MA 486-1E (Statistics), Dr Chernov Final exam Mon, May 3, 2004

Student's name ____________________

1. (10 pts) Given a random sample

$$4.0,\ 6.7,\ 3.5,\ 6.1,\ -4.2,\ 6.1,\ 28.3,\ -1.9,\ 19.0,\ 5.8,\ 3.9,\ 3.88,\ 6.0,$$

(a) find a point estimate for the median, m;

Answer: 5.8.

(b) find point estimates for the quartiles, $\pi_{0.25}$ and $\pi_{0.75}$;

Answers: 3.69 and 6.4.

(c) find an approximate 98% confidence interval for the median, m;

Answer: $(y_3, y_{11}) = (3.5, 6.7)$.

(Bonus) Find $P(4.0 < \pi_{0.7} \leq 19.0)$;

Answer: 0.9181.

(Bonus) Find an upper and lower bound on the probability $P(0 < m < 10)$.

Answers: $P < P(y_2 < m < y_{12}) = 0.9966$ and $P > P(y_3 < m < y_{11}) = 0.9776$.

2. (10 pts) Let X have an exponential distribution function $F(x) = 1 - e^{-\lambda x}$ with parameter $\lambda > 0$. Let $X_1, X_2, \ldots, X_n$ be a random sample from this distribution. Use Neyman-Pearson lemma to find the best critical region for testing H_0: $\lambda = 2$ against H_1: $\lambda = 8$. The best critical region must be given by an inequality that contains $X_1, \ldots, X_n$ (but not λ) and an unknown critical value c. Explain how to find c if the level of the hypothesis, α, is given.

Answer: the likelihood function is

$$L(\lambda) = \lambda^n e^{-\lambda \sum x_i}$$

hence the best critical region is

$$\frac{L(2)}{L(8)} \leq \text{const}$$

or

$$e^{6 \sum x_i} \leq \text{const}$$

or

$$\sum_{i=1}^{n} x_i \leq c$$

To find c, one needs o solve the equation

$$P(\sum x_i \leq c;\ \lambda = 2) = \alpha$$

3. (15 pts) In a linear regression problem $y = \alpha + \beta(x - \bar{x})$, 10 data points are observed and the following accumulated values are obtained:

$$\sum x_i = 20, \ \sum y_i = 15, \ \sum x_i^2 = 50, \ \sum y_i^2 = 75, \ \sum x_i y_i = 60$$

(a) Find $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma^2}$.
(b) Find 90% confidence intervals for α, β, and σ^2.
(c) Test the hypothesis H_0: $\beta = 0$ against H_1: $\beta > 0$ at the level 99%. What is the approximate p-value?
(Bonus) Find a 95% prediction interval for $E(Y)$ when $x = 4$.

Note: the data in this problem contained an error; the resulting estimate for σ^2 turned out to be negative (which must not happen). The student were required to take its absolute value and proceed.

Answers: point estimates are

$$\hat{\alpha} = 1.5, \qquad \hat{\beta} = 3, \qquad \hat{\sigma^2} = -3.75 \mapsto 3.75$$

Confidence intervals are

$$\alpha \in (0.23, 2.77), \qquad \beta \in (1.73, 4.27), \qquad \sigma^2 \in (2.42, 13.72)$$

Test statistic is $T = 4.38$, the p-value is ≈ 0.
The prediction interval is $(3.97, 11.03)$.

4. (15 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	21	26	24	21	23
2	17	24	17	19	18
3	19	19	19	17	16

Apply the two factor analysis of variance. 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

$$H_B : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$$

(i.e. the brand of gasoline makes no difference) at the level $\alpha = 5\%$.

Answers: $SS(A) = 70$, $SS(B) = 36$, $SS(E) = 24$, $F_A = 11.67$, $F_B = 3$. We reject H_A but accept H_B.

5. (15 pts) The following numbers were generated by a computer program:

$$-0.7,\ 2.9,\ 2.2,\ -2.6,\ 3.4,\ 0.2,\ -3.1,\ 3.2,\ 4.8,\ 2.5.$$

We need to test the hypothesis that the program generates a standard normal random variable $N(0,1)$, i.e. $X = N(0,1)$.
(a) Use the Kolmogorov-Smirnov test at the level $1-\alpha = 80\%$.

Answer: $D = 0.5861 > 0.32 = d_{10}$, hence we reject the hypothesis.

(b) Sketch an empirical distribution function.

(c) If you had $n = 100$ random values in this problem, what would be the critical value for the test?

Answer: $c = 1.07/\sqrt{100} = 0.107$.

(Bonus) Is there any formula for computing the p-value of the Kolmogorov-Smirnov test?

Answer: yes, p-value= $Q(D_n\sqrt{n})$, where

$$Q(s) = 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2s^2}.$$

6. (10 pts) Two samples from random variables X and Y were recorded:

$$X: \ 6, \ 9, \ 3, \ 8, \ 10, \ 9, \ 6.5, \ 4, \ 12, \ 3$$

$$Y: \ 5, \ 7, \ 1, \ 4.5, \ 2, \ 8.7$$

Use the Wilcoxon test (for two samples) to test the hypothesis $H_0 : m_X = m_Y$ against $H_1 : m_X > m_Y$. Compute the p-value.

Answers: $W = 38$, $Z = -1.41$, p-value $=0.0793$.

7. (10 pts) Two samples from random variables X and Y were observed:

$$X: \ 14, \ 11, \ 3, \ 12, \ 2, \ 11, \ 5, \ 17, \ 12, \ 4, \ 2, \ 11$$

$$Y: \ 6, \ 18, \ -2, \ 7, \ 0, \ 19, \ -5, \ 9, \ 6, \ 20, \ -1, \ 7$$

Use the run test to test the hypothesis $H_0: F_X = F_Y$ against $H_1: F_X \neq F_Y$. Compute R, find $E(R)$ and $\text{Var}(R)$, compute the Z-value and finally the p-value. Which hypothesis would you accept?

Answers: $R = 5$, $Z = -3.34$, p-value=0.0004.

[Bonus] Compute the exact probabilities $P(R = 6)$ and $P(R = 7)$.

Answers: 0.00224 and 0.00671.

8. (15 pts) The median score in freshmen biology tests is used to be 60 (out of 100). A new method of teaching the biology course is proposed. It is used in an experimental class of 20 students, and the scores in this class are

$$72,\ 61,\ 43,\ 55,\ 88,\ 34,\ 67,\ 99,\ 62,\ 54,$$
$$62,\ 81,\ 49,\ 93,\ 80,\ 72,\ 50,\ 79,\ 55,\ 89.$$

Use the Wicoxon test to test the hypothesis $H_0 : m = 60$ against $H_1 : m > 60$. Find the p-value. Which hypothesis would you accept?

Answers: $W = 90$, $Z = 1.68$, p-value=0.0465.