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

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

Mark
___ 40

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

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

Attempt 5 questions.

Question 1:

A company specializing in fast food products offers a sandwich in five different styles (A, B, C, D, and E). A random sample of $n = 500$ sales has produced the following data:

Style	A	B	C	D	E
Number sold (observed counts)	124	96	112	78	90

(Hint: If there is no style preference, then $p_1 = p_2 = p_3 = p_4 = p_5 = 1/5$).

(1) If there is no style preference, calculate the expected number sold (expected count) for each style, and fill in the following table:

Style	A	B	C	D	E
Observed counts (O_i)	124	96	112	78	90
Expected counts (E_i)					

(2) Test the hypothesis that there is no style preference at the $\alpha = 0.05$ level of significance. (Your discussion should include the null and the alternative hypotheses, the value of the test statistic, the rejection region, and your conclusion.)

Question 2:

900 people were interviewed to study the relationship between their sex and smoking habit. The study has produced the following data:

smoking habit

		Daily	Sometimes	never
Sex	Male	150	100	200
	Female	50	100	300

(1) If the smoking habit is independent of the sex of the person interviewed, calculate the expected number of people (expected count) for each cell, and fill in the following table:

smoking habit

		Daily	Sometimes	never
Sex	Male			
	Female			

(2) Is there sufficient evidence to indicate that the smoking habit is dependent on (not independent of) the sex of the person interviewed? Use $\alpha = 0.05$ level of significance. (Your discussion should include the null and the alternative hypotheses, the value of the test statistic, the rejection region, and your conclusion.)

Question 3:

A study was conducted to determine the effects of absence on students' achievement in a particular Math course. A total of 10 students participated in the study. At the end of the semester, the number of absences (X) and the final mark (Y) of each student were recorded. These results were obtained:

X	0	0	2	3	3	4	5	5	6	7	8	8	9	10	12	12
Y	95	80	85	70	80	75	65	70	65	75	55	60	55	50	55	40

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

Use Appendix (A) to solve the following questions.

(a) Find the least-squares estimate of α and β.

(b) Write down the estimated least-squares line (prediction equation).

(c) Use the prediction equation to predict the final mark for a student who has been absent for 11 times.

(d) Do the data present sufficient evidence to indicate that Y and X are linearly related? Explain your answer.

(e) Test $H_o: \beta = 0$ against $H_a: \beta \neq 0$ (use $\alpha = 0.05$).

(f) Find a 95% confidence interval for β.

(g) Calculate the coefficient of correlation (r) between X and Y.

(h) Calculate and interpret the coefficient of determination (R^2).

Question 4:

In order to study the relationship of advertising and capital investment on corporate profits, the following data, recorded in thousands of Riyals collected for ten medium-sized firms within the same year. The variable Y represents profit for the year, X_1 represents capital investment, and X_2 represents advertising expenditure.

Y	150	160	100	30	120	10	160	180	130	50
X_1	250	10	60	300	290	200	120	150	60	160
X_2	40	50	30	10	20	0	40	50	40	20

We analyzed these data using Excel based on two models; model (1): $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, and model (2): $y = \beta_0 + \beta_2 X_2$. The Excel printout is given in Appendix (B: B1 and B2).

(a) For model (1), find and interpret the value of the coefficient of determination.

(b) For model (1), test $H_o : \beta_1 = \beta_2 = 0$ against $H_a : \beta_j \neq 0, \textit{ for some } j$. (use $\alpha = 0.05$).

(c) For model (1), test $H_o : \beta_1 = 0$ against $H_a : \beta_1 \neq 0$. (use $\alpha = 0.05$).

(d) For model (1), test $H_o : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$. (use $\alpha = 0.05$).

(e) Which model do you prefer; model (1) or model (2)? Why?

(f) For the model you have chosen in part (e), write the prediction equation (estimated equation) relating Y and the predictor variable(s).

(g) Using the model you have chosen in part (e), estimate the yearly corporate profits for a medium-sized firm whose capital investment was 220 thousands Riyals and whose advertising expenditure was 40 thousands Riyals.

Question 5:

(I) Standards set by government agencies indicate that Americans should not exceed an average daily sodium intake of 3300 milligram (mg). To find out whether Americans are exceeding (i.e., to have more than) this limit, a sample of $n = 200$ Americans is selected, and the mean and standard deviation of daily sodium intake are found to be $\overline{X} = 3500$ mg and $S = 1000$ mg, respectively. Use $\alpha = 0.05$ to conduct a test of hypothesis.
Arrange your answer as follows:

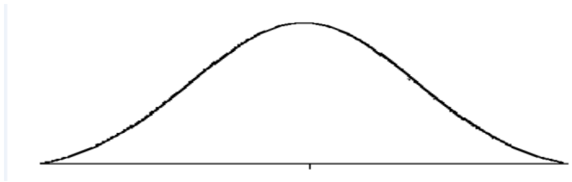
(a) The null and the alternative hypotheses are:

H_o :

H_a :

(b) The value of the test statistic is:

(c) The rejection region for the test is:



(d) The conclusion is:

(II) The records of a company show that $X_1 = 30$ men in a sample of $n_1 = 100$ male employees versus $X_2 = 10$ women in a sample of $n_2 = 50$ female employees hold high-ranked positions in the company. Do these data present sufficient evidence to indicate that the proportion of men holding high-ranked positions is different than the proportion of women holding high-ranked positions? Use $\alpha = 0.05$ to conduct a test of hypothesis. Arrange your answer as follows:

