Roll No. – 18058568015

Examination: B.Sc.(H)

Subject: Statistics

Paper: Time Series Analysis

Code: 32377905

Ques 1.

THEORY (Formula Used)

METHOD OF PARTIAL SUMS:

Equation for modified exponential curve: $y_t = a + bc^t$ (1)

The given time series data is split up into 3 equal parts, each containing n consecutive values of y_t corresponding to t=1,2,…,n ; t=n+1,n+2,…,2n ; t=2n+1,2n+2,…,3n. Let S1, S2 and S3 represent the partial sums of the 3 parts respectively, such that,

- $S_1 = \sum_{t=1}^{n} yt$
- $S_2 = \sum_{t=n+1}^{2n} yt$
- $S_3 = \sum_{t=2n+1}^{3n} yt$

Substituting for y_t from equation (1), we get the values of a, b and c.

- $c = \left(\frac{S3-S2}{S2-S1}\right)^{1/n}$
- $b = \frac{(c-1)((S2-S1)^3)}{c((S3-2S2+S1)^\wedge 2)}$
- $a = \frac{1}{n}\left[\frac{S1S3-(S2^\wedge 2)}{S3-2S2+S1}\right]$

GROMPERTZ CURVE:

The Grompertz curve describes a trend in which the growth increments of the logarithms are declining by a constant percentage. Thus the natural values of the trend would show a declining ratio of increase, but the ratio does not decrease by either a constant amount or a constant percentage.

Equation: $y_t = a + b^{c\wedge t}$

Taking log both sides,

$\log y_t = \log a + c^t * \log b$

Let, $\log y_t = Y_t$, $\log a = A$, $\log b = B$

Therefore, we get

$Y_t = A + Bc^t$

The above equation is comparable to the equation of modified exponential curve.

CALCULATIONS:

Table 1.1

t	Time(hours)	Deaths (y)	Y_t	Yt(estimated)	Estimated Deaths(y)
1	0	59	4.077537444	3.921351612	50.46861266
2	0.1	97	4.574710979	4.620081927	101.5023476
3	0.2	102	4.624972813	5.127285229	168.5588975
4	0.6	291	5.673323267	5.495460452	243.5836588
5	0.7	394	5.976350909	5.762716193	318.21148
6	0.9	432	6.068425588	5.95671521	386.3389953
7	1.3	456	6.12249281	6.097537699	444.7612846
8	1.4	491	6.196444128	6.199759729	492.6306621
9	1.7	500	6.214608098	6.273961964	530.5753397
10	2.1	520	6.253828812	6.327824832	559.9373107
11	2.4	536	6.284134161	6.3669235	582.2637362
12	2.6	568	6.342121419	6.395304941	599.0259631
13	2.9	604	6.403574198	6.415906825	611.4950285
14	3.3	650	6.476972363	6.430861584	620.7085102
15	3.7	695	6.543911846	6.441717136	627.4833493
16	3.8	725	6.586171655	6.449597103	632.44743
17	3.9	800	6.684611728	6.455317114	636.0754025

