PART I (MCQ Code 2):- Answer the following questions by circling the correct option or writing the correct numeric answer. For some of the questions, Scenarios are given in the formula sheet.

(1.01) Which of the following table values would be appropriate for a 95% confidence interval for the mean of y from a simple linear regression problem if the sample size is 7?

a. 1.895
b. 2.015
c. 2.365
d. 2.571
e. 1.960

(1.02) A regression model has 3 independent variables and 20 observations, and the calculated Durbin-Watson *d* statistic is 0.91. What conclusion will be reached in testing for positive autocorrelation of the residuals at the $\alpha = 0.05$ level?

a. There is no positive autocorrelation
b. There is positive autocorrelation
c. Test is inconclusive
d. The value of *d* is incorrect
e. None of the others

(1.03) In multiple regression, the _____ procedure permits variables to enter and leave the model at different stages of its development.

a. forward selection
b. residual analysis
c. backward elimination
d. stepwise regression
e. a and b.

(1.04) Referring to Scenario 1, if this series is smoothed using exponential smoothing with a smoothing constant of 1/3, how many terms would it have?

a. 3
b. 4
c. 5
d. 6

(1.05) Referring to Scenario 1, if a three-term moving average is used to smooth this series, what would be the second calculated term?

a. 36
b. 40.5
c. 54
d. 72

(1.06) Referring to Scenario 1, if a three-term moving average is used to smooth this series, what would be the last calculated term?

a. 72
b. 93
c. 114
d. 126

(1.07) The regression line $\hat{y} = 3 + 2x$ has been fitted to the data points (4,8), (2,5), and (1,2). The residual sum of squares will be:

a. 10.
b. 15.
c. 13.
d. 22.
e. None of the others

(1.08) Referring to Scenario 7, the best interpretation of the coefficient of Q_3 (0.617) in the regression equation is:

a. the number of contracts in the third quarter of a year is approximately 62% 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 it would be during the fourth quarter.
c. the number of contracts in the third quarter of a year is approximately 314% higher than the average over all 4 quarters.
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. None of the others

(1.09) Referring to Scenario 7, the best interpretation of the constant 3.37 in the regression

a. the fitted value for the first quarter of 2011, prior to seasonal adjustment, is log(3.37).
b. the fitted value for the first quarter of 2011, after to seasonal adjustment, is log(3.37).
c. the fitted value for the first quarter of 2011, prior to seasonal adjustment, is 103.37.
d. the fitted value for the first quarter of 2011, after to seasonal adjustment, is 103.37.
e. None of the others

(1.10) Referring to Scenario 7, which of the following values is the best forecast for the number of contracts in the third quarter of 2014?

a. 49,091
b. 133,352
c. 421,697
d. 1,482,518
e. None of the others

(1.11) Referring to Scenario 7, to obtain a forecast for the first quarter of 2014 using the model, which of the following sets of values should be used in the regression equation?

a. $X = 12, Q_1 = 0, Q_2 = 0, Q_3 = 0$
b. $X = 12, Q_1 = 1, Q_2 = 0, Q_3 = 0$
c. $X = 13, Q_1 = 0, Q_2 = 0, Q_3 = 0$
d. $X = 13, Q_1 = 1, Q_2 = 0, Q_3 = 0$
e. None of the others

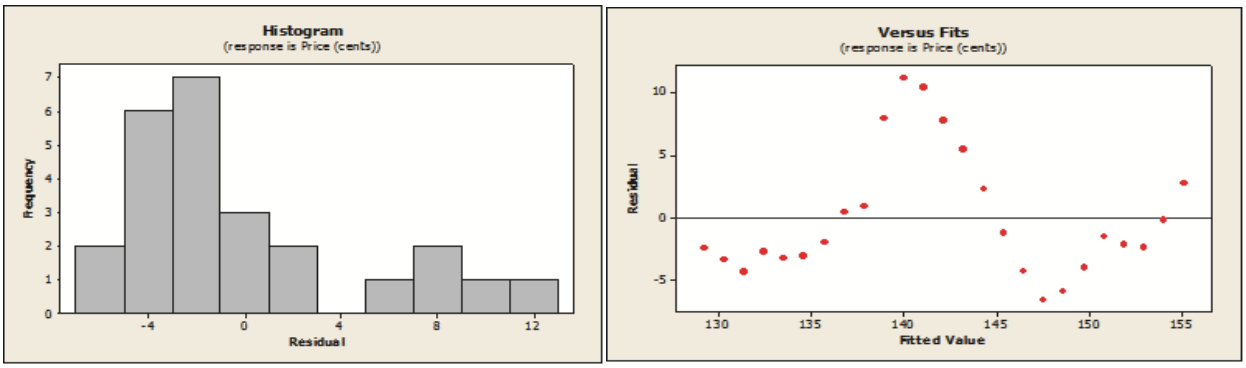
(1.12) Referring to Scenario 7, the best interpretation of the coefficient of X (0.117) in the regression equation is:

