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

STAT 212: BUSINESS STATISTICS II

Semester 133
Final Exam
Thursday August 18, 2014
7:00 pm – 9:30 pm

Name: ID#: Section: **Serial:**

Part	Question No	Full Marks	Marks Obtained
I	**1**	6	
	2	14	
	3	10	
	4	12	
	5	13	
II	**MCQ**	25	
Total		80	

Note: For each question: Clearly state your hypotheses, assumptions and your conclusions. Use 5% level of significance unless specified otherwise in the problem.

Part I (written questions)

Q1. (4+2 points) A study was recently performed in which it attempted to develop a regression model to explain variation in the mileage ratings of new cars. At one stage of the analysis, the estimated regression took the following form:

$$\hat{y} = 34.2 - .003X_1 + 4.56X_2$$

where X_1 = Vehicle weight, and X_2 =0 if standard transmission and X_2 =1 if automatic transmission.

a. Interpret the regression coefficient for the variables X_1 and X_2.

b. Provide an estimate of the average highway mileage you would expect to obtain from a Cadillac with automatic transmission weighs 4012 pounds.

Q2. (1+1+12 points) Recently a survey was conducted involving customers of a fitness center. Participants were asked to indicate how often they use the club by checking one of the following categories: 0–1 time per week; 2-3 times per week; 4-5 times per week. The following data show how males and females responded to this question.

	0-1	**2-3**	**4-5**
Males	41	61	50
Females	109	89	60

One of the purposes of the survey was to determine whether there is a relationship between the gender of the customer and the number of visits made each week.

a. State the appropriate null and alternative hypothesis.

b. What test procedure is appropriate to use to conduct this test?

c. Conduct the hypothesis test using $\alpha = 0.01$.

Q3. (10 points) A company claims that its bulbs are **superior** to those of its main competitor. If a study showed that a sample of 16 of its bulbs had a mean lifetime of 647 hours of continuous use with a standard deviation of 27 hours, while a sample of 16 bulbs made by its main competitor had a mean lifetime of 638 hours of continuous use with standard deviation of 31 hours. Do you think that the company claim is justified?

Q4. (6+3+3 points) A major developer of housing communities in a city kept a record of the relative cost of labor and materials in its market areas for the past three years. These data are as follows:

Years	**1998**	**1999**	**2000**
Average Labor cost	SR 49000	57000	63000
Average Material Cost	SR 95000	104000	110000
% Material Cost	67	68	66

a. Determine the simple index for the Average Labor Cost every year using 1999 as the base year.

$I_{1998} =$

$I_{1999} =$

$I_{2000} =$

b. Find the unweighted aggregate index for the two components in the construction of the house in 1998 using 1999 as the base year and interpret it.

$I_{1998} =$

Interpretation:

c. Construct a Paasche index number of 2000 using 1999 as the base year and interpret it.

$I_{2000} =$

Interpretation:

Q5. (3+4+6 points) The following data are average 1992-1997 (based on the first quarter) market shares of Japanese cars in the market.

27, 25.5, 26.4, 27.3, 28.2, 30.1

a. Using your calculator, do a linear trend analysis to predict the value of 1998.

b. Using a 3-year moving average smoothing, calculate the second and the last moving averages.

c. Using exponential smoothing with $w = 0.3$, calculate the smoothed value for 1994.

Part II (multiple choice questions)

1. When testing the estimate of a linear regression coefficient based on a sample of 14 pairs, the calculated value of the *F*-statistic was 6.92. What is the smallest significance level at which you would reject the hypothesis H_0: $\beta_1 = 0$?

a. 0.10
b. 0.01
c. 0.05
d. 0.02
e. 0.025

2. If the correlation between age of an auto and money spent for repairs is +0.90

a. 90% of the money spent for repairs is explained by the age of the auto
b. 81% of money spent for repairs is unexplained by the age of the auto
c. 81% of the variation in the money spent for repairs is explained by the age of the auto
d. 90% of the money spent for repairs is unexplained by the age of the auto
e. none of the above

3. Suppose pairs of data values (*X*i,*Y*i) have been gathered and are plotted on a diagram (given below). A statistician asserts that the product-moment correlation between *X* and *Y* in this case is very low.

Which of the following comments makes the most sense?

a. We cannot tell
b. The statistician is wrong, because knowing X lets us predict the average value of Y quite accurately.
c. The statistician is wrong, because X and Y increase and decrease together.
d. The statistician is wrong, because the relationship between X and Y is not linear.
e. The statistician is right, because the correlation measures how much linear relationship there is, and the relationship is certainly not linear.

