Math 218 Mathematical Statistics

Final Exam

May 2009

For this take-home final exam, you may use class notes, the text, and other books and articles. Do not talk to anyone about the test except me. If you have any questions about the test, ask or email me.

Write the answer to each of the five questions on a separate page of paper, then staple the pages together.

You may leave your answers as expressions such as $\binom{8}{4}\frac{e^{1/3}}{\sqrt{2\pi}}$ if you like. Points for each problem are in square brackets.

Problem 1. [24; 6 points each part] Suppose that you're interested in determining whether a particular dangerous substance, like mercury or a pesticide, is entering the food chain and concentrating further along the food chain. You might sample the amounts of chemical residues in two different species—one predator species and one prey species—to see if the prey species has a lower concentration of the substance and the predator species that eats the prey species has a higher concentration of the substance.

a. Suppose that you are designing this experiment. Answer each of these questions with a sentence, or two at most: (1) Would this be a comparative or a noncomparative study? (2) What confounding factors might there be? (3). What can you do to assure that the sampling is random?

b. Determine what the null hypothesis H_0 and the alternative hypothesis H_1 should be. Should this be a one-sided test or a two-sided test; why? Explain why why you chose what you did to be H_0 instead of H_1.

c. The experiment will involve measuring the substance levels in n_1 animals of the predator species and n_2 animals in the prey species. Suppose that for economic reasons you can only have small samples of the two populations. What is the test statistic and the rejection region for the null hypothesis H_0 you gave in part b?

d. Suppose that you test $n_1 = 10$ predators and $n_2 = 13$ prey and you find the mean concentration of the substance in the predator sample is $\overline{x}_1 = .041$ while that in the prey sample is $\overline{x}_2 = .026$, and the standard deviation for the predator sample is $s_1 = .017$ while that for the prey is $s_2 = .006$. Is H_0 rejected at the $\alpha = .05$ significance level? (Show your work.)

Problem 2. [24; 8 points each part] Suppose that the median sales prices for new singlefamily homes in a particular locale are as follows

year	median sales price in thousands
1992	207
1993	223
1994	235
1995	249
1996	264
1997	278
1998	295
1999	312

a. Given this data, find the least squares line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$. (You may also want to graph the data and visually compare it to this line, but that's not required.)

b. Compute the error sum of squares (SSE), the total sum of squares (SST), and the regression sum of squares (SSR). Also find the coefficient of determination r^2 and the sample correlation coefficient r.

c. Test the significance of the linear relationship between year and median sales price using a significance level of $\alpha = .05$. (Refer to the section on analysis of variance for simple linear regression.)

Problem 3. [24; 6 points each part] Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with mean λ. This distribution has the density function $f(x) = \dfrac{1}{x!}\lambda^x e^{-\lambda}$, for $x = 0, 1, 2, \ldots$. Furthermore, the mean of this Poisson distribution equals λ, and its variance also equals λ.

a. Write down the likelihood function

$$L(\lambda | x_1, \ldots, x_n) = f(x_1, \ldots, x_n | \lambda) = \prod_{i=1}^{n} f(x_i | \lambda),$$

then find its logarithm and simplify that.

b. Compute the derivative of the logarithm of the likelihood function.

c. Determine the maximum likelihood estimator $\hat{\lambda}$ for λ by finding the critical point of the logarithm of the likelihood function.

d. Determine the mean and variance of $\hat{\lambda}$.

Problem 4. [10] In your own words, briefly explain the difference between a 95% confidence interval as used by a classical statistician and a 95% probability interval as used by a Bayesian statistician. (Try to keep your answer down to half a handwritten page.)

Problem 5. [18; 9 points each part] A publication for college students is planning to survey students to estimate the average amount students spend on living expenses, excluding rent, per month. The publication expects to sample a large number of students, at least $n = 40$, and has done surveys like this in the past and has the ability to select random samples of students.

a. Describe a process, which may depend on n, that will give a 95% confidence interval for the average living expenses.

b. Suppose that the publication wants the margin of error (half the length of the confidence interval) to no more than \$20. Explain how you could determine how large n has to be to achieve this margin of error.