a. the quarterly compound growth rate in contracts is around 30.92%.
b. the annually compound growth rate in contracts is around 30.92%.
c. the quarterly compound growth rate in contracts is around 11.7%.
d. the annually compound growth rate in contracts is around 11.7%.
e. None of the others

(1.13) In testing the hypotheses $H_0: \pi = 0.40$, $H_1: \pi > 0.40$ at the 5% significance level, if the sample proportion is 0.45, and the standard error of the sample proportion is 0.035, the appropriate conclusion would be:

a. to reject H_0.
b. not to reject H_0.
c. to reject H_1.
d. to reject both H_0 and H_1.
e. None of the others

(1.14) According to the residual plots shown below, which linear regression assumptions appear to be violated?



a. Linearity
b. Normality
c. Equal Variance
d. Both "a" and "b"
e. All of the above

(1.15) Referring to Scenario 3, for the cell with 1 to 4 years of training time and a high defect rate, what is the contribution to the overall χ^2 statistic for the independence test?

a. 0.36
b. 0.1296
c. 0.015
d. 0.0144
e. None of the others

(1.16) A mortgage broker is offering home mortgages at a rate of 9.5% but is fearful that this value is higher than what many others are charging. A sample of 40 mortgages filed in the county court house shows an average of 9.25% with a standard deviation of 8.61%. Does this sample indicate a smaller average? Use $\alpha = 0.05$ and assume a normally distributed population.

a. Yes, the test statistic falls in the rejection region.
b. No, the test statistic falls in the acceptance region.
c. Yes, because the test statistic is greater than -1.645.
d. No, because the test statistic is -1.85 and falls in the rejection region.
e. Yes, because the sample mean of 9.25 is below 9.5.

(1.17) If a hypothesis is rejected at the 0.025 level of significance, it:

a. must be rejected at any level.
b. must be rejected at the 0.01 level.
c. must not be rejected at the 0.01 level.
d. may be rejected or not rejected at the 0.01 level.
e. None of the others

(1.18) Referring to Scenario 4, to test whether aggregate price index has a negative impact on consumption, the *p*-value is:

a. 0.0001
b. 0.4165
c. 0.8330
d. 0.8837
e. None of the others

(1.19) Consider testing the hypothesis $H_0: \mu = 800$ vs. $H_1: \mu \neq 800$. If the value of the test statistic z equals 1.75, then the p-value is:

a. 0.0401.
b. 0.0802.
c. 0.4599.
d. 0.9198.
e. None of the others

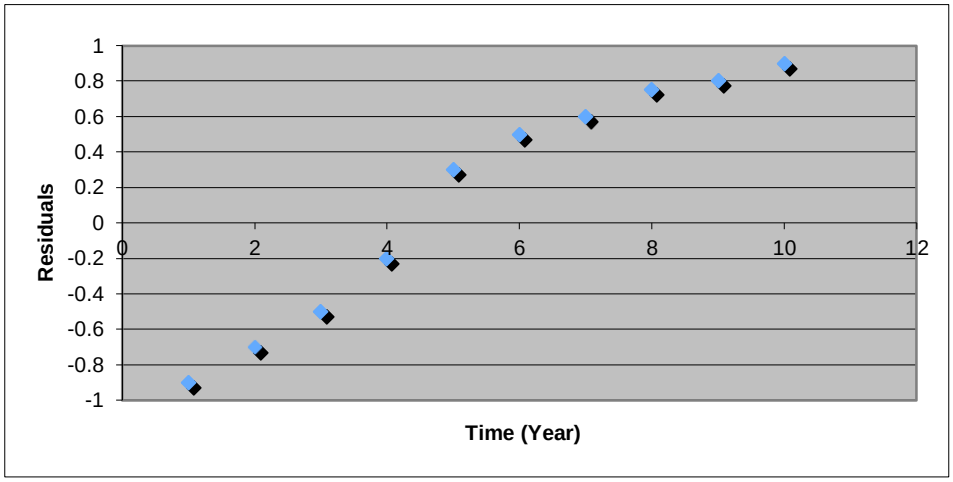
(1.20) The model $E(y) = \beta_0 + \beta_1 x$ is not suitable when