4. The McNemar test is

a. F-distributed.
b. Chi-square distributed.
c. Normal distributed.
d. T-student distributed.
e. Binomial distributed.

Molybdenum rods produced on a production line are supposed to average 2.2 inches in length. It is desired to check whether the process is in control. Let X = length of such a rod. Assume X is approximately normally distributed. Suppose a sample of size 400 rods is taken and yields a sample average length of 2 inches, and $\sum(x - \bar{x})^2 = 399$.
To test H_1: $\mu \neq 2.2$ at level $\alpha = 8\%$
Use the above information's to solve the **questions from 5 to 6**

5. One would use a _____ confidence interval for μ and hence a table value of _____.

a. 92%, 1.75
b. 92%, 1.67
c. 92%, 1.41
d. 96%, 2.06
e. 96%, 1.75

6. In testing H_0, one should _____ the H_0 since the value _____ lies _____ the confidence interval.

a. Not reject, 2, within.
b. Not reject, 2.2, within.
c. Reject, 2, outside of.
d. Reject, 2.2, outside of.
e. Cannot tell, we need more information's.

7. Suppose a t-test for the hypothesis that H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$ is carried out and we find $t_{STAT} = 1.8$. The observed significance level of the test is:

a. The probability of getting a *t*-value > 1.8.
b. The Type I error probability of the test.
c. The probability of getting a *t*-value > 1.8 or < –1.8.
d. The Type II error probability of the test.
e. We cannot tell

8. A sample of size 36 is taken from a population with unknown mean μ and standard deviation equal to 3. In a test of H_1: $\mu \neq 5$ at $\alpha = 0.01$, we would reject H_0 if:

a. $\bar{x} - 5 < 1.29$ or $5 - \bar{x} < 1.29$
b. $\bar{x} - 5 > 7.74$ or $5 - \bar{x} > 7.74$
c. $\bar{x} - 5 > 1.29$ or $5 - \bar{x} < 7.74$
d. $\bar{x} - 5 > 1.29$ or $5 - \bar{x} > 1.29$
e. $\bar{x} - 5 < 7.74$ or $5 - \bar{x} < 7.74$

9. A result was said to be statistically significant at the 5% level. This means:

a. The null hypothesis is probably wrong.
b. The result would be unexpected if the null hypothesis were true.
c. The null hypothesis is probably true.
d. The alternative hypothesis is probably true.
e. None of the above.

10. The critical value of a test statistic is determined from:

a. The sampling distribution of the statistic assuming H_0.
b. Calculations from the data.
c. Calculations based on many actual repetitions of the same experiment.
d. The sampling distribution of the statistic assuming H_1.
e. None of the above.

11. Indicate which assumptions are needed to use the sample mean and normal tables to test a hypothesis about a population mean, and unknown variance, to test a hypothesis about μ?

I. The data are a random sample.
II. The population distribution is normal.
III. The sample size is large.

a. I, II, and III.
b. I and either II or III.
c. II and III.
d. only II.
e. only I.

12. A home owner claims that the current market value of his house is at least $40,000. Sixty real estate agents were asked independently to estimate the house's value. The hypothesis test that followed ended with a decision of "reject H_0". Which of the following statements accurately states the conclusion?

a. The home owner is wrong; he should not sell his home.
b. The home owner is right; the house is worth $40,000.
c. The home owner is right; the house is worth less than $40,000.
d. The home owner is wrong; the house is worth more than $40,000.
e. The home owner is wrong; the house is worth less than $40,000.

13. Suppose a test was taken by 36 students and the variance of the distribution of scores was 100. To test H_0: $\mu \geq 80$ vs. H_1: $\mu < 80$, using $\alpha = 0.05$. Assume the population of test scores is normally distributed. What (to the nearest tenth) is the starting point of the region of rejection in terms of $\bar{x}$ values?

a. 77.6
b. 76.1
c. 76.7
d. 77.7
e. 77.3

An Industrial company undertook a training program to improve perceived job satisfaction. Its goal was to increase the perceived job satisfaction score of its employees to a mean level above 70 on a scale from 0 to 100. After the program the company sends letters to a random sample of 36 of its employees and finds that the mean satisfaction score is 74 with a sample standard deviation of 9.

