STA 312f07 Assignment 4

Do this assignment in preparation for the quiz on Friday, Oct. 12th. The questions are practice for the quiz, and are not to be handed in.

1. Recall from an earlier assignment that the *reliability* of a measurement is the squared correlation between the observed score and the true score. Suppose we have two equivalent measurements:

$$\begin{aligned} X_1 &= T + e_1 \\ X_2 &= T + e_2, \end{aligned}$$

   where $E(T) = \mu$, $Var(T) = \sigma_T^2$, $E(e_1) = E(e_2) = 0$, $Var(e_1) = Var(e_2) = \sigma_e^2$, and T, e_1 and e_2 are all independent. Suppose we were to measure the true score T by adding the two imperfect measurements together. Would the result be more reliable?

   (a) Let $S = X_1 + X_2$. Calculate the reliability of S. Is there any harm in assuming $\mu = 0$?

   (b) Suppose you take k independent measurements (in psychometric theory, these would be called equivalent test items). What is the reliability of $S = \sum_{i=1}^k X_i$? Show your work.

   (c) What happens as $k \rightarrow \infty$?

2. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean θ_1 and variance $\theta_2 + \theta_3$, where $-\infty < \theta_1 < \infty$, $\theta_2 > 0$ and $\theta_3 > 0$. Is this model identified? Answer Yes or No and prove your answer.

3. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean θ and variance θ^2, where $-\infty < \theta < \infty$. Is this model identified? Answer Yes or No and justify your answer. You need not show the actual calculation.

4. Consider the following simple regression through the origin with measurement error in both the independent and dependent variables. This is the model for one observation. Implicitly, it holds for $i = 1, \ldots, n$, but the subscript i on all the random variables is invisible.

$$
\begin{aligned}
\eta &= \gamma\xi + \zeta \\
X_1 &= \xi + \delta_1, \\
X_2 &= \xi + \delta_2, \\
Y &= \eta + \epsilon,
\end{aligned}
$$

where ξ and η are latent variables, ζ, δ_1, δ_2, ϵ and ξ and are independent normal random variables with expected value zero, $Var(\xi) = \phi$, $Var(\zeta) = \psi$, $Var(\delta_1) = Var(\delta_2) = \theta_\delta$, and $Var(\epsilon) = \theta_\epsilon$. The regression coefficient γ is a fixed constant. The observable variables are X_1, X_2 and Y.

(a) Is this model identified? Answer Yes or No and prove your answer.

(b) Is just the parameter γ (a function of the parameter vector) identified?

(c) Suppose we were to *re-parameterize* the model by letting $\sigma^2 = \psi + \theta_\epsilon$. Would the re-parametrized model be identified? Does this seem like a good idea?

5. Consider the following multivariate regression model with no measurement error. This is the model for one observation. Implicitly, it holds for $i = 1, \ldots, n$, but the subscript i on all the random variables is invisible.

$$\mathbf{Y} = \boldsymbol{\Gamma}\mathbf{X} + \boldsymbol{\zeta}$$

where

$\mathbf{Y}$ is an $m \times 1$ random vector of observable dependent variables, so the regression can be multivariate; there are m dependent variables.

$\mathbf{X}$ is a $p \times 1$ observable random vector; there are p independent variables. $\mathbf{X}$ has expected value zero and variance-covariance matrix $\boldsymbol{\Phi}$, a $p \times p$ symmetric and positive definite matrix of unknown constants.

$\boldsymbol{\Gamma}$ is an $m \times p$ matrix of unknown constants. These are the regression coefficients, with one row for each dependent variable and one column for each independent variable.

$\boldsymbol{\zeta}$ is the error term of the latent regression. It is an $m \times 1$ random vector with expected value zero and variance-covariance matrix $\boldsymbol{\Psi}$, an $m \times m$ symmetric and positive definite matrix of unknown constants. $\boldsymbol{\zeta}$ is independent of $\mathbf{X}$.

Is this model identified? Answer Yes or No and show your work.