NAME ______________________________

Honors Exam in Statistics

May 2006

Please read carefully:

- This three hour exam consists of 6 questions of equal weight.
- Answer all 6 questions.
- This is a closed book/closed notes exam.
- You may use a calculator that does not do algebra or calculus.
- Use additional answer sheets as needed.
- Label each sheet with the question number and your name at the top.
- Do not write the answer to more than one question on a single page.
- Do not write on the back of any sheet.
- Normal, t, χ^2 and F tables should be supplied with this exam.

Question 1

A roulette wheel in Las Vegas has 38 slots. At one spin of the wheel, a slot is selected at random. Of the 38 slots, 18 are red, 18 are black, and there are two 'zeros'. If you bet 1 dollar on red before a spin and a red slot appears, you get your original dollar and one dollar more. If a black number or a zero is selected, you lose your dollar.

In Atlantic City, the rules are the same except that there are only 37 slots, 18 red, 18 black and only 1 zero. Winnings are the same for a one dollar bet on red.

The gambler in Atlantic City is AC, the gambler in Las Vegas is LV. You may leave answers as a sum as long as the answer is clear.

(a) For a single one dollar bet, how much more does LV expect to lose compared to AC?

(b) In one bet, AC bets 2 dollars, while LV bets 1 dollar. Who expects to lose more money, and by how much?

(c) AC bets 1 dollar twice in a row, while LV bets 1 dollar once. Who expects to lose more money and by how much?

(d) LV decides on a series of bets on red starting with one dollar. If LV wins any bet, then LV quits and goes home. If LV loses, then LV makes another bet on red for twice the amount of the previous bet.

  (i) What is LV's expected winning?

  (ii) What is the probability that LV wins money?

  (iii) Unfortunately LV has only 1023 dollars available. If LV loses 1023 dollars LV must go home. Now what is the probability that LV wins money, and

  (iv) Now what is LV's overall expected winning?

Question 2

Consider random variables

$$\begin{aligned} X &\sim \text{Binomial}(n, \pi), \\ Y &\sim \text{Poisson}(\lambda), \\ Z &\sim \text{N}(\mu, \tau^2). \end{aligned}$$

where $0 < \pi < 1$ is the probability of success in a Bernoulli trial, $n > 0$ is an integer and λ is the mean of the Poisson. Let

$$\lambda = n\pi = \mu = \tau.$$

All three random variables X, Y, and Z have the same mean.

(a) Give the variances of the three random variables. Order the variances from smallest to largest.

(b) Give the probability of being greater than zero for all three random variables. Simplify each probability. Order the three probabilities.

(c) Suppose $n = 1/\pi$. Show that as n gets large $P(X = k)$ approaches $P(Y = k)$ for fixed k.

(d) A possible normal approximation to the Poisson approximates $P(Y = k)$ by $\Phi(k+.5; \mu, \sigma^2) - \Phi(k-.5; \mu, \sigma^2)$ where $\Phi(\cdot; \mu, \sigma^2)$ is the cumulative distribution function for a normal with mean μ and variance σ^2. (i) What choice for μ and σ^2 should you make for this to be reasonable, and (ii) what condition (state, don't prove) on λ makes this reasonable?

Question 3

You plan to survey n Pennsylvanian pregnant women (PPW) and determine whether they consumed alcohol during pregnancy (Yes or No). It is known that nationally 30% of women drink during pregnancy. Use $\alpha = .05$ in the following. We wish to accurately estimate the alcohol consumption proportion in PPW.

(a) We want our estimate to be within .03 of the true value of the proportion. What sample size n is required if PPW are thought to (i) be like pregnant women nationally? (ii) drink more than pregnant women nationally? (iii) drink less than pregnant women nationally?

(b) What sample size n is required to test, with 90% power, whether PPW drink alcohol at a rate 5% more than the national rate?

(c) We collect data from PPW and nationally as well and observe the following results:

	Alcohol	No Alcohol
PPW	11	26
Nationally	6	31

Test whether the proportion of PPW consuming alcohol is greater than that of women nationally. State hypotheses, assumptions, decision rules, conclusions and p-values. Are the assumptions of your test met by this study? Use $\alpha = .05$.

(d) For the above data, what is the 99% confidence interval for the difference in proportions?

Question 4

You study serum (blood) nutrient level in two groups of 10 randomly assigned children, one group untreated, and one group is given a radically new formulation vitamin W pill that is supposed to metabolize into vitamin Z. The following information about Z vitamin serum levels from the study is available.

Group	n	mean	sd
no pill	10	2.20	.96
W pill	10	3.21	1.01

(a) Construct a 99% confidence level for the difference in means.

(b) Given your result in (a), does the vitamin W pill significantly increase Z vitamin serum levels? (yes/no and one very short sentence justification).

(c) It is decided to do a new study where n children have their Z vitamin serum levels tested before and after taking the vitamin Z pill. Use $\alpha = .05$ and power of .9. The correlation of vitamin Z serum levels before and after taking the pill is .5. How large an n do you need? Assume a Z vitamin serum level difference in means equal to one, and a population sd of Z vitamin serum level before or after taking the pill of one.

Question 5

You have $i = 1, \dots, n$ independent bivariate observations (x_i, y_i). We are interested in the regression of y_i on x_i

$$y_i = a + bx_i + \text{error}$$

where the error is assumed to have constant variance. Derive all your answers.

(a) Assuming a is known, what is the least squares estimate of b?

(b) Assuming b is known, what is the least squares estimate of a?

(c) Suppose in the population, $E[x_i] = 7$, $E[y_i] = 3$, $\text{var}[x_i] = \text{var}[y_i] = 2$, and the correlation between x_i and y_i is .75. What are a and b?

(d) Suppose in a sample of observations, $n = 10$, $n\bar{x} = \sum_{i=1}^{n} x_i = 20$, $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 20$, $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 20$, where $\hat{y}_i = \hat{a} + \hat{b}x_i$ is the fitted value from the regression. What are the standard errors of the least squares estimates of a and b?

Question 6

A light bulb from Home Despot has a lifetime y that is distributed exponentially with mean μ and density

$$f(y|\mu) = \frac{1}{\mu}\exp(-\frac{y}{\mu}).$$

We test a sample of light bulbs of size n and find their lifetimes.

(a) Write down $L(\mu|\text{data})$, the likelihood of μ given the data. Derive the maximum likelihood estimate of μ. Derive the asymptotic standard error of your estimate.

(b) Derive the least squares estimate of μ.

(c) Suppose that we collect five observations 1, 2, 2, 4, 6 (in years). What is the likelihood of μ, the maximum likelihood estimate of μ and what is its standard error (i.e. standard deviation estimated from the data)?

(d) Suppose that we originally had tested 10 light bulbs, but after 6 years, and observations of 1, 2, 2, 4, and 6 years, the study ends due to a loss of funds, leaving 5 light bulbs still burning at $y_i = 6$. Let $Z = (1, 2, 2, 4, 6)$ be the life times that we do know.

1. What is the probability $P(y > 6|\mu)$, as a function of μ, that a single light bulb would still be burning at 6 years given a mean of μ?
2. The likelihood of μ given this data on ten light bulbs is $L(\mu|Z) * [P(y > 6|\mu)]^5$. Now what is the maximum likelihood estimate of μ?