Use the above information's to solve the **questions from 14 to 15**

14. Given the sample data, the company should use what estimated standard error?

a. 2.667
b. 10
c. 9
d. 6
e. 1.5

15. If the company made a type-I error this would translate to

a. Concluding the mean is above 70 when it is not above.
b. Concluding incorrectly that the training program had failed.
c. Concluding incorrectly that the training program was a success.
d. Concluding the mean is not above 70 when it is above.
e. We cannot tell

Given below are EXCEL outputs for various estimated autoregressive models for Coca-Cola's real operating revenues (in billions of dollars) from 1975 to 1998. From the data, we also know that the real operating revenues for 1996, 1997, and 1998 are 11.7909, 11.7757 and 11.5537, respectively.

AR(1) Model:

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.1802077	0.39797154	0.452815546	0.655325119
XLag1	1.011222533	0.049685158	20.35260757	2.64373E-15

AR(2) Model:

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.30047473	0.4407641	0.681713257	0.503646149
X Lag 1	1.17322186	0.234737881	4.998008229	7.98541E-05
X Lag 2	-0.183028189	0.250716669	-0.730020026	0.474283347

AR(3) Model:

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.313043288	0.514437257	0.608515972	0.550890271
XLag1	1.173719587	0.246490594	4.761721601	0.000180926
XLag2	-0.069378567	0.373086508	-0.185958391	0.854678245
XLag3	-0.122123515	0.282031297	-0.433014053	0.670448392

Use the above information to solve the **questions from 16 to 17**

16. What is the appropriate AR model for Coca-Cola's real operating revenue?

a. AR(1)
b. AR(2)
c. AR(3)
d. AR(4)
e. None of the above

17. If one decides to use AR(3), what will the predicted real operating revenue for Coca-Cola be in 1999?

a. $11.59 billion
b. $11.62 billion
c. $11.84 billion
d. $12.47 billion
e. None of the above

A contractor developed a multiplicative time-series model to forecast the number of contracts in future quarters, using quarterly data on number of contracts during the 3-year period from 1996 to 1998. The following is the resulting regression equation:

$\widehat{\log Y} = 3.37 + 0.117\,X - 0.083\,Q1 + 1.28\,Q2 + 0.617\,Q3$

Where $\hat{Y}$ is the estimated number of contracts in a quarter.
X is the coded quarterly value with X = 0 in the first quarter of 1996.
Q1 is a dummy variable equal to 1 in the first quarter of a year and 0 otherwise.
Q2 is a dummy variable equal to 1 in the second quarter of a year and 0 otherwise.
Q3 is a dummy variable equal to 1 in the third quarter of a year and 0 otherwise.
Use the above information to solve the **questions from 18 to 20**

18. The best interpretation of the coefficient of Q3 (0.617) in the regression equation is:

a. The number of contracts in the third quarter of a year is approximately 314% higher than the average over all 4 quarters.
b. The number of contracts in the third quarter of a year is approximately 62% higher than the average over all 4 quarters.
c. The number of contracts in the third quarter of a year is approximately 62% higher than it would be during the fourth quarter.
d. The number of contracts in the third quarter of a year is approximately 314% higher than it would be during the fourth quarter.
e. There is no meaning.

19. Using the regression equation, which of the following values is the best forecast for the number of contracts in the second quarter of 2000?

a. 144212
b. 4355119
c. 391742
d. 1238797
e. We cannot find the forecast

20. In testing the coefficient for *Q*1 in the regression equation (– 0.083), the results were a t-statistic of – 0.66 and an associated p-value of 0.530. Which of the following is the best interpretation of this result?

a. The number of contracts in the first quarter of the year is significantly different than the number of contracts in an average quarter.
b. The number of contracts in the first quarter of the year is not significantly different than the number of contracts in the fourth quarter for a given coded quarterly value of X.
c. The number of contracts in the first quarter of the year is not significantly different than the number of contracts in an average quarter.
d. The number of contracts in the first quarter of the year is significantly different than the number of contracts in the fourth quarter for a given coded quarterly value of X.
e. We cannot test.

For Part II

Shade (fully) your choices in the table below:

1	a	b	c	d	e
2	a	b	c	d	e
3	a	b	c	d	e
4	a	b	c	d	e
5	a	b	c	d	e
6	a	b	c	d	e
7	a	b	c	d	e
8	a	b	c	d	e
9	a	b	C	d	e
10	a	b	C	d	e

11	a	b	c	d	e
12	a	b	c	d	e
13	a	b	c	d	e
14	a	b	c	d	e
15	a	b	c	d	e
16	a	b	c	d	e
17	a	b	c	d	e
18	a	b	c	d	e
19	a	b	c	d	e
20	a	b	c	d	e

With My Best Wishes