a. the relationship between y and x is linear.
b. the relationship between y and x is nonlinear.
c. the coefficients β_0 and β_1 are independent of one another.
d. the coefficients β_0 and β_1 are dependent of one another.
e. None of the others.

(1.21) Lily Energy Systems manufactures wood-burning heaters and fireplace inserts. One of its systems has an electric blower, which is thermostatically controlled. The blower is designed to automatically turn on when the temperature in the stove reaches 125 degrees F. and turn off at 85 degrees F. Complaints from customers indicate that the thermostat control is not working properly. The company feels that the thermostat is acceptable if the variance in the cutoff temperature is less than or equal to 175. The company takes a sample of 24 thermostats and finds that the variance equals 289. The calculated chi-square test statistic and the table value for a 0.05 significance level are:

a. 35.172, 38.99.
b. 37.983, 35.172.
c. 38.076, 38.99.
d. 37.983, 38.076.
e. none of these.

(1.22) After estimating a trend model for annual time-series data, you obtain the following residual plot against time. The problem with your model is that



a. the cyclical component has not been accounted for.
b. the seasonal component has not been accounted for.
c. the trend component has not been accounted for.
d. the irregular component has not been accounted for.

(1.23) Referring to Scenario 3, what is the expected number of employees with less than 1 year of training time and a high defect rate?

a. 4.17
b. 4.60
c. 5.28
d. 9.17
e. None of the others

(1.24) The cyclical component of a time series

a. represents periodic fluctuations which reoccur within 1 year.
b. represents periodic fluctuations which usually occur in 2 or more years.
c. is obtained by adding up the seasonal indexes.
d. is obtained by adjusting for calendar variation.

(1.25) A Marine drill instructor recorded the time in which each of 11 recruits completed an obstacle course both before and after basic training. To test whether any improvement occurred, the instructor would use:

a. a t-distribution with 10 degrees of freedom
b. a t-distribution with 11 degrees of freedom
c. a t-distribution with 20 degrees of freedom
d. a chi-square distribution with 10 degrees of freedom
e. None of the others

(1.26) Referring to Scenario 5, what is the correct interpretation for the estimated coefficient for X_2?

a. Holding the effect of the distance from a quad constant, the estimated mean costs for parking on campus is $0.1552 per hour more than parking off campus.
b. Holding the effect of the distance from a quad constant, the estimated mean costs for parking on campus is $0.1552 per hour more than parking on campus.
c. Holding the effect of the distance from a quad constant, the estimated mean costs for parking on campus is $0.1552 per hour more than parking off campus for each additional block away from the quad.
d. Holding the effect of the distance from a quad constant, the estimated mean costs for parking on campus is $0.1552 per hour more than parking on campus for each additional block away from the quad.
e. None of the others.

(1.27) Using the confidence interval when conducting a two-tail test for the population mean μ we do not reject the null hypothesis if the hypothesized value for μ:

a. is to the left of the lower confidence limit.
b. is to the right of the upper confidence limit.
c. falls between the lower and upper confidence limits.
d. falls in the rejection region.
e. None of the others

(1.28) A spouse stated that the average amount of money spent on Christmas gifts for immediate family members is above $1200. The correct set of hypotheses is:

a. $H_0: \mu = 1200, H_1: \mu < 1200$
b. $H_0: \mu > 1200, H_1: \mu = 1200$
c. $H_0: \mu = 1200, H_1: \mu > 1200$
d. $H_0: \mu < 1200, H_1: \mu = 1200$
e. None of the others

(1.29) The amount of time required to reach a customer service representative has a huge impact on the customer satisfaction. For this study, data were collected on the amount of time required to reach a customer service representative for two different hotels. State the null and alternative hypotheses for testing if there is evidence of a difference in the variabilities of the amount of time required to reach a customer service representative between the two hotels.

