INET Conference Preliminary Materials

These questions are based off of Chapters 1-3 of circulated reading 'Advanced Data Analysis From an Elementary Point of View' by Cosma Shalizi.

1. What are two reasons R^2 is a poor indicator of model fit?
2. What is the optimal linear predictor? How does formulating and estimating it differ from the usual way of specifying linear regression?
3. Why are the following statements, in general, wrong:
   a. A variable that has a non-zero regression coefficient must influence the response.
   b. A variable that has a zero regression coefficient must not influence the response.
   c. If the independent variables change, we can predict how much the response will change by plugging its value in to the regression.
4. Suppose $\mathbb{X}$ is an $n \times p$ matrix and Y is an $n \times 1$ column vector. Derive the *ordinary least squares* estimator for predicting Y from X. That is, find

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}} \, ||\mathbb{Y} - \mathbb{X}\beta||_2^2 = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - X_i\beta)^2,$$

   where Y_i is the i^{th} element of $\mathbb{Y}$ and X_i is the i^{th} row of $\mathbb{X}$. Did you have to make any assumptions to find $\widehat{\beta}$?
5. Now that you've read the preliminary material and answered a few questions, what do you hope to learn from this course?