KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICAL SCIENCES
DHAHRAN, SAUDI ARABIA
STAT 212: BUSINESS STATISTICS II

Semester 052
Mid Term Exam No.2B
Sunday April 16, 2006
8:05 – 9:35 pm

Please **circle** your:

Instructor's name	&	**section number**	
Raid Anabosi		1	2
Mohammad F. Saleh		3	4
Mohammad H. Omar		5	

Name: Student ID#: Serial #:

Directions:

1) You must **show all work** to obtain full credit for questions on this exam.
2) DO NOT round your answers at each step. Round answers only if necessary at **your final step to 4 decimal** places.

Question No	Full Marks	Marks Obtained
1	**14**	
2	**13**	
3	**17**	
4	**16**	
Total	**60**	

This page is intentionally left blank

1. (14 marks = 2+1+2+6+1+1+1) The Real Estate Association compiles data on U.S. property sales by type of buyers. Here is the1995 distribution:

Type of Buyer	Consumer	Business	Government
Probability	0.500	0.475	0.025

A random sample of last year's U.S. property sales gave the following data

Type of Buyer	Consumer	Business	Government
Frequency	1422	1521	57
Expected	**1500**	**1425**	**75**

At the **5%** significant level, do the data provide sufficient evidence to conclude that last year's type-of-buyers distribution of U.S. property sales is different from the 1995 distribution? (Answer the following sub questions to arrive at this conclusion).

a. The test hypotheses are:

H_0: Last year distribution is as the same as 1995

H_A: Last year distribution is NOT the same as 1995

b. The critical value is:

$$k = 3,\ \alpha = 0.05 \Rightarrow \chi^2_{0.05,2} = 5.9915$$

c. The decision rule is:

Reject H_0 if $\chi^2_{cal} > \chi^2_{\alpha,df} = 5.9915$

d. The test statistic is:

$$\chi^2_{cal} = \frac{(1422-1500)^2}{1500} + \frac{(1521-1425)^2}{1425} + \frac{(57-75)^2}{75} = 14.8434$$

e. The decision is:

Since $\chi^2_{cal} = 14.8434 > \chi^2_{\alpha,df} = 5.9915 \Rightarrow Reject\ \ H_0$

f. Your conclusion is:

Last year distribution is NOT as the same as 1995's distribution

g. What type of error you might have committed in (e) above.

Type I error (Reject H_0 given that H_0 is True)

2. (13 marks = 2+2+1+6+1+1) A publication of the international revenue service contains data on top wealth holders by marital status. A random sample of 490 top wealth holders yielded the following contingency table

Net Worth		Married	Others	*Total*
	$ 100,000 - $ 499,999	227	117	***344***
		222.5469	**121.4531**	
	$ 500,000 and above	90	56	***146***
		94.4531	**51.5469**	
	Total	***317***	***173***	***490***

At the **2.5%** significance level, do the data provide sufficient evidence to conclude that net worth and marital status are statistically independent for top wealth holders? (Answer the following sub questions to arrive at this conclusion).

a. The test hypotheses are:
H_0: Net worth and marital status are independent

H_A: Net worth and marital status are NOT independent

b. The critical value is:

$$k=(r-1)(c-1)=(2-1)(2-1)=1 \Rightarrow \chi^2_{0.025,1}=5.0239$$

c. The decision rule is:

Reject H_0 if $\chi^2_{cal} > \chi^2_{\alpha,df} = 5.0239$

d. The test statistic is:

$$\chi^2_{cal}=\frac{(227-222.55)^2}{222.55}+\frac{(117-121.54)^2}{121.54}+\frac{(90-94.45)^2}{94.45}+\frac{(56-51.55)^2}{51.55}$$
$$=0.089104+0.163271+0.209943+0.384693=0.8470$$

e. The decision is:

Since $\chi^2_{cal}=0.8470<\chi^2_{\alpha,df}=5.0239 \Rightarrow DO\ NOT\ reject\ H_0$

f. Your conclusion is:
The two variables are independent (NOT related)

