Online Academic Data Analysis Bootcamp Using Open-Access Program R: Essentials. Session 6, Correlations

Patrick Njage

March 24, 2022

There is some evidence that students tend to pick courses of lecturers they perceive to be enthusastic and good communicators. In a fascinating study, Tomas Chamorro-Premuzic and his colleagues (Chamorro-Premuzic, Furnham, Christopher, Garwood, & Martin, 2008) tested the hypothesis that students tend to like lecturers who are like themselves. The authors measured students’ own personalities using a very well-established measure (the NEO-FFI) which measures five fundamental personality traits: neuroticism, extroversion, openness to experience, agreeableness and conscientiousness. Students also completed a questionnaire in which they were given descriptions (e.g., ‘warm: friendly, warm, sociable, cheerful, affectionate, outgoing’) and asked to rate how much they wanted to see this in a lecturer from -5 (I don’t want this characteristic at all) through 0 (the characteristic is not important) to +5 (I really want this characteristic in my lecturer). The characteristics were the same as those measured by the NEO-FFI. As such, the authors had a measure of how much a student had each of the five core personality characteristics, but also a measure of how much they wanted to see those same characteristics in their lecturer. These are the data from that study. The data contains the following variables:

* age: participant age (years)
* sex: participant’s biological sex
* stu\_neurotic: Student neuroticism score on the NEO-FFI
* stu\_extro: Student extroversion score on the NEO-FFI
* stu\_open: Student openness to experience score on the NEO-FFI
* stu\_agree: Student agreeableness score on the NEO-FFI
* stu\_consc: Student conscientiousness score on the NEO-FFI
* lec\_neurotic: Student rating of how much they wanted the characteristic of neuroticism in their lecturers from -5 (I don’t want this characteristic at all) through 0 (the characteristic is not important) to +5 (I really want this characteristic in my lecturer)
* lec\_extro: Student rating of how much they wanted the characteristic of extroversion in their lecturers from -5 (I don’t want this characteristic at all) through 0 (the characteristic is not important) to +5 (I really want this characteristic in my lecturer)
* lec\_open: Student rating of how much they wanted the characteristic of openness to experience in their lecturers from -5 (I don’t want this characteristic at all) through 0 (the characteristic is not important) to +5 (I really want this characteristic in my lecturer)
* lec\_agree: Student rating of how much they wanted the characteristic of agreeableness in their lecturers from -5 (I don’t want this characteristic at all) through 0 (the characteristic is not important) to +5 (I really want this characteristic in my lecturer)
* lec\_consc: Student rating of how much they wanted the characteristic of conscientiousness in their lecturers from -5 (I don’t want this characteristic at all) through 0 (the characteristic is not important) to +5 (I really want this characteristic in my lecturer)

1. Load the dataset “chamorro\_premuzic.csv” into R as “personalityData”
2. Create a dataframe called “personalityData2” consisting of the following variables “stu\_neurotic”, “stu\_extro”, “stu\_open”, “stu\_agree”, “stu\_consc”, “lec\_neurotic”, “lec\_extro”, “lec\_open”, “lec\_agree”, “lec\_consc”. The dataset contains missing values. Select complete cases listwise by running the code personalityData2 <- personalityData2[complete.cases(personalityData2), ]

c ) Create a scatter plot.

* Add a linear modeæ regression line, a title plus x and y labels

1. Run the correlation analysis using personalityData2.

* Mention and interprete a few significant correlations, report and interprete them and their confidence intervals.

1. Do a Shapiro-Wilks test to see whether the data are normal and apply alternative correlation approaches where deviation from normality is indicated.
2. Conduct two point-biserial correlations featuring gender versus two of the continuous variables in personalityData
3. Conduct a partial correlation between “stu\_neurotic” and “stu\_extro” while controlling for the effect of “stu\_agree”, “stu\_consc” and “lec\_neurotic”. How does this correlation compare with the value when the effect of “stu\_agree”, “stu\_consc” and “lec\_neurotic” are not controlled for.