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

STAT 213: STAT213 STATISTICS METHODS FOR ACTUARIES

Final Exam, Term 121
Tuesday January 1, 2013

Student Surname: ID#

You are allowed to use electronic calculators and other reasonable writing accessories that help write the exam. Try to define events, formulate problem and solve.

Do not keep your **mobile** with you during the exam, turn off your mobile and leave it aside.

Question No	Full Marks	Marks Obtained
1	30	
2	20	
3	10	
4	20	
5	10	
6	10	
7	10	
Total	110	

Note: You may assume $\alpha = 0.05$ for testing and 95% for confidence interval estimation if not otherwise stated.

Q1. The owner of restaurant **A** wants to study the characteristics of his customers. He decides to focus on two variables:

- The amount of money spent by customers
- Whether the customer's order dessert.

The results from a sample of 62 customers are as follows:

- Amount spent: the mean = $40.11 and the standard deviation = $ 7.41
- 20 customers purchased dessert

a. (*7 points*) Is there evidence that the population mean amount spent per customer in the restaurant is less than $42.11?

1. The hypothesis:

2. The test statistics

3. The decision rule and the critical value (s)

4. The decision

5. The conclusion

b. (*1 point*) Do you need any assumptions? If yes, what? If no, why?

c. *(5 points)* Construct and interpret a 90% confidence interval estimate of the population proportion of customers who purchase dessert.

The owner of a competing restaurant **B** wants to conduct a similar survey in his restaurant.

d. (*2 points*) Based on his competitors information he tries to find the sample size he needs to have 95% confidence to estimate the population mean amount spent in his restaurant to within $\pm$\$2.7. What is the sample size?

He decides to sample the number he found above, and the results are

- Amount spent: the mean = \$35.32 and the standard deviation = \$ 15.65
- 25 customers purchased dessert

e. (*8 points*) Do you think that there is no difference between the percentages of the customers who purchased dessert in both restaurants? Test using the p – value approach

1. The hypothesis:

2. The test statistics

3. The decision rule

4. The decision

5. The conclusion

f. (*7 points*) Construct a 99% confidence interval for the difference in the mean spent in both restaurants. Based on this confidence interval, what conclusion can you draw about the mean amount spent in restaurant **B** compared to the mean amount spent in restaurant **A**?

Q2. A candy bar manufacturer is interested in trying to estimate how sales are influenced by the price of their product. To do this, the company randomly chooses 6 small cities and offers the candy bar at different prices. Using candy bar sales as the dependent variable, the company will conduct a simple linear regression on the data below:

City	River Falls	Hudson	Ellsworth	Prescott	Rock Elm	Stillwater
Price ($)	1.3	1.6	1.8	2	2.4	2.9
Sales	100	90	90	40	38	32

You may use the following information's

$n = 6, \sum x = 12, \ \sum x^2 = 25.66, \ \sum y = 390, \sum y^2 = 30268, \sum xy = 700$

1. (*8 points*) If the price of the candy bar is set at $2, the estimated average sales will be?

2. (*3 points*) What is the coefficient of correlation for these data? What does this tell you about the relation between the price and the sales

3. (*9 points*) Find a 99% confidence interval for the true slop. Use you answer to test whether a change in price will have any impact on average sales

Q3. A financial analyst wanted to examine the relationship between salary (in $1,000) and 4 variables: age (X_1 = Age), experience in the field (X_2 = Experience), number of degrees (X_3 = Degrees), and number of previous jobs in the field (X_4 = Previous jobs). He took a sample of 20 employees and obtained the following MINITAB output:

```
Predictor        Coef      SE Coef     T      P
Constant       -9.6111    2.77988  -3.457  0.0035
Age             1.3276    0.11491  11.553  0.0001
Experience     -0.1067    0.14265  -0.748  0.4660
Degrees         7.3113    0.80324   9.102  0.0001
Previous jobs  -0.5041    0.44771  -1.126  0.2778


S = 2.267   R-Sq = 98.4%   R-Sq(adj) = 97.9%

Analysis of Variance

Source          DF     SS       MS      F       P
Regression       4  4609.83  1152.45  224.16  0.0001
Residual Error  15    77.11     5.14
Total           19  4686.94
```

1. (*8 points*) Which independent variables **not** significant? Test at 5% level of significant, write the hypotheses.

