M. S. Comprehensive Exam
Friday, August 24, 2007

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. Suppose a researcher wants to design a 2^{7-2} experiment with factors A, B, C, D, E, F, G. Upon further consideration, it was determined that the last two factors, F and G, must be assigned to interactions involving the first five factors.

(a) Suppose the factor assignment of $F = ABC$ and $G = BCD$ is made. Ignoring three-factor interactions and higher, list all of the two-factor interaction aliases.

(b) What is the resolution of the design in (a) and explain how it can be determined.

(c) Instead, suppose the factor assignment of $F = ABCD$ and $G = ABDE$ was made. Ignoring three-factor interactions and higher, list all of the two factor interaction aliases.

(d) In general, would either of the designs of (a) or (c) be preferred over the other? Explain why or why not.

Problem 2. The number of mishaps each week on an assembly line was collected.

Number of Mishaps	Observed Frequencies
0	14
1	28
2	32
3	25
4	9
5	6
6+	2

(a) You need to test if the data follows a Poisson distribution with mean $\lambda = 2$. Use level of significance $\alpha = .05$. Give the hypotheses, the value of the test statistic, and the criteria for rejecting the null hypotheses.

(b) If the mean of the Poisson distribution is not known, but estimated to be $\lambda = 2$, what changes are made to the answer to part (a)?

Problem 3. Let $y = \beta_0 + \beta_1 x + \epsilon$ be a simple linear regression, and the observations are $x_i, y_i, i = 1, \ldots, n$. Let $\bar{y}$ denote the mean of $\{y_i\}$ and $\hat{y}_i$ denote the least squares estimate of y_i (i.e., the fitted value) by the linear model. Prove that:

$$R^2 = r^2$$

where R^2, the coefficient of determination, and r^2, the correlation coefficient, are defined as

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

and

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \, .$$

Problem 4. Let the failure time of a component be T, a positive, continuous random variable. Let F be the distribution (cdf) for T and suppose it may be written as $F(t) = 1 - e^{-H(t)}, t > 0$, for some differentiable function H.

(a) Show the density f of T satisfies

$$\frac{f(t)}{1 - F(t)} = H'(t) \, .$$

The function $h(t) \equiv H'(t)$ is referred to as the failure rate, or hazard function, for T.

(b) Show for positive t and u,

$$P(T > t + u | T > t) = e^{-\int_t^{t+u} h(y)dy}$$

(c) Suppose the hazard function h for some random variable T is $h(t) = \theta$ for all $t > 0$. Find and name the distribution of T.

Problem 5. A *parallel* system is one that functions as long as at least one of its components functions. A parallel system has q independent components, the i-th component having a life length with an exponential distribution with mean β_i $(i = 1, \ldots, q)$. The life length of the system is the maximum of the individual life lengths of its components. What is the probability density function of the life length of the system?

Problem 6. Answer the following. (The parts are not related.)

(a) If Y is uniformly distributed over $(0, 5)$, what is the probability that the roots of the equation $4x^2 + 4xY + Y + 2$ are both real?

(b) If U is uniformly distributed over $(0, 1)$, find the density function of $Z = e^U$.

Problem 7. Let $X_1, \dots, X_n$ be an i.i.d. sample from an exponential distribution with probability density function

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) We want to test $H_0 : \lambda = 1$ versus $H_A : \lambda = 2$. Show that the Neyman-Pearson test rejects the null hypothesis for small values of $2\sum_{i=1}^n X_i$.

(b) Find the distribution of $2X_i$ under the null hypothesis $H_0 : \lambda = 1$.

(c) It can be shown that $2\sum_{i=1}^n X_i$ has a χ^2_{2n} distribution under $H_0 : \lambda = 1$.

Using the table for the percentiles of the χ^2-distribution, obtain the most powerful test for testing $H_0 : \lambda = 1$ versus $H_A : \lambda = 2$ at level $\alpha = .05$. Compute the rejection region of this test.

(d) We observed $n = 20$ observations with sample mean $\bar{x} = 1.22$. Test (at level .05) $H_0 : \lambda = 1$ versus $H_A : \lambda = 2$.

Problem 8. Let $X_1, \dots, X_n$ be iid Poisson(μ).

(a) Compute $p = P\{X_1 \geq 1\}$.

(b) Compute the maximum likelihood estimator $\widehat{p}$ of p.

(c) Compute the asymptotic variance of $\widehat{p}$.

Problem 9. Find a sufficient statistic for θ in each of the following situations. (Your statistics should also be minimal, but you are **not** required to show this.)

(a) You observe $X_1, \dots, X_n$ iid from the density

$$f(x \mid \theta) = \begin{cases} e^{x-\theta} & \text{for } x \leq \theta \\ 0 & \text{for } x > \theta\,. \end{cases}$$

(b) You observe independent random variables $X_1, \dots, X_n$ where X_i has the density

$$f_{X_i}(x \mid \theta) = \begin{cases} e^{x-i\theta} & \text{for } x \leq i\theta \\ 0 & \text{for } x > i\theta\,. \end{cases}$$