Table 1.2

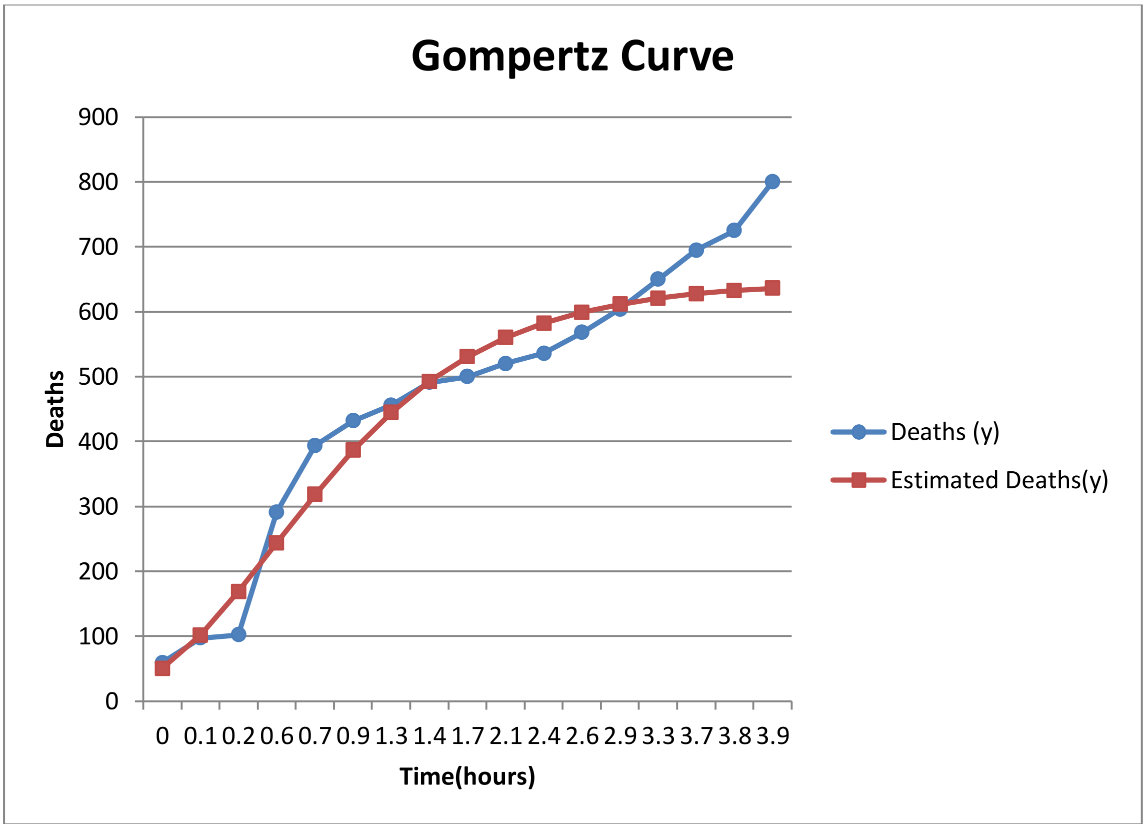
S1=	24.92689541
S2=	30.85579944
S3=	32.05071399

Table 1.3

c=	0.725892795
B=	-3.51169387
A=	6.470464891

R=	0.965275084

Graph 1.1



RESULT:

- Trend values using Grompertz Curve (method of partial sums) have been calculated and shown in Table 1.1.
- The trend values have been plotted along with the given values in Graph 1.1.
- Correlation coefficient between the given amount and the estimated amount is: $R^2 = 0.965275084$.

CONCLUSION:

- The R^2 value calculated is almost equal to 1 (0.965275084). This indicates that the values estimated are almost equal to the given values.

Ques 3.

THEORY (Formula Used)

Correlogram: A correlogram is a visual way to show autocorrelation in the data that changes over time (i.e. time series data).

Correlogram of Moving Average: For a moving average of extent m, with weights ($a_1, a_2, \ldots a_m$) of random components (ε_i ; i=1,2,…), the generated series is given by:

$$y_i = a_1\varepsilon_{i+1} + a_2\varepsilon_{i+2} + \cdots + a_{k+1}\varepsilon_{i+k+1} + \cdots + a_m\varepsilon_{i+m}$$

$$y_{i+k} = a_1\varepsilon_{i+k+1} + a_2\varepsilon_{i+k+2} + \cdots + a_{m-k}\varepsilon_{i+m-k} + \cdots + a_m\varepsilon_{i+k+m}$$

Where ε_i's are iid $N(0, \sigma^2)$. Thus,

$E(y_i) = 0 = E(y_{i+k})$

And $\mathrm{Var}(y_i) = E(y_i^2) = (a_1^2 + a_2^2 + \ldots + a_m^2)\sigma^2 = \sigma^2 \sum_{j=1}^{m} a_j^2 \;\forall\, i = 1, 2, \ldots$