3. (17 marks = 7+1+9) The *Kelley Blue Book* provides information on wholesale and retail price of cars. The following age and price data for 10 randomly selected Ford Mustang between 1 and 6 years old. Here x denoted age (in years) and y denotes price (in hundreds of dollars

x	6	6	6	2	2	5	4	5	1	3
y	175	165	180	310	269	200	240	213	310	210

You have calculated **some** of the necessary summary information to carry out the analyses as follows:

$$\sum x = 40,\ \sum x^2 = 192,\ \sum y = 2272,\ \sum y^2 = 541880,\ \sum xy = 8243$$

a. Obtain the correlation coefficient

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}\sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}} = \frac{8243 - \frac{(40)(2272)}{10}}{\sqrt{192 - \frac{(40)^2}{10}}\sqrt{541880 - \frac{(2272)^2}{10}}}$$

$$r = -0.932$$

b. Interpret the value of the correlation coefficient in terms of the linear relationship between the two variables.

Strong inverse (negative) linear relationship between x and y

c. (9 marks = 2+2+2+1+2) At **1%** level of significance, do the data provide sufficient evidence to conclude that the age and the price of Ford Mustang are negatively linear correlated?

I. State the hypotheses:

$$H_0 : \rho \geq 0 \quad vs. \ H_A : \rho < 0$$

II. The critical value(s) is (are):

$$-t_{\alpha,n-2} = -t_{0.01,8} = -2.8965$$

III. The test statistic is:

$$T_{cal} = r\sqrt{\frac{n-2}{1-r^2}} = (-0.932)\sqrt{\frac{8}{1-(-0.932)^2}} = -7.273$$

IV. The decision rule is:

Reject H_0 if $T_{cal} < -t_{\alpha,n-2} = -2.8965$

V. Your decision and your conclusion are:

Since $T_{cal} = -7.273 < -t_{\alpha,n-2} = -2.8965 \Rightarrow Reject\ H_0$

Age and price are NEGATIVELY (inversely) linear correlated

4. (16 marks = 7+1+2+5+1) A manufacturing company is interested in predicting the number of defects that will be produced each hour on the assembly line. The managers believe that there is a relationship between the defect rate and the production rate per hour. The managers also believe that they can use production rate to predict the number of defects. The following data were collected for 10 randomly selected hours.

Defects (y)	40	60	20	40	60	50	60	40	20	80
Production Rate Per Hour (x)	800	900	700	750	800	800	900	600	590	820

Also, to predict the number of defects (**y**) using production rate (**x**), the manager obtained the following summary statistics.

$$n = 10,\ \sum x = 7660,\ \sum x^2 = 5973000,\ \sum y = 470,\ \sum y^2 = 25300,$$
$$\sum xy = 373400, \text{ and } SSE = 1512.121$$

Assuming that **x** is the independent variable and **y** is the dependent variable then

a. Determine the fitted regression equation for the data.

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{373400 - \frac{(7660)(470)}{10}}{5973000 - \frac{(7660)^2}{10}} = 0.1269$$

$$b_0 = \bar{y} - b_1\bar{x} = 47 - (0.1269)(766) = -50.203$$

$$\hat{Y} = -50.203 + 0.1269\,X$$

b. What does the slope of the regression line represent in terms of the *number of defects*?

If the production time INCREASES by ONE hour, then the number of defects will *INCREASES* by 0.1269 defects.

c. The standard error of the regression model is:

$$S_\varepsilon = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{1512.121}{8}} = 13.748$$

d. Calculate the **coefficient of determination** and **interpret** its value.

$$SST = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 25300 - \frac{(470)^2}{10} = 3210$$

$$SSR = SST - SSE = 3210 - 1512.121 = 1697.879$$

$$R^2 = \frac{SSR}{SST} = \frac{1697.879}{3210} = 0.5298$$

This means that only 52.98% of the variation in the defect rate is explained by the variation in the production rate

e. Use the regression equation that you obtained in part (a) to predict **the number of defects** when the production rate is **590** per hour.

$$\hat{Y}\Big|_{x=590} = -50.203 + 0.1269(590) = 24.727 \ defects$$