Mathematics 376 – Probability and Statistics II
Final Examination
May 8, 2010

Directions

Do all work in the blue exam booklet, and *include all work necessary to justify your answers.* Be sure to read each question carefully before starting to work. There are 200 regular points and 20 Extra Credit points.

General Note: A "random sample" always consists of independent measurements from an indicated distribution.

I. Let Y_1, Y_2, Y_3, Y_4, Y_5 be a random sample from a normal distribution with mean $\mu = 4$ and standard deviation $\sigma = 10$.

A) (10) What is the distribution of $\overline{Y} = \frac{1}{5}(Y_1 + Y_2 + Y_3 + Y_4 + Y_5)$? Why?
B) (10) What is the distribution of $U = \frac{(Y_1-4)^2+(Y_2-4)^2+(Y_3-4)^2}{100}$? Why?
C) (10) What is the distribution of $V = \frac{\sqrt{3}(Y_4-4)}{10\sqrt{U}}$, where U is as in part B? Why?
D) (10) How would you determine the PDF for the sample maximum $Y_{(5)}$? (Note: The CDF for a normal random variable is not an elementary function; just give a "recipe" for how it might be computed.)

II. A random variable Y is said to have a *log-normal* distribution with parameters μ, σ if its pdf has the form

$$f(y) = \begin{cases} \frac{1}{y\sqrt{2\pi\sigma^2}} e^{-(\ln(y)-\mu)^2/(2\sigma^2)} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

A) (10) Compute $E(\ln(Y))$ for a log-normal random variable. (Hint: Set up the integral, then change variables.)
B) (10) Let $Y_1, \ldots, Y_n$ be a random sample from a log-normal distribution with unknown μ and known $\sigma = 1$. Find the maximum-likelihood estimator for μ.
C) (5) Is your estimator from part B biased or unbiased? Why?

III. (15) Solid copper produced by melting powdered ore is tested for "porosity" (the volume fraction due to air bubbles). A sample of $n_1 = 40$ porosity measurements made in one lab has $\overline{y}_1 = .25$ and $s_1^2 = .001$. A second set of $n_2 = 50$ measurements is made using identical ore in a second lab, yielding $\overline{y}_2 = .17$ and $s_2^2 = .002$. Find a 95% confidence interval for $\mu_1 - \mu_2$, the difference of the population mean porosity measurements from the two labs. What conclusion can you draw from your interval?

IV. The Rockwell hardness index for steel is determined by pressing a diamond point into the metal with a specified force and measuring the depth of penetration. Out of a random sample of 50 ingots of a certain grade of steel from one manufacturer, the Rockwell hardness index was greater than 62 in 31 of the ingots.

A) (15) The manufacturer claims that at least 75% of the ingots of this type of it produces steel will have Rockwell hardness index greater than 62. Is there sufficient evidence to refute this claim? Use a test at the $\alpha = .01$ level.
B) (15) Using the rejection region you found in part A, compute the Type II error probability β of your test if it is actually true that 65% of the ingots have hardness index greater than 62.

V. Consider the following measurements of the weights of yields of two breeds of apple trees (in kilograms):

Breed 1 :	80.6	80.3	81.5	80.7	80.4	
Breed 2 :	79.5	79.9	81.0	79.4	79.2	81.4

Assume the measurements come from normal populations.

A) (20) Estimate the *variances* σ_1^2 and σ_2^2 of the yields of the two breeds. Is there evidence to suspect that $\sigma_2^2 > \sigma_1^2$? Explain.
B) (15) Test the null hypothesis $H_0 : \mu_1 = \mu_2$ against the alternative $H_a : \mu_1 \neq \mu_2$. Estimate the p-value of your test and state your conclusion clearly and succinctly.

VI. The following table gives measurements of the amount of sodium chloride that dissolved in 100 grams of water at various temperatures in a chemistry experiment.

x (degrees C)	y (grams)
0	7.3
15	13.0
30	23.3
45	30.7
60	39.7
75	47.7

A) (20) Find the equation of the least squares regression line for this data set.
B) (15) Is there sufficient evidence to say that $\beta_1 > .45$? Explain, using the p-value of an appropriate test.

VII. (20) (A "Thought Question") Does a "statistically significant" result where we reject some H_0 mean that H_0 is *far from being true*? Answer intuitively first. Then answer the following: Suppose each individual we draw from a population has either property A or property B (but not both). Let p be the proportion of the population that has property A. We test $H_0 : p = 1/2$ versus $H_a : p > 1/2$ with large-sample tests with $n = 100, 1000, 10000, 100000$. What must the observed proportion of sampled individuals with property A be in order to reject H_0 at the $\alpha = .05$ level in each case? How does this square with what you said at first?

Extra Credit (20) Recall that in the "Big Theorem" in the multiple regression case, we said that in the entries $c_{ij}\sigma^2$ of the covariance matrix of the least squares estimators, the c_{ij} were the entries of the matrix $(X^t X)^{-1}$. Show this is true by direct computation for the X from a simple linear model of the form $Y = \beta_0 + \beta_1 x + \epsilon$.

Have a safe, enjoyable, and productive summer!