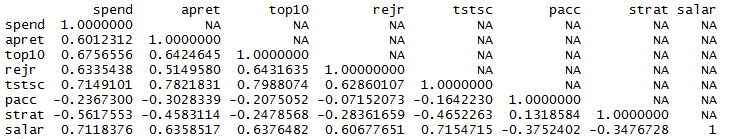
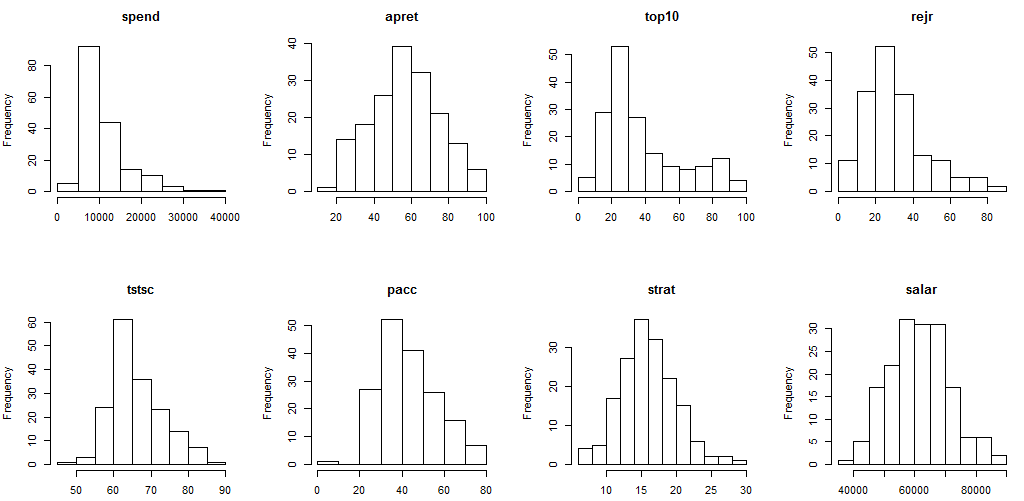
**Assignment 5**

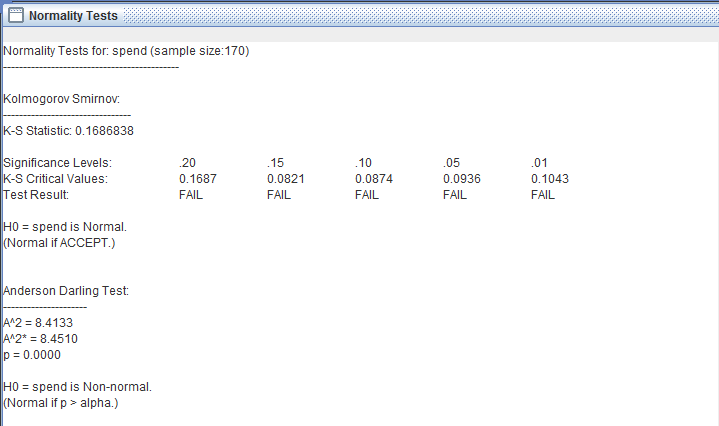
**Katherine Rodgers and John Merranko**

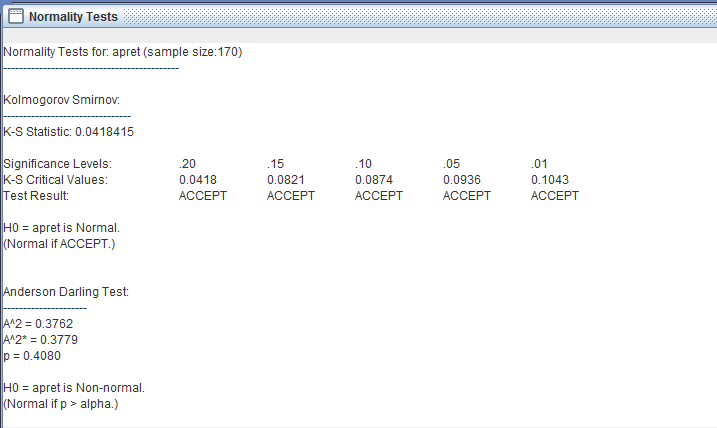
First off, even though the assignment description said we are using data from 1993, we feel after reading chapter 19 that we are actually using data from 1992. The correlation matrix below matches the 1992 data from the chapter perfectly. Also, the number of measurements (n=170) also corresponds to the 1992 data.

