MA 486/586-1G (Statistics), Dr Chernov   Final exam   Fri, Apr 25, 2008

Student's name ____________________

Each problem is worth 16 points; all 7 problem are worth 112 points. To get full credit for the test you need to earn 100 points. If you get more than 100 points, it will be for extra credit.

1. Five experimental data points are observed:

$$(7, 6),\ (3, 5),\ (-3, 0),\ (-5, -1),\ (3, 0)$$

Estimate the parameters α and β of the regression line $y = \alpha + \beta(x - \bar{x})$. Compute the sample variances s_x^2 and s_y^2, the sample covariance c_{xy} and the sample correlation coefficient r.
Draw a scatter plot, marking the data points and the regression line.

Answers: $\hat{\alpha} = 2$, $\hat{\beta} = 0.542$, $s_x^2 = 24$, $s_y^2 = 10.5$, $c_{xy} = 13$, $r = 0.82$.

2. (a) Let $x_1, \ldots, x_n$ be a random sample from the distribution with probability density function

$$f(x; \mu, \theta) = \sqrt{\theta/\pi}\; e^{-\theta(x-\mu)^2}$$

where μ and $\theta > 0$ are unknown parameters. Find sufficient statistics.

Answer: $u_1 = \sum x_i$ and $u_2 = \sum x_i^2$.

(b) Let $x_1, \ldots, x_n$ be a random sample from the distribution with probability density function

$$f(x; \theta) = \sqrt{\theta/\pi}\; e^{-\theta x^2}$$

where $\theta > 0$ is an unknown parameter. Find the Rao-Cramer lower bound on the variance of unbiased estimators of θ.

Answer: $\mathrm{Var}\,\hat{\theta} \geq \frac{2\theta^2}{n}$.

3. Daily changes in a stock market have been recorded over a period of 20 days as follows:

$$+15,\ +3,\ +20,\ -12,\ -1,\ -6,\ +13,\ +5,\ +27,\ 0,$$

$$+19,\ -6,\ +6,\ -1,\ +20\ +1,\ +44,\ -16,\ +30,\ ,+2$$

Use the run test to test two hypotheses: one about a trend effect (that toward the end of the period the market drifts upward or downward), and the other hypothesis about a cyclic effect (that advances and declines tend to alternate). Use normal approximation. Find the p-value in both tests.

Answers: $R = 16$, $\mu = 11$, $\sigma^2 = 4.74$, $Z = 2.3$, p-value=0.0107.

4. A computer program supposedly generates an exponential random variable with mean $\mu = 1/2$. The following numbers were produced by this program:

$$1.1,\ 0.2,\ 0.2,\ 1.0,\ 1.9,\ 0.3,\ 0.5,\ 0.4,\ 0.1,\ 0.7$$

Use the Kolmogorov-Smirnov test to test the hypothesis that the program works right. Let $\alpha = 5\%$. Sketch an empirical distribution function. Indicate how you would construct a 95% confidence band around the empirical distribution function.

Answers: $D_n = 0.23$, critical region is $D_n > 0.41$, accept H_0.

[Bonus] Give a formula for computing the p-value of the test and find the p-value approximately.

5. Estimate the main effect, the three two-factor interactions, and the three-factor interaction in a 2^3 factorial design experiment. The data, in the canonical order (see Section 20.6), are: $x_1 = 2$, $x_2 = 2.5$, $x_3 = 2.3$, $x_4 = 1$, $x_5 = 2$, $x_6 = 3$, $x_7 = 2$, and $x_8 = 3$. Construct an approximate $q - q$ plot to see if any of these effects seem to be significantly larger than the others.

Answers: $[A] = 0.15$, $[B] = -0.15$, $[C] = 0.275$, $[AB] = -0.225$, $[AC] = 0.35$, $[BC] = 0.15$, $[ABC] = 0.225$.

6. In the following table, 800 individuals are classified by gender and by whether they answer Yes, No, or Not Sure in a certain poll. Test the null hypothesis that the probabilities of the answers are independent of the gender. Let $\alpha = 0.01$.

Gender	Yes	No	Not Sure
Male	206	131	23
Female	194	189	57

.

Answers: $Q = 17.5$, critical region is $Q > 9.210$, accept H_1.

7. Each of five cars is driven each of three different brands of gasoline. The number of miles per gallon driven for each of $5 \times 3 = 15$ combinations is recorded in the table below.

	Gasoline			
Car	1	2	3	
1	20	14	20	
2	26	25	27	
3	22	16	22	
4	20	16	24	
5	22	19	22	

Compute $SS(E)$, $SS(A)$, $SS(B)$. Test the hypotheses about the relevance of the car and the brand of gasoline for gas mileage, at the level $\alpha = 5\%$.

Answers: $SS(E) = 18$, $SS(A) = 108$, $SS(B) = 70$, $F_A = 12$, $F_B = 15.56$. Critical regions are $F_A > 3.84$ and $F_B > 4.46$. Reject both H_A and H_B.