M. S. Comprehensive Exam (Part I of Written Exam)
Sunday, January 7, 2001

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 draw 52 cards without replacement from a well-shuffled deck of 52 cards. We place them in a line from left to right. We say there is a Jack-Queen contact in positions $(i, i+1)$ if position i has a Jack and position $i+1$ has a Queen or if position i has a Queen and position $i+1$ has a Jack.

(a) What is the expected number of Jack-Queen contacts?

(b) Use Markov's inequality to give an upper bound for the probability that there is at least one Jack-Queen contact.

Problem 2. Consider the simple regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

with $\mathrm{Var}(\epsilon_i) = \sigma^2$. The least squares estimates for β_0 and β_1 are 1.84 and 1.09, respectively. The estimated variances for $\hat{\beta}_0$ and $\hat{\beta}_1$ are

$$s^2(\hat{\beta}_0) = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{SXX} \right) = 0.124, \qquad s^2(\hat{\beta}_1) = \hat{\sigma}^2 / SXX = 0.025,$$

where SXX is the corrected sum of squares for X, $\bar{x}$ is a positive number, $\hat{\sigma}^2 = 1.64$, and $n = 35$.

(a) Based on the information given above, estimate $\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1)$.

(b) Let $\hat{Y}$ be the fitted value at $X = 4$. Estimate the variance of $\hat{Y}$ if $\hat{Y}$ is being used to estimate the mean value of Y at $X = 4$.

Problem 3. Suppose that when measuring the **radius** of a circle, an error is made that has a $N(0, \sigma^2)$ distribution with σ^2 unknown. If n independent measurements are made, find the best unbiased estimator of the **area** of the circle (and prove it is best unbiased).

Problem 4. A sociologist is examining the personalities of students in different major fields. She has randomly selected ten students from each of five majors, and plans to do an analysis of variance on each of the subscores of the personality test. The five fields are Math, Computer Science, English, History, and Art.

Suppose she gets the ANOVA results and the overall F test from the first subscore, a measure of intraversion/extraversion, is significant. (An introvert is primarily concerned with his/her own mental life, not the lives of others.) Assuming the usual ANOVA assumptions are satisfied, what further tests would you do to examine the differences among the five majors? Justify your choice.

Problem 5. A certain community is composed of m families, n_i of which have i children, $\sum_{i=1}^{r} n_i = m$ (r is the number of children in the largest family). If one of the families is randomly chosen, let X denote the number of children in that family. If one of the children is randomly chosen, let Y denote the total number of children in the family of that child. Show $E[Y] \geq E[X]$.

Problem 6. Consider a sample of $n = 68$ randomly selected individual tax returns for 1999 prepared by a local branch of a national tax preparation service. The following multiple linear regression models are run :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_1 X_4 + \beta_6 X_2 X_4 + \beta_7 X_3 X_4 + \epsilon$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

with the sums of squared residuals equal 401.61 and 504.88, respectively, where X_1 represents total income; X_2 is earned income; X_3 is federal itemized or standard deductions; X_4 represents marital status ($X_4 = 1$ if married, $X_4 = 0$ if single); Y represents total tax paid as a percent of total income. Using a 5% level of significance, test the hypothesis that $\beta_5 = \beta_6 = \beta_7 = 0$ versus the alternative that at least one of these beta coefficients is not zero.

Problem 7. Suppose that $X_n \sim AsN(\theta, \sigma_n^2(\theta))$ (see definition below), where θ is a real-valued parameter. A differentiable function g is called a variance stabilizing transformation for X_n if $g(X_n) \sim AsN(g(\theta), \xi_n^2)$, where ξ_n does not depend on θ.

(a) Assume that $\sigma_n(\theta) = h(\theta) \cdot \nu_n$, where $\nu_n \to 0$. Show that a variance stabilizing transformation satisfies the differential equation $g'(x) = c/h(x)$, where c is a constant.

(b) Find **a** variance stabilizing transformation for X_n, where $X_n = Y_n/n$ and $Y_n \sim$ Binomial(n, p).

Definition: AsN stands for "asymptotically normal". To be precise, $X_n \sim AsN(\mu, \sigma_n^2)$ means that the sequence $\{X_n\}$ satisfies

$$\frac{X_n - \mu}{\sigma_n} \xrightarrow{D} N(0,1) \quad \text{as } n \to \infty\,.$$

Problem 8. Let $X_1, X_2, X_3, \ldots$ be i.i.d. random variables. Let N be a nonnegative integer-valued random variable which is independent of the X_i's. Show that

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = E(N)\mathrm{Var}(X_1) + (EX_1)^2\mathrm{Var}(N)\,.$$

(Assume that both EN^2 and EX_1^2 are finite.)

Problem 9. Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution on the interval $(\theta, 1/\theta)$ where $0 < \theta < 1$. Find the maximum likelihood estimator of θ.