STAT516 Final Exam Spring 2006

Name:
Student ID:
Show all work. Explain your answers. Partial credit will be given. There are 5 questions, and a total of 100 points. Points are shown for each question.

1. Let $(X_1, X_2, \cdots, X_k)$ be the vector of sample observations representing a multinomial random variable with parameters n, p_1, p_2, $\cdots$, and p_k.

   (a) (7 points) Write out the joint likelihood of X_1, X_2, $\cdots$, X_k

   (b) (7 points) Show that the maximum likelihood estimate for p_1 is X_1/n

   (c) (7 points) Show that X_1/n is an unbiased estimator for p_1

2. The owner of Fit Forever Health Club is considering adding an indoor swimming pool to his facility. The manager decided to take a survey to determine whether member opinion about the addition of a pool was independent of the age of the member. Two hundred members were selected at random and asked to state their opinion. The following data was recorded:

   Table 1: Opinion Concerning Pool

Age	Favor	Undecided	Oppose
≤ 30	40	20	18
> 30 & ≤ 50	30	25	20
> 50	10	16	21

   (a) (7 points) Under the assumption that a member's opinion concerning the addition of an indoor pool is independent of the member's age, find the probability that a randomly sampled member is under age 30 and favors the addition.

   (b) (9 points) Perform an appropriate hypothesis test for the independence between a member's opinion and the member's age.

3. Let (x_1, Y_1), (x_2, Y_2), $\cdots$, and (x_n, Y_n) be a set of points satisfying the assumptions of the simple linear model.

   (a) (7 points) Show that the estimated regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ at $x = \bar{x}$ is $\bar{Y}$. *Hint: plug in the formula for $\hat{\beta}_0$ and leave $\hat{\beta}_1$ as it is.*

   (b) (7 points) Find the 95% confidence interval for the regression line at $x = \bar{x}$.

Table 2: The regression data

x	30	25	90	60	50	35	75	110	45	80
y	0.3	0.2	5.0	3.0	2.0	0.5	4.0	6.0	1.5	4.0

Table 3: The regression equation is $y = -1.74 + 0.0731x$

	Estimate	Standard deviation	T-ratio	p-value
β_0	-1.7359	0.1882	-9.22	0.000
β_1	0.073099	0.002867	25.50	0.000

4. A scientist is studying the relationship between $x =$ inches of annual rainfall and y=inches of shoreline erosion. One study reported data as shown in Table 2. Table 3 contains part of the regression analysis results by MINITAB, a statistical software.

   (a) (7 points) Write out the formula that is used to compute the estimate of β_1.

   (b) (7 points) Write out the formula that is used to compute the standard deviation of β_1.

   (c) (7 points) The T-ratio for β_1 is the T test statistic for the hypothesis: $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$. Write out the formula that is used to compute the T-ratio.

   (d) (7 points) For the same hypothesis as in (c), the p-value is used to determine whether to reject H_0. Explain how the p-value for β_1 is obtained and draw your conclusion for the hypothesis based on the p-value.

   (e) (7 points) Let σ^2 be the variance of the random error in the simple linear regression model. The estimate of σ based on the above data is $S = 0.2416$. Using the fact that $(n-2)S^2/\sigma^2 \sim \chi^2_{n-2}$, perform an appropriate test for the hypothesis $H_0 : \sigma^2 = 1$ vs $H_1 : \sigma^2 \neq 1$. State your conclusion.

5. Suppose that X and Y have a bivariate normal pdf with $\mu_X = 3$, $\mu_Y = 6$, $\sigma_X^2 = 4$, $\sigma_Y^2 = 10$, and $\rho = 0.5$.

   (a) (7 points) Find $P(5 < Y < 6.5)$.

   (b) (7 points) Find $P(5 < Y < 6.5 | X = 2)$.