Chapter 20: Partial and Semi Partial Correlations

Open the Academic Pay.csv file and see if years in service is related to salary among professors. Next, analyze the data further to find out if rank has an effect on the association of these two variables. First of all, find out if the data meets assumptions for parametric tests.

file = read.csv("Academic Pay.csv")

dualData = with(file, data.frame(yrs.service, salary))

tripleData = with(file, data.frame(yrs.service, salary, rank))

You will have noticed that there are serious violations of the normality assumption, for example, using the Shapiro-Wilk test, so a non-parametric test should be used. In this example, Kendall's *tau* is used.

If, as in the book, you use the pcor function from the ppcor library, with the $estimate and "kendall" options, you should see these results for the dualData and tripleData data frames respectively:



What initially appeared to be a low correlation between years of service and salary disappears when rank is controlled for. (If you try *p* values, the partial correlation of years of service and salary has a high *p* value.) The partial correlation for salary and rank is of a moderate strength.

Here is a 'honed-down' version of the relationship of the relationship between years of service and salary, controlling for rank (at the expense of more detailed coding):