a. $H_0: \sigma_1^2 - \sigma_2^2 \geq 0$ versus $H_1: \sigma_1^2 - \sigma_2^2 < 0$
b. $H_0: \sigma_1^2 - \sigma_2^2 \leq 0$ versus $H_1: \sigma_1^2 - \sigma_2^2 > 0$
c. $H_0: \sigma_1^2 - \sigma_2^2 = 0$ versus $H_1: \sigma_1^2 - \sigma_2^2 \neq 0$
d. $H_0: \sigma_1^2 - \sigma_2^2 \neq 0$ versus $H_1: \sigma_1^2 - \sigma_2^2 = 0$
e. None of the others

(1.30) Referring to Scenario 6, what is the value of the test statistic for testing whether there is an upward curvature in the response curve relating the demand (Y) and the price (X)?

a. -5.14
b. 0.95
c. 373
d. 0.3647
e. 29.64

(1.31) Referring to Scenario 6, does there appear to be significant upward curvature in the response curve relating the demand (Y) and the price (X) at 10% level of significance?

a. Yes, since the p-value for the test is less than 0.10.
b. No, Since the p-value of β_2 is near 0.
c. No, since the p-value for the test is greater than 0.10.
d. Yes, since the value of β_2 is positive.
e. None of the others.

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS & STATISTICS
DHAHRAN, SAUDI ARABIA

STAT 212: Statistics for Business II
Semester 161
Final Exam (Code 2)
Wednesday January 18, 2016
8:00 am

Please circle your instructor's name:

Abbas Al-Sawi

Name: ID #: Section #: Serial #:

Question No	Full Marks	Marks Obtained
Part I: MCQ	**62**	
PART II		
2	**04**	
3	**12**	
4	**10**	
5	**06**	
Total	**94**	

Question	ANSWER				
	a	b	c	d	e
1				d	
2		b			
3				d	
4				d	
5				d	
6			c		
7				d	
8				d	
9					e
10			c		
11		b			
12	a				
13		b			
14				d	
15			c		
16		b			
17				d	
18		b			
19		b			
20		b			
21		b			
22			c		
23			c		
24		b			
25	a				
26	a				
27			c		
28			c		
29			c		
30		b			
31			c		

Q.No.2:- The actual number of help-wanted ads in a local newspaper was 1682, 1963, 2451, and 3205 for quarters I through IV, respectively, of the preceding year. The corresponding deseasonalized values are 2320, 2362, 2205, and 2414, respectively. Based on this information, what seasonal index is associated with each quarter?

Q.No.3:- Referring to Scenario 8, answer the following questions.

a. Fit a linear trend forecasting equation.

b. Fit a quadratic trend forecasting equation.

c. Fit an exponential trend forecasting equation

d. Using the forecasting equations in (a) through (c), what are your annual forecasts of the number of stores open for 2008?

Linear Trend Forecast	Quadratic Trend Forecast	Exponential Trend Forecast

e. What forecast from (d) do you think you should use? Why?

f. Fit a third-order autoregressive model to the number of stores and test for the significance of the third-order autoregressive parameter. (Use $\alpha = 0.05$.)

g. If necessary, fit a second-order autoregressive model to the number of stores and test for the significance of the second-order autoregressive parameter. (Use $\alpha = 0.05$.)

h. If necessary, fit a first-order autoregressive model to the number of stores and test for the significance of the first-order autoregressive parameter. (Use $\alpha = 0.05$.)

i. Forecast the number of stores open in 2008 using the appropriate autoregressive model.

Q.No.4:- Given below are the average prices for three types of energy products in the United States from 1992 to 1995.

Year	Electricity	Natural Gas	Fuel Oil
1992	43.205	25.893	0.892
1993	16.959	28.749	0.969
1994	47.202	28.933	1.034
1995	48.874	29.872	0.913
1996	48.693	28.384	0.983

a. What is the simple price indexes for electricity in 1994 using 1993 as the base year? Interpret your result.

b. What is the unweighted aggregate price index for the group of three energy items in 1994 using 1993 as the base year? Interpret your result.

c. What is the Paasche price index for the group of three energy items in 1994 for a family that consumed 13 units of electricity, 26 unites of natural gas and 235 units of fuel oil in 1994 using 1992 as the base year? Interpret your result.

d. What is the Laspeyres price index for the group of three energy items in 1994 for a family that consumed 15 units of electricity, 24 units of natural gas and 200 units of fuel oil in 1992 using 1992 as the base year? Interpret your result.

