MA 486-1E (Statistics), Dr Chernov Final exam Fri, May 2, 2003

Student's name ______________________________

Full credit is given for 100 points. The maximum total in this test is 108 points.

1. (12 pts) Five experimental data points are observed:

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

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.

Partial answers: $\hat{\alpha} = 2.2$, $\hat{\beta} = -0.331$, $s_x^2 = 14.8$, $s_y^2 = 1.7$, $c_{xy} = -4.9$, and $r = -0.977$.

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

$$-0.1,\ 0.9,\ -1.2,\ -0.6,\ 2.6,\ 0.5,\ -3.1,\ 1.2$$

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.

Partial answers: $D_n = \max|\Phi(x) - F_n(x)| = 0.1915$ occurs at $x = 0.5$. Since $0.1915 < d_n = 0.46$ (found from Table VIII), we accept H_0.

3. (12 pts) (a) Given a random sample

$$2.0,\ 11.2,\ 1.5,\ -6.0,\ 1.7,\ 2.3,\ -8.5,\ 4.6,\ 1.4,\ 2.1.\ 1.2$$

Find an approximate 94% confidence interval for the median, m, of the corresponding distribution. What is the exact confidence level of your interval?

Answer: CI is $[y_3, y_9] = [1.2, 2.3]$.

(b) For the previous sample, find the probability

$$P(1.4 < \pi_{0.65} < 11.2)$$

Answer:

$$P(y_4 < \pi_{0.65} < y_{11}) = P(4 \leq b(11, 0.65) < 11) = 0.979$$

4. (12 pts) A random sample of 200 students were classified by gender and by the instrument that they played. Test whether the selection of instrument is independent of the gender of the respondent. Let $\alpha = 0.05$.

Gender	Yes	No	Not Sure
Female	40	30	50
Male	20	30	30

Answer:

$$Q = \frac{(40 - 0.4 \times 120)^2}{0.3 \times 120} + \cdots + \frac{(30 - 0.4 \times 80)^2}{0.4 \times 80} = 3.82$$

This is less than $\chi^2_{0.05}(2) = 5.991$, hence we accept H_0.

5. (12 pts) Daily changes in a stock market have been recorded over a period of 16 days as follows:

$$-5,\ +11,\ +26,\ -1,\ +4,\ -2,\ +7,\ +1,$$
$$+10,\ -5,\ +2,\ -11,\ +22,\ 0,\ +15,\ +1$$

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.

Partial answers: $R = 15$ and

$$Z = \frac{15-9}{\sqrt{3.733}} = 3.105$$

For the trend test, the p-value is $\Phi(3.105) \approx 0.999$.
For the cyclic effect test, the p-value is $1 - \Phi(3.105) \approx 0.001$.

6. (12 pts) Test the hypothesis $H_0 : m_X = m_Y$ against $H_1 : m_X \neq m_Y$. The following data were observed:

$$X : \; 8, \; 15, \; 7, \; 11, \; 20, \; 18, \; 10, \; 5$$

$$Y : \; 13, \; 6, \; 9, \; 15, \; 5, \; 6$$

Use the Wilcoxon test (for two samples). Let $\alpha = 10\%$.

(Bonus) Compute the p-value. Sketch the q-q plot.

Partial answers: $W = 37$ and

$$Z = \frac{37 - 45}{\sqrt{60}} = -1.033$$

The critical region for the test is $|Z| > z_{\alpha/2} = 1.645$. We accept H_0.
The p-value is $P(|Z| > 1.033) = 2(1 - \Phi(1.033))$.

7. (12 pts) Estimate the main effect, the three two-factor interactions, and the three-factor interaction in a 2^3 factorial design experiment. The data, in standard order, are: $x_1 = 6$, $x_2 = 4$, $x_3 = 2$, $x_4 = 5$, $x_5 = 3$, $x_6 = 4$, $x_7 = 1$, 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.

Partial answers: $[A] = 0.5$, $[B] = -0.75$, $[C] = -0.75$, $[AB] = 0.75$, $[AC] = 0.25$, $[BC] = 0$, and $[ABC] = -0.5$.

8. (12 pts) In a regression problem, $n = 20$ data points are observed and the following accumulated values are found:

$$\sum x_i = 40,\ \sum y_i = 25,\ \sum x_i^2 = 120,\ \sum x_i y_i = 60,\ \sum y_i^2 = 80$$

Find $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma^2}$. Find 95% confidence intervals for α, β, and σ^2.

Find a 98% prediction interval for $E(Y)$ when $x = 8$.

Partial answers: $\hat{\alpha} = 1.25$, $\hat{\beta} = 0.25$, $\hat{\sigma^2} = 2.31$. The confidence intervals are: for α [0.5,2.0], for β [-0.28,0.78], and for σ^2 [1.5,5.6].

9. (12 pts) Let X_1, X_2, X_3 be three independent random variables that have some distributions with mean values μ_1, μ_2, μ_3 respectively. Test the hypothesis

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

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

X_1 :	15	12	15			
X_2 :	9	12	10	9		
X_3 :	8	6	9	8	8	9

Construct an ANOVA table (without p-value) and state your conclusion.

Partial answers: SS(TO)=90, SS(T)=72, SS(E)=18. The F-ratio is $F = 20$. This is greater than $F_{0.05}(2, 10) = 4.10$, hence we accept H_1.