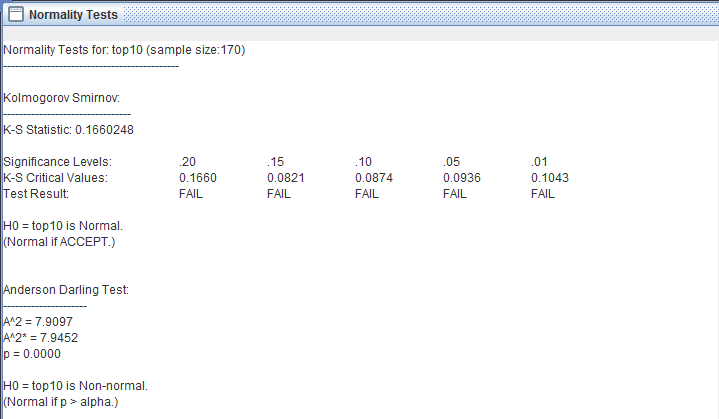


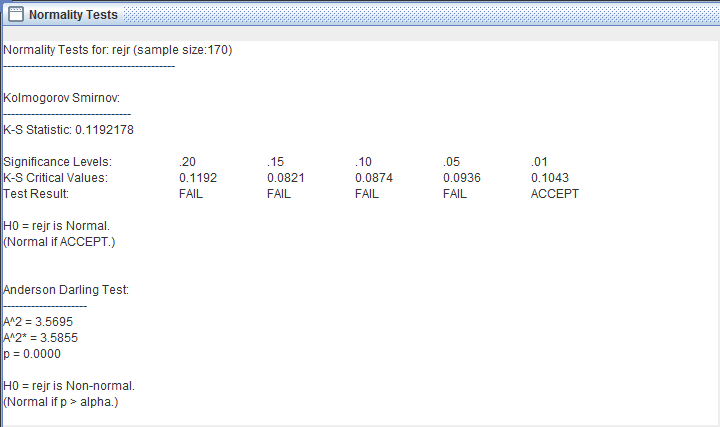
We used R to make the histograms, due to the ease of displaying them together, and used the normality tests provided in Tetrad V to look at the normality of the data. Spend, top10, and rejr are all skewed and didn’t pass the normality test. The rest of the variables passed across the board except for tstsc, which passed up to a significance level of 0.1. We also used R to look at the linearity of the data. Because of the results, it seems that the assumption of normality and linearity for all variables for the PC algorithm in the paper may be stretched a little bit.