Q.No.5:- Referring to Scenario 2,

a. the data provide the information for testing the hypothesis that the proportion of unacceptable is the same for the four employees. Calculate the test statistic.

b. what are the values of the absolute difference and the critical range for the Marascuilo procedure to test for the difference in proportions between Employee 2 and Employee 4?

Scenario 1: The following table contains the number of complaints received in a department store for the first 6 months of last year.

Month	Complaints
January	36
February	45
March	81
April	90
May	108
June	144

Scenario 2: Four employees are monitored to determine whether there is any difference in the proportions of acceptable parts produced by the employees. The sample of parts produced is given below.

Expected counts are printed below observed counts

	Employee1	Employee2	Employee3	Employee4	Total
Acceptable	265 261.25	270 261.25	225 228.00	190 199.50	950
Unacceptable	10 13.75	5 13.75	15 12.00	20 10.50	50
Total	275	275	240	210	1000

Scenario 3: One criterion used to evaluate employees in the assembly section of a large factory is the number of defective pieces per 1,000 parts produced. The quality control department wants to find out whether there is relationship between years of experience and defect rate. Since the job is repetitious, after the initial training period any improvement due to a learning effect might be offset by a loss of motivation. A defect rate is calculated for each worker in a yearly evaluation. The results for 100 workers are given in the table below.

		Years Since Training Period		
		< 1 Year	1 – 4 Years	5 – 9 Years
Defect Rate:	High	6	9	9
	Average	9	19	23
	Low	7	8	10

Scenario 4: An economist is interested to see how consumption for an economy (in $ billions) is influenced by gross domestic product ($ billions) and aggregate price (consumer price index). The output of this regression is partially reproduced below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.991
R Square	0.982
Adjusted R Square	0.976
Standard Error	0.299
Observations	10

ANOVA

	df	SS	MS	*F*	*Signif F*
Regression	2	33.4163	16.7082	186.325	0.0001
Residual	7	0.6277	0.0897		
Total	9	34.0440			

	Coeff	StdError	*t Stat*	*P-value*
Intercept	– 0.0861	0.5674	– 0.152	0.8837
GDP	0.7654	0.0574	13.340	0.0001
Price	– 0.0006	0.0028	– 0.219	0.8330

Scenario 5: As a project for his business statistics class, a student examined the factors that determined parking meter rates throughout the campus area. Data were collected for the price ($) per hour of parking, blocks to the quadangle, and whether the parking is on or off campus. The population regression model hypothesized is $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon$

where
Y is the meter price per hour
X_1 is the number of blocks to the quad
X_2 is a dummy variable that takes the value 1 if the meter is located on campus and 0 otherwise.

The following results are obtained.

Regression Statistics	
Multiple R	0.5536
R Square	0.3064
Adjusted R Square	0.2812
Standard Error	0.4492
Observations	58

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	4.9035	2.4518	12.1501	0.0000
Residual	55	11.0984	0.2018		
Total	57	16.0019			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 99%*	*Upper 99%*
Intercept	1.6500	0.2028	8.1359	0.0000	1.1089	2.1912
Block	-0.2504	0.0529	-4.7355	0.0000	-0.3915	-0.1093
Campus	0.1552	0.1297	1.1966	0.2366	-0.1908	0.5011

Scenario 6: A certain type of rare gem (an expensive, polished and engraved stone) serves as a status symbol for many of its owners. In theory, for low prices, the demand increases and it decreases as the price of the gem increases. However, experts hypothesize that when the gem is valued at very high prices, the demand increases with price due to the status owner believe they gain in obtaining the gem. Thus, the model proposed to best explain the demand for the gem by its price is a quadratic model:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$$

where Y = demand (in thousands) and X = retail price per carat.

