Chapter 3

A linear regression model for the fish catches weight

In this chapter we present a preliminary analysis and fitting a model of the time series of monthly fish catches weight over the period from January 1977 to December 2006. The study is based on the general assumption that any decline of fish population magnitude in the Lake Basin will generate a decline in fish catch weights caught by fishermen.

The analysis is based on the graphical method and well known statistical methods. First, the characteristics of the data are described, second, assessing the need for data transformation, and third fitting a linear regression model to the data. The results presented in this chapter also appear in Chesoh and Lim (2008).

Prince of Songkla University Pattani Campus

3.1 Characteristics of the data

Data comprised 30 years, 360 months, 360 records of catch in tonnage over the period from January 1977 to December 2006. Since original data obtained from various sources and over a 30 year period, data cleaning, rechecking and standardization were carried on prior to analysis. An overall distribution of total catch weight in the Lake Basin over a 30 year period from 1977 to 2006 is shown in Table 3.1.

During 1977 to 1986, the Lake annual fish catches was highest (3,639 tonnes) in 1980 and lowest (1,816.6 tonnes) in 1986. From 1987 to 1996, the Lake annual fish catches was highest (4,817.2 tonnes) in 1996 and lowest (1,954.2 tonnes) in 1987. From 1997 to 2006, the Lake annual fish catches was highest (4,013.0 tonnes) in 1997 and lowest

(1,622.3 tonnes) in 2006. Total annual catch fluctuated between 1,622 and 4,817 tonnes (average 1977–2006, 2,638.85 tonnes). The catch slightly increased over the preceding decade to peak at 3,639 tonnes in 1980, gradually declined to 1,817 tonnes in 1986, then dramatically increased to a second peak at 4,817 tonnes at 1996, before entering a steady decline to 1,622 tonnes in 2006, which was shown in Figure 3.1.

