MA 486-25 (Statistics), Dr Chernov — Final exam — Thu, May 25, 2000

Student's name ____________________

Full credit is given for 100 points. The maximum total in this test is 110 points.

1. (14 pts) Six experimental data points are observed:

$$(-2, 1),\ (1, 0),\ (1, -2),\ (2, -2),\ (2, -1),\ (6, -5)$$

Compute the Gaussian brackets $[X]$, $[Y]$, $[X^2]$, $[XY]$, $[Y^2]$. Estimate the parameters α and β of the regression line $y = \alpha + \beta x$. Compute the sample variances s_x^2 and s_y^2, the sample covariance c_{xy} and the sample correlation coefficient r. Compute the RSS (Residual Sum of Squares).

Draw a scatter plot, marking the data points and the regression line.

Some answers:

$$[X] = 10,\ [Y] = -9,\ [X^2] = 50,\ [XY] = -40,\ [Y^2] = 35$$

$$\hat{\alpha} = -0.75, \quad \hat{\beta} = -0.25, \quad RSS = 2.75$$

$$s_x^2 = 6.67, \quad s_y^2 = 4.3, \quad c_{xy} = -5, \quad r = -0.9338$$

2. (12 pts) The following numbers were generated by a computer program:

$$-1.12,\ 0.43,\ -2.61,\ 0.02,\ 1.46,\ -0.13,\ 0.57,\ 0.14,$$

Test the hypothesis that the program generates a standard normal random variable $N(0,1)$, i.e. $X = N(0,1)$. Use the Kolmogorov-Smirnov test with Table VIII at the level $1-\alpha = 95\%$. Also, sketch an empirical distribution function.

Indicate how you would construct a 95% confidence band around the empirical distribution function.

Some answers:

The distribution function of $N(0,1)$ is $\Phi(x)$, whose values are given by Table V. So, we find $\Phi(-2.61) = 0.0045$, $\Phi(-1.12) = 0.1314$, etc.

The test statistic $D_n = 0.1983$. Since $D_n < d_{0.05} = 0.46$, we accept H_0.

3. (12 pts) (a) For 8 randomly selected adult persons in Alabama, the annual income (in thousands of dollars) is given by:

$$28.9,\ 45.3,\ 104.6,\ 51.4,\ 13.7,\ 39.2,\ 28.4,\ 51.6$$

Find an approximate 93% confidence interval for the median, m, of the annual income of Alabama residents. Use Table 10.1. What is the exact confidence level of your interval?

(b) For the previous sample, find the probability

$$P(13.7 < \pi_{0.25} < 45.3)$$

(c) For a random sample of size $n = 1200$ from an unknown distribution, find an approximate 99% confidence interval for the third quartile, $\pi_{0.75}$. Use normal approximation. Give the answer in the form $(y_{\dots}, y_{\dots})$. Describe in words how you would find this interval if you were given a sample of 1200 random numbers.

Some answers:

(a) $(y_2, y_7) = (28.4, 51.6)$. The exact level is 92.96%.

(b) $P(y_1 < \pi_{0.25} < y_5) = P(1 \le b(8, 0.25) \le 4) = 0.9727 - 0.1001 = 0.8726$

(c) $900 + 0.5 \pm 2.576\sqrt{1200 \times 0.75 \times 0.25} = 862, 939$, so the interval is (y_{862}, y_{939}).

4. (12 pts) Each of three cars is driven each of five different brands of gasoline. The number of miles per gallon driven for each of $3 \times 5 = 15$ combinations is recorded in the table below.

	Gasoline					
Car	1	2	3	4	5	
1	22	21	20	26	21	
2	22	19	25	27	22	
3	19	14	15	25	17	

Compute $SS(E)$, $SS(A)$, $SS(B)$. Test the hypotheses

$$H_A : \alpha_1 = \alpha_2 = \alpha_3$$

(i.e. the car makes no difference) and

$$H_B : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

(i.e. the brand of gasoline makes no difference) at the level $\alpha = 1\%$.

Some answers:

$$SS(E) = 28, \quad SS(A) = 70, \quad SS(B) = 108$$

$$F_A = \frac{70/2}{28/8} = 10$$

$$F_B = \frac{108/4}{28/8} = 7.71$$

Since $F_A > F_{0.01}(2, 8) = 8.65$, we reject H_A.

Since $F_B > F_{0.01}(4, 8) = 7.01$, we reject H_B.

5. (12 pts) Test the hypothesis $H_0 : m_X = m_Y$ against $H_1 : m_X < m_Y$. The following data were observed:

$$X : \ 2, \ -1, \ 4, \ 0, \ 6, \ -1, \ 2, \ -3$$

$$Y : \ 3, \ -2, \ 7, \ 1, \ 5, \ 3$$

(a) Use the median test, compute the p-value (use the exact formula for the probabilities of the V-statistic).
(b) Use the Wilcoxon test (for two samples), again compute the p-value.

