Formulas

Stephanie Ranft S2459825

November 4, 2019

Statistics 2
PSBE2-07

Exercises

Semi-partial and partial correlation

Description:

This fictional data set, "Exam Anxiety", provides questionnaire scores by students prior to an exam (the variables are anxiety, preparedness, and grade).

`Revise` Time spent studying for the exam (in hours).

`Exam` Performance in the exam (percentages).

`Anxiety` Anxiety prior to the exam as measured by the Exam Anxiety Questionnaire.

We wish to know which independent variable (of revise and anxiety) best describes the dependent variable (exam performance).

For convenience, Y = `exam`, X_1 = `revise`, and X_2 = `anxiety`, then $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ is the population regression equation. In order to compute the partial and semi-partial correlations, we need *other* regression equations. First, we regress X_1 and X_2 (separately) on Y, and also regress X_1 and X_2 on each other:

$$Y = \alpha_0 + \alpha_1 X_1 + e_{Y.X_1}; \qquad Y = \gamma_0 + \gamma_1 X_2 + e_{Y.X_2};$$
$$X_1 = \delta_0 + \delta_1 X_2 + e_{X_1.X_2}; \qquad X_2 = \kappa_0 + \kappa_1 X_1 + e_{X_2.X_1}.$$

Then we can calculate the partial correlation coefficients as

$$pr_1 = \operatorname{Cor}\left(e_{Y.X_2}, e_{X_1.X_2}\right) = \frac{r_{Y,X_1} - r_{Y,X_2} \cdot r_{X_1,X_2}}{\sqrt{\left(1 - r^2_{Y,X_2}\right) \cdot \left(1 - r^2_{X_1,X_2}\right)}}$$

$$pr_2 = \operatorname{Cor}\left(e_{Y.X_1}, e_{X_2.X_1}\right) = \frac{r_{Y,X_2} - r_{Y,X_1} \cdot r_{X_1,X_2}}{\sqrt{\left(1 - r^2_{Y,X_1}\right) \cdot \left(1 - r^2_{X_1,X_2}\right)}}$$

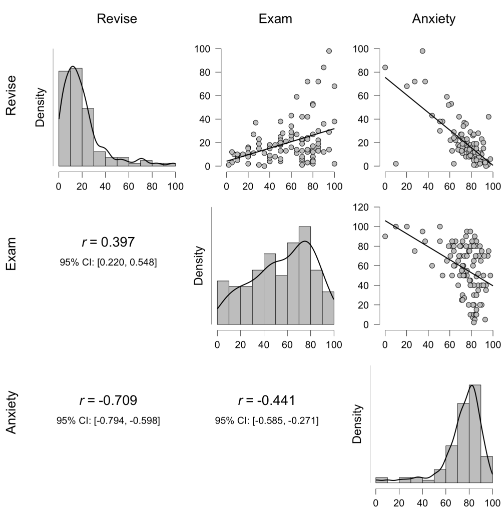

Figure 1

Using fig. 1, calculate the partial correlation coefficients for `revise` and `anxiety`. Describe what these values represent.

The semi-partial correlation coefficients are calculated as

$$sr_1 = pr_1 \cdot \sqrt{1 - r^2_{Y,X_2}} = \frac{r_{Y,X_1} - r_{Y,X_2} \cdot r_{X_1,X_2}}{\sqrt{1 - r^2_{X_1,X_2}}}; \qquad sr_2 = pr_2 \cdot \sqrt{1 - r^2_{Y,X_1}} = \frac{r_{Y,X_2} - r_{Y,X_1} \cdot r_{X_1,X_2}}{\sqrt{1 - r^2_{X_1,X_2}}}$$

Using fig. 1, calculate the semi-partial correlation coefficients for `revise` and `anxiety`. Describe what these values represent.

If we square the partial and semi-partial correlation coefficients, we can use the Ballentine Venn Diagram to compute their values.

$$pr_1^2 = \frac{R^2 - r^2_{Y,X_2}}{1 - r^2_{Y,X_2}}; \qquad pr_2^2 = \frac{R^2 - r^2_{Y,X_1}}{1 - r^2_{Y,X_1}};$$
$$sr_1^2 = pr_1^2 \cdot \left(1 - r^2_{Y,X_2}\right) = R^2 - r^2_{Y,X_2}; \qquad sr_2^2 = pr_2^2 \cdot \left(1 - r^2_{Y,X_1}\right) = R^2 - r^2_{Y,X_1}.$$

The circles represent the variances of each variable, and how they overlap. As none of our variables are uncorrelated, all circles overlap.
In words, what do the letters *a, b, c, e* represent?
Calculate the values of the squared partial and semi-partial correlations, and explain their meaning.
Use the values *a, b, c, e* to represent the squared partial and semi-partial correlations.
Which is a better predictor of the variable `exam`, `revision` or `anxiety`? Summarise your findings in a one or two sentences.

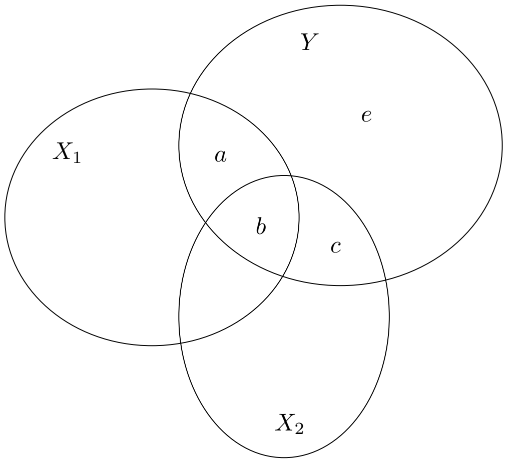

Figure 2