Student project for NEAS VEE- Time Series

Death caused by Cancer in Thailand

Areerat Maiam

Introduction

In Thailand during past 5 years, cancer becomes number 1 of reason caused Thai people die. In year 2011, cancer caused about 15% of all deaths in Thailand and 13% of all human deaths worldwide. Death rate from cancers increases from 85 to 95 (per 100K) in 5 years.

The Cancers is a broad group of diseases involving unregulated cell growth which are over 200 different cancers. Rates are rising as more people live to an old age and as mass lifestyle changes occur in the developing world. This project attempts to perform a time series analysis of the death by cancer in Thailand

The table below show no of deaths by leading cause in 2007 -2011.

Table 2.3.6 Number of Deaths and Death Rates per 100,000 Population by Leading Causes of Death, 2007 - 2011 in Thailand

List	Cause	No of Deaths by Leading cause					Death Rates per 100,000 Population				
		2007	2008	2009	2010	2011	2007	2008	2009	2010	2011
	Total Death	**393,255**	**397,327**	**393,916**	**411,331**	**414,670**	**625**	**629**	**621**	**646**	**646**
1	Cancer	53,434	55,403	56,058	58,076	61,082	85	88	88	91	95
		14%	14%	14%	14%	15%					
2	Accident	35,661	34,851	35,304	32,861	33,868	57	55	56	52	53
		9%	9%	9%	8%	8%					
3	Hypertension and cerebrovascular disease	15,286	15,596	15,648	20,018	22,947	24	25	25	31	36
		4%	4%	4%	5%	6%					
4	Disease of the heart	18,452	18,820	18,375	18,399	20,130	29	30	29	29	31
		5%	5%	5%	4%	5%					
5	Pneumonia and other diseases of lung	14,179	14,542	14,542	16,369	16,884	23	23	23	26	26
		4%	4%	4%	4%	4%					

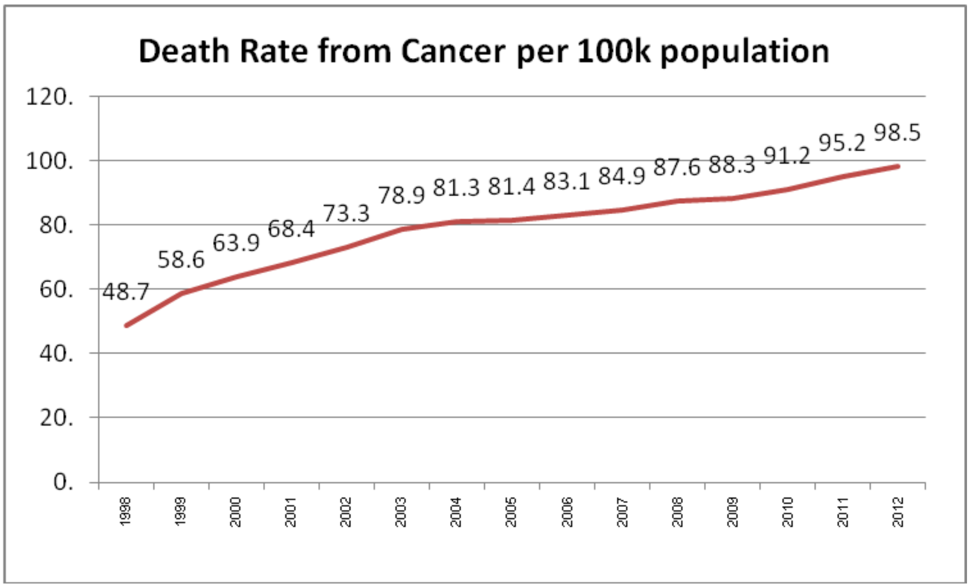
Source : Health Information Unit, Bureau of Health Policy and Strategy

Table 2.3.6_54_Death by L

Data Exploration

The data for this project is obtained from National Economic and Social Development Board (NESDB) (web address: http://social.nesdb.go.th/SocialStat/StatReport_Final.aspx?reportid=441&template=2R1C&yeartype=M&subcatid=15)

The following Graph below displays death rate by cancer from 1998 – 2012. It is shown upward direction which implied non-stationary.