This model was fit to data collected for a sample of 12 rare gems of this type. A portion of the computer analysis is shown below:

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.994
R Square	0.988
Standard Error	12.42
Observations	12

ANOVA

	df	SS	MS	*F*	*Signif F*
Regression	2	115145	57573	373	0.0001
Residual	9	1388	154		
Total	11	116533			

	Coeff	StdError	*t Stat*	*P-value*
Intercept	286.42	9.66	29.64	0.0001
Price	−0.31	0.06	−5.14	0.0006
Price Sq	0.000067	0.00007	0.95	0.3647

Scenario 7: 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 2011 to 2013. The following is the resulting regression equation:

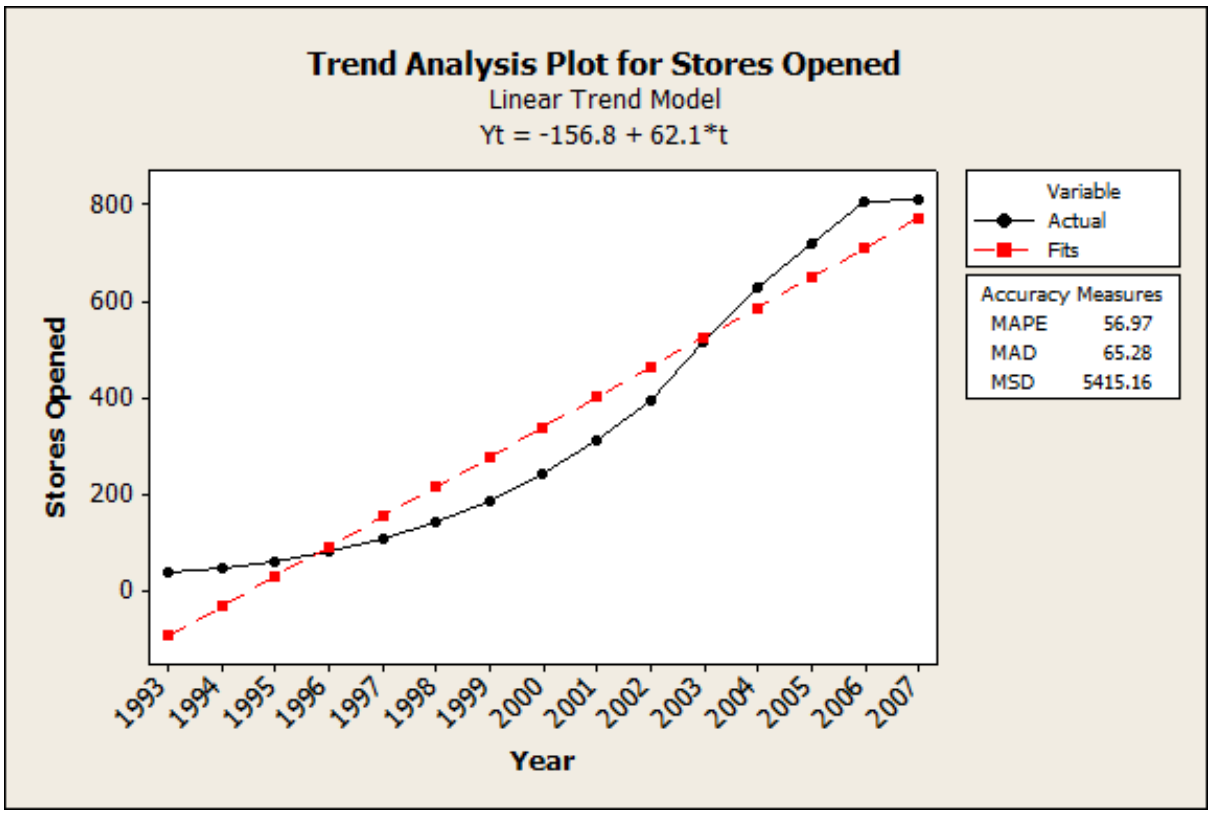
$$\log \hat{Y} = 3.37 + 0.117X - 0.083Q_1 + 1.28Q_2 + 0.617Q_3$$