Which hypothesis would you accept at the 10% level?

Some answers:

$n_1 = 8, n_2 = 6, n_1 + n_2 = 14, \ k = 7.$

(a) $V = 5$ and the p-value is

$$\frac{C_{8,5}C_{6,2} + C_{8,6}C_{6,1} + C_{8,7}C_{6,0}}{C_{14,7}} = 0.296$$

(b) $W = 53$ and

$$Z = \frac{53 - 6(8+6+1)/2}{\sqrt{48 \times 15/12}} = 1.03$$

so the p-value is

$$1 - \Phi(1.03) = 0.15$$

Accept H_0.

6. (12 pts) Daily changes in a stock market have been recorded over a month period (i.e., during 20 working days) as follows:

$$+28,\ -6,\ +40,\ +2,\ +15,\ -38,\ -2,\ -16,\ -52,\ +16,$$

$$+1,\ +35,\ -14,\ +10,\ +4,\ +27,\ -12,\ -1,\ +3,\ +23$$

Use the run test to test two hypotheses: one about a trend effect (that toward the end of the month 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.

Compute the exact probability $P(R = 16)$, where R is the number of runs.

Some answers:

$R = 13$, $ER = 11$, $\operatorname{Var} R = 4.74$

$$Z = \frac{13 - 11}{\sqrt{4.74}} = 0.91$$

The p-values for the trend test is

$$\Phi(0.91) = 0.8186$$

The p-values for the cycle test is

$$1 - \Phi(0.91) = 0.1814$$

$$P(R = 16) = \frac{2\, C_{9,7}\, C_{9,7}}{C_{20,10}} = 0.014$$

7. (10 pts) Let X_1, X_2, X_3, X_4 be four independent random variables that have some normal distributions $N(\mu_i, \sigma^2)$. Test the hypothesis

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

at the level $\alpha = 5\%$. The observed data are given in the table below:

X_1 :	4	7	7	
X_2 :	13	8	9	
X_3 :	5	7	3	
X_4 :	6	5	9	8

(Note that not all the samples have the same length!) Construct an ANOVA table (without p-value) and state your conclusion.

Some answers:

The grand mean is 7. Then, $SS(E) = 38$, $SS(T) = 42$,

$$F = \frac{42/3}{38/9} = 3.3$$

Since $F < F_{0.05}(3, 9) = 3.86$, we accept H_0.

8. (12 pts) The customer service department of a telephone company knows that the median waiting time of incoming customer calls is 210 (in 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)

$$330, 140, 20, 215, 85, 255, 5, 140, 200, 380, 160, 120$$

Test the hypothesis $H_0 : m = 210$ against $H_1 : m < 210$.

(a) First, use the sign test, find the p-value. [Use the exact binomial distribution.]

(b) Then use the Wilcoxon test, again find the p-value.

Which hypothesis would you accept?

Some answers:

(a) $R = 8$, the p-value is

$$P(b(12, 0.5) \geq 8) = 0.1938$$

(b) $W = -34$, $Z = -1.33$, the p-value is

$$1 - \Phi(1.33) = 0.0982$$

9. (14 pts) In a regression problem, $n = 32$ data points are observed and the following values are found for the Gaussian brackets:

$$[X] = 4, \ [Y] = 2, \ [X^2] = 13, \ [XY] = 4, \ [Y^2] = 5$$

Find $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma^2}$. Find 98% confidence intervals for α, β, and σ^2. Test the hypothesis $\beta = 0$ against the alternative $\beta > 0$ at 99% level.

Find a 90% prediction interval for Y when $x = 0.125$.

Some answers:

$$\hat{\alpha} = 0.025, \qquad \hat{\beta} = 0.3, \qquad \hat{\sigma^2} = 0.117$$

Confidence interval for α:

$$0.025 \pm 2.457\sqrt{0.117/30} = [-0.128, 0.178]$$

Confidence interval for β:

$$0.3 \pm 2.457\sqrt{3.74/30(13 - 16/32)} = [0.054, 0.54]$$

Confidence interval for σ^2:

$$[32 \times 0.117/50.89, 32 \times 0.117/14.95] = [0.074, 0.25]$$

Test statistic: $T = 3.00$. Since $T > 2.457$, we accept the alternative $\beta > 0$.
The prediction interval is

$$0.025 + 0.3 \times 0.125 \pm 1.697 \cdot \sqrt{32 \times 0.117/30} \cdot \sqrt{1 + 1/32 + 0} = [-0.546, 0.671]$$