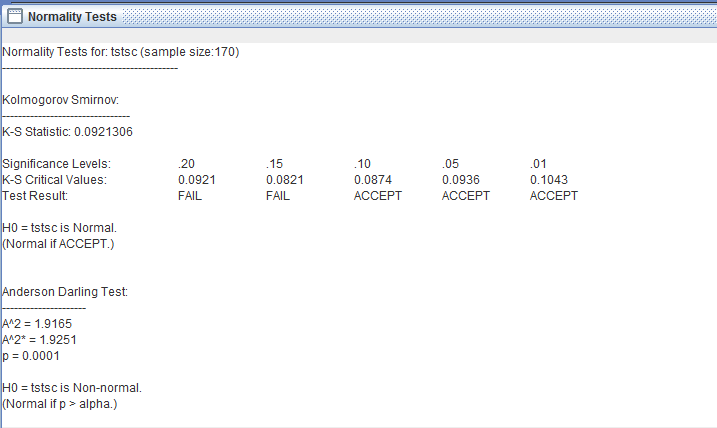


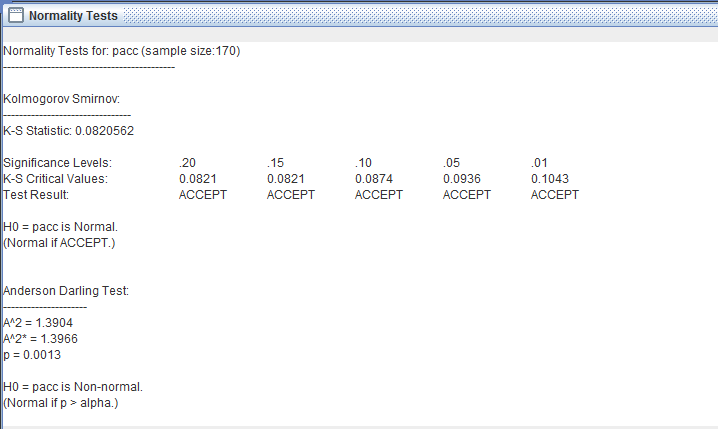


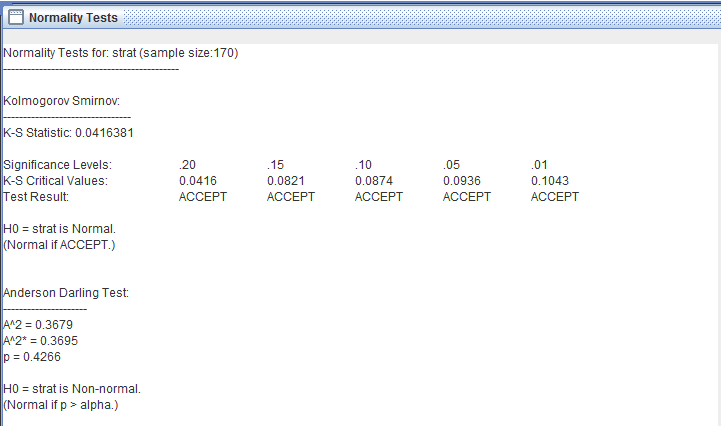


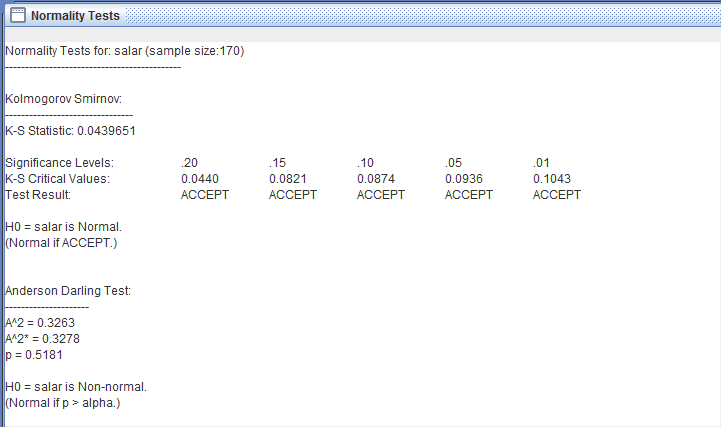


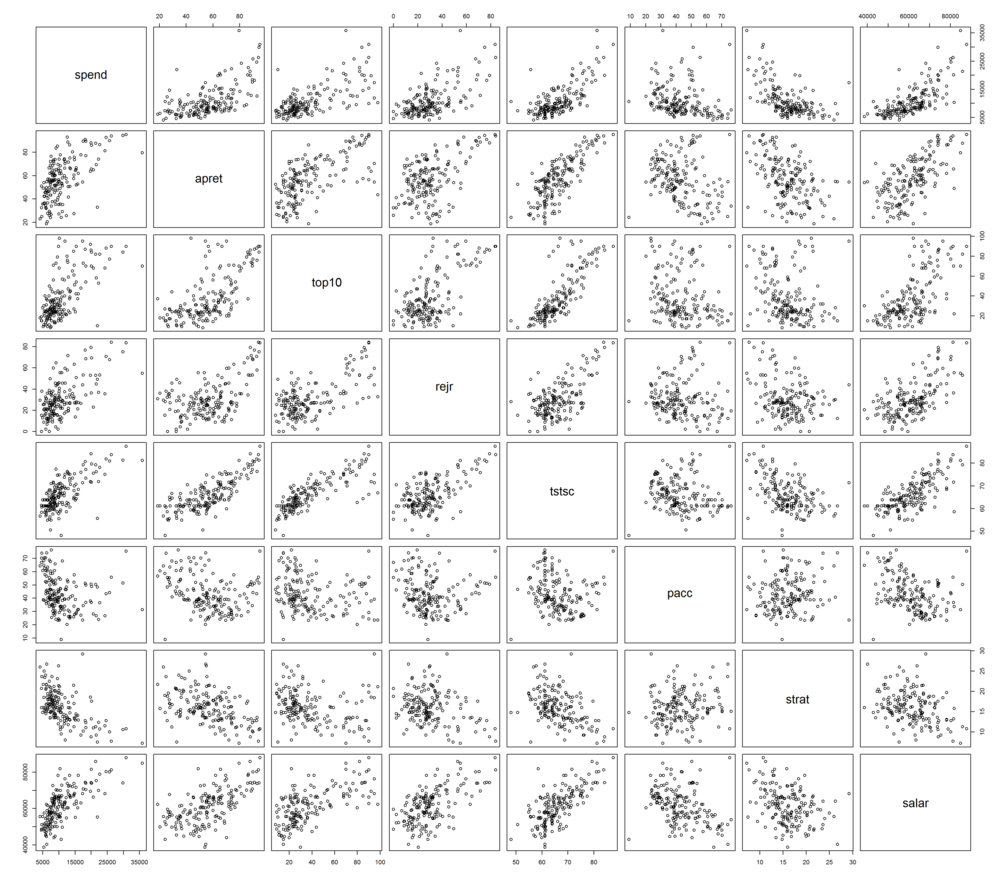




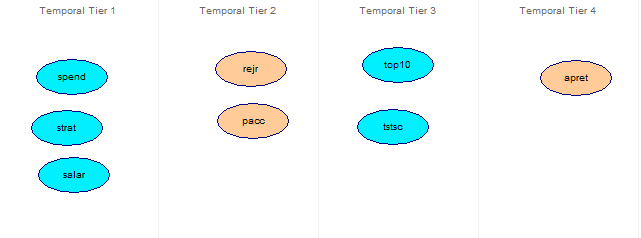




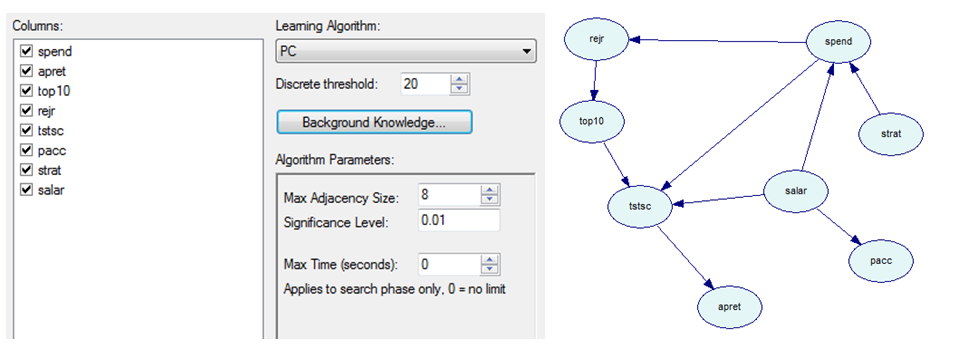


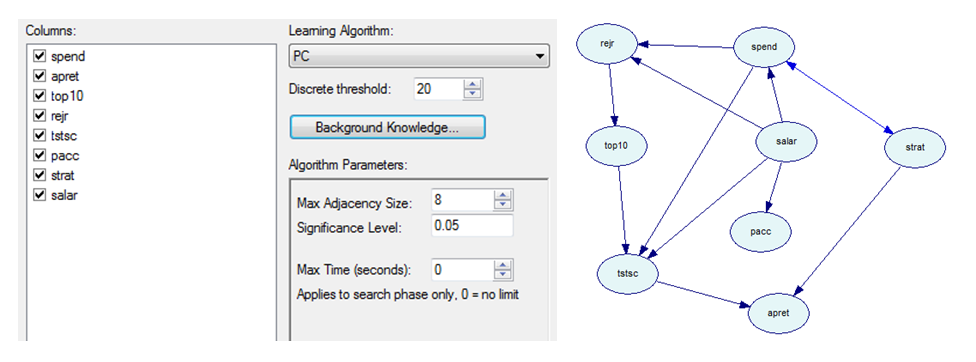


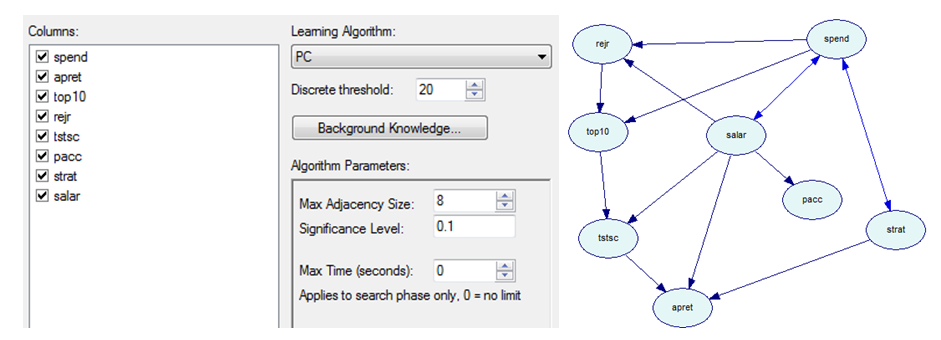
We also agreed with the author’s assumptions for temporal tier ordering as prior knowledge. Within any given year, the average spending per student, student-teacher ratio, and average faculty salary have already been determined from the budget and therefore cannot be caused by the remaining variables. Also, the average tests scores of incoming freshmen and the percentage of incoming freshmen who were among the top 10% students in their high schools are only calculated for those students who have accepted and enrolled in the school and therefore do not influence the rejection rate or the percent of admitted applicants who accept the university’s offer. Lastly, we agreed with their assumption that the average retention rate (i.e., percentage of students making it