KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICAL SCIENCES
DHAHRAN, SAUDI ARABIA

STAT 212: BUSINESS STATISTICS II

Semester 052
Mid Term Exam No.2A
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**	

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

Type of Buyer	Consumer	Business	Government
Probability	0.497	0.485	0.018

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

Type of Buyer	Consumer	Business	Government
Frequency	1422	1521	57
Expected	**1491**	**1455**	**54**

At the **2.5%** significant level, do the data provide sufficient evidence to conclude that last year's type-of-buyers distribution of U.S. cars 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(s) is(are):

$$k = 3,\ \alpha = 0.025 \Rightarrow \chi^2_{0.025,2} = 7.3778$$

c. The decision rule is:

Reject H_0 if $\chi^2_{cal} > \chi^2_{\alpha,df} = 7.3778$

d. The test statistic is:

$$\chi^2_{cal} = \frac{(1422-1491)^2}{1491} + \frac{(1521-1455)^2}{1455} + \frac{(57-54)^2}{54} = 6.354$$

e. The decision is:

Since $\chi^2_{cal} = 6.354 < \chi^2_{\alpha,df} = 7.3778 \Rightarrow DO\ NOT\ reject\ \ H_0$

f. Your conclusion is:

Last year distribution is as the same as 1995's distribution

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

Type I error (Do not reject H_0 given that H_0 is False)

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 487 top wealth holders yielded the following contingency table

		Married	**Others**	***Total***
Net Worth	**$ 100,000 - $ 249,999**	227	117	***344***
		223.9179	**120.0821**	
	$ 300,000 and above	90	53	***143***
		93.0821	**49.9179**	
	Total	***317***	***170***	***487***

At the **1%** 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(s) is (are):

$$k = (r-1)(c-1) = (2-1)(2-1) = 1 \Rightarrow \chi^2_{0.01,1} = 6.6349$$

c. The decision rule is:

Reject H_0 if $\chi^2_{cal} > \chi^2_{\alpha,df} = 6.6349$

d. The test statistic is:

$$\chi^2_{cal} = \frac{(227-223.92)^2}{223.92} + \frac{(117-120.08)^2}{120.08} + \frac{(90-93.08)^2}{93.08} + \frac{(53-49.92)^2}{49.92}$$
$$= 0.042424 + 0.079109 + 0.102056 + 0.190304 = 0.4139$$

e. The decision is:

Since $\chi^2_{cal} = 0.4133 < \chi^2_{\alpha,df} = 6.6349 \Rightarrow DO\ NOT\ reject\ H_0$

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 Corvettes 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	4
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:

$$\sum x = 41,\ \sum x^2 = 199 = 199,\ \sum y = 2272,\ \sum y^2 = 541880,\ \sum xy = 8453$$

a. Obtain the correlation coefficient

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}\sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}} = \frac{8453 - \frac{(41)(2272)}{10}}{\sqrt{199 - \frac{(41)^2}{10}}\sqrt{541880 - \frac{(2272)^2}{10}}}$$

$$r = -0.968$$

b. Interpret the value of the linear 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 **5%** level of significance, do the data provide sufficient evidence to conclude that the age and the price of Corvettes are negatively linear correlated?

I. State the hypotheses:

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

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

$$-t_{\alpha,n-2} = -t_{0.05,8} = -1.8595$$

III. The test statistic is:

$$T_{cal} = r\sqrt{\frac{n-2}{1-r^2}} = (-0.968)\sqrt{\frac{8}{1-(-0.968)^2}} = -10.91$$

IV. The decision rule is:

Reject H_0 if $T_{cal} < -t_{\alpha,n-2} = -1.8595$

V. Your decision and your conclusion are:

Since $T_{cal} = -10.91 < -t_{\alpha,n-2} = -1.8595 \Rightarrow Reject\ H_0$

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 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)	20	30	10	20	30	25	30	20	10	40
Production Rate Per Hour (x)	400	450	350	375	400	400	450	300	295	410

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

$$n = 10,\ \sum x = 3830,\ \sum x^2 = 1493250,\ \sum y = 235,\ \sum y^2 = 6325,$$
$$\sum xy = 93350, \text{ and } SSE = 378.0302$$

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

a. Determine the fitted regression equation for the data.

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{93350 - \frac{(3830)(235)}{10}}{1493250 - \frac{(3830)^2}{10}} = 0.1269$$

$$b_0 = \bar{y} - b_1\bar{x} = 23.5 - (0.1269)(383) = -25.1$$

$$\hat{Y} = -25.1 + 0.1269\, X$$

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:

$$S_\varepsilon = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{378.0302}{8}} = 6.87414$$

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

$$SST = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 3625 - \frac{(235)^2}{10} = 802.5$$

$$SSR = SST - SSE = 802.5 - 378.0302 = 424.4698$$

$$R^2 = \frac{SSR}{SST} = \frac{424.4698}{802.5} = 0.5298$$

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 (c) to predict the **number of defects** if the production rate is **300** per hour.

$$\widehat{Y}\Big|_{x=300} = -25.1 + 0.1269(300) = 12.97\ defects$$

Page is intentionally left blank

Turn Page Over
Turn Page Over