2. (*2 points*) What does the coefficient of ***Degrees*** tell you about the model?

Q4. The superintendent of a school district wanted to predict the percentage of students passing a sixth-grade proficiency test. She obtained the data on percentage of students passing the proficiency test (*% Passing*), daily average of the percentage of students attending class (*% Attendance*), average teacher salary in dollars (*Salaries*), and instructional spending per pupil in dollars (*Spending*) of 47 schools in the state.

Following is the multiple regression output with *Y = % Passing* as the dependent variable, X_1 =:*% Attendance*, X_2 = *Salaries* and X_3 = *Spending*:

```
Predictor         Coef        SE Coef      T      P
Constant        -753.4225    101.1149    -7.4511  2.88E-09
% Attendance       8.5014      1.0771     7.8929  6.73E-10
Salary          6.85E-07       0.0006     0.0011  0.9991
Spending          0.0060       0.0046     1.2879  0.2047


S = _______   R-Sq = _________   R-Sq(adj) = _______

Analysis of Variance

Source          DF     SS       MS        F         P
Regression       3   7965.08   2655.03   24.2802  2.3853E-09
Residual Error  43   4702.02   109.35
Total           46  12667.11
```

a. (*2 points*) What is the average change in the percentage of students passing the proficiency test when daily average of the percentage of students attending class increases by 1% holding constant the effects of all the remaining independent variables?

b. (*3 points*) How much of the total variation in the percentage of students passing the proficiency test can be explained by daily average of the percentage of students attending class, average teacher salary, and instructional spending per pupil after adjusting for the number of predictors and sample size?

c. (*3 points*) Predict the percentage of students passing the proficiency test for a school which has a daily average of 95% of students attending class, an average teacher salary of 40,000 dollars, and an instructional spending per pupil of 2000 dollars.

d. (*3 points*) Test whether there is a significant relationship between percentage of students passing the proficiency test and the entire set of explanatory variables? Write the hypotheses, and the decision role and the conclusion

e. (*3 points*) Test whether instructional spending per pupil has any effect on percentage of students passing the proficiency test? Write the hypotheses, and the decision role and the conclusion

f. (*3 points*) Calculate and interpret a 95% confidence interval estimate for the effect of a one dollar increase in average teacher salary on average percentage of students passing the proficiency test?

g. (*3 points*) Based on your solution in part (f), Can you conclude that average teacher salary has no impact on average percentage of students passing the proficiency test? Why?

Q5. (*10 points***)** Marketers know that tastes differ in various regions of the country. In the rental car business, an industry expert has given the opinion that there are strong regional preferences of size of car and quotes. The following data in support of the view:

Region of Country

Preferred Car Type	Northeast	Southeast	Northwest	Southwest	
Full – size	105	120	105	70	
Intermediate	120	100	130	150	
All other	25	30	15	30	

Do the data support the expert's opinion at the 0.05 significance level?

1. The hypothesis:

2. The test statistics

3. The decision rule and the critical value

4. The decision

5. The conclusion

Q6. (*10 points***)** In 1990, the average weekly food cost, for suburban family of four, was estimated to be $154.40. The following table presents the retail prices of selected food items from 1990 to 1997:

Year	Bread	Eggs	Eggs %
1990	0.7	1	32%
1991	0.72	1.01	28%
1992	0.74	0.93	31%
1993	0.76	0.87	36%
1994	0.75	0.87	34%
1995	0.84	1.16	38%
1996	0.87	1.31	33%
1997	0.88	1.17	37%

Referring to the above table, find the following

a. The **simple price index** number for the Bread in 1996 based on 1993, and **interpret** it	
Interpretation	
b. The **Laspeyres** price index number for 1996 based on 1993, and **interpret** it	
Interpretation	
c. The **Unweighted price index** number for 1996 based on 1993, and **interpret** it	
Interpretation	

If the **adjusted Potato price** in 1993 is 5.77 and in 1996 is 5.87 and the CPI in 1993 is 100 and 1996 is 109.6, then

d. Find the **actual prices** for the Potato in 1993 and 1996	Actual price in 1993 is
	Actual price in 1996 is

Q6. (*10 points)* Select the true answer

1. If, as a result of a hypothesis test, you reject the null hypothesis when it is false, then you have committed
   a. No error.
   b. An acceptance error
   c. A Type I error
   d. A Type II error
   e. You can't tell depend the information given.

