

STAT 578 - Sample Midterm Exam

Instructor: Professor D. Wiens

Time: 80 minutes

Instructions: Answer any four of these 5 equally weighted questions in the answer booklets supplied.

1. Consider the linear model

$$Y_i = \mathbf{x}_i^T \boldsymbol{\theta} + \varepsilon_i, \quad i = 1, \dots, n,$$

where the ε_i are independent and identically distributed random variables. Suppose that the Y_i each have a *Laplace* distribution, with density

$$p(y_i | \boldsymbol{\theta}, \sigma_\varepsilon) = \frac{1}{\sqrt{2}\sigma_\varepsilon} e^{-\frac{\sqrt{2}|y_i - \mathbf{x}_i^T \boldsymbol{\theta}|}{\sigma_\varepsilon}}.$$

- (a) Write down the likelihood function $L(\boldsymbol{\theta}, \sigma_\varepsilon)$.
 - (b) Show that the MLE of $\boldsymbol{\theta}$ is the L1 estimate.
 - (c) Derive the MLE of σ_ε .
2. Consider the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$. Assume that the $n \times p$ matrix \mathbf{X} has full column rank.
 - (a) Define the ‘hat’ matrix.
 - (b) Assume that the model has an intercept, so that \mathbf{X} has a column of ones. Show that then the sum of all elements of the hat matrix is n .
 - (c) What does the sum of *squares* of the elements of the hat matrix equal?
 3. An economist is studying the relationship between amount of savings and level of income for middle-income families from urban and rural areas, based on independent samples from the two populations. Each of the two relations - urban and rural - can be modelled by linear regression. She wishes to determine whether, at given income levels, urban and rural families tend to save the same amount; i.e. whether the two regression lines are the same. If they are not, she wishes to explore whether at least the amounts of savings out of an additional dollar of income are the same for the two groups.
 - (a) Write out a multiple linear regression model which may be fitted to the combined data from the two groups, and which allows one to answer both the economist’s questions.

- (b) What hypothesis would be tested, to see if the two regression lines are the same?
- (c) In the hypothesis in b) is rejected, how would you analyze the economist's second question?
- (d) How would the p-value of the test in c) be calculated? (Give an answer in the form $P(A > b)$, where A is a particular random variable whose distribution you specify, and b is a number whose calculation you describe.)
4. Suppose that you have fitted a straight line regression model $Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i$, and that an examination of the residuals then leads you to believe that the variance of ε_i is proportional to x_i (> 0). How would you proceed?
5. Some measures of influence are based on the change in the estimated covariance matrix of the regression coefficients, when the i^{th} case (\mathbf{x}_i, y_i) is removed from the sample. One suggested measure is the "covariance ratio"

$$CVR_i = \frac{\det \left[S_{(i)}^2 \left(\mathbf{X}_{(i)}^T \mathbf{X}_{(i)} \right)^{-1} \right]}{\det \left[S^2 \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \right]}.$$

Show that this reduces to

$$CVR_i = \left(\frac{S_{(i)}^2}{S^2} \right)^p / (1 - h_{ii}).$$