EASTERN UNIVERSITY, SRI LANKA

THIRD EXAMINATION IN SCIENCE - 2005/2006

(Mar./Apr.' 2008)

SECOND SEMESTER

ST 304 - DATA ANALYSIS

(Proper and Repeat)

Answer all questions

(Q1. and Q2. using MINITAB) and

(Q3. and Q4. using SAS)

Time : Two hours

Q1. (a) The average number of annual trips per family to amusement parks in the United States is Poisson distributed, with a mean of 0.6 trips per year. What is the probability that randomly selected American family

    i. did not make a trip to an amusement park last year?

    ii. took exactly one trip to an amusement park last year?

    iii. took two or more trips to amusement parks last year?

    iv. took three or fewer trips to amusement parks over a three - year period?

    v. took exactly four trips to amusement parks during a six - year period?

(b) Enter the names of 10 people recorded as $a, b, c, d, e, f, g, h, i$ and j.

    i. Assign numbers 1 to 10 for the names a to j respectively. Separate names that have received odd numbers.

    ii. Arrange 10 names in a random sequence.

1

(c) i. Create 25 random integers (0 to 9).

ii. Calculate proportion of integers greater than 5 in the above 25 integers.

Q2. (a) Current Construction Reports from the U.S. Census Bureau contain data on new privately owned housing units. Data on new privately owned housing units (1,000s) built in the West between 1970 and 1997 as follows. Use these time - series data to develop an autoregression model with a one - period lag. Now try an autoregression model with a two - period lag. Discuss the results and compare the two models.

Year	Housing Starts (1,000)	Year	Housing Starts (1,000)
1970	311	1984	436
1971	486	1985	468
1972	527	1986	483
1973	429	1987	420
1974	285	1988	404
1975	275	1989	396
1976	400	1990	329
1977	538	1991	254
1978	545	1992	288
1979	470	1993	302
1980	306	1994	351
1981	240	1995	331
1982	205	1996	361
1983	382	1997	364

(b) The mean price of corn sold in September 2005 was $2.00 per bushel. A sample of 22 transactions for the sale of corn in September 2006 has a mean $2.08 and standard deviation $0.20. Assume the price of corn is normally distributed. Has the mean price of corn changed over the past year at a 5% level of significance?

2

Q3. A survey was conducted and data were collected and coded. The data layout is shown below (all values are numeric):
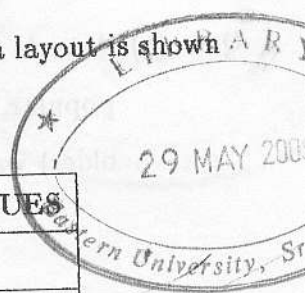
VARIABLE	DESCRIPTION	COLUMNS	CODING VALUES
ID	Subject identifier	1 - 3	
SEX		4	1=Male 2=Female
PARTY	Political Party	5	1=Republican 2=Democrat 3=Not registered
VOTE	Did you vote in the last election?	6	0=No 1=Yes
FOREIGN	Do you agree with the government's foreign policy?	7	0=No 1=Yes
SPEND	Should we increase domestic spending?	8	0=No 1=Yes

Collected data are shown below:

00711110

01322101

13721001

11711111

42813110

01723101

03712101

(a) Create a SAS data set, complete with labels and formats for this questionnaire.

(b) Generate frequency counts for the variables SEX, PARTY, VOTE, FOREIGN and SPEND.

(c) Test if there is a relationship between voting in the last election and agreement with spending and foreign policy.

3

Q4. (a) A taste test was conducted to rate the preference between brands C and P of a popular soft drink. In addition, the age category (1 = less than 20, 2 = 20 or older) was recorded. Preference data (on a scale of 1-10) are displayed below:

		Brand	
		C	P
		7	9
		6	8
	1	6	9
		5	9
		6	8
AGE CATEGORY			
		9	6
		8	7
	2	8	6
		9	6
		7	5

i. Write a SAS program to analyze these data with a two - way analysis of variance.

ii. Draw an interaction graph.

iii. Follow up with a t-test comparing brand C to brand P for each age group separately.

(b) For the data given below, perform a t-test in order to test $H_0 : \mu = 70$ vs $H_a : \mu \neq 70$ and interpret your results.

72, 69, 80, 73, 76, 68, 71, 75, 74, 73

4