Lab 6: Correlations and Regressions in Personality Data

Introduction

In this lab, we are going to look at personality data and how it relates to conformity. This will use correlation and regression. Much of the code in this lab is based on the previous lab, so use that lab as a basis to answer the questions. I will not give you explanations for the code to type in many cases, so use that lab to figure out what you should type.

We will use some data that some students collected for a research project previously. In this project, the researchers gave students a series of questionnaires looking how personality traits affected a participant's likelihood to conform.

In this case, the variables are named as follows:

- Participant: a participant number
- E: extraversion score
- A: agreeableness score
- C: conscientiousness score
- N: Neuroticism score
- O: Openness score
- Gender: gender coded F and M
- Age: participant age
- Conformity: their score on a survey about how much they would conform
- RATING: a participant's rating to the question "how much do I think I conform to others' behaviors"
- SelfEsteem: a score on a self-esteem inventory with higher scores indicating more self-esteem

One last note about data sharing. In this case, the individuals who shared this data did not use a consistent capitalization scheme. I left the data as is in order to illustrate why this is not a good idea. You should develop a consistent way you name your variables when you enter data. Here are my suggestions:

- Never use capital letters, unless it's a normally capitalized abbreviation, like "IQ". Or you could just capitalize the first letter all the time.
- Never, ever use spaces. I use underscores to divide spaces. Some others will bunch all the words together, capitalizing the first letter of each word. Either way is fine, but be consistent.

Now let's get started:

Step 1: Create a folder on your computer and download the "lab6.csv" file into that folder. Create an R notebook and save it in that folder.

Step 2: Create a code chunk to load the tidyverse library. Create a second code chunk and import your data using the `read.csv()` command as the dataframe named `d`.

Step 3: Use the ggplot() command to create histograms of each of the personality variables. Please note in your annotations whether you see any outliers and whether these look like normal distributions.

Step 4: Use the summary command to output the means and ranges of the personality variable.

Correlation between personality variables

One of the theories about the five factor theory of personality is that the variables are largely independent: that is, they are not correlated. If they were correlated, they may be measuring the same underlying construct. In practice, there is some slight correlation between some variables, and some correlation can happen by random chance or in a non-random sample (like college students).

What we are going to do is look at the correlation between each of the personality variables.

Step 5: Use `cor.test()` to examine the correlation of all the personality variables. This will be 10 possible correlations. Write each of these in APA format. Are there any that are significant?

Here's an example.

```
cor.test(d$E, d$A)
```

```
##
##  Pearson's product-moment correlation
##
## data:  d$E and d$A
## t = 1.2305, df = 58, p-value = 0.2235
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.09840162  0.39733983
## sample estimates:
##       cor
## 0.1595087
```

In this case, I would write $r(58) = .16$, $p = .22$.

Question 1: Is there a problem with running 10 correlations, like this? (Hint: there is). What is the problem? What might we do about that problem?

Regression with R

The researchers were interested in whether there was a relationship between the personality variables and conformity. In this case, we will use the `lm()` command to test whether each personality variable predicts conformity.

For example, if we wanted to test whether Extraversion predicts Conformity, we would type:

```
x = lm(Conformity~E, data = d)
summary(x)
```

```
##
## Call:
## lm(formula = Conformity ~ E, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48888 -0.11615  0.01112  0.14974  0.43605
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.357231   0.097139   3.678 0.000517 ***
## E           0.004540   0.003648   1.245 0.218310
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2006 on 58 degrees of freedom
## Multiple R-squared:  0.02601,    Adjusted R-squared:  0.009217
## F-statistic: 1.549 on 1 and 58 DF,  p-value: 0.2183
```

Step 6: Use the `lm()` command to create equations for each of the variables predicting conformity. Also note the R-squared value. Which personality variable is the best predictor of conformity? That is, which one

has the highest R-squared value?

Summary

In the last part of this lab, I want you to find a hypothesis of interest and explore it using correlation or regression. You should think of how you think one of the personality variables or conformity relates to the other variables in the dataset that we did not study, such as RATING or SelfEsteem.

Step 7: Write your prediction about the correlational or regression analysis, set up a null and alternative hypothesis, and then create a code chunk to test your prediction.

Step 8: After you run the analysis, interpret the results to see whether your prediction worked. Make sure to frame this in terms of your null hypothesis.

Step 9: Now you're all done! Knit the notebook and submit.