STAT 450

Final Examination

Richard Lockhart 3 December 1990

Instructions: This is an open book exam. You may use notes, books and a calculator. The exam is out of 60. Each question is worth 10 marks. Since some questions are much harder than others you would be well advised to make sure you find the ones you can do quickly and do them well. In Q 1 part a) is worth 8, part b) 2. In Q 2, the parts are worth 3, 4 and 3; you may well be able to do c) without being able to do a) or b). In this case a well done version of part c) might be worth as much as 5 out of 10. In Q 5 the parts are worth 5, 2, 2 and 1. In Q 6 the parts are worth 3 each. The extra 2 marks are bonus marks for part d) of that question. **DON'T PANIC.**

1. Suppose U and V are independent random variables each with the same density

$$f(t) = \exp(-t)1(t > 0).$$

   (a) Find the density of $T = U/V$.

   (b) Suppose a test statistic has the distribution of T. Derive a formula for the upper α critical point of the test statistic.

2. Suppose $\mathbf{Y}$ has a multivariate normal distribution with mean vector $\mathbf{X}\beta$ and variance covariance matrix $\sigma^2\mathbf{D}$ where $\mathbf{D} = diag(d_1, \ldots, d_n)$ is a known diagonal matrix.

   (a) Give a general formula for the log-likelihood $\ell(\beta, \sigma)$. Your answer should be written in matrix form.

   (b) Let $\mathbf{B} = diag(d_1^{-1/2}, \ldots, d_n^{-1/2})$. Let $\mathbf{Z} = \mathbf{BY}$. What is the distribution of $\mathbf{Z}$? Show that $\mathbf{Z}$ satisfies a linear model and identify the design matrix $\mathbf{X}^*$ for this linear model.

   (c) For the special case $\mathbf{D} = diag(1, 1/4, 1/9)$ and

$$\mathbf{X} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

   derive the likelihood equations.

3. Suppose that Z_1, Z_2, Z_3 are independent standard normal variables. What is the distribution of $(5Z_1^2 + 2Z_2^2 + 5Z_3^2 + 4Z_1Z_2 + 4Z_2Z_3 - 2Z_1Z_3)/6$?

4. Suppose that $Y_1, \ldots, Y_n$ are independent random variables and that $x_1, \ldots, x_n$ are the corresponding values of some covariate. Suppose that the density of Y_i is

$$f(y_i) = \exp\left(-y_i \exp(-\beta x_i) - \beta x_i\right) 1(y_i > 0)$$

   where β is an unknown parameter. Find the log-likelihood, the score function and the Fisher information. You may use without proof the fact that $\mathrm{E}(Y_i) = \exp(x_i\beta)$.

5. Consider the salmon data discussed in class. Imagine that the researcher begins with the model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2$. Fitting this model by least squares, the error sum of squares is 2.231177. For the model $Y_i = \beta_0 + \beta_1 X_i$ the error sum of squares is 2.59839. For the model $Y_i = \beta_0 + \beta_2 X_i^2$ the error sum of squares is 2.641787. For the model $Y_i = \beta_0 + X_i$ the error sum of squares is 3.144104. There are 20 data points.

   (a) Test the hypothesis that $\beta_1 = 1$ and $\beta_2 = 0$ using this information.

(b) Now consider the model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Let $ESS_u = 2.59839$ be the error sum of squares when this model is fitted by least squares. Let $\hat{\beta}_i$ be the corresponding least squares estimates. Show that the MLE of σ^2 is $\hat{\sigma^2} = ESS_u/20$ and evaluate the log-likelihood, $\ell(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})$. (The answer is a number which can be calculated from the information given in this question.)

(c) Use the idea in b) to calculate the profile likelihood $\ell(\hat{\beta}_0(1), 1, \hat{\sigma}(1))$ which is the maximum of the log likelihood for the model $Y_i. = \beta_0 + x_i + \epsilon_i$.

(d) Now compute a P-value for the likelihood ratio test statistic $2(\ell(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}) - \ell(\hat{\beta}_0(1), 1, \hat{\sigma}(1)))$.

6. A regression model $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ is fitted. It is found that

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}^{-1} = \begin{matrix} 2 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{matrix}$$

and that

$$\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \begin{matrix} 3 \\ 2 \\ 1 \end{matrix}$$

There are 23 data points. The error sum of squares is 80.

(a) Give a 90% confidence interval for $\beta_1 - 2\beta_3$.

(b) Test the hypothesis that $\beta_1 = \beta_2$.

(c) A new observation is planned; the row in the design matrix for this new observation would be 1,-1,2. Give a 95% prediction interval for the new Y.

(d) Assume that in fact $\beta_1 = \beta_3 = 0$. What is the distribution of $\hat{\beta}_1/\hat{\beta}_3$?