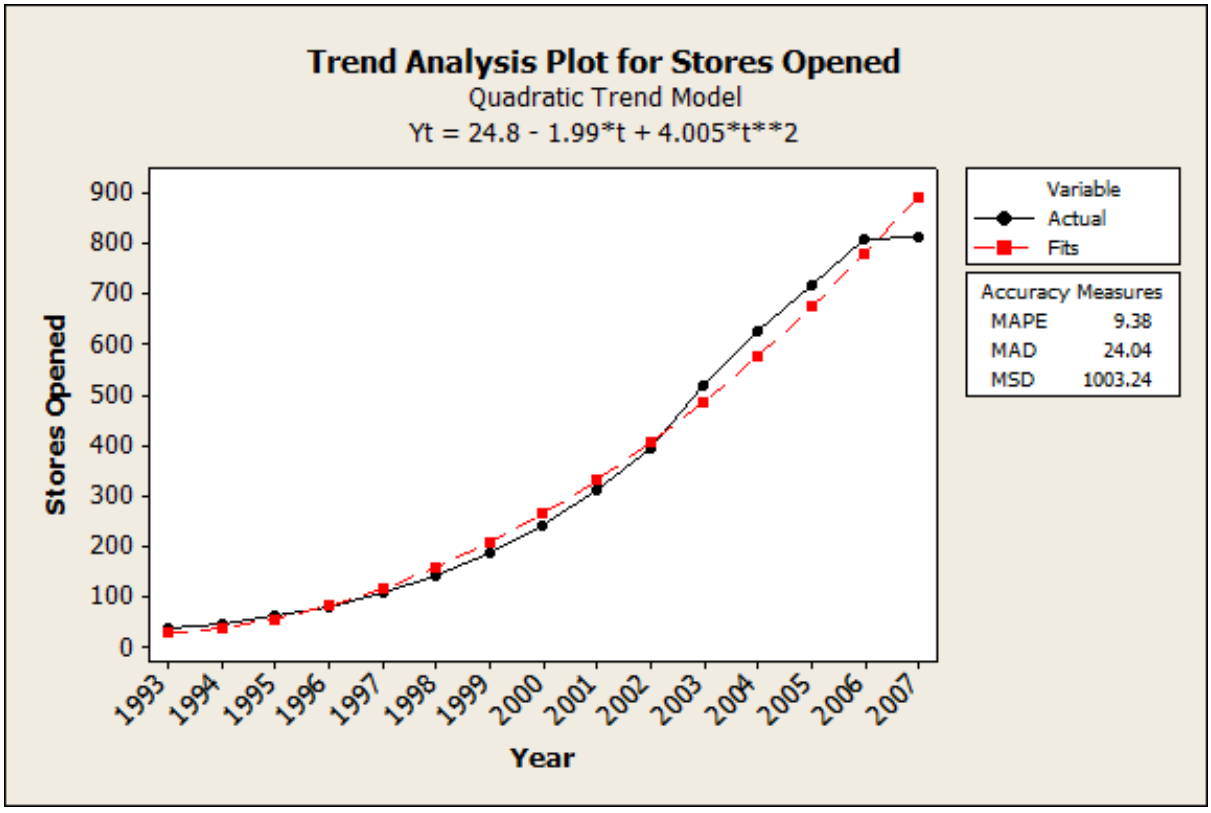
where $\hat{Y}$ is the estimated number of contracts in a quarter, X is the coded quarterly value with $X = 0$ in the first quarter of 2011. Q_1 is a dummy variable equal to 1 in the first quarter of a year and 0 otherwise. Q_2 is a dummy variable equal to 1 in the second quarter of a year and 0 otherwise. Q_3 is a dummy variable equal to 1 in the third quarter of a year and 0 otherwise.