In this study, the data used for this project is annual cancer incident rate from 1998 to 2012 (15 years) as below:-

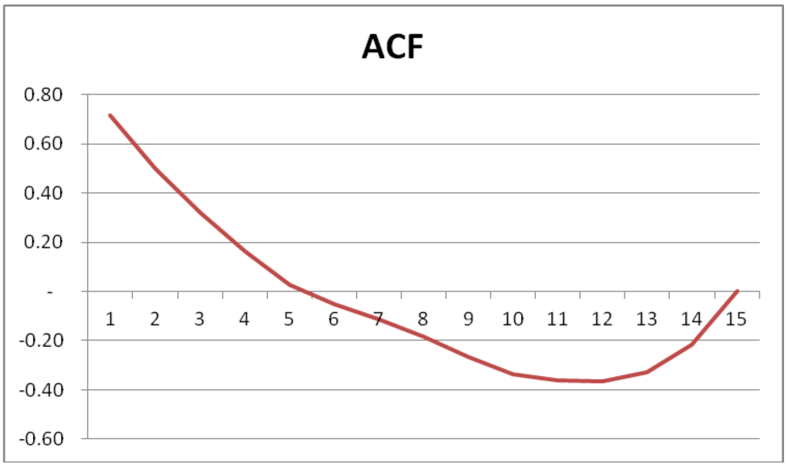
Cause of Death	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Cancer	48.7	58.6	63.9	68.4	73.3	78.9	81.3	81.4	83.1	84.9	87.6	88.3	91.2	95.2	98.5

Analysis & Model

We can observe from the graph that the death caused by cancer demonstrates an upward trend over time and therefore non-stationary. To confirm this, we will compute the sample autocorrelation function ("ACF") at different lags using the following formula:

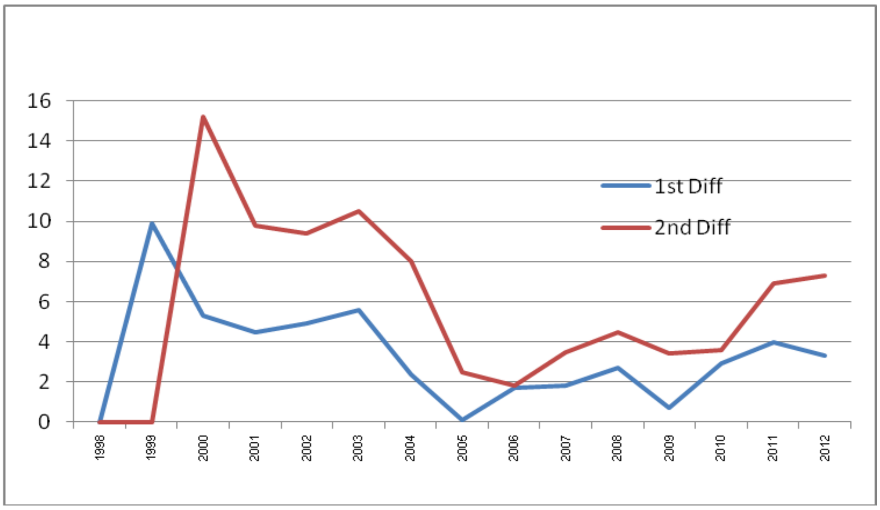
$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} \qquad \text{for } k = 1, 2, \ldots$$

The subsequent correlogram plots the calculated ACF versus lag k:



The above illustration shows that ACF is fall below zero starting from lag 6, proving non-stationary. We now proceed with obtaining stationary through differencing.

Graph of 1st and 2nd different



Comparing the R-Square for AR(1) and AR(2), we see that they are all quite close to 1. AR(1) has better R-Square than AR(2).

SUMMARY OUTPUT AR(1)

Regression Statistics	
Multiple R	0.990009556
R Square	0.980118921
Adjusted R Square	0.978462164
Standard Error	1.956052689
Observations	14

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2263.503437	2263.503437	591.5889598	1.40503E-11
Residual	12	45.91370545	3.826142121		
Total	13	2309.417143			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-14.80881648	3.830441308	-3.866086252	0.002243742	-23.15463114	-6.463001825	-23.15463114	-6.463001825
X Variable 1	1.138836093	0.046822133	24.32260183	1.40503E-11	1.036819429	1.240852756	1.036819429	1.240852756

SUMMARY OUTPUT AR(2)

Regression Statistics	
Multiple R	0.97129654
R Square	0.943416969
Adjusted R Square	0.938273057
Standard Error	3.184514929
Observations	13

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1859.930588	1859.930588	183.4045722	3.32525E-08
Residual	11	111.5524887	10.14113534		
Total	12	1971.483077			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-26.80066766	7.651095647	-3.502853564	0.004945392	-43.64061562	-9.960719691	-43.64061562	-9.960719691
X Variable 1	1.24350249	0.091820907	13.54269442	3.32525E-08	1.041406036	1.445598945	1.041406036	1.445598945

Conclusion

AR(1) mode of Y(t) = -14.8088164802837 + 1.13883609265289 Y(t-1) is adopted. With this model, see below graph of actual compared with prediction.