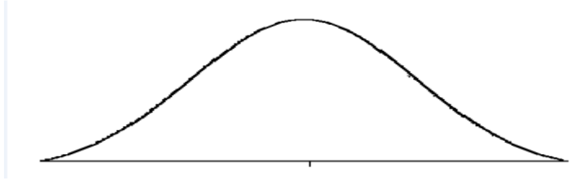
(a) The null and the alternative hypotheses are:

H_o :

H_a :

(b) The value of the test statistic is:

(c) The rejection region for the test is:



(d) The conclusion is:

Question 6:

(I) Two random samples of sizes n_1=10 and n_2=10 observations were selected independently from two normal populations with equal variances. The following results were obtained:

	1st Sample	2nd Sample
Sample mean ($\overline{X}$)	60	50
Sample variance (S^2)	20	24

(1) Do these data indicate that there is a difference between the means of the two normal populations? Use α=0.05. Arrange your answer as follows:

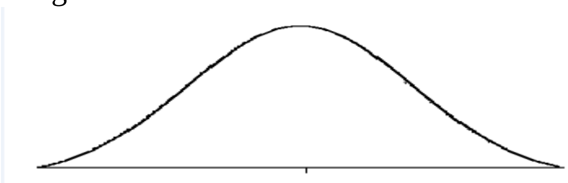
(a) The null and the alternative hypotheses are:

H_o :

H_a :

(b) The value of the test statistic is:

(c) The rejection region for the test is:



(d) The conclusion is:

(2) Construct a 95% confidence interval for $\mu_1-\mu_2$.

(II) A study has been made to compare the mean monthly salary of the employees in 4 companies (A, B, C, and D). A completely randomized design has been used for this study. The ANOVA table of this study follows (with α=0.05):

Source of Variation	SS	df	MS	F	P-value	F crit
Treatment	188000	.	62666.667		0.68338	3.23887
Error	1980000	16	123750	XXXXXXX	XXXXX	XXXXX
Total		19	XXXXX	XXXXXXX	XXXXX	XXXXX

(a) Complete the ANOVA table above.

(b) Do the data provide sufficient evidence to indicate a difference in the mean monthly salaries among the companies? Use α=0.05.

Appendix (A) for Question 3:

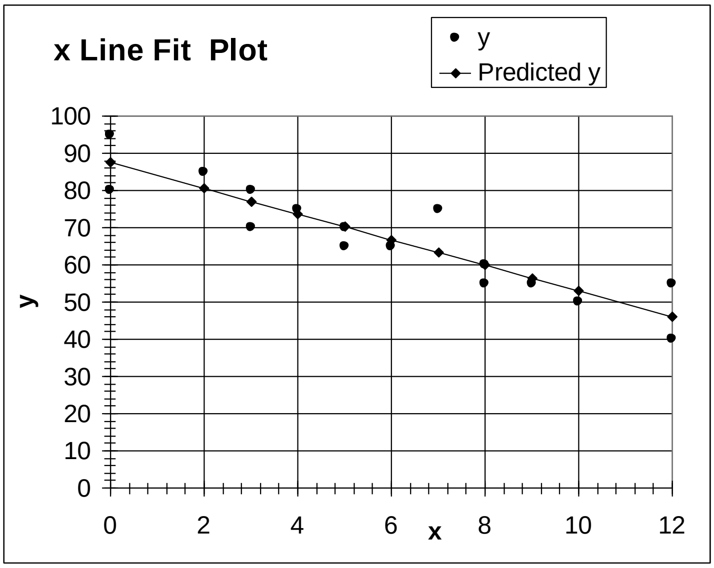
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.913845069
R Square	0.83511281
Adjusted R Square	0.823335154
Standard Error	6.04088837
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2587.544848	2587.544848	70.9065353	0.000000750
Residual	14	510.8926521	36.49233229		
Total	15	3098.4375			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	87.43972445	2.839926497	30.78943224	0.000000000000029	81.34868793	93.53076098
x	-3.447187141	0.409375388	-8.420601837	0.000000750482888	-4.32521002	-2.569164262



Appendix (B) for Question 4:

(B1) Output for model (1): $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.960424827
R Square	0.922415849
Adjusted R Square	0.900248949
Standard Error	18.85915999
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	29600.32459	14800.1623	41.61230645	0.000130079
Residual	7	2489.675409	355.6679156		
Total	9	32090			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-27.05686835	24.01240896	-1.126786921	0.296972196	-83.83719289	29.7234562
X1	0.139574137	0.077480473	1.801410485	0.114647528	-0.043638069	0.322786342
X2	3.79083355	0.456867016	8.29745509	0.000072	2.710514724	4.871152376

(B2) Output for model (2): $y = \beta_0 + \beta_2 X_2$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.941514296
R Square	0.886449169
Adjusted R Square	0.872255316
Standard Error	21.34199544
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	28446.15385	28446.15385	62.45302934	0.0000477
Residual	8	3643.846154	455.4807692		
Total	9	32090			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	9.769230769	14.25533223	0.685303619	0.512516646	-23.10362428	42.64208582
X2	3.307692308	0.418550966	7.902722907	0.0000477	2.34251205	4.272872566