Tutorial letter 201/1/2018

Forecasting II

STA2604

Semester 1

Department of Statistics

Solutions to Assignment 1 and 2,
Trial exam paper and solutions

UNISA university of south africa

STA2604
Forecasting II
Solutions to Assignment 01

QUESTION 1

(1.1) Forecasting is the act of predicting future events and conditions. Here are five examples:

- Education: forecasting the number of new graduates in order to plan job opportunities.
- Human Resources: forecasting the number of retirees in an institution for replacement plan.
- Demography: forecasting the size of a population in a certain region for infrastructure planning.
- Business: forecasting the demand for a certain product in a shop for supply planning.
- Environment: forecasting weather and climatic conditions in a certain region for planning safety measures.

(1.2) The four components of a time series are the following:

- Trend: describes the upward or downward change for a time series during a given period of time.
- Seasonal variations: describes periodic patterns in a time series during a year and repeated in subsequent years.
- Cyclical variations: describes upward or downward changes around the trend that may occur after several years (e.g. five years).
- Irregular variations: describes changes in time series that follow no known (predictable) pattern.

(1.3) The difference between cross-sectional and time series data is that cross-sectional data are observed at a single time point while times series data are observed in a sequence of time points (i.e. two or more time points).

(1.4) Let y_t and $\widehat{y}_t$ denote the actual and predicted number of customers, respectively, in the table below:

Day	y_t	$\widehat{y}_t$	e_t	$\lvert e_t \rvert$	e_t^2
Monday	100	125	-25	25	625
Tuesday	120	110	10	10	100
Wednesday	130	140	-10	10	100
Thursday	123	136	-13	13	169
Friday	150	155	-5	5	25
Total				63	1019

(1.4.1) The forecast error e_t for each day is given in the fourth column of the table.

(1.4.2) MAD $= \dfrac{\sum_{t=1}^{5} |e_t|}{5} = \dfrac{63}{5} = 12.6.$

(1.4.3) MSE $= \dfrac{\sum_{t=1}^{5} e_t^2}{5} = \dfrac{1019}{5} = 203.8.$

(1.4.4) We have to compare the MAD and the root mean square error since MAD and MSE are not in the same unit. $\sqrt{MSE} = \sqrt{203.8} = 14.3$ which is slightly larger than the MAD. Therefore, the MAD would be recommended. The MAD is often recommended for skewed data as in this case where prediction errors are so disproportional.

QUESTION 2

(2.1) First construct a table with necessary information for all parts of the question.

t	y_t	ty_t	t^2	$\widehat{y}_t$	$e_t = y_t - \widehat{y}_t$	e_t^2	$d_t = (e_t - e_{t-1})^2$	$(y_t - \bar{y}_t)^2$
1	298	298	1	297.5	0.5	0.25		6872.41
2	300	600	4	306.3	-6.3	39.69	46.24	6544.81
3	300	900	9	315.1	-15.1	228.01	77.44	6544.81
4	350	1400	16	323.9	26.1	681.21	1697.44	954.81
5	340	1700	25	332.7	7.3	53.29	353.44	1672.81
6	360	2160	36	341.5	18.5	342.25	125.44	436.81
7	400	2800	49	350.3	49.7	2470.09	973.44	364.81
8	350	2800	64	359.1	-9.1	82.81	3457.44	954.81
9	360	3240	81	367.9	-7.9	62.41	1.44	436.81
10	346	3460	100	376.7	-30.7	942.49	519.84	1218.01
11	356	3916	121	385.5	-29.5	870.25	1.44	620.01
12	370	4440	144	394.3	-24.3	590.49	27.04	118.81
13	400	5200	169	403.1	-3.1	9.61	449.44	364.81
14	390	5460	196	411.9	-21.9	479.61	353.44	82.81
15	420	6300	225	420.7	-0.7	0.49	449.44	1528.81
16	410	6560	256	429.5	-19.5	380.25	353.44	846.81
17	460	7820	289	438.3	21.7	470.89	1697.44	6256.81
18	458	8244	324	447.1	10.9	118.81	116.64	5944.41
19	470	8930	361	455.9	14.1	198.81	10.24	7938.81
20	480	9600	400	464.7	15.3	234.09	1.44	9820.81
210	7618	85828	2870			8255.8	10712.16	59523.8

The values of $\widehat{\beta}_1$ and $\widehat{\beta}_0$ are the following:

$$\widehat{\beta}_1 = \frac{20\sum_{t=1}^{20} ty_t - \sum_{t=1}^{20} t \sum_{t=1}^{20} y_t}{20\sum_{t=1}^{20} t^2 - \left(\sum_{t=1}^{20} t\right)^2} = \frac{20 \times 85828 - 210 \times 7618}{20 \times 2870 - (210)^2} = \frac{116780}{13300} = 8.7805 \simeq 8.8$$

$$\widehat{\beta}_0 = \bar{y}_t - \widehat{\beta}_1 \bar{t} = \frac{\sum_{t=1}^{20} y_t - \widehat{\beta}_1 \sum_{t=1}^{20} t}{20} = \frac{7618 - 8.7805 \times 210}{20} = \frac{5774.095}{20} = 288.7048 \simeq 288.7$$

Thus, the fitted model is: $\widehat{y}_t = 288.7 + 8.8t$.

(2.2) The point forecast at month 21 is: $\widehat{y}_{21} = 288.7 + 8.8 \times 21 = 473.5$. If there was a seasonal index, this result should be multiplied by that index. The computation of the 95% prediction interval needs some extra calculations. The fitted model and the values of t in the first column were used to fill in the elements of the fifth column. The residuals e_t in the sixth column are $e_t = y_t - \widehat{y}_t$. The squared residuals e_t^2 may now be used to calculate the mean square residual as: $MSE = \frac{e_t^2}{n-1}$ and thus its square root is $s = \sqrt{\frac{e_t^2}{n-1}} = \sqrt{\frac{8255.8}{19}} = 20.85$. The D distance is: $D = \frac{1}{n} + \frac{(t_0-\bar{t})^2}{SS_{tt}}$ where $t_0 = 21$, $\bar{t} = \frac{210}{20} = 10.5$, and

$$SS_{tt} = \sum_{t=1}^{20}(t-\bar{t})^2 = \sum_{t=1}^{20} t^2 - \frac{\left(\sum_{t=1}^{20} t\right)^2}{20} = 2870 - \frac{210^2}{20} = 665.$$

Hence,

$$D = \frac{1}{n} + \frac{(t_0-\bar{t})^2}{SS_{tt}} = \frac{1}{20} + \frac{(21-10.5)^2}{665} = 0.2158.$$

The 95% prediction interval is: $[\widehat{y}_t \pm t_{[0.025]}^{(20-2)} s\sqrt{1+D}]$. That is: $[473.5 \pm t_{[0.025]}^{(18)} 20.85\sqrt{1+0.2158}]$ where $t_{[0.025]}^{(18)} = 2.101$. Hence, the required 95% prediction interval is: $[425.20; 521.80]$.

(2.3) The null and alternative hypotheses are:
H_0: The error terms are not autocorrelated.
H_a: The error terms are positively correlated.
The Durbin-Watson test statistic for positive autocorrelation is:

$$d = \frac{\sum_{t=2}^{20}(e_t - e_{t-1})^2}{\sum_{t=1}^{20} e_t^2} = \frac{10712.16}{8255.8} = 1.30.$$

We will reject H_0 if $d < d_{L,0.05}$, do not reject H_0 if $d > d_{L,0.05}$ and fail to conclude H_0 if $d_{L,0.05} < d <_{U,0.05}$. For $n = 20$, we have: $d_{L,0.05} = 1.20$ and $d_{U,0.05} = 1.41$. Since $d_{L,0.05} = 1.20 < d = 1.30 <_{U,0.05} = 1.41$, we fail to conclude.

(2.4) The mean sales is $\bar{y}_t = \frac{\sum_{t=1}^{20} y_t}{20} = 380.9$, and the total variation is: $\sum_{t=1}^{20}(y_t - \bar{y}_t)^2 = 59523.8$.

We found that the total explained variation was $\sum_{t=1}^{20}(y_t - \widehat{y}_t)^2 = \sum_{t=1}^{20} e_t^2 = 8255.8$. Thus, the coefficient of determination is:

$$R^2 = \frac{\sum_{t=1}^{20}(y_t - \bar{y}_t)^2}{\sum_{t=1}^{20}(y_t - \widehat{y}_t)^2} = \frac{8255.8}{59523.8} = 0.1387.$$

The adjusted coefficient of determination is:

$$R^2_{\text{adj}} = \bar{R}^2 = \left(R^2 - \frac{k}{n-1}\right)\left(\frac{n-1}{n-(k+1)}\right) = \left(0.1387 - \frac{1}{19}\right)\frac{19}{18} = 0.0909.$$

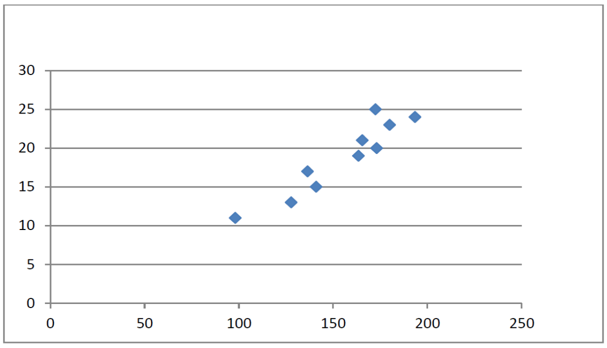
(2.5) There is no point of calculating a VIF here since there is only one predictor variable, and thus we are not able to regress it on another predictor variable. If we assume that y_t is also a predictor variable (which is not the case), then

$$VIF = \frac{1}{1-R^2} = \frac{1}{1-0.1387} = 1.1610.$$

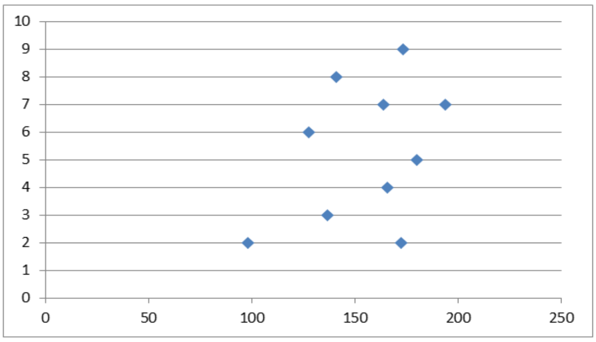
Since VIF is close to one (unit), then y_t is not related to t.

QUESTION 3

(3.1) Plots of y versus x_1, then y versus x_2.



Plot of y versus x1



Plot of y versus x2

(3.2) The plots in part (3.1) indicate that y is linearly related with x_1 and somehow linearly (weakly) related with x_2. With this information model (E1) is a reasonable model for these data.

(3.3) Model (obtained using Excel).

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9952
R Square	0.9905
Adjusted R Square	0.9877
Standard Error	3.1851
Observations	10

Note that overall, the model fits the data since R^2 and R^2_{adj} are large. However, this information is often misleading. Residual analysis is a more suitable approach.

ANOVA

	df	SS	MS	F	Significance F
Regression	2	7368.0699	3684.0349	363.1375	0.0000
Residual	7	71.0151	10.1450		
Total	9	7439.0850			

Here, the ANOVA table again confirms that overall, the model fits the data since F is large ($F = 363.1375$) with $p < 0.0001$. However, again this information is often misleading. Residual analysis is a more suitable approach. The table of parameter estimates is given below, and indicates that both x_1 (size) and x_2 (rating) have positive significant effects on y (price) since all p-values are very small ($\alpha = 0.05$ is often used when no significance level is indicated).

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	29.3419	4.8062	6.1050	0.0005	17.9771	40.7067
X Variable 1	5.6082	0.2245	24.9768	0.0000	5.0773	6.1392
X Variable 2	3.8441	0.4256	9.0311	0.0000	2.8376	4.8506

Hence the predictive model is:

$$\widehat{y} = 29.3419 + 5.6082x_1 + 3.8441x_2.$$

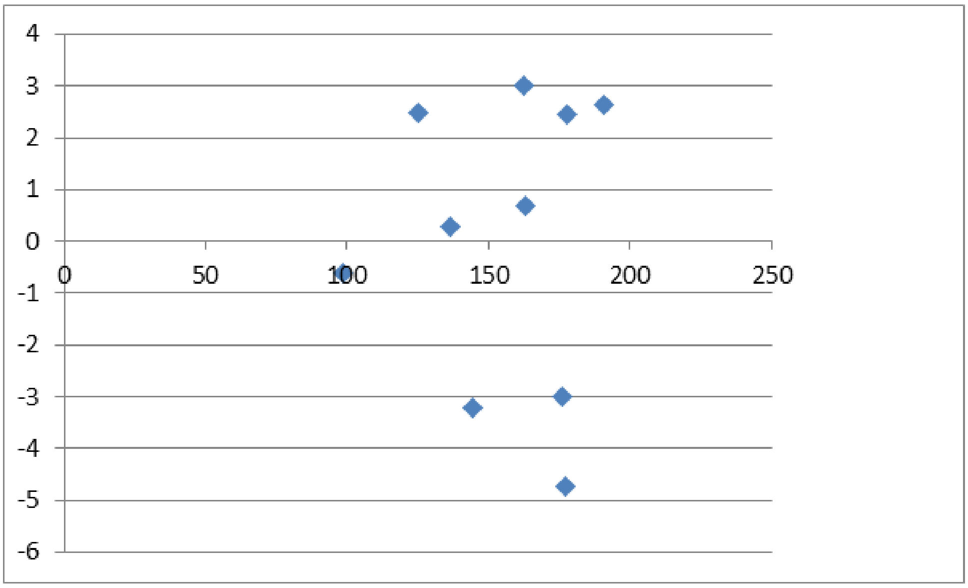
(3.4) Residuals are calculated as $e = y - \widehat{y}$ where $\widehat{y} = 29.3419 + 5.6082x_1 + 3.8441x_2$.
Manual calculation in Excel (as requested in the question) gives the following table:

y	x_1	x_2	$\widehat{y}$	Residuals ($e = y - \widehat{y}$)
180	23	5	177.551	2.449
98.1	11	2	98.7203	-0.6203
173.1	20	9	176.1028	-3.0028
136.5	17	3	136.2136	0.2864
141	15	8	144.2177	-3.2177
165.5	21	4	162.4905	3.0095
193.5	24	7	190.8474	2.6526
127.8	13	6	125.3131	2.4869
163.5	19	7	162.8064	0.6936
172.5	25	2	177.2351	-4.7351

If an option to calculate residuals was ticked in Excel, we could obtain the following output (correct answer but the aim of the question was to assess if you know how the residuals are computed):

Observation	Predicted Y	Residuals
1	177.5513	2.4487
2	98.7205	-0.6205
3	176.1029	-3.0029
4	136.2138	0.2862
5	144.2177	-3.2177
6	162.4908	3.0092
7	190.8477	2.6523
8	125.3132	2.4868
9	162.8066	0.6934
10	177.2355	-4.7355

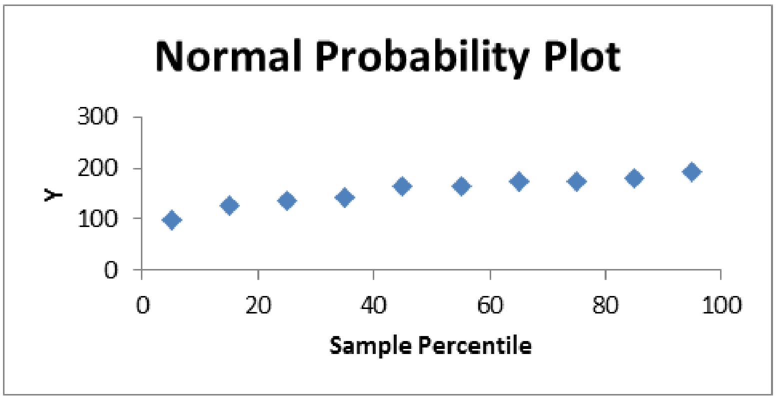
Plot of residuals versus fitted values:



Residuals (y axis) vs fitted values

(3.5) Clearly, the residuals are randomly scattered around the line $e = 0$. Thus, the assumption made in part (3.2) that model (E2) fits the data is supported by residual analysis.

(3.6) The following normal probability plot was obtained using Excel:



This graph indicates that the assumption of normality is not violated since the dots are approximately located on a straight line.

(3.7) The following output was obtained by regressing x_2 on x_1.

Regression Statistics	
Multiple R	0.0433
R Square	0.0019
Adjusted R Square	-0.1229
Standard Error	2.6456
Observations	10

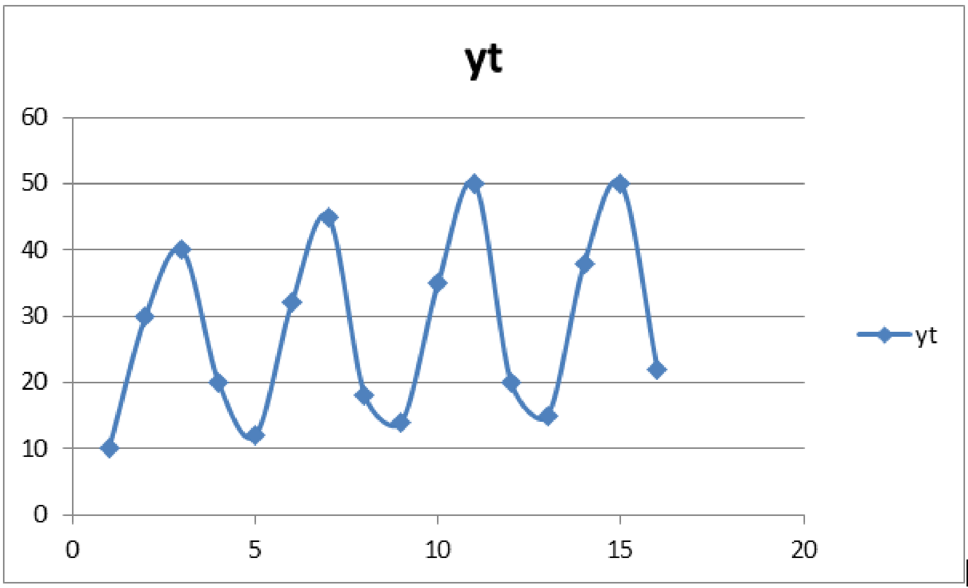
$VIF_2 = \frac{1}{1-R^2} = \frac{1}{1-0.0019} = 1.0019036 \simeq 1$. Since $VIF_2 \simeq 1$, then the variable x_2 is not related to x_1. We conclude that there is no evidence on multicollinearity (colinearity) in these data.

QUESTION 4

(4.1) The dummy variables are defined in the following table:

Quarter	Q_2	Q_3	Q_4
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

(4.2) The plot of y_t versus t (with $t = 1, 2, \ldots, 16$) is given below:



(4.3) It appears to be an increasing seasonal variation.

(4.4) The output for the fitted model is the following:

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.9939
R Square	0.9879
Adjusted R Square	0.9835
Standard Error	1.7487
Observations	16

ANOVA					
	df	SS	MS	F	Significance F
Regression	4	2744.8	686.2	224.3984	0.0000
Residual	11	33.6375	3.0580		
Total	15	2778.4375			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	8.9438	1.1103	8.0553	0.0000	6.5000	11.3875
t	0.5438	0.0978	5.5624	0.0002	0.3286	0.7589
Q_2	20.4563	1.2404	16.4920	0.0000	17.7262	23.1863
Q_3	32.4125	1.2519	25.8911	0.0000	29.6571	35.1679
Q_4	5.6188	1.2708	4.4214	0.0010	2.8217	8.4158

This table indicates that sales significantly increase with time and there is increase seasonal variations for quarters 2, 3, 4 compared to quarter 1 (all p-values very small). The prediction model is:

$$\widehat{y} = 8.9438 + 0.5438t + 20.4563Q_2 + 32.4125Q_3 + 5.6188Q_4.$$

(4.4) $t = 17$ corresponds to the first quarter of year 4, that is $Q_2 = Q_3 = Q_4 = 0$, and thus the predicted model can be written as: $\widehat{y} = 8.9438 + 0.5438t$. Hence, the point forecast is $\widehat{y}_{17} = 8.9438 + 0.5438 \times 17 = 18.1884$.
From the regression ANOVA table, $s = \sqrt{MSE} = \sqrt{3.0580} = 1.7487$, and since there were 5 parameters to estimate, we have

$$t^{(n-p)}_{[\alpha/2]} = t^{(16-5)}_{[0.025]} = t^{(16-5)}_{[0.025]} = t^{(11)}_{[0.025]} = 2.201.$$

The mean value of t is:

$$\bar{t} = \frac{\sum_{t=1}^{16} y_t}{16} = \frac{1 + 2 + \cdots + 16}{16} = 8.5.$$

Then $(17-\bar{t})^2 = (17-8.5)^2 = 72.25$. Calculations for $\sum_{t=1}^{16}(t - \bar{t})^2$ can be read from the following table.

Year	Quarter	y_t	t	Q_2	Q_3	Q_4	$t-\bar{t}$	$(t-\bar{t})^2$
1	1	10	1	0	0	0	-7.5	56.25
1	2	30	2	1	0	0	-6.5	42.25
1	3	40	3	0	1	0	-5.5	30.25
1	4	20	4	0	0	1	-4.5	20.25
2	1	12	5	0	0	0	-3.5	12.25
2	2	32	6	1	0	0	-2.5	6.25
2	3	45	7	0	1	0	-1.5	2.25
2	4	18	8	0	0	1	-0.5	0.25
3	1	14	9	0	0	0	0.5	0.25
3	2	35	10	1	0	0	1.5	2.25
3	3	50	11	0	1	0	2.5	6.25
3	4	20	12	0	0	1	3.5	12.25
4	1	15	13	0	0	0	4.5	20.25
4	2	38	14	1	0	0	5.5	30.25
4	3	50	15	0	1	0	6.5	42.25
4	4	22	16	0	0	1	7.5	56.25
Total								340

Hence, the 95% prediction interval of y_{17} is thus: $\left[\widehat{y}_{17} \pm s t_{[0.025]}^{(11)} \sqrt{1+\frac{1}{16}+\frac{(17-\bar{t})^2}{\sum_{t=1}^{16}(t-\bar{t})^2}}\right]$, that is:

$$\left[18.1884 \pm 1.7487 \times 2.201\sqrt{1+\frac{1}{16}+\frac{72.25}{340}}\right] = [18.1884 \pm 4.3460] = [13.8424; 22.5344].$$

STA2604
Forecasting II
Solutions to Assignment 02

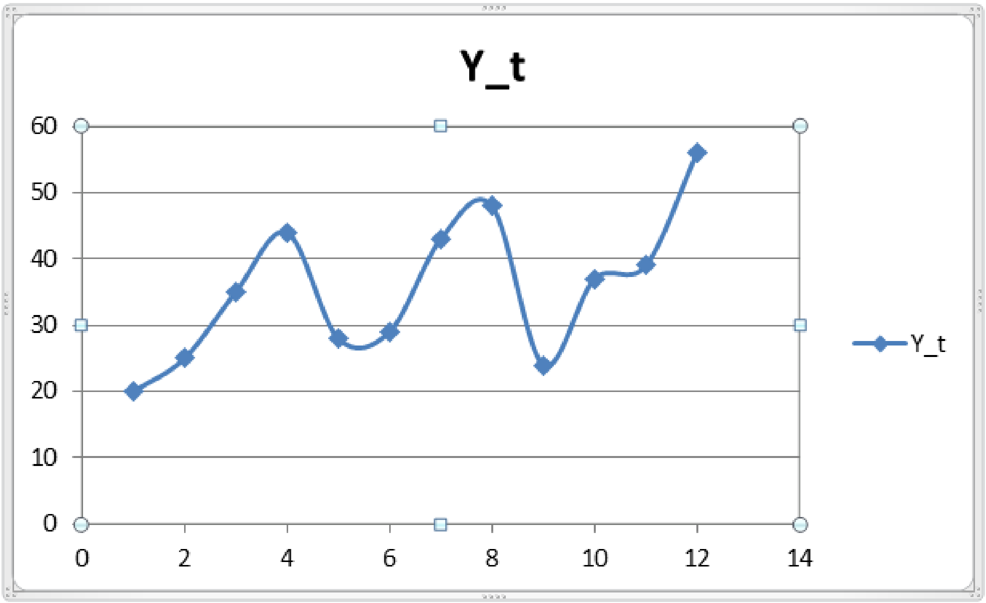
QUESTION 1

The multiplicative decomposition model is

$$y_t = TR_t \times SN_t \times CL_t \times IR_t$$

where TR_t, SN_t, CL_t and $IR-T$ are the trend, seasonal, cyclical and irregular factors, respectively.

(1.1) The plot of sales versus time is given by the following graph:



This graph indicates that the trend for the data appears to be increasing.

(1.2) The four-point moving averages are in given in the fifth column of Table 1.4.

Table 1.1: Analysis of Oligopoly data using multiplicative decomposition of a time series

Year	Quarter	t	y_t	4MA	CMA_t	$sn_t \times ir_t$	sn_t	d_t	tr_t	$\widehat{y}_t$	$cl_t \times irt_t$	cl_t	ir_t
1	1	1	20				0.71	28.09	25.86	18.41	1.09		
1	2	2	25				0.87	28.60	27.64	24.16	1.03	1.06	0.98
1	3	3	35	31	32	1.0938	1.12	31.16	29.43	33.05	1.06	1.06	1.00
1	4	4	44	33	33.5	1.3134	1.29	34.09	31.21	40.28	1.09	1.11	0.98
2	1	5	28	34	35	0.8	0.71	39.33	32.99	23.49	1.19	1.08	1.10
2	2	6	29	36	36.5	0.7945	0.87	33.17	34.78	30.40	0.95	1.06	0.90
2	3	7	43	37	36.5	1.1781	1.12	38.28	36.56	41.06	1.05	0.99	1.06
2	4	8	48	36	37	1.2973	1.29	37.19	38.34	49.49	0.97	0.95	1.02
3	1	9	24	38	37.5	0.64	0.71	33.71	40.12	28.57	0.84	0.94	0.89
3	2	10	37	37	38	0.9737	0.87	42.33	41.91	36.63	1.01	0.88	1.15
3	3	11	39	39			1.12	34.72	43.69	49.07	0.79	0.92	0.86
3	4	12	56				1.29	43.39	45.47	58.69	0.95		

(1.3) The centered moving averages, $CMA_t = tr_t \times cl_t$, are given in the sixth column of Table 1.4.

(1.4) The $sn_t \times ir_t = y_t/CMA_t$ values are given in the seventh column of Table 1.4.

(1.5) Steps for the calculation of seasonal indices are given in the following tables:

Year	Quarter			
	1	2	3	4
1	-	-	1.0938	1.3134
2	0.8	0.7945	1.1781	1.2973
3	0.64	0.9737	-	-
Total	1.44	1.7682	2.2719	2.6107
Mean (unadjusted seasonal indices)	0.72	0.8841	1.1359	1.3054

The total index is: $0.72 + 0.8841 + 1.1359 + 1.3054 = 4.0454$.

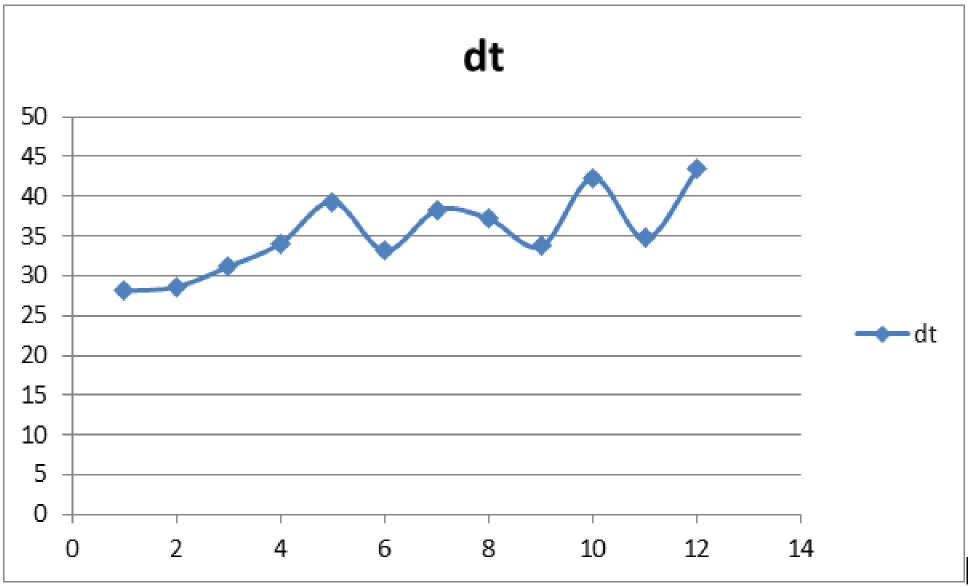
Adjusted seasonal indices (must sum to 4, the number of seasons; quarters here):

Quarter	Value
1	$\frac{0.72\times4}{4.0454} = 0.7119$
2	$\frac{0.8841\times4}{4.0454} = 0.8742$
3	$\frac{1.1359\times4}{4.0454} = 1.1232$
4	$\frac{1.3054\times4}{4.0454} = 1.2907$

These adjusted seasonal indices are copied in the eighth column of Table 1.4

(1.6) The deseasonalised observations, $d_t = y_t/sn_t$, are given in the ninth column of Table 1.4.

(1.7) The plots of these observations versus time is given in the following figure. The figure indicates an overall increasing trend.



(1.8) This can easily be done using Excel, but must be done by hand in an exam. We need to use a different table here. The last column of the following table will be used in part (1.14), not here.

Year	Quarter	t	y_t	ty_t	t^2	$(t-\bar{t})^2$
1	1	1	20	20	17	30.25
1	2	2	25	50	4	20.25
1	3	3	35	105	9	12.25
1	4	4	44	176	16	6.25
2	1	5	28	140	25	2.25
2	2	6	29	174	36	0.25
2	3	7	43	301	49	0.25
2	4	8	48	384	64	2.25
3	1	9	24	216	81	6.25
3	2	10	37	370	100	12.25
3	3	11	39	429	121	20.25
3	4	12	56	672	144	30.25
Total		78	428	3037	650	143
Mean		6.5	35.6667			

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^{12} ty_t - \frac{1}{12}\left(\sum_{t=1}^{12} t\right)\left(\sum_{t=1}^{12} y_t\right)}{\sum_{t=1}^{12} t^2 - \frac{\left(\sum_{t=1}^{12} t\right)^2}{12}} = \frac{3037 - \frac{78\times 428}{12}}{650 - \frac{78^2}{12}} = 1.7832.$$

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1\bar{t} = \frac{1}{12}\sum_{t=1}^{12} y_t - \frac{1}{12}\widehat{\beta}_1\sum_{t=1}^{12} t = \frac{1}{12}\times 428 - \frac{1}{12}\times 1.7832\times 78 = 24.0758.$$

Hence the estimate of of the trend is: $tr_t = 24.0758 + 1.7832t$.

(1.9) The trend values tr_t are given in the tenth column of Table 1.4, then the values $\widehat{y}_t = tr_t \times sn_t$ are given in the eleventh column of Table 1.4. Finally, the values $cl_t \times ir_t$ are given in the twelfth column of Table 1.4.

(1.10) The cyclical factors calculated as the three-point moving averages

$$cl_t = \frac{cl_{t-1}ir_{t-1} + cl_t ir_t + cl_{t+1}ir_{t+1}}{3}$$

(see textbook, page 335) are given in the thirteenth column of Table 1.4.

(1.11) The irregular values, $ir_t = \dfrac{cl_t \times ir_t}{cl_t}$, are given in the fourteenth column of Table 1.4.

(1.12) All values of cl_t are close to 1, thus we cannot determine any well-define cycle for these data.

(1.13) At year 4, the values of t are 13, 14, 15 and 16 for the four quarters, respectively. The trend was found to be $tr_t = 24.0758 + 1.7832t$, and the four seasonal indices were found to be 0.7119, 0.8742, 1.1232 and 1.2907. Now using the prediction equation $\widehat{y}_t = tr_t \times sn_t$, we obtain:
$\widehat{y}_{13} = (24.0758 + 1.7832 \times 13) \times 0.7119 = 33.6$
$\widehat{y}_{14} = (24.0758 + 1.7832 \times 14) \times 0.8742 = 42.9$
$\widehat{y}_{15} = (24.0758 + 1.7832 \times 15) \times 1.1232 = 57.1$
$\widehat{y}_{16} = (24.0758 + 1.7832 \times 16) \times 1.2907 = 67.9$

(1.14) Here we first calculated prediction intervals of the trend at the four time points then we used the formula $[\widehat{y} \pm B_t(1-\alpha)]$ where $B_t(1-\alpha)$ is the error band in the prediction interval prediction

$[tr_t \pm B_t(1-\alpha)]$ (see textbook on pages 335-336). This time, we fit the trend using Excel to obtain the standard error.

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.5913
R Square	0.3496
Adjusted R Square	0.2846
Standard Error	9.1975
Observations	12

ANOVA					
	df	SS	MS	F	Significance F
Regression	1	454.7203	454.7203	5.3753	0.04289
Residual	10	845.9464	84.5946		
Total	11	1300.6667			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	24.0758	5.6607	4.2531	0.0017	11.4630	36.6886
X Variable 1	1.7832	0.7691	2.3185	0.0429	0.0695	3.4970

The mean value of t is: $\bar{t} = 6.5$.
For time 13, the error band is:

$$\begin{aligned} B_{13}(0.95) &= s \times t^{(12-2)}_{[0.25]} \sqrt{1 + \frac{1}{12} + \frac{(13-\bar{t})^2}{\sum_{t=1}^{12}(t-\bar{t})^2}} \\ &= 9.1975 \times 2.228 \sqrt{1 + \frac{1}{12} + \frac{(13-6.5)^2}{143}} \\ &= 24.0621 \end{aligned}$$

Therefore, the prediction interval is: $[33.6 \pm 24.1] = [9.5; 57.7]$. For time 14, the error band is:

$$\begin{aligned} B_{14}(0.95) &= s \times t^{(10)}_{[0.25]} \sqrt{1 + \frac{1}{12} + \frac{(14-\bar{t})^2}{\sum_{t=1}^{12}(t-\bar{t})^2}} \\ &= 9.1975 \times 2.228 \sqrt{1 + \frac{1}{12} + \frac{(14-6.5)^2}{143}} \\ &= 24.9017 \end{aligned}$$

Therefore, the prediction interval is: $[42.9 \pm 24.9] = [18; 67.8]$. For time 15, the error band is:

$$\begin{aligned} B_{15}(0.95) &= s \times t^{(10)}_{[0.25]} \sqrt{1 + \frac{1}{12} + \frac{(15-\bar{t})^2}{\sum_{t=1}^{12}(t-\bar{t})^2}} \\ &= 9.1975 \times 2.228 \sqrt{1 + \frac{1}{12} + \frac{(15-6.5)^2}{143}} \\ &= 25.8279 \end{aligned}$$

Therefore, the prediction interval is: $[57.1 \pm 25.8] = [31.3; 82.9]$.

For time 16, the error band is:

$$B_{16}(0.95) = s \times t^{(10)}_{[0.25]} \sqrt{1 + \frac{1}{12} + \frac{(16 - \bar{t})^2}{\sum_{t=1}^{12}(t - \bar{t})^2}}$$
$$= 9.1975 \times 2.228 \sqrt{1 + \frac{1}{12} + \frac{(16 - 6.5)^2}{143}}$$
$$= 26.8317$$

Therefore, the prediction interval is: $[67.9 \pm 26.8] = [41.1; 94.7]$.

Notice that these confidence intervals are very wide, and thus this approach cannot be reliable.

QUESTION 2

(2.1) The additive decomposition model is

$$y_t = TR_t + SN_t + CL_t + IR_t$$

where TR_t, SN_t, CL_t and $IR - T$ are the trend, seasonal, cyclical and irregular terms, respectively. All calculations will be done as in Question 1 by replacing the sign $\times$ by the sign $+$, and the sign / by the sign $-$.

Table 1.2: Analysis of Oligopoly data using multiplicative decomposition of a time series

Year	Quarter	t	y_t	4MA	CMA_t	$sn_t + ir_t$	sn_t	d_t	tr_t	$\widehat{y}_t$	$cl_t + irt_t$	cl_t	ir_t
1	1	1	20				-10.5	30.5	25.86	15.36	4.64		
1	2	2	25				-4.5	29.5	27.64	23.14	1.86	2.52	-0.67
1	3	3	35	31	32	3	4.5	30.5	29.43	33.93	1.07	1.74	-0.67
1	4	4	44	33	33.5	10	10.5	33.5	31.21	41.71	2.29	2.96	-0.67
2	1	5	28	34	35	-7	-10.5	38.5	32.99	22.49	5.51	2.17	3.33
2	2	6	29	36	36.5	-7.5	-4.5	33.5	34.78	30.28	-1.28	2.06	-3.33
2	3	7	43	37	36.5	6.5	4.5	38.5	36.56	41.06	1.94	-0.06	2.00
2	4	8	48	36	37	11	10.5	37.5	38.34	48.84	-0.84	-1.51	0.67
3	1	9	24	38	37.5	-13.5	-10.5	34.5	40.12	29.62	-5.62	-2.29	-3.33
3	2	10	37	37	38	-1	-4.5	41.5	41.91	37.41	-0.41	-5.07	4.67
3	3	11	39	39			4.5	34.5	43.69	48.19	-9.19	-3.19	-6.00
3	4	12	56				10.5	45.5	45.47	55.97	0.03		

(2.1.1) The plot of sales versus time is the same as in Question 1, part (1.1).

(2.1.2) The four-point moving averages are given in the fifth column of Table 1.2, but same as those in Table 1.4.

(2.1.2) The centered moving averages are given in the sixth column of Table 1.2, but same as those in Table 1.4.

(2.1.4) The $sn_t \times ir_t = y_t - CMA_t$ are in given in the seventh column of Table 1.2.

(2.1.5) Steps for the calculation of seasonal indices are given in the following tables:

Year	Quarter			
	1	2	3	4
1	-	-	3	10.5
2	-7	-7.5	6.5	11
3	-13.5	-1	-	-
Total	-20.5	-8.5	9.5	21.5
Mean (unadjusted seasonal indices)	-10.25	-4.25	4.75	10.75

The total index is: $-10.25 - 4.25 + 4.75 + 10.75 = 1$.

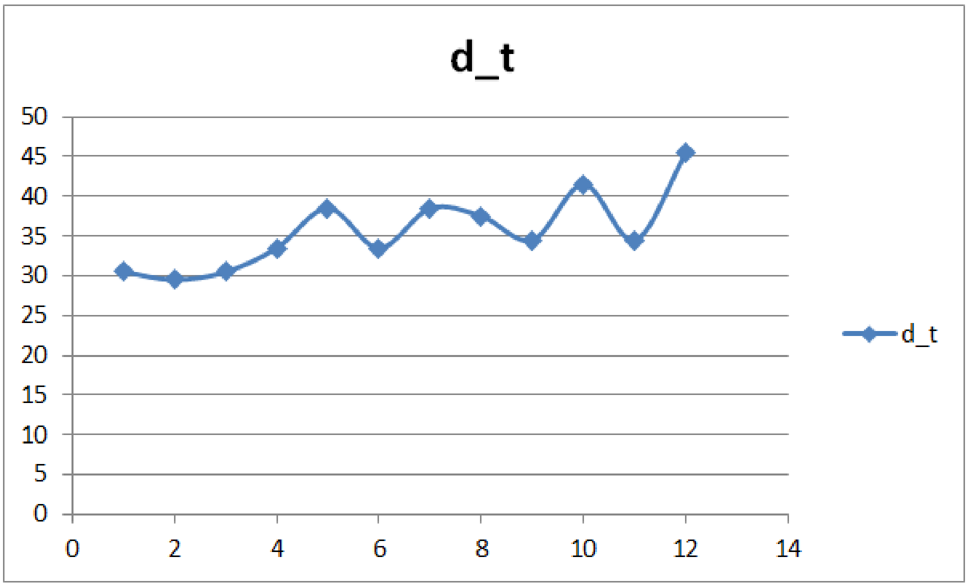
Adjusted seasonal or normalised indices (sum must be equal to zero):

Quarter	Value
1	$-10.25 - \frac{1}{4} = -10.5$
2	$-4.25 - \frac{1}{4} = -4.5$
3	$4.75 - \frac{1}{4} = 4.5$
4	$10.75 - \frac{1}{4} = 10.5$

These adjusted seasonal indices are copied in the eighth column of Table reft2

(2.1.6) The deseasonalised observations $d_t = y_t - sn_t$ are given in the ninth column of Table 1.2.

(2.1.7) The plots of these observations versus time is given in the following figure. The figure indicates an overall increasing trend.



(2.1.8) The estimated equation of the trend is the same as the one found in Question 1, part (1.8) since we are using the same model. That is: $tr_t = 24.0758 + 1.7832t$.
The trend values tr_t are given in the tenth column of Table 1.2.

(2.1.9) The point forecast values $\widehat{y}_t = tr_t + sn_t$ are given in the eleventh column of Table 1.2. Finally, the values $cl_t + ir_t$ are given in the twelfth column of Table 1.2.

(2.1.10) The cyclical terms calculated as the three-point moving averages

$$cl_t = \frac{(cl_{t-1} + ir_{t-1}) + (cl_t + ir_t) + (cl_{t+1} + ir_{t+1})}{3}$$

(see textbook, page 339) are given the thirteenth column of Table 1.2.

(2.1.11) The irregular values $ir_t = (cl_t + ir_t) - cl_t$ are given in the fourteenth column of Table 1.2.

(2.1.12) There is a well defined cycle in these data since the cl_t values largely vary, and also the irregular variations are in general different from zero for these data.

(2.1.13) At year 4, the values of t are 13, 14, 15 and 16 for the four quarters, respectively. The trend was found to be $tr_t = 24.0758 + 1.7832t$, and the four seasonal indices were found to be -10.5, -4.5, 4.5 and 10.5. Now using the prediction equation $\widehat{y}_t = tr_t \times sn_t$, we obtain:
$\widehat{y}_{13} = (24.0758 + 1.7832 \times 13) - 10.5 = 36.8$
$\widehat{y}_{14} = (24.0758 + 1.7832 \times 14) - 4.5 = 42.9$
$\widehat{y}_{15} = (24.0758 + 1.7832 \times 15) + 4.5 = 44.5$
$\widehat{y}_{16} = (24.0758 + 1.7832 \times 16) + 10.5 = 63.1$

(1.14) The error bands $B_t(1-\alpha)$ were calculated in Question 1, part (1.14). Hence the 95% prediction interval forecasts for year 4 are the following:
For time 13, we have: $[33.8 \pm 24.1] = [9.7; 57.9]$.
For time 14, we have: $[42.9 \pm 24.9] = [18; 67.8]$.
For time 15, we have: $[44.5 \pm 25.8] = [18.7; 70.3]$.
For time 16, we have: $[63.1 \pm 26.8] = [36.3; 89.9]$.
Again these intervals are very wide.

(2.2) The additive decomposition model does not seem more appropriate than the multiplicative decomposition model. The latter found no evidence of irregular variation while the additive found that irregular variation in the data is evident, probably because of lack-of-fit.

QUESTION 3

(3.1) The model will be same as in Question 1, part (1.8). That is: $\widehat{y}_t = 24.0758 + 1.7832t$.

(3.2) $l_0 = \bar{y} = \frac{1}{12}\sum_{t=1}^{12} y_t = \frac{20+25+35+44+28+29+43+48+24+37+39+56}{12} = 35.6667$.
The smoothed estimates for $\alpha = 0.1$ are given by the recursive equation:
$l_T = \alpha y_T + (1-\alpha)l_{T-1}$. For example:
$l_1 = \alpha y_1 + (1-\alpha)l_0 = 0.1 \times 20 + 0.9 \times 35.6667 = 34.1000$
$l_2 = \alpha y_2 + (1-\alpha)l_1 = 0.1 \times 25 + 0.9 \times 34.1000 = 33.1900$.
The other values are calculated in the similar manner, and these are reported in the third column of Table 1.3.

(3.3) The forecasts made last period ($\widehat{y}_t$) are reported in the forth column of Table 1.3. They include the smoothed estimate from time zero to time $t - 1 = 11$.

(3.4) The forecast errors $e_t = y_t - \widehat{y}_t$ are given in the fifth column of Table 1.3.

(3.5) The squared forecast errors e_t^2 are given in the sixth column of Table 1.3.

(3.6) The mean squared forecast error (MSFE) is the average of the squared forecast errors, that is:
$MSFE = \frac{\sum_{t=1}^{12} e_t^2}{12} = 115.4432$.

Table 1.3: Simple exponential smoothing for the Oligopoly data

t	Actual sales	Smoothed estimates	Forecast made last period	Forecast errors	Square forecast errors
0		35.6667			
1	20	34.1000	35.6667	-15.6667	245.4455
2	25	33.1900	34.1000	-9.1000	82.8105
3	35	33.3710	33.1900	1.8100	3.2760
4	44	34.4339	33.3710	10.6290	112.9751
5	28	33.7905	34.4339	-6.4339	41.3954
6	29	33.3115	33.7905	-4.7905	22.9492
7	43	34.2803	33.3115	9.6885	93.8675
8	48	35.6523	34.2803	13.7197	188.2294
9	24	34.4871	35.6523	-11.6523	135.7760
10	37	34.7384	34.4871	2.5129	6.3148
11	39	35.1645	34.7384	4.2616	18.1616
12	56	37.2481	35.1645	20.8355	434.1171
Mean	35.6667				115.4432

STA2604
FORECASTING II
TRIAL EXAM PAPER: SEMESTER 1
JUNE 2018

Read the following instructions carefully:

1) Attempt all questions.
2) Show all relevant computations and steps.
3) You may use a non-programmable calculator.

QUESTION 1 **[17]**

(1.1) Consider the following data: 15 59 34 80 100 129 189 167.
An observer stated that there is no evidence that these data can be classified as time series data. What is the reason of his/her statement? (2)

(1.2) Consider the monthly mean temperatures (in degrees Celsius) for a given city during the twelve months of a certain year: 25 26 24 20 10 14 15 18 19 22 24 21.
Why these data can be classified as time series data? (2)

(1.3) A market researcher stated that the price of food has been depending on the time of the year during the last 15 years. Give the time series concept being referred to in such a statement. (2)

(1.4) Give the main difference between (a) seasonal variations and (b) cyclical variations in time series. (6)

(1.5) Define irregular variations in time series and give two examples. (5)

QUESTION 2 [20]

The following table presents actual and predicted mean number of visitors at a certain place during the last eight months.

Month	Actual number of visitors, y_t	Predicted number of visitors, $\widehat{y}_t$
Jul	160	170
Aug	125	135
Sep	125	130
Oct	130	125
Nov	140	150
Dec	130	125
Jan	110	120
Feb	106	100

Calculate:

(2.1) the forecast error for each month (3.5)

(2.2) the MAD (5.5)

(2.3) the MSE (5.5)

(2.4) the MAPE (5.5)

QUESTION 3 [39]

The data in the following table give quarterly sales of mountain bike for two consecutive years by a bicycle shop in Switzerland (in \$100). Assume a multiplicative decomposition time series model.

Year	Quarter	bike Sales
1	1	10
	2	31
	3	43
	4	16
2	1	11
	2	33
	3	45
	4	17

(3.1) Assume that a linear trend $TR_t = \beta_0 + \beta_1 t$ describes the deseasonalized observations where $t = 1$ for quarter 1 of year $1, \ldots, 8$ for quarter 4 of year 2. Calculate the least squares estimates of β_0 and β_1. (8)

(3.2) Compute the appropriate four-point moving averages (4-MA) for the data. (2)

(3.3) Compute the centered moving averages (CMA) for the data. (2)

(3.4) Compute the adjusted seasonal indices sn_t (not in percentages) for these data. (11)

(3.5) Compute the deseasonalized observations $d_t = y_t/sn_t$ for these data. Use two-decimal places for the final results. (4)

(3.6) Use the results in part (3.1) to compute the estimated trend values tr_t for these data. (4)

(3.7) Compute the estimated values (point forecasts) $\widehat{y}_t = tr_t \times sn_t$. Use two-decimal places for the final results. (4)

(3.8) Compute the point sales forecasts for the four quarters of Year 3. Use two-decimal places for the final results. (4)

QUESTION 4 **[24]**

Consider the bike sales data presented in Question 3.

(4.1) Using the results in part (3.7), compute the eight residuals. (8)

(4.2) Explain why we cannot use the Durbin-Watson test to determine whether or not there are first-order positive or negative autocorrelations for these data. (2)

(4.3) Using the simple exponential smoothing with $\alpha = 0.2$, determine the smoothed levels for the first three observations. (6)

(4.4) Now assume that the Holt's trend corrected exponential smoothing, with $\alpha = 0.2$ and $\gamma = 0.1$, is more appropriate than the simple exponential smoothing.

(a) Give the initial estimates of the level l_0 and the growth rate b_0. (2)

(b) Calculate the value of l_1 for $y = 10$ (observation for the first quarter of Year 1). (2)

(c) Estimate the growth rate for $y = 10$. (2)

(4.5) Based on the results for l_1 in parts (4.3) and (4.4), would you recommend the Holt's trend corrected exponential smoothing or the simple exponential smoothing? Briefly explain. (2)

Formulae

(1) $e_t = y_t - \widehat{y}_t$

(2) $MAD = \dfrac{\sum_{t=1}^{n} |e_t|}{n}$

(3) $MSE = \dfrac{\sum_{t=1}^{n} (e_t)^2}{n}$

(4) $APE = \dfrac{100|e_t|}{y_t}$

(5) $MAPE = \dfrac{\sum_{t=1}^{n} APE_t}{n}$

(6) $\widehat{\beta}_1 = \dfrac{n\sum_{t=1}^{n} ty_t - \sum_{t=1}^{n} t \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} t^2 - (\sum_{t=1}^{n} t)^2}$

(7) $$\widehat{\beta}_0 = \frac{\sum_{t=1}^{n} y_t - \widehat{\beta}_1 \sum_{t=1}^{n} t}{n}$$

(8) $$l_T = \alpha y_T + (1-\alpha) l_{T-1}$$

(9) $$l_T = \alpha y_T + (1-\alpha)[l_{T-1} + b_{T-1}]$$

(10) $$b_T = \gamma(l_T - l_{T-1}) + (1-\gamma) b_{T-1}$$

Solutions to the Trial Examination Paper

QUESTION 1

(1.1) The chronological pattern of the times at which the data were collected is not given.

(1.2) The mean temperatures were collected monthly in a certain year. Therefore, the chronological pattern of the times at which the data were collected is given.

(1.3) The time series concept referred to in this case is the seasonal variation.

(1.4) The main difference between seasonal and cyclical variations is that seasonal variations are regular, generally during each year, while cyclical variations may appear after a long period, generally after more than one year, and are not necessarily regular.

(1.5) Irregular variations are fluctuations that do not follow a recognisable or a regular pattern, and thus are unpredictable. Two examples are: floods, droughts.

QUESTION 2

Let y_t and $\widehat{y}_t$ denote the actual and predicted number of visitors, respectively, in the table below:

Month	y_t	$\widehat{y}_t$	e_t	$\lvert e_t\rvert$	e_t^2	$APE_t = \frac{\lvert e_t\rvert}{y_t} \times 100$
Jul	160	170	-10	10	100	6.25
Aug	125	135	-10	10	100	8
Sep	125	130	-5	5	25	4
Oct	130	125	5	5	25	3.8462
Nov	140	150	-10	10	100	7.1429
Dec	130	125	5	5	25	3.8462
Jan	110	120	-10	10	100	9.0909
Feb	106	100	6	6	36	5.6604
Tot				51	511	47.8366

(2.1) The forecast error e_t for each month is given in the fourth column of the table.

(2.2) MAD $= \dfrac{\sum_{t=1}^{8} |e_t|}{8} = \dfrac{51}{8} = 6.375.$

(2.3) MSE $= \dfrac{\sum_{t=1}^{8} e_t^2}{8} = \dfrac{511}{8} = 63.875.$

(2.4) MAPE $= \dfrac{\sum_{t=1}^{8} APE_t}{8} = \dfrac{47.8366}{8} = 5.9796.$

QUESTION 3

(3.1) First construct a table with necessary information for estimating β_1 and β_0.

t	y_t	ty_t	t^2
1	10	10	1
2	31	62	4
3	43	129	9
4	16	64	16
5	11	55	25
6	33	198	36
7	45	315	49
8	17	136	64
36	206	969	204

The values of $\widehat{\beta}_1$ and $\widehat{\beta}_0$ are the following:

$$\widehat{\beta}_1 = \frac{8\sum_{t=1}^{8} ty_t - \sum_{t=1}^{8} t \sum_{t=1}^{8} y_t}{8\sum_{t=1}^{8} t^2 - \left(\sum_{t=1}^{8} t\right)^2} = \frac{8 \times 969 - 36 \times 206}{8 \times 204 - (36)^2} = \frac{336}{336} = 1.$$

$$\widehat{\beta}_0 = \bar{y}_t - \widehat{\beta}_1 \bar{t} = \frac{\sum_{t=1}^{8} y_t - \widehat{\beta}_1 \sum_{t=1}^{8} t}{8} = \frac{206 - 1 \times 36}{8} = \frac{170}{8} = 21.25.$$

Thus, the fitted model is: $\widehat{y}_t = 21.25 + t$.

(3.2) The four-point moving averages are given in the fifth column of Table 1.4.

Table 1.4: Analysis of bike sales data using the multiplicative decomposition of a time series.

Year	Quarter	t	y_t	4MA	CMA_t	$sn_t \times ir_t$	sn_t	d_t	tr_t	$\widehat{y}_t = tr_t \times sn_t$
1	1	1	10				0.42	23.81	22.25	9.35
1	2	2	31				1.25	24.80	23.25	29.06
1	3	3	43	25	25.125	1.7114	1.71	25.15	24.25	41.47
1	4	4	16	25.25	25.5	0.6275	0.63	25.40	25.25	15.91
2	1	5	11	25.75	26	0.4231	0.42	26.19	26.25	11.03
2	2	6	33	26.25	26.375	1.2512	1.25	26.40	27.25	34.06
2	3	7	45	26.5			1.71	26.32	28.25	48.31
2	4	8	17				0.63	26.98	29.25	18.43

(3.3) The centered moving averages, $CMA_t = tr_t \times cl_t$, are given in the sixth column of Table 1.4.

(3.4) First calculate the $sn_t \times ir_t = y_t/CMA_t$ values. These are reported in the seventh column of Table 1.4. Steps for the calculation of seasonal indices are given in the following tables:

Year	Quarter			
	1	2	3	4
1	-	-	1.7114	0.6275
2	0.4231	1.2512	-	-
Total	0.4231	1.2512	1.7114	0.6275
Mean (unadjusted seasonal indices)	0.4231	1.2512	1.7114	0.6275

The total index is: $0.4231 + 1.2512 + 1.7114 + 0.6275 = 4.0132$.

Adjusted seasonal indices (must sum to 4, the number of seasons; quarters here):

Quarter	Value
1	$\frac{0.4231\times 4}{4.0132} = 0.4217$
2	$\frac{1.2512\times 4}{4.0132} = 1.2471$
3	$\frac{1.7114\times 4}{4.0132} = 1.7058$
4	$\frac{0.6275\times 4}{4.0132} = 0.6254$

These adjusted seasonal indices, rounded to the nearest two-decimal place numbers, are copied in the eighth column of Table 1.4.

(3.5) The deseasonalised observations $d_t = y_t/sn_t$ are given in the ninth column of Table 1.4.

(3.6) The trend values tr_t, obtained using $\widehat{y}_t = 21.25 + t$, are given in the tenth column of Table 1.4.

(3.7) The point estimates $\widehat{y}_t = tr_t \times sn_t$ are given in the eleventh column of Table 1.4.

(3.8) The four quarters of year 4 correspond to $t = 9$, $t = 10$, $t = 11$ and $t = 12$, respectively.
We use $\widehat{y}_t = (21.25 + t) \times sn_t$.
For $t = 9$, we have $\widehat{y}_9 = (21.25 + 9) \times 0.42 = 12.71$.
For $t = 10$, we have $\widehat{y}_{10} = (21.25 + 10) \times 1.25 = 39.06$.
For $t = 11$, we have $\widehat{y}_{11} = (21.25 + 11) \times 1.71 = 55.15$.
For $t = 12$, we have $\widehat{y}_{12} = (21.25 + 12) \times 0.63 = 20.95$.

QUESTION 4

(4.1) The eight residuals are given in the third column of the following table:

y_t	$\widehat{y}_t$	$e_t = y_t - \widehat{y}_t$
10	9.35	0.65
31	29.06	1.94
43	41.47	1.53
16	15.91	0.09
11	11.03	-0.03
33	34.06	-1.06
45	48.31	-3.31
17	18.43	-1.43

(4.2) We cannot use the Durbin-Watson test since the number of residuals is less than 15.

(4.3) $l_0 = \frac{1}{8}\sum_{t=1}^{8} y_t = \frac{206}{8} = 25.75.$
We use $l_T = \alpha t_T + (1-\alpha)l_{T-1}$ with $\alpha = 0.2$ and $l_0 = 25.75$.
We obtain:
$l_1 = 0.2 \times 10 + 0.8 \times 25.75 = 22.6.$
$l_2 = 0.2 \times 31 + 0.8 \times 22.26 = 24.01.$
$l_3 = 0.2 \times 43 + 0.8 \times 24.01 = 27.81.$

(4.3) Use of the Holt's trend corrected exponential smoothing with $\alpha = 0.2$ and $\gamma = 0.1$.

(a) $l_0 = \widehat{\beta}_0 = 21.25$ and $b_0 = \widehat{\beta}_1 = 1$.

(b) We use $l_T = \alpha y_T + (1-\alpha)[l_{T-1} + b_{T-1}]$ where $T = 1$, $l_0 = 21.25$ and $\alpha = 0.2$. We obtain:

$$\begin{aligned} l_1 &= \alpha y_1 + (1-\alpha)[l_0 + b_0] \\ &= 0.2 \times 10 + 0.8 \times (21.25 + 1) \\ &= 19.8 \end{aligned}$$

(c) We use $b_T = \gamma[l_T - l_{T-1} + (1-\gamma)b_{T-1}]$ where $T = 1$, $\gamma = 0.1$ and $b_0 = 1$. We obtain:

$$\begin{aligned} b_1 &= \gamma[l_1 - l_0 + (1-\gamma)b_0] \\ &= 0.1(19.8 - 21.25) + 0.9 \times 1 \\ &= 0.76. \end{aligned}$$

(4.4) The forecast error using the simple exponential smoothing is: $10 - 22.6 = -12.6$ and the forecast error using the Holt's trend corrected exponential smoothing is $10 - 19.8 = -9.8$. Therefore, based on this single observation, the Holt's trend corrected exponential smoothing should be recommended since it gives the smallest forecast error.