STA 431s09 Assignment 5

Do this assignment in preparation for the quiz on Friday, Feb. 13th. Answers to Question 1 are practice for the the quiz, and are not to be handed in. For Question 2, bring both your log file and your list file to the quiz; they may (or may not) be handed in.

1. Recall from lecture the *invariance principle* of maximum likelihood estimation. Let the parameter of a model be θ_1, and $\theta_2 = g(\theta_1)$; then $\widehat{\theta_2} = g(\widehat{\theta_1})$. For structural equation models where $\boldsymbol{\Sigma}$ is not restricted by the model (that is, for "saturated" models) and $\boldsymbol{\theta} = g(\boldsymbol{\Sigma})$, one can use the invariance principle to obtain $\widehat{\boldsymbol{\theta}}$ in closed form, with no need for numerical approximation.

   So, consider the simple regression model

   $$Y = \gamma X + \zeta,$$

   where γ is an unknown constant, $X \sim N(0, \phi)$, $\zeta \sim N(0, \psi)$ and the random variables X and ζ are independent. X and Y are observable (manifest) variables.

   (a) What is the parameter vector $\boldsymbol{\theta}$ for this model? It has three elements.

   (b) What is the distribution of the data vector $(X, Y)'$? Of course the expected value is zero; obtain the covariance matrix in terms of $\boldsymbol{\theta}$ values. Show your work.

   (c) Now solve three equations in three unknowns to express the three elements of $\boldsymbol{\theta}$ in terms of $\sigma_{i,j}$ values. This gives you the function g in $\boldsymbol{\theta} = g(\boldsymbol{\Sigma})$.

   (d) For a sample of size n, give the MLE $\widehat{\boldsymbol{\Sigma}}$. Your answer is a matrix containing three scalar formulas (or four formulas, if you write down the same thing for $\widehat{\sigma}_{1,2}$ and $\widehat{\sigma}_{2,1}$). Write your answer in terms of X_i and Y_i quantities. You are *not* being asked to derive anything. Just translate the matrix MLE into scalar form.

   (e) Obtain the formula for $\widehat{\gamma}$ and simplify. Show your work.

   (f) Give the formula for $\widehat{\phi}$.

   (g) Obtain the formula for $\widehat{\psi}$ and simplify. Show your work.

2. Use `proc calis` to fit the following model to the SAT data of Assignments Three and Four.

   $$Y = \gamma_1 X_1 + \gamma_2 X_2 + e,$$

   where the covariance matrix of $(X_1, X_2)'$ is $\boldsymbol{\Phi}$, $V(e) = \psi$, e is independent of X_1 and X_2, all the expected values are zero (say the data have been "centered"), and all the random variables are normally distributed. Be ready to answer questions like the following:

(a) Draw a path diagram for this model.

(b) Why does "Chisquare" equal zero? Is something wrong? Answer Yes or No and explain in your own words.

(c) Compare the "Independence Model Chisquare" to your numerical answer to 6g from Assignment 4. Close, right? Now multiply by 199 instead of 200. Hummm. What is going on here?

(d) Do high Verbal scores lead to better GPA? (Remember, these are causal models.)

   i. What test statistic would you use to answer this question? Your answer is a single number from the printout.

   ii. What is the null hypothesis for this question? Answer in symbols from the model.

   iii. Do you reject H_0 at $\alpha = 0.05$? Answer Yes or No.

   iv. What (if anything) do you conclude about GPA? Answer in English. No Greek letters or statistical terminology is allowed.

(e) Do high Math scores lead to better GPA?

   i. What test statistic would you use to answer this question? Your answer is a single number from the printout.

   ii. What is the null hypothesis for this question? Answer in symbols from the model.

   iii. Do you reject H_0 at $\alpha = 0.05$? Answer Yes or No.

   iv. What (if anything) do you conclude about GPA? Answer in English. No Greek letters or statistical terminology is allowed.

(f) Compare numerical values of $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ from this assignment to $\widehat{\beta}_1$ and $\widehat{\beta}_2$ from Assignment 3. Comment.

(g) Compare the numerical values of the test statistics (for γ_1 and γ_2) from this assignment to those for β_1 and β_2 from Assignment 3. Comment.

(h) Of course the idea of scores on a standardized test *causing* good or poor performance in university is silly. Draw a path diagram for a more reasonable model for these data, one with two latent variables. One could make a case for correlated errors, but for simplicity please make the errors uncorrelated with the (latent) exogenous variables and with each other. The latent exogenous variables should be correlated.