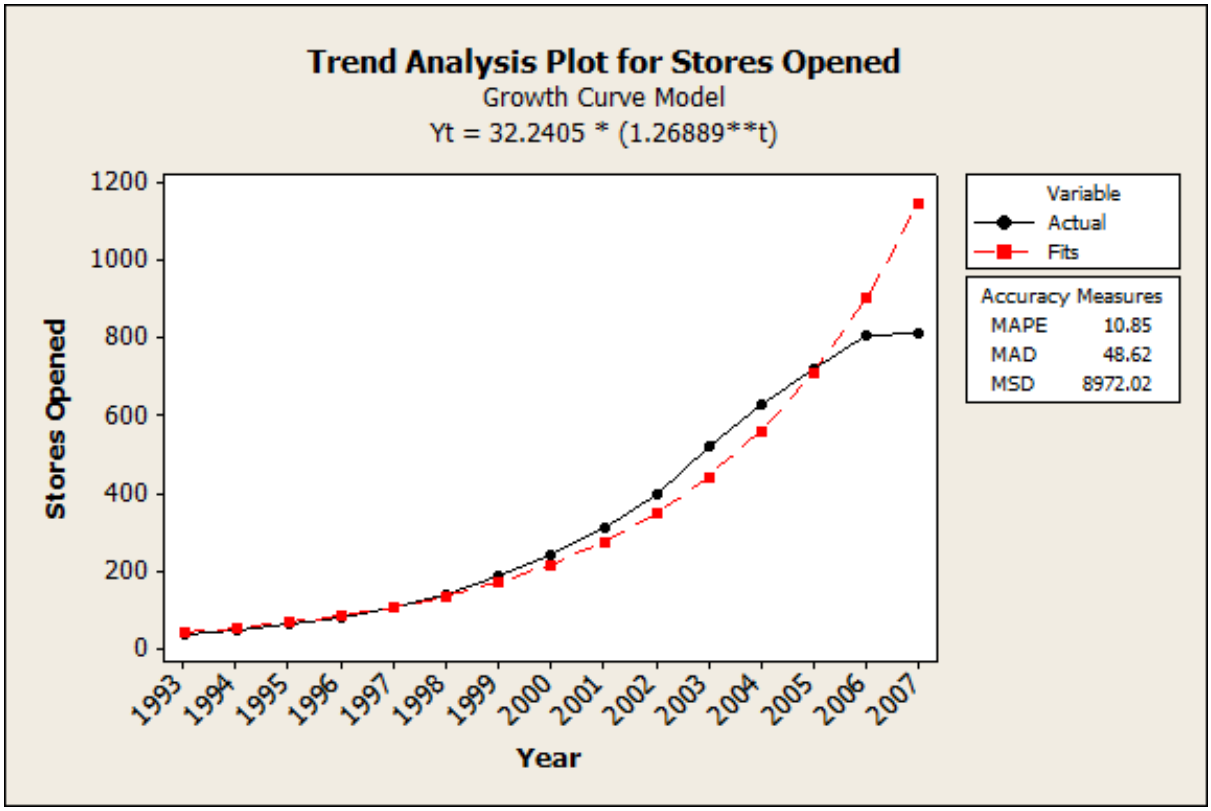
Scenario 8: Bed, Bath & Beyond is a nationwide chain of retail stores that sell a wide assortment of merchandise, including domestics merchandise and home furnishings, as well as food, giftware, and health and beauty care items. The data show the number of stores open at the end of the fiscal year from 1993 to 2007:

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Coded Year	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Stores Opened	38	45	61	80	108	141	186	241	311	396	519	629	721	809	815

A portion of the computer analysis is shown below:







Regression Analysis: Stores Opened versus Yi-1, Yi-2, Yi-3, Yi-4

The regression equation is
Stores Opened = 14.6 + 2.01 Yi-1 - 0.900 Yi-2 + 1.20 Yi-3 - 1.80 Yi-4

11 cases used, 4 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	14.59	11.14	1.31	0.238
Yi-1	2.0137	0.3698	5.45	0.002
Yi-2	-0.9002	0.8569	-1.05	0.334
Yi-3	1.199	1.082	1.11	0.310
Yi-4	-1.7996	0.7876	-2.28	0.062

S = 15.8925 R-Sq = 99.8% R-Sq(adj) = 99.6%

Regression Analysis: Stores Opened versus Yi-1, Yi-2, Yi-3

The regression equation is
Stores Opened = 12.9 + 1.95 Yi-1 - 0.182 Yi-2 - 0.999 Yi-3

12 cases used, 3 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	12.91	11.70	1.10	0.302
Yi-1	1.9466	0.4377	4.45	0.002
Yi-2	-0.1825	0.9526	-0.19	0.853
Yi-3	-0.9986	0.6109	-1.63	0.141

S = 18.9551 R-Sq = 99.7% R-Sq(adj) = 99.5%

Regression Analysis: Stores Opened versus Yi-1, Yi-2

The regression equation is
Stores Opened = 10.8 + 2.50 Yi-1 - 1.65 Yi-2

13 cases used, 2 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	10.76	10.45	1.03	0.328
Yi-1	2.4994	0.2706	9.24	0.000
Yi-2	-1.6507	0.3088	-5.35	0.000

S = 19.5823 R-Sq = 99.6% R-Sq(adj) = 99.5%

Regression Analysis: Stores Opened versus Yi-1

The regression equation is
Stores Opened = 34.9 + 1.07 Yi-1

14 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	34.86	15.15	2.30	0.040
Yi-1	1.06743	0.03795	28.13	0.000

S = 36.3824 R-Sq = 98.5% R-Sq(adj) = 98.4%