Similarly, $\mathrm{Var}(y_{i+k}) = E(y_{i+k}^2) = \sigma^2 \sum_{j=1}^{m} a_j^2$

$$E(y_i y_{i+k}) = (a_1 a_{k+1} + a_2 a_{k+2} + \cdots + a_{m-k} a_m)\sigma^2 = \sigma^2 \sum_{j=1}^{m-k} a_j a_{j+k}, k < m$$

$$\left.\begin{aligned} r_k &= \frac{E(y_i\, y_{i+k})}{\sqrt{\mathrm{Var}\,(y_i)\,\mathrm{Var}\,(y_{i+k})}} \\ &= \frac{\sum_{j=1}^{m-k} a_j\, a_{j+k}}{\sum_{j=1}^{m} a_j^2}, \text{ if } k < m \\ r_k &= 0, \text{ if } k \geq m \end{aligned}\right\}$$

CALCULATIONS:

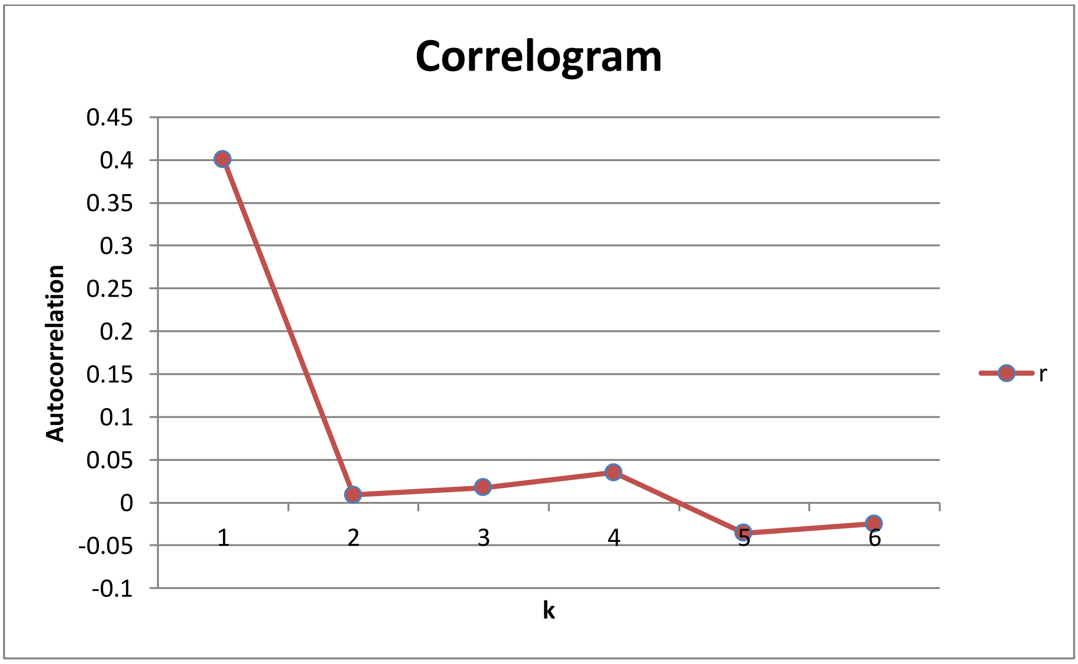
Table 3.1

m	a	a2	k					
			1	2	3	4	5	6
1	0.25	0.0625						
2	0.125	0.015625	0.03125					
3	-0.9375	0.878906	-0.11719	-0.23438				
4	-0.46875	0.219727	0.439453	-0.05859	-0.11719			
5	-0.23438	0.054932	0.109863	0.219727	-0.0293	-0.05859		
6	-0.11719	0.013733	0.027466	0.054932	0.109863	-0.01465	-0.0293	
7	-0.125	0.015625	0.014648	0.029297	0.058594	0.117188	-0.01563	-0.03125
SUM	-1.50781	1.261047	0.505493	0.010986	0.021973	0.043945	-0.04492	-0.03125

Table 3.2

k	1	2	3	4	5	6
r	0.400852	0.008712	0.017424	0.034848	-0.03562	-0.02478

Graph 3.1



RESULT:

- The value of autocorrelation (r_k)for the order k have been shown in Table 3.2
- Graph 3.1 shows the correlogram plotted for the moving average.

Ques 4.