through the studies) does not influence anything else. Because Mark mentioned in class that is better not to restrict the model with too much prior knowledge, we only used this temporal ordering. The following is a screenshot from Genie showing this ordering:

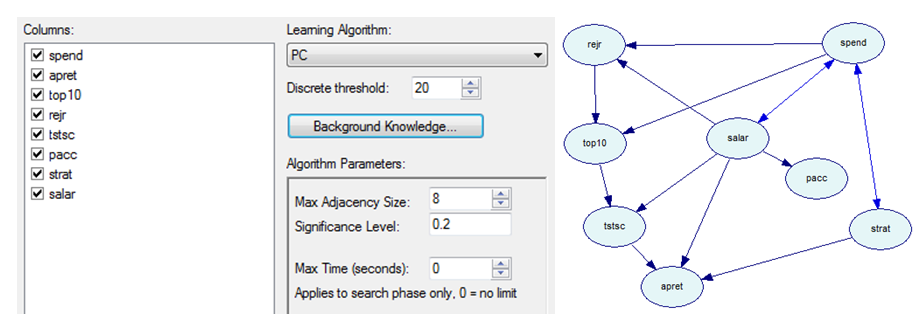


We decided to use the same significance levels as the authors to compare results in Genie compared to those obtained in the Tetrad II system they used. Here are our results:









All in all, the models from the chapter and from Genie are similar but have some differences. For the 0.01 significance level, the model in Genie resembled their findings. The main difference is that their model didn’t include a causal relationship between spend and tstsc. Also, the directionality of some of the causal relationships is different. For a significance level of 0.05, the Genie model also differed from their model. Along with the spend-tstsc relationship mentioned before, other differences in the Genie model include causal relationships from sala to rejr and from strat to apret. Their results don’t find these relationships until they run the model at p=0.1. For p=0.1 and p=0.2, the Genie models stay the same just as they do in their results. One difference of these Genie models to the two previous lower p value Genie graphs is the relationship from spend to tstsc is not shown. Instead, a causal relationship from spend to top10 exists. This corresponds to a relationship between spend and top10 in their models at these significance levels. Also, the Genie models show a causal relationship from salar to apret which doesn’t exist in their models.

We think the main observation to make for the assignment is that the primary causal factor on average percentage of graduation is average standardized test scores of incoming students. This is the only causal relationship that was detected at every significance level we tried. In terms of actionable intelligence from this finding, these results indicate that while there is a complex chain of dependency between all modeled factors, perhaps the simplest, most directly effective strategy for a college to improve its graduation percentage would be for the college to focus on applicants' standardized test scores when determining eligibility for admission.