MA 486/586-2D (Statistics), Dr Chernov — Final exam — Tue, May 2, 2006

Student's name ______________________________

Each problem is 15 points ($7 \times 15 = 105$ points total); full credit is given for 100 points.

1. In a college, students are classified according to the period of the day when they took calculus (morning, afternoon, evening) and the success or failure in the course:

	Morning	Afternoon	Evening
Pass	60	70	20
Fail	20	20	10

Test the hypothesis that the success rate is independent of the time period of the class. Let $\alpha = 0.05$.

Answers: $Q = 1.48$, critical region is $Q > \chi^2_{0.05}(2) = 5.991$, so we accept H_0.

2. 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	25	15	20	
2	27	25	26	
3	22	17	21	
4	19	14	21	
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: $SS(E) = 28$, $SS(A) = 108$, $SS(B) = 70$. Then $F_A = 7.7$ and $F_B = 10$. Critical regions are $F_A > F_{0.05}(4, 8) = 3.84$ and $F_B > F_{0.05}(2, 8) = 4.46$. We reject both H_A and H_B.

3. (a) Describe how you construct an approximate 95% confidence interval for the median given a sample of size $n = 20$. What is the exact confidence level?

(b) Given a random sample of size $n = 25$ one can construct an approximate 90% confidence interval for the third quartile in the form $y_i < \pi_{0.75} < y_j$. Determine i and j via normal approximation. Find the exact confidence level by using Table II.

Answers: (a) the CI is (y_6, y_{15}) at 95.86% confidence level; (b) the CI is (y_{15}, y_{23}) at 93.82% confidence level.

4. The customer service department of a telephone company knows that the median waiting time of incoming customer calls is 70 seconds. The new manager restructures the department in order to reduce the waiting time. After restructuring, a random sample of 12 incoming calls gives the waiting times (in seconds)

$$40, 110, 47, 7, 72, 28, 85, 3, 46, 66, 128, 52$$

Use the Wilcoxon test to test the hypothesis $H_0 : m = 70$ against $H_1 : m < 70$. Determine the p-value. What does the p-value mean? How would you conclude the test depending on the significance level α?

Answers: $W = -34$ and $Z = W/\sqrt{12 \cdot 13 \cdot 25/6} = -1.33$, the p-value is $\Phi(-1.33) = 0.0918$. We accept H_0 if $\alpha < 0.0918$ and we accept H_1 if $\alpha > 0.0918$.

5. Daily changes in a stock market have been recorded over a period of 18 days as follows:

$$+2,\ +16,\ +1,\ +24,\ -10,\ +3,\ -4,\ +11,\ +2$$

$$+8,\ -1,\ +5,\ 0,\ +28,\ +12,\ -4, -7, +3$$

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: $R = 16$, $Z = (16 - 10)/\sqrt{4.235} = 2.92$. For the 'trend effect' the p-value is $\Phi(2.92) = 0.9982$, for the 'cyclic effect' the p-value is $1 - \Phi(2.92) = 0.0018$.

6. Eight experimental data points are recorded along an unknown line:

$$(3,0),\ (0,3),\ (-3,5),\ (-6,10),\ (1,3),\ (5,-3),\ (2,2),\ (6,-4)$$

Estimate the parameters α_1 and β of the regression line $y = \alpha_1 + \beta x$. Draw the scatter plot. Compute the sample variances s_x^2 and s_y^2, the sample covariance c_{xy} and the sample correlation coefficient r.

[Bonus] Test the hypothesis $H_0\colon \beta = -2$ against $H_1\colon \beta > -2$.

Answers: $\hat{\alpha} = 2$, $\hat{\beta} = -1.0982$, the line equation is $y = 3.0982 - 1.0982\,x$, and also $s_x^2 = 16$, $s_y^2 = 20$, $c_{xy} = -17.57$, $r = -0.9823$.

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

$$2.1,\ 0.4,\ 0.8,\ 3.0,\ 1.6,\ 0.2,\ 0.4,\ 0.5,\ 1.2,\ 0.7$$

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.

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

Answers: $D = 0.2297$, the critical region is $D > d = 0.41$, so we accept H_0. The p-value is ≈ 0.67.