THEORY (Formula Used)

VARIATE DIFFERENCE METHOD:

1. We first have to find the values of Δyt, Δ^2yt, Δ^3yt … for the given values of t and yt.

2. Then compute: V1 $= \frac{\mu_2'(\Delta yt)}{C_1^2}$, V2 $= \frac{\mu_2'(\Delta^2 yt)}{C_2^4}$

3. If, let us say, V1 and V2 do not differ significantly than either of them can be regarded as an estimate of V. Otherwise, calculate V3, V4, … till two successive estimates of V are homogeneous.

Significance of $(V_k - V_{k+1})$**.** Homogeneity of two successive estimates of V cannot be tested by Variance Ratio Test (F-test) since the consecutive terms are not independent. *O.* Andersen obtained the standard error of $(V_k - V_{k+1})$ and found that for large samples,

$$R_k = \frac{V_k - V_{k+1}}{V_k} \cdot H_{kN} \sim N(0, 1) \qquad \ldots(2\cdot67)$$

where V_k and V_{k+1} are consecutive estimates of V from the kth and $(k+1)$th differences of y_t and H_{kN} ; a function of k and N (the total number of observations in the given time series), corresponds to the variance of the ratio and has been tabulated in the book "*Variate Difference Method*" by **G. Titner**. Thus if $|R_k| > 1\cdot96$, the difference is significant (otherwise not) at 5% level of significance.

CALCULATIONS:

Table 4.1

t	yt	Δyt	Δ^2yt	Δ^3yt	Δ^4yt	Δ^5yt	Δ^6yt
1	1006						
		112					
2	1118		-106				
		6		-30			
3	1124		-136		184		
		-130		154		-238	
4	994		18		-54		67
		-112		100		-171	
5	882		118		-225		530
		6		-125		359	
6	888		-7		134		-605
		-1		9		-246	

7	887		2		-112		642
		1		-103		396	
8	888		-101		284		-791
		-100		181		-395	
9	788		80		-111		373
		-20		70		-22	
10	768		150		-133		-66
		130		-63		-88	
11	898		87		-221		739
		217		-284		651	
12	1115		-197		430		-1253
		20		146		-602	
13	1135		-51		-172		1089
		-31		-26		487	
14	1104		-77		315		-1437
		-108		289		-950	
15	996		212		-635		1995
		104		-346		1045	
16	1100		-134		410		-1561
		-30		64		-516	
17	1070		-70		-106		1140
		-100		-42		624	
18	970		-112		518		-2054
		-212		476		-1430	
19	758		364		-912		3035
		152		-436		1605	
20	910		-72		693		-3482
		80		257		-1877	
21	990		185		-1184		5315
		265		-927		3438	
22	1255		-742		2254		-6796
		-477		1327		-3358	
23	778		585		-1104		2620
		108		223		-738	
24	886		808		-1842		5098
		916		-1619		4360	
25	1802		-811		2518		-7888
		105		899		-3528	
26	1907		88		-1010		5202
		193		-111		1674	
27	2100		-23		664		-4333
		170		553		-2659	
28	2270		530		-1995		7372
		700		-1442		4713	
29	2970		-912		2718		-8943

		-212		1276		-4230	
30	2758		364		-1512		4635
		152		-236		405	
31	2910		128		-1107		3718
		280		-1343		4123	
32	3190		-1215		3016		-8685
		-935		1673		-4562	
33	2255		458		-1546		6204
		-477		127		1642	
34	1778		585		96		-3109
		108		223		-1467	
35	1886		808		-1371		3870
		916		-1148		2403	
36	2802		-340		1032		-2025
		576		-116		378	
37	3378		-456		1410		
		120		1294			
38	3498		838				
		958					
39	4456						
SUM=		3450	846	944	1324	1226	

Table 4.2

	$(\Delta yt)^2$	$(\Delta^2 yt)^2$	$(\Delta^3 yt)^2$	$(\Delta^4 yt)^2$	$(\Delta^5 yt)^2$	$(\Delta^6 yt)^2$
	12544					
		11236				
	36		900			
		18496		33856		
	16900		23716		56644	
		324		2916		4489
	12544		10000		29241	
		13924		50625		280900
	36		15625		128881	
		49		17956		366025
	1		81		60516	
		4		12544		412164
	1		10609		156816	
		10201		80656		625681
	10000		32761		156025	
		6400		12321		139129
	400		4900		484	
		22500		17689		4356

