# M. S. Comprehensive Exam (Part I of Written Exam)
# Saturday, January 2, 1999

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.** From an ordinary deck of 52 cards, cards are drawn at random, one by one, and with**out** replacement until a heart is drawn. What is the expected value of the number of cards drawn?

---

**Problem 2.** An investigator was interested in comparing two drug products (A and B) in overweight female volunteers. Twenty volunteers were randomly selected. Ten of the 20 women were randomly assigned to A and the remaining 10 to B. The response of interest $y$ is a score on a rating scale used to measure the mood of a subject. On the study day, all 20 women are required to complete a checklist at 8 A.M. Then each subject was given the prescribed medication (A or B). Each subject is required to complete the checklist again at 12 A.M. Part of the data are listed in the following table:

| Drug A | | Drug B | |
|---|---|---|---|
| 8 A.M. | 12 A.M. | 8 A.M. | 12 A.M |
| x | y | x | y |
| 5 | 20 | 7 | 19 |
| 10 | 23 | 12 | 26 |
| ... | ... | ... | ... |

**(a)** To conduct an analysis of covariance for this experiment, you may compare several models. Write the models you will use and state your assumptions.

**(b)** One main objective of an analysis of covariance is to compare the treatment means after adjusting for the effect of covariate $x$. Which two models do you want to compare in order to achieve this purpose? How to make the comparison?

---

**Problem 3.** Suppose $Y_1, \cdots, Y_n$ are independent observations and $x_1, \cdots, x_n$ are fixed positive known constants (levels). $Y_i$ is assumed to have a Poisson distribution with mean $e^{\beta x_i}$, i=1, $\cdots$, n, where $\beta$ is unknown. Show that the maximum likelihood estimate $\hat{\beta}$ of $\beta$ exists if and only if not all the $Y_i$ are zero.

---

**Problem 4.** You are being interviewed by an engineer in an organization in which you want to work. Answer the engineer's questions in everyday English comprehensible to your interviewer who is very bright and has not taken any statistics courses.

**(a)** I do not understand why you statisticians use techniques that you know contain errors. For example, in the reporting of every poll, a "margin of error" is mentioned. If you know you have made errors in the poll, why don't you correct them? That's what an engineer would do!

**(b)** I also do not understand all the emphasis you folks put on "simple random sampling." If I want to know how good the paper towels are that we manufacture, I will go to the warehouse and pick cartons from several places and test some towels from each carton. Why is a "simple random sample" better than this? (In your answer you may wish to describe a demonstration using physical objects or computer simulations that would help the engineer to understand you.)

---

**Problem 5.** The model $E(Y) = X\beta$ is fit by least squares to a set of data $(Y_{n\times 1}, X_{n\times p})$, where $n > p$ and $\text{rank}(X) = p$. Let $H = (h_{i,j})_{n\times n}$ be the hat matrix, or the projection matrix such that the fitted values $\hat{Y} = HY$.

**(a)** Show that $\sum_{i=1}^{n} h_{i,i} = p$.

**(b)** Suppose that all the design points happen to have the same leverage ($h_{1,1} = h_{2,2} = \ldots = h_{n,n}$). Show that the residuals $\hat{\varepsilon}_i$'s satisfy

$$|\text{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j)| \le \frac{n-p}{n}\sigma^2$$

where we assume $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are i.i.d. with $\text{Var}(\varepsilon_i) = \sigma^2$.

---

**Problem 6.** Let $X$ and $Y$ be independent and identically distributed $N(0, \sigma^2)$ random variables.

**(a)** Show that $X^2 + Y^2$ and $\frac{X}{\sqrt{X^2+Y^2}}$ are independent.

**(b)** Let $\theta = \sin^{-1} \frac{X}{\sqrt{X^2+Y^2}}$. Show that $\theta$ is uniformly distributed on $(-\frac{\pi}{2}, \frac{\pi}{2})$.

---

**Problem 7.** Suppose that $X_1$ and $X_2$ are two independent exponential random variables with means $EX_1 = 1/3$ and $EX_2 = 1/2$.

**(a)** Let $Y = X_1 + X_2$. Find the probability density function of $Y$.

**(b)** Let $Z = \min(X_1, X_2)$. Find the probability density function of $Z$.

---

**Problem 8.** Let $A_1, A_2, \ldots$ be a finite or countably infinite sequence of events. Define $p_k = P(A_1 \cap A_2 \cap \cdots \cap A_k)$ for all $k$. If, for any value of $k$ and any choice of subscripts $1 \leq i_1 < i_2 < \cdots < i_k$, we have

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = p_k \,,$$

the events are said to be exchangeable.

**(a)** We have two urns. Urn 1 contains two red and three white balls. Urn 2 contains one red and one white ball. We choose urn 1 with probability 2/3 and urn 2 with probability 1/3. After we choose the urn, we draw $n$ balls from it **with** replacement. Let $A_j$ be the event that the $j$-th draw is a red ball. Show that $A_1, A_2, \ldots, A_n$ are exchangeable.

**(b)** Let $X$ be a random variable with $P(0 \leq X \leq 1) = 1$. Let the events $A_1, A_2, \ldots$ be conditionally independent given $X$, and let $P(A_j|X) = X$ for all $j$. Show that $A_1, A_2, \ldots$ are exchangeable.

---

**Problem 9.** Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. from the probability density $f(x|\theta) = 2x/\theta^2$ for $0 < x < \theta$ (and $f(x|\theta) = 0$ otherwise).

**(a)** Find a sufficient statistic for $\theta$. (Your statistic should also be minimal, but you are not required to show this.)

**(b)** Find an ancillary statistic (and show that it is ancillary).