2. You have created a 95% confidence interval for μ with result $(10, 15)$. What decision will you make if you test $H_0 : \mu = 16$ *vs* $H_A : \mu \neq 16$ at $\alpha = 0.1$?
   a. Don't reject H_0 in favor of H_A
   b. Reject H_0 in favor of H_A
   c. Reject H_0 in favor of H_A
   d. Fail to reject H_A in favor of H_0
   e. We can't tell what our decision will be from the information given

3. Suppose you wish to test $H_0 : \mu \leq 47$ *vs* $H_A : \mu > 47$. What will the result if we conclude that the mean is greater than 47 when the true value is really 52?
   a. We have made a Type II error.
   b. We have made a Type I error.
   c. We have made a correct decision.
   d. You can't tell depend the information given.
   e. None of the above true.

4. When testing $H_0 : \mu_1 - \mu_2 = 0$ *vs* $H_A : \mu_1 - \mu_2 \neq 0$, the observed value of z – score was – 2.13. The p – value for the test would be
   a. 0.9834
   b. 0.0166
   c. 0.9668
   d. 0.0332
   e. None of the above

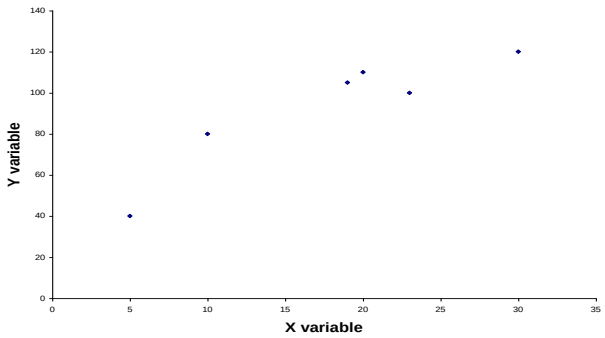
5. The sample correlation coefficient between X and Y is 0.375. It has been found out that the p – value is 0.256 when testing $H_0 : \rho = 0$ *vs* $H_A : \rho \neq 0$. To test $H_0 : \rho \leq 0$ *vs* $H_A : \rho > 0$ at significance level of 20%, the p – value is?
   a. We can't find the p – value depend the information that we have.
   b. 0.256.
   c. 0.872.
   d. 0.375.
   e. 0.128.

6. When testing for the independence in a contingency table with 3 rows and 4 columns, the degrees of freedom for the test statistic would be
   a. 12
   b. 7
   c. 20
   d. 6
   e. None of the above.

7. If the dependent variable increases as the independent variable increases in an estimating equation, the coefficient of correlation will be in the range:
   a. – 0.05 to 0
   b. – 1 to 0
   c. 0 to +1
   d. – 0.05 to + 0.5
   e. None of the above.

8. After taking a sample and computing $\bar{x}$, a statistician says," I am 95 percent confident that the population mean is between 106 and 122." What does he really mean? :
   a. The probability is 0.95 that μ is between 106 and 122.
   b. 95 percent of the intervals calculated from samples of this size will contain the population mean.
   c. The probability is 0.95 that $\mu = 114$, the midpoint of the interval.
   d. None of the above.

9. Consider the following chart. Which of the following statements is most correct?
   a. There is a positive linear relationship between the two variables.
   b. There is a negative linear relationship between the two variables.
   c. There is no apparent relationship between the two variables.
   d. There is a perfect linear relationship between the two variables.
   e. None of the above.



10. A major retail store has studied customer behavior and found that the distribution of time customers spend in a store per visit is symmetric with a mean equal to 17.3 minutes. Based on this information, which of the following is true?
    a. The distribution is bell-shaped.
    b. The median is approximately 17.3 minutes.
    c. The median is to the right of the mean.
    d. None of the above.
    e. a and c but not b