Table 3.1 Monthly tonnes of fish catch weights in Songkhla Lake during 1977–2006.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
1977	424.3	306.1	145.5	93.9	140.8	185.5	130.8	88.0	116.2	221.8	327.7	483.5	**2,664.1**
1978	354.7	551.0	247.6	69.6	167.9	82.6	102.9	60.8	88.3	102.6	123.4	432.8	**2,384.2**
1979	244.8	255.0	221	148.9	243.3	257.3	257.9	84.6	97.1	160.6	112.2	494.8	**2,577.5**
1980	274.6	159.5	261.2	123.1	237.4	390.2	402.5	263.7	252.5	395.8	474.2	404.3	**3,639.0**
1981	398.6	563.7	317.1	209.7	258.3	265.8	195.0	132.6	119.7	131.5	455.4	340.1	**3,387.5**
1982	236.9	287.3	184.9	145.6	155.7	166.2	126.1	81.6	121.3	168.3	290.7	114.1	**2,078.7**
1983	283.2	107.5	148.3	112.1	162.4	208.1	166.3	133.6	146.6	110.1	196.8	317.6	**2,092.6**
1984	250.3	98.6	118.7	114.2	171	211.4	162.2	96.2	121.7	172.9	179.8	358.4	**2,055.4**
1985	200.5	172.5	179.3	157.9	174.5	164.8	151.8	144.6	156.2	172.1	178.4	140.1	**1,992.7**
1986	285.1	240.3	201.4	121.4	161.1	167.3	89.4	64.0	61.1	79.1	62.8	283.6	**1,816.6**
1987	306.7	258.5	216.6	130.6	173.3	180	96.2	68.8	65.7	85.1	67.6	305.1	**1,954.2**
1988	338.0	284.9	238.8	143.9	191	198.4	106.0	75.8	72.4	93.8	74.5	336.3	**2,153.8**
1989	389.1	327.9	274.8	165.7	219.8	228.3	122.0	87.3	83.3	107.9	85.7	387.0	**2,478.8**
1990	201.5	117.0	191.7	90.4	174.3	286.4	295.4	193.5	185.3	290.5	348.0	296.7	**2,670.7**
1991	213.8	124.1	203.3	95.9	184.8	203.8	313.3	205.3	196.6	308.1	369.1	314.7	**2,732.8**
1992	319.3	126.7	295.4	138.3	172.8	195.3	148.2	140.2	227.5	370	174.1	376.9	**2,684.7**
1993	341.3	135.4	215.8	154.7	184.7	208.8	158.4	149.9	243.2	395.5	186.1	402.8	**2,776.6**
1994	373.4	110.1	345.5	178.7	202.1	228.4	173.3	164.0	266.1	432.8	201.6	380.7	**3,056.7**
1995	414.2	164.3	383.3	309.2	224.2	253.4	192.3	181.9	295.2	480.1	225.9	488.9	**3,612.9**
1996	552.3	219.1	511	412.3	298.9	337.9	256.4	242.5	393.6	640.1	301.2	651.9	**4,817.2**
1997	460.5	182.6	426	343.7	249.2	281.7	213.8	202.2	328.1	533.6	251.1	543.5	**4,016.0**
1998	529.9	446.7	374.3	225.7	299.4	311.0	166.2	118.9	113.5	147.0	116.8	527.2	**3,376.6**
1999	295.4	117.2	273.3	220.5	159.9	180.7	137.1	129.7	210.5	342.4	161.1	348.7	**2,576.5**
2000	193.5	112.4	184.0	86.8	167.3	275.0	283.6	185.8	177.9	278.9	334.1	284.9	**2,564.2**
2001	241.7	95.9	223.7	180.4	130.8	147.9	146.9	181.9	205.8	329.9	294.7	349.9	**2,529.6**
2002	173.4	109.8	199.8	152.3	179.1	288.2	288.1	201.5	196.7	261.3	319.8	258.9	**2,628.7**
2003	149.0	102.9	171.7	157.3	127.0	152.6	125.1	173.0	122.3	277.1	188.1	304.5	**2,050.5**
2004	277.4	172.1	254.8	153.2	202.7	207.9	170.7	175.7	177.4	153.1	171.2	215.0	**2,331.3**
2005	81.2	151.1	178.3	141.0	163.8	190.6	162.4	127.0	161.0	137.6	179.8	169.8	**1,843.6**
2006	135.2	115.6	137.0	147.9	150.6	146.4	178.5	109.4	138.0	109.1	127.9	126.7	**1,622.3**
Mean	**298.0**	**207.2**	**244.1**	**164.2**	**190.9**	**220.1**	**184.0**	**142.1**	**171.4**	**249.6**	**219.3**	**348.0**	**2,638.9**

