises 11 and 12, Group 01"**, where `01` once again designates your group number. That information should also appear at the top of your script, in the code header.

Make sure to complete the rest of the code header with your feedback.

Thanks a lot for your work!