HONG KONG STATISTICAL SOCIETY

2015 EXAMINATIONS – SOLUTIONS

HIGHER CERTIFICATE – MODULE 7

The Society is providing these solutions to assist candidates preparing for the examinations in 2017.

The solutions are intended as learning aids and should not be seen as "model answers".

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

1 (i) (a) $Sm(y_t) = \sum_{r=-q}^{q} a_r y_{t+r}$ with $a_{-r} = a_r$, $\sum_{r=-q}^{q} a_r = 1$

½ or equivalently $\sum w_r y_{t+r} / \sum w_r$.

(b) $m_t = \alpha y_t + (1-\alpha) m_{t-1}$.

½ (or) $= m_{t-1} + \alpha(y_t - m_{t-1})$

(ii)

Date	y_t	$Sm(y_t)$	m_t
June, 10	51	—	51
Dec	25	41·67	43·20
June, 11	49	35·00	44·94
Dec	31	44·33	40·76
June, 12	53	37·67	44·43
Dec	29	45·67	39·80
June, 13	55	36·33	44·36
Dec	25	—	38·55

1/6

3 point MA – end point problem – no forecasts
– over-reacts to seasonal effect here.
– weight $\frac{1}{4}$ $\frac{1}{2}$ $\frac{1}{4}$ would be better or iteration.

EWMA – does deal with end of series & can be used for forecasting.
– over influenced by initial level (51 here)
– slow to react to trend or seasonality

1/6

(iii) 1. Trend can be estimated using MA and removed by subtracting from series.
2. Seasonal effect can then be estimated (eg seasonal average) and removed.
Then 1. and 2. repeated.

1/4

2 (i) The plot may have exhibited a trend or else a tendency to random changes of level.

/2

(ii) ε_t are uncorelated, mean zero, constant variance (possibly also assumed Normal, so $N(0, \sigma^2)$)

/2

(iii) $x_t - x_{t-1} = 0.4(x_{t-1} - x_{t-2}) + \varepsilon_t + 0.7\varepsilon_{t-1}$

$\therefore \quad x_t = 1.4x_{t-1} - 0.4x_{t-2} + \varepsilon_t + 0.7\varepsilon_{t-1}$

/4 Model for x_t is ARIMA $(1,1,1)$, $p=d=q=1$.

/2 (iv) $\hat{\varepsilon}_t = x_t - 1.4x_{t-1} + 0.4x_{t-2} - 0.7\hat{\varepsilon}_{t-1}$

(v) The residuals should be approximately white noise so the racf should be zero. To test this they can be compared with lines drawn at $\pm 2/\sqrt{n}$ ie at $\pm$ 0.16. Significant values might

/4 suggest improvements (eg extra MA terms).

(vi) * Time series plot of residuals - outlies? seasonal?
* Residual partial acf - maybe extra AR terms?

/3 * Normal qq plot or histogram - Normality, outlies.

(vii) The chi-square statistic is based on the cumulative sum of squares of residual autocorelations. Here it is significant for lags up to 12, 24 and 36 suggesting the model is

/3 not adequate.

3(i) $\sum_i N_{i0} R_{i0} = \sum_i D_{i0} = D_0$ (say), the total deaths in the standard population. So

$$P_L(0,A) = \sum_i N_{i0} R_{iA} / D_0.$$

This compares the total deaths if A's death rates had applied to the standard population with the actual total in that standard population

/4

(ii) $$P_p(0,A) = \sum_i N_{iA} R_{iA} / \sum_i N_{iA} R_{i0}$$
$$= D_A / \sum_i N_{iA} R_{i0}.$$

Here the actual deaths in A are compared to an estimate where standard population rates are applied to A's population structure.

/4

(iii)

Age, i	Death rate, R_{iA}	$N_{i0} \times R_{iA}$
≤ 64	$342/128 = \underline{2.67}$	$1322 \times 2.67 = 3530$
65-74	$373/14 = \underline{26.64}$	$131 \times 26.64 = 3490$
≥ 75	$825/9 = \underline{91.67}$	$121 \times 91.67 = \underline{11092}$
		18112

So $P_L(0,A) = 18112 / (1322 \times 2 + 131 \times 21.8 + 121 \times 87.2)$

$= 18112 / 16051 = \underline{1.12814}$

(or $\underline{113}$ as %)

/5

(iv)

Age, i	Death rate, R_{iB}	$N_{i0} \times R_{iB}$
≤ 64	$210/78 = 2.69$	$1322 \times 2.69 = 3556$
65-74	$249/9 = 27.67$	$131 \times 27.67 = 3625$
≥ 75	$651/7 = 93.00$	$121 \times 93 = \underline{11253}$
		18434

So $P_L(0,B) = 18434/16051 = \underline{1.1485}$ (or $\underline{115}$%)

Death rates for B are higher than A's for all age groups and $P_L(0,B) > P_L(0,A)$ so "Yes!"

/4

(vi) $$P_L(0,B) - P_L(0,A) = \left(\sum_i N_{i0} R_{iB} - \sum N_{i0} R_{iA}\right) / D_0$$
$$= \sum_i N_{i0} (R_{iB} - R_{iA}) / D_0.$$

Now $R_{iB} > R_{iA}$ implies $P_L(0,B) > P_L(0,A)$.

/3

4 (i) a. Chain-linked Laspeyres: $P_L^*(2;0) = P_L(0,1)P_L(1,2)/100$

" " Paasche: $P_P^*(2;0) = P_P(0,1)P_P(1,2)/100$

Fisher: $P_F(i,j) = \sqrt{P_L(i,j) \times P_P(i,j)}$

$$\therefore P_F^*(2;0) = P_F(0,1) \times P_F(1,2) / 100$$
$$= \sqrt{P_L(0,1)\times P_P(0,1)} \cdot \sqrt{P_L(1,2)\times P_P(1,2)} / 100$$
$$= \sqrt{P_L(0,1)\times P_L(1,2)/100} \times \sqrt{P_P(0,1)\times P_P(1,2)/100}$$
$$= \underline{\underline{\sqrt{P_L^*(2;0)\, P_P^*(2;0)}}} \quad \text{as required}$$

/6

b. $P_L(11,12) = \dfrac{20\times 101{\cdot}2 + 40\times 97{\cdot}3 + 70\times 103{\cdot}9}{20+40+70} = 101{\cdot}45$ (2dp)

$$P_L(12,13) = \frac{21\times 100{\cdot}7 + 40\times 100{\cdot}2 + 75\times 105{\cdot}4}{21+40+75} = 103{\cdot}14$$

$$P_L^*(13;11) = 101{\cdot}45 \times 103{\cdot}14/100 = 104{\cdot}64$$

If $P_P^*(13;11) = 100{\cdot}7\times 102{\cdot}4/100 = 103{\cdot}12$

/7 then $P^*(13;11) = \sqrt{104{\cdot}64\times 103{\cdot}12} = \underline{103{\cdot}9}$

(ii) a. Suppose the Laspeyres index is written as

$$P_L = \Sigma_i\, p_{it_1} q_{it_0} / \Sigma_i\, p_{it_0} q_{it_0}$$

while Paasche is

$$P_P = \Sigma_i\, p_{it_1} q_{it_1} / \Sigma_i\, p_{it_0} q_{it_1}$$

If prices p_{it} are tending to increase a corresponding decrease in q_{it} is to be expected. This will cause P_L to overstate inflation

/4 while P_P understates it.

b. In the Fisher-index the over and under estimation effects in (a) cancel (to a degree.) Laspeyres is however more convenient as prices p_{it} are generally readily established

/3 before accurate volumes q_{it} become known.