PRINCE SULTAN UNIVERSITY
Department of Mathematical Sciences
Final Examination
Second Semester 2007–2008 (072)
STAT 271

Student Name:			
Student Number:		Section Number:	
Teacher Name:		Attendance Number:	

Mark
___
100

- Time allowed is 2½ hours.
- Write down your answer in the space provided underneath the question.
- You may use a programmable calculator and/or your formula sheet.
- Use $\alpha = 0.05$ if not specified.
- Some critical values of Z-distribution:

$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.285	1.645	1.96	2.325	2.575

- Some critical values of t-distribution:

df =5	
$t_{0.05}$	$t_{0.025}$
2.015	2.571

df =7	
$t_{0.05}$	$t_{0.025}$
1.895	2.365

- Some critical values of F-distribution:

df_1=2	df_2=9
$F_{0.05}$=4.26	

df_1=1	df_2=7
$F_{0.05}$=5.59	

- Some critical values of Chi-Square distribution:

df =3
$\chi^2_{0.05}$=7.81473

Question:	**Mark:**
1	
2	
3	
4	
5	
Total	

Term Mark Out of 60	Final Exam Mark Out of 40	Total Mark Out of 100	Grade

Question 1: (20 marks)

Six water specimens taken from a river at a specific location during the low-water season gave readings of 4.9, 5.1, 4.9, 5.0, 5.0, and 4.7 (ppm) of dissolved oxygen. Assume that this data represents a random sample from a normal distribution with unknown variance σ^2.

(I) Does the data provide sufficient evidence to indicate that the mean of dissolved oxygen content (μ) is less than 5.0 ppm? Use $\alpha = 0.05$ to conduct a small-sample test of hypotheses. Arrange your answer as follows:

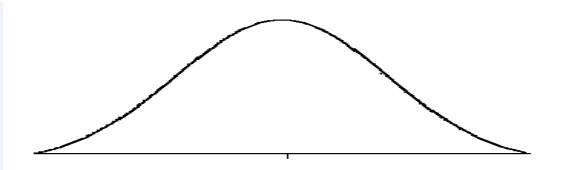
(a) The null and the alternative hypotheses: [4 marks]

H_o :

H_1 :

(b) The test statistic: [4 marks]

(c) The rejection region: [4 marks]



(d) The conclusion: [4 marks]

(II) Construct a 95% small-sample confidence interval for the mean of dissolved oxygen content (μ). [4 marks]

Question 2: (24 marks)

(I) Two random samples of sizes $n_1 = 50$ and $n_2 = 60$ were selected independently from two populations with variances $\sigma_1^2 = 16$ and $\sigma_2^2 = 25$, respectively. The sample means of theses random samples are $\overline{X}_1 = 55$ and $\overline{X}_2 = 53$, respectively. Does this data indicate that there is a difference between the means of the two populations? Use α=0.05. Arrange your answer as follows:

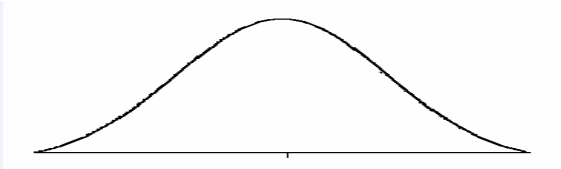
(a) The null and the alternative hypotheses: [4 marks]

H_o :

H_1 :

(b) The test statistic: [4 marks]

(c) The rejection region: [4 marks]



(d) The conclusion: [4 marks]

(II) It is claimed that 35% of the households in a certain country own at least one car. In a random sample of 250 households, 80 households own at least one car. Does this data provide sufficient evidence to indicate that the proportion of households with at least one car is different from 0.35? Use α=0.05. Explain and justify your answer. [8 marks]

Question 3: (12 marks)

The Graduate Record Examination (GRE) scores were recorded for four students admitted to each of three graduate programs (A, B, and C).

