STA 431s11 Final Exam Reference Sheet

$Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)}$

$f(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{k}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$

$L(\boldsymbol{\mu},\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-n/2}(2\pi)^{-nk/2}\exp -\frac{n}{2}\left\{tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu})\right\}$

$$\begin{aligned}
\mathbf{Y}_i &= \boldsymbol{\beta}\mathbf{Y}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\epsilon}_i \\
& \quad V(\mathbf{X}_i) = \boldsymbol{\Phi}_{11} \text{ and } V(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi}
\end{aligned}$$

$$\begin{aligned}
\mathbf{F}_i &= \left(\begin{array}{c} \mathbf{X}_i \\ \mathbf{Y}_i \end{array}\right) \\
& \quad V(\mathbf{F}_i) = \boldsymbol{\Phi} = \left(\begin{array}{cc} V(\mathbf{X}_i) & C(\mathbf{X}_i,\mathbf{Y}_i) \\ C(\mathbf{Y}_i,\mathbf{X}_i) & V(\mathbf{Y}_i) \end{array}\right) = \left(\begin{array}{cc} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{12}' & \boldsymbol{\Phi}_{22} \end{array}\right)
\end{aligned}$$

$$\begin{aligned}
\mathbf{D}_i &= \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i \\
& \quad V(\mathbf{e}_i) = \boldsymbol{\Omega}
\end{aligned}$$

Please see Identifiability Rules on the reverse

Rules for Parameter Identifiability

Note: All the rules listed here assume that errors are independent of exogenous variables that are not errors, and that all variables have expected value zero.

1. *Counting Rule:* If a model has more parameters than covariance structure equations, the parameter vector can be identifiable on at most a set of volume zero in the parameter space. This applies to all models.

2. *Measurement model* (Factor analysis) In these rules, latent variables that are not error terms are described as "factors."

   (a) *Double Measurement Rule*: Parameters of the double measurement model are identifiable. All factor loadings equal one. Correlated measurement errors are allowed within sets of measurements, but not between sets.

   (b) *Three-Variable Rule for Standardized Factors*: The parameters of a measurement model will be identifiable if

   - The errors are independent, and
   - Each observed variable is caused by only one factor, and
   - The variance of each factor equals one, and
   - There are at least 3 variables with non-zero loadings per factor, and
   - The sign of one non-zero loading is known for each factor.

   Factors may be correlated.

   (c) *Three-Variable Rule for Unstandardized Factors*: The parameters of a measurement model will be identifiable if

   - The errors are independent, and
   - Each observed variable is caused by only one factor, and
   - For each factor, at least one factor loading equals one, and
   - There are at least 2 additional variables with non-zero loadings per factor.

   Factors may be correlated.

   (d) *Two-Variable Rule for Standardized Factors*: A factor with just two variables may be added to a measurement model whose parameters are identifiable, and the parameters of the combined model will be identifiable if

   - The two new variables are caused only by the new factor, and
   - The errors for the two additional variables are independent of one another and of those already in the model, and
   - The variance of the new factor equals one, and
   - Both new factor loadings are non-zero, and
   - The sign of one new loading is known, and
   - The new factor has a non-zero covariance with at least one factor already in the model.

   (e) *Two-Variable Rule for Unstandardized Factors*: A factor with just two variables may be added to a measurement model whose parameters are identifiable, and the parameters of the combined model will be identifiable if

   - The two new variables are caused only by the new factor, and
   - The errors for the two additional variables are independent of one another and of those already in the model, and
   - At least one new factor loading equals one, and
   - The other new factor loading is non-zero, and
   - The new factor has a non-zero covariance with at least one factor already in the model.

   (f) *Four-variable Two-factor Rule*: The parameters of a measurement model with two factors and four observed variables will be identifiable if

   - Two variables are caused by each factor, and
   - All errors are independent, and
   - All factor loadings are non-zero, and
   - The covariance of the two factors does not equal zero, and
   - For each factor, either the variance of the factor equals one and the sign of one loading is known, or at least one factor loading equals one.

   (g) *Model Combination Rule*: Two models may be combined into a single model, and the parameters of the combined model will be identifiable if

   - The parameters of two measurement models are identifiable when the models are considered separately, and
   - The error terms of the first model are independent of the error terms in the second model.

   The additional parameters of the combined model are the covariances between the two sets of factors.

   (h) *Cross-over Rule*: Any number of new observable variables may be added to a measurement model, and the result is a model whose parameters are all identifiable if

   - The parameters of the original model are identifiable, and
   - The error terms associated with the new variables are independent of the error terms in the existing model.

   Each new variable may be caused by any or all of the factors, potentially resulting in a cross-over pattern in the path diagram. The error terms associated with the new set of variables may be correlated with one another.

   (i) *Error-Free Rule*: A vector of observable variables may be added to the factors of a measurement model, and the result is a model whose parameters are all identifiable if

   - The parameters of the original model are identifiable, and
   - The new observable variables are independent of the errors in the measurement model and that in the measurement model.

   The main application is that variables assumed to be measured without error may be included in the latent component of a structural equation model, provided that the parameters of the measurement model for the other variables are identifiable.

3. *Latent variable model*: Here, Identifiability means that the parameters involved are functions of $V(\mathbf{F}) = \mathbf{\Phi}$.

   (a) *Regression Rule:* If no endogenous variables cause other endogenous variables, the model parameters are identifiable.

   (b) *Acyclic Rule*: Parameters of the Latent Variable Model are identifiable if

   - The model is acyclic (no feedback loops through straight arrows), and
   - $V(\boldsymbol{\epsilon}) = \mathbf{\Psi}$ has the following block diagonal structure. Organize the variables into sets. Set 0 consists of the exogenous variables. For $j = 1, \ldots, k$, each variable in set j is caused by at least one variable in set $j-1$, and also possibly by variables in earlier sets. Error terms for the variables in a set may have non-zero covariances. All other covariances between error terms are zero. These conditions are satisfied if $\mathbf{\Psi}$ is diagonal.

4. *Two-Step Rule*: This applies to models with both a measurement component and a latent variable component, including the full two-stage structural equation model.

   1: Consider the measurement model as a factor analysis model, ignoring the structure of $V(\mathbf{F})$. Check identifiability.

   2: Check identifiability of the latent model separately, usually using the Counting Rule and the Acyclic Rule.

   If both identification checks are successful, the parameters of the combined model are identifiable.