Path Analysis Assignment

The following dataset includes many demographic and motivational variables predicting student’s withdrawal rates for their first year of college. You will set up and test a path model predicting withdrawal for students.

Some example ideas:

* Demographics 🡪 satisfaction / commitment 🡪 intentions to quit 🡪 withdrawal
* Demographics 🡪 student success 🡪 intentions to quit 🡪 withdrawal
* Demographics 🡪 semester / career goals 🡪 intentions to quit 🡪 withdrawal
* Other combinations of included variables. Try at least four variables predicting withdrawal with at least 3 points (something 🡪 something 🡪 withdrawal … they cannot all predict withdrawal separately).

Variables:

* Subject number
* Age
* ACT self-report
* Gender (1 = male, 2 = female)
* Declared a major (0 = no, 1 = yes)
* High School GPA
* Married (0 = no, 1 = yes, exclude the 3s)
* Instate/out (0 = outstate, 1 = instate, exclude the 3s)
* Class Year (freshman through senior+)
* Campus Job (0 = no, 1 = yes)
* Academic scholarship (0 = no, 1 = yes)
* Other scholarship (0 = no, 1 = yes)
* Number of older siblings
* Number of younger siblings
* Parents married (no, yes, divorces)
* Combined parent income (keep as continuous, ranges from 15K to 200K)
* Father education (keep as continuous, ranges from some high school to Ph.D.)
* Mother education (keep as continuous, ranges from some high school to Ph.D.)
* Academic Self Efficacy
* Intrinsic Motivation
* Integration
* Identification
* Introjection
* External Regulation
* Amotivation
* Relative Autonomy
* Affective commitment + Calculative Commitment
* Search Behavours + Intent to Search + Intent to Quit
* Satisfaction with MSU + Satisfaction Extra Curriculars
* Hours attempted semester 1
* Hours completed semester 1
* SEM GPA semester 1
* Cumulative GPA semester 1
* Withdrawal (THE DV!, 0 – stayed, 1 – withdrew, exclude all others)
* Semester goals – a combination of goals about the student’s semester
* Career goals – a combination of goals about the student’s career path

1. First, pick the variables you want to use, and then do data screening for just those variables.
   1. Accuracy: you can assume the data set is accurate.
      1. Exclude the precollege students (i.e., anything not 0 or 1) from the withdrawal variable.
   2. Missing data:
      1. Include a table that shows the missing data by person (summary of missingness by participant).
         1. Exclude all people who are missing more than 5% for the variables you are interested in.
      2. Include the percent/proportion of data missing by variable.
         1. Because you are likely using a small number of variables, you will not have any participants you can replace.
   3. Outliers
      1. Create Mahalanobis distance scores.
      2. Include the *df* for your cut off score.
      3. Incude the cut off score.
      4. Include a summary of the number of outliers in your dataset.
      5. Exclude all multivariate outliers.
   4. Multicollinearity
      1. Are any of the variables too correlated?
      2. Run a correlation table (do not need to include).
      3. Use the symnum function to look at the data to see if any of the correlations are too high.
   5. Normality
      1. Include the multivariate histogram.
      2. Is the data normal?
   6. Linearity
      1. Include the QQ plot for linearity.
      2. Is the data linear?
   7. Homogeneity and Homoscedasticity
      1. Include a residual scatterplot of the data.
      2. Is the data homogeneic?
      3. Is the data homoscedastic?
2. Sample size: do you have an adequate sample size for the data?
3. Estimate the degrees of freedom before you run 🡪 did you get it correct?
4. Model testing:
   1. Test your first model idea.
   2. Look at the path estimates to see if any of your paths are non-significant.
   3. Try deleting those paths (one at a time!) OR try an alternative model (same variables).
   4. You should have at least three models total.
   5. Include fit indices for all models in a table.
      1. Include important fit indices discussed (X2,RMSEA, SRMR, CFI).
      2. Include for each change/step you took with the model (i.e. first model, then each modification change you tried, one path at a time).
   6. Note which model was the best by using model comparison rules.
   7. Did you see any Heywood cases?
   8. Include a picture of all models with estimates on the picture.
5. For your best model only, which estimates were significant?

Include an APA style write up with all of the above information and this document with your notes. It does not have to be super formal, but be sure to answer every question in some way.