Graduate Program		
A	B	C
619	710	549
548	690	542
627	640	607
509	590	524

A completely randomized design has been used for this study. The ANOVA table of this study follows:

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups		2	11664.75		0.0397	4.2565
Within Groups		9	2474.0833	XXXXXXX	XXXXXXX	XXXXX
Total			XXXXXXX	XXXXXXX	XXXXXXX	XXXX

(a) Complete the ANOVA table above. (SST, SSE, SSTOT, df_{SSTOT}, and F–ratio) [4 marks]

(b) Does this data provide sufficient evidence to indicate a difference in the mean GRE scores for applicants admitted to the three programs? Use α=0.05. Explain and justify your answer. [8 marks]

Question 4: (24 marks)

To test a subject's ability to estimate sizes, he was shown 9 different objects and asked to estimate their length or diameter. The objects were then measured, and the results were recorded in the table below.

Object	1	2	3	4	5	6	7	8	9
X (Actual)	9.25	4.25	6.75	3.75	15.75	5.0	6.0	41.5	10.25
Y (Estimated)	10.0	4.0	7.5	2.75	14.5	3.75	7.0	42.0	9.5

Summary statistics:

$$\sum X = 102.5 \ , \ \sum Y = 101 \ , \ \bar{X} = 11.3889 \ , \ \bar{Y} = 11.2222$$

$$\sum XY = 2299.94 \ , \sum X^2 = 2299.63 \ , \ \sum Y^2 = 2307.38$$

$$S_{XY} = 1149.662 \ , \ S_{XX} = 1132.269 \ , \ S_{YY} = 1173.936$$

Assume the relationship between Y and X is given by the following simple linear regression model: $Y = \alpha + \beta X$. The ANOVA table is given below:

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression		1167.323		1236.620	0.000000004
Residual		6.60773		XXXXXX	XXXXXX
Total	8	1173.931	XXXXXX	XXXXXX	XXXXXX

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	??????	0.4615	-0.7402	0.4832	-1.4330	0.7497
X	??????	0.0289	35.1656	0.000000004	0.9471	1.0836

(a) Complete the ANOVA table above. (df_{SSR}, df_{SSE}, MSR, and MSE) [4 marks]

(b) Find the least-squares estimate of α and β. [4 marks]

(c) Write down the estimated least-squares line (prediction equation). [2 marks]

(d) Use the prediction equation to predict the estimated length (Y) if the actual length is X=8. [2 marks]

(e) Find a 95% confidence interval for β. [2 marks]

(f) Test $H_o : \beta = 0$ against $H_1 : \beta \neq 0$ (use $\alpha = 0.05$). [2 marks]

(g) Does the data present sufficient evidence to indicate that Y and X are linearly related? Use α=0.05. Explain and justify your answer. [2 marks]

(h) Calculate the value of the coefficient of determination (R^2). [2 marks]

(i) Interpret the value of the coefficient of determination (R^2) obtained in part (h). [2 marks]

(j) Calculate the value of the coefficient of correlation (r) between X and Y. [2 marks]

Question 5: (20 marks)

A freeway with four lanes (A, B, C, and D) in each direction was studied to see whether drivers prefer to drive on the inside lanes (lanes B and C). A total of 500 automobiles were observed during heavy early-morning traffic, and the number of cars in each lane was recoded.

Lane	A	B	C	D
Number of cars (observed counts)	119	147	138	96

(Hint: If there is no lane preference, then $p_1 = p_2 = p_3 = p_4 = 1/4$).

(1) If there is no lane preference, calculate the expected number of cars (expected count) for each lane, and fill in the following table: [4 marks]

Lane	A	B	C	D
Observed counts (O_i)	119	147	138	96
Expected counts (E_i)				

(2) Does the data present sufficient evidence to indicate that drivers have lane preference (i.e., test the hypothesis that there is a lane preference) at the $\alpha = 0.05$ level of significance. Arrange your answer as follows: [4 marks for each part]

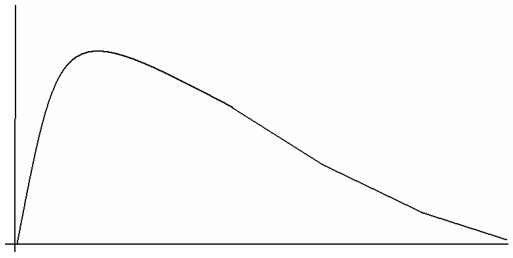
(a) The null and the alternative hypotheses:

H_o :

H_1 :

(b) The test statistic:

(c) The rejection region:



(d) The conclusion: