M. S. Comprehensive Exam
Saturday, August 27, 2005

You have three hours. Do as many problems as you can. No one is expected to answer all the problems correctly. Partial credit will be given. All problems are worth an equal amount of credit.

Put your solution to each problem on a separate sheet of paper.

Problem 1. We consider measurements on rock samples from a petroleum reservoir (`rock` data). Each sample was measured for permeability (`perm`, in milli-Darcies), total area of pores (`area`, in pixels out of 256 by 256), total perimeter of pores (`peri`, in pixels) and shape (`shape`, perimeter/sqrt(area)).

We fitted a linear regression with `perm` as the dependent variable and `area` and `peri` as the independent variables:

```
g<-lm(perm~area+peri,data=rock)
```

The anova table of `rock` is as follows:

```
> anova(g)

Analysis of Variance Table

Response: perm
          Df  Sum Sq Mean Sq F value    Pr(>F)
area       1 1417333 1417333  22.352 2.263e-05 ***
peri       1 4738469 4738469  74.729 4.052e-11 ***
Residuals 45 2853383   63409
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(a) How many observations does the data set contain?

(b) Explain (1) in words what the number 4738469 under “Sum Sq” means and (2) how do you compute it?

(c) Compute the regression sum of squares.

(d) Next we fit the regression in a different order:

```
gg<-lm(perm~peri+area,data=rock)
```

Describe the changes in the anova table given by

```
>anova(gg)
```

Problem 2. Consider the linear regression model

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

with $X_1^T X_2 = 0$. Find the least squares estimates of β_1 and β_2. Justify your answer.

[Note: Y, X_1, X_2, ϵ have dimensions $n \times 1$, $n \times p_1$, $n \times p_2$, $n \times 1$, respectively, with $n \geq p_1 + p_2$. X_1 and X_2 have full rank.]

Problem 3. Consider a Factorial design with two factors A and B. Suppose the factors A and B have a and b levels, respectively and there are r replications for each treatment.

(a) Write the ANOVA model for the above factorial design *with interaction*. Explain the meaning of each symbol that you use. Write down the distributional assumptions about the random variables in the model. If any constraints or assumptions regarding the parameters are used, also write them down.

(b) Write down the null and alternative hypotheses in terms of the model parameters for testing no interaction between the two factors. How would you test this by using the ANOVA table?

(c) State a graphical method for testing no interaction between the factors. What are the advantages and disadvantages of the graphical method compared to the method based on the ANOVA table?

Problem 4. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each with mean μ and variance σ^2. Let N be a nonnegative integer-valued random variable with finite variance that is independent of the X_i's. Find the variance of $\sum_{i=1}^{N} X_i$.

Problem 5. Let X be distributed as a Poisson random variable with mean λ and Y be distributed as a Poisson random variable with mean μ. Assume X and Y are independent.

For a fixed positive integer w, find the distribution of X given $X + Y = w$.

Problem 6. For a continuous random variable Y, show that

$$E[Y] = \int_0^\infty P\{Y > y\}dy - \int_0^\infty P\{Y < -y\}dy\,.$$

Problem 7. Suppose a random sample of size n is taken from a population with density

$$f(x) = \begin{cases} \frac{2x}{2\theta+1} & \text{for } \theta \le x \le \theta + 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 1$. Find the maximum likelihood estimate of θ. Is it unique?

Problem 8. Let $X_1, \ldots, X_n$ be an i.i.d. sample from an $N(\mu, \sigma^2)$ distribution. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Let $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

(a) Find the maximum likelihood estimators of μ and σ^2, respectively.

(b) State the distribution of

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S}.$$

(You do **not** need to derive this distribution.)

(c) Use (b) to *derive* a $(1 - \alpha)\%$ confidence interval for μ.

Problem 9. Suppose that X_1 and X_2 are iid observations from the pdf

$$f(x|\theta) = \theta x^{\theta-1} e^{-x^{\theta}} \quad \text{for } x > 0 \text{ and } \theta > 0.$$

Show that $(\log X_1)/(\log X_2)$ is an ancillary statistic.