16900		3969		7744	
	7569		48841		546121
47089		80656		423801	
	38809		184900		1570009
400		21316		362404	
	2601		29584		1185921
961		676		237169	
	5929		99225		2064969
11664		83521		902500	
	44944		403225		3980025
10816		119716		1092025	
	17956		168100		2436721
900		4096		266256	
	4900		11236		1299600
10000		1764		389376	
	12544		268324		4218916
44944		226576		2044900	
	132496		831744		9211225
23104		190096		2576025	
	5184		480249		12124324
6400		66049		3523129	
	34225		1401856		28249225
70225		859329		11819844	
	550564		5080516		46185616
227529		1760929		11276164	
	342225		1218816		6864400
11664		49729		544644	
	652864		3392964		25989604
839056		2621161		19009600	
	657721		6340324		62220544
11025		808201		12446784	
	7744		1020100		27060804
37249		12321		2802276	
	529		440896		18774889
28900		305809		7070281	
	280900		3980025		54346384
490000		2079364		22212369	
	831744		7387524		79977249
44944		1628176		17892900	
	132496		2286144		21483225
23104		55696		164025	
	16384		1225449		13823524
78400		1803649		16999129	
	1476225		9096256		75429225
874225		2798929		20811844	

		209764		2390116		38489616
	227529		16129		2696164	
		342225		9216		9665881
	11664		49729		2152089	
		652864		1879641		14976900
	839056		1317904		5774409	
		115600		1065024		4100625
	331776		13456		142884	
		207936		1988100		
	14400		1674436			
		702244				
	917764					
SUM=	5304150	7570320	18751974	52956958	166229342	568108286

Table 4.3

$\mu'_2(\Delta yt)$	$\mu'_2(\Delta^2 yt)$	$\mu'_2(\Delta^3 yt)$	$\mu'_2(\Delta^4 yt)$	$\mu'_2(\Delta^5 yt)$	$\mu'_2(\Delta^6 yt)$
139582.895	204603.243	520888.167	1513055.943	4889098.294	17215402.606

Table 4.4

V1	V2	V3	V4	V5	V6
69791.447	34100.541	26044.408	21615.085	19401.184	18631.388

Table 4.5

H(1,39)	H(2,39)	H(3,39)	H(4,39)	H(5,39)
11.995	15.818	18.489	20.351	21.617

Table 4.6

R1	R2	R3	R4	R5
6.134	3.737	3.144	2.084	0.858

RESULT:

We have:
$|R_1| = 6.134$, $|R_2| = 3.737$, $|R_3| = 3.144$, $|R_4| = 2.084$, $|R_5| = 0.858$

Now,

- As $|R_1| > 1.96$, therefore difference between V_1 and V2 is significant.
- As $|R_2| > 1.96$, therefore difference between V_2 and V_3 is significant.
- As $|R_3| > 1.96$, therefore difference between V_3 and V_4 is significant.
- As $|R_4| > 1.96$, therefore difference between V_4 and V_5 is significant.
- As $|R_5| < 1.96$, therefore difference between V_5 and V_6 is not significant

CONCLUSION:

As the difference between V5 and V6 is insignificant, hence we can say that any of the two can be taken as the variance of the random component.

Therefore,

Variance of the random component = 19401.184
OR
Variance of the random component = 18631.388

Ques 5.

THEORY (Formula Used)

- Exponential smoothing is a time series forecasting method for univariate data. Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older.
- Equation $F_{t+1} = \alpha D_t + (1-\alpha)F_t$, is used to forecast the values for the next months. Here, F is forecasted value, D is demand and α is smoothing constant.
- For t=1, F_t is the first forecasting value, obtained by calculating median of the first 6 observations.

WHAT WOULD BE THE SIZE OF α :

A measure of effectiveness of exponential smoothing can be obtained under the assumption that the process is completely stable, so that $X_1, X_2, \ldots$ are independent, identically distributed random variables with variance σ^2. It then follows that (for large t)

$$\text{var}[F_{t+1}] \approx \frac{\alpha\sigma^2}{2-\alpha} = \frac{\sigma^2}{(2-\alpha)/\alpha},$$

so that the variance is statistically equivalent to a moving average with $(2-\alpha)/\alpha$ observations. For example, if α is chosen equal to 0.1, then $(2-\alpha)/\alpha = 19$. Thus, in terms of its variance, the exponential smoothing method with this value of α is *equivalent* to the moving-average method that uses 19 observations.

Or $\frac{(2-\alpha)}{\alpha} = n$ (no. of observations)

And we can find the value of α .

CALCULATIONS

Table 5.1

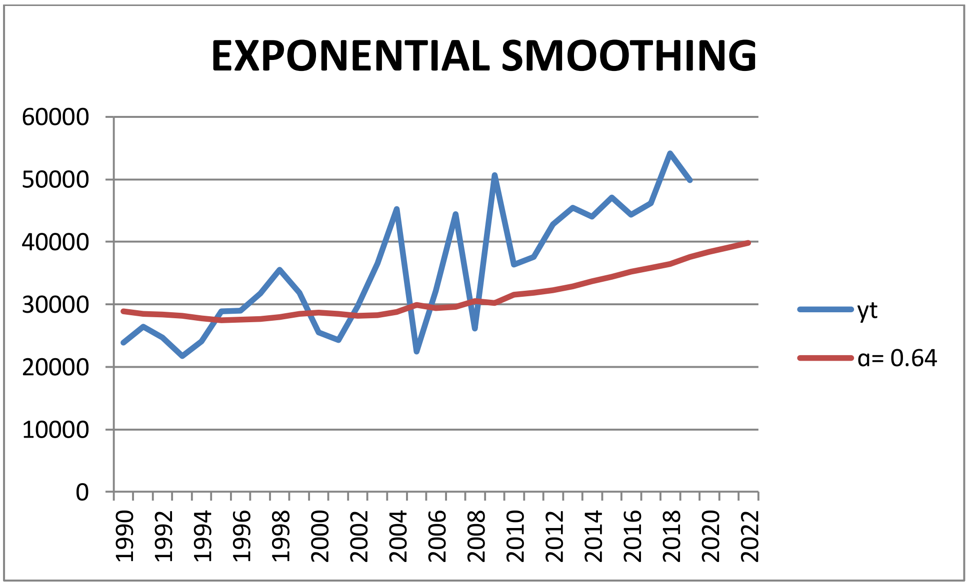
Given Data	
t	yt
1990	23832
1991	26410
1992	24735
1993	21778
1994	24104
1995	28846
1996	29010
1997	31749
1998	35508
1999	31830
2000	25510
2001	24333
2002	29709
2003	36543
2004	45218
2005	22432
2006	32146
2007	44452
2008	26111
2009	50675
2010	36313
2011	37552
2012	42794
2013	45476
2014	44069
2015	47052
2016	44333
2017	46169
2018	54110
2019	49850
2020	
2021	
2022	

Table 5.2

Exponential Smoothing
α= 0.64
28846
28522.51613
28386.22477
28150.66188
27739.5224
27504.97257
27591.49047
27683.00721
27945.32933
28433.24356
28652.38914
28449.65436
28184.06375
28282.44674
28815.38566
29873.61884
29393.5144
29571.09412
30531.15256
30245.98143
31563.98263
31870.37084
32236.92756
32918.02901
33728.22069
34395.36774
35211.92466
35800.38113
36469.32429
37607.4324
38397.27547
39136.16092
39827.37635

α=	0.064516
1-α=	0.935484

Graph 5.1



RESULT:

- Table 5.2 shows the exponential smoothing done for the data given in Table 5.1
- Alpha was calculated to be 0.64
- Graph 5.1 has been plotted to compare the values obtained for different smoothing constants.
- Value forecasted for the next 3 years is:

2020	3216.129
2021	6224.766
2022	9039.297

CONCLUSION:

- The logical weights are assigned to all the variables, means more weight has been given to recent values and less to previous.
- Hence we can say that the value of α should be minimum and logical to get the best forecast value.