Prince of Songkla University Pattani Campus

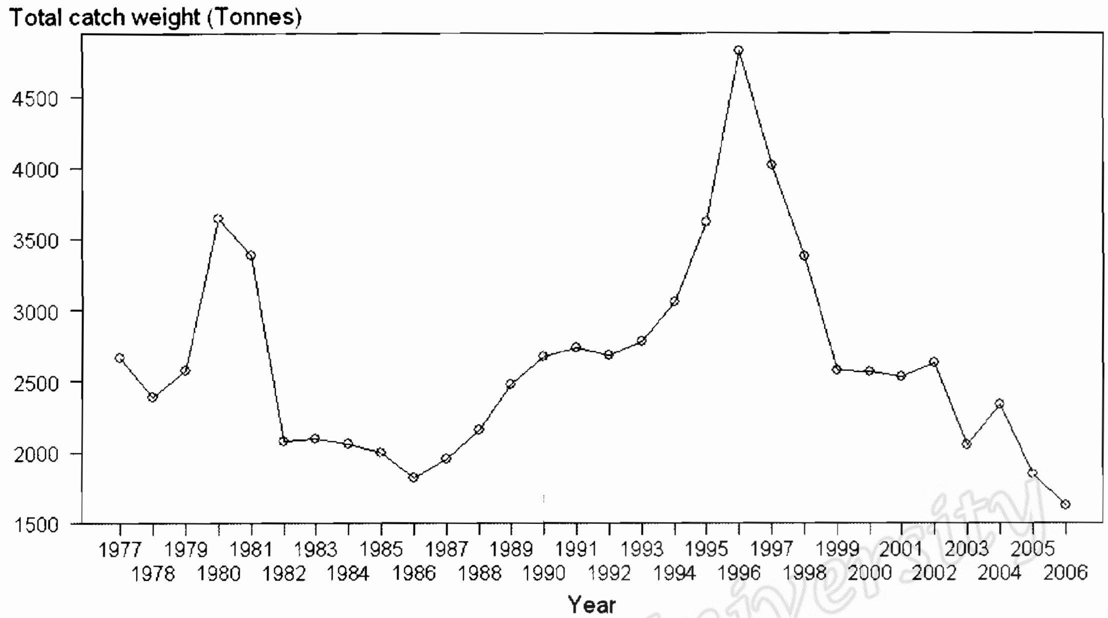

Figure 3.1 Annual fish catch weights in Songkhla Lake during 1977–2006

Prince of Songkla University Pattani Campus

Figure 3.1 and 3.2 show that the variance of the fish catch weights over time is not constant. This indicates that the data may need to be transformed before fitting a model.

In addition, the average monthly catch was 219.9 tonnes (range 60.8 – 651.9), and since 2000 there seems to be less regular seasonal periodic fluctuation. The average monthly catch was typically lowest in August (142.1 tonnes) while the highest was in December (348 tonnes) year round, which was shown in Figure 3.2.

As described in Chapter 2, a method of assessing the need to transform data involves plotting sample means against sample standard deviations on a logarithmic scale. We grouped the data into 20 one and half year periods. The relationship between the mean and standard deviation of each sample is shown in Figure 3.3.

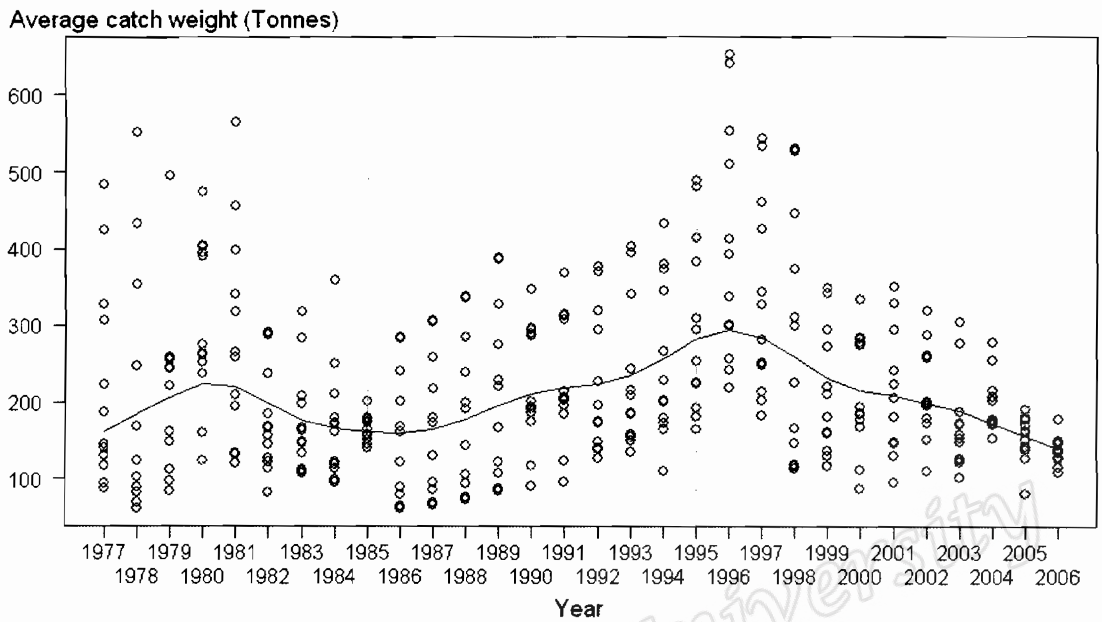

Figure 3.2 Average monthly fish catch weights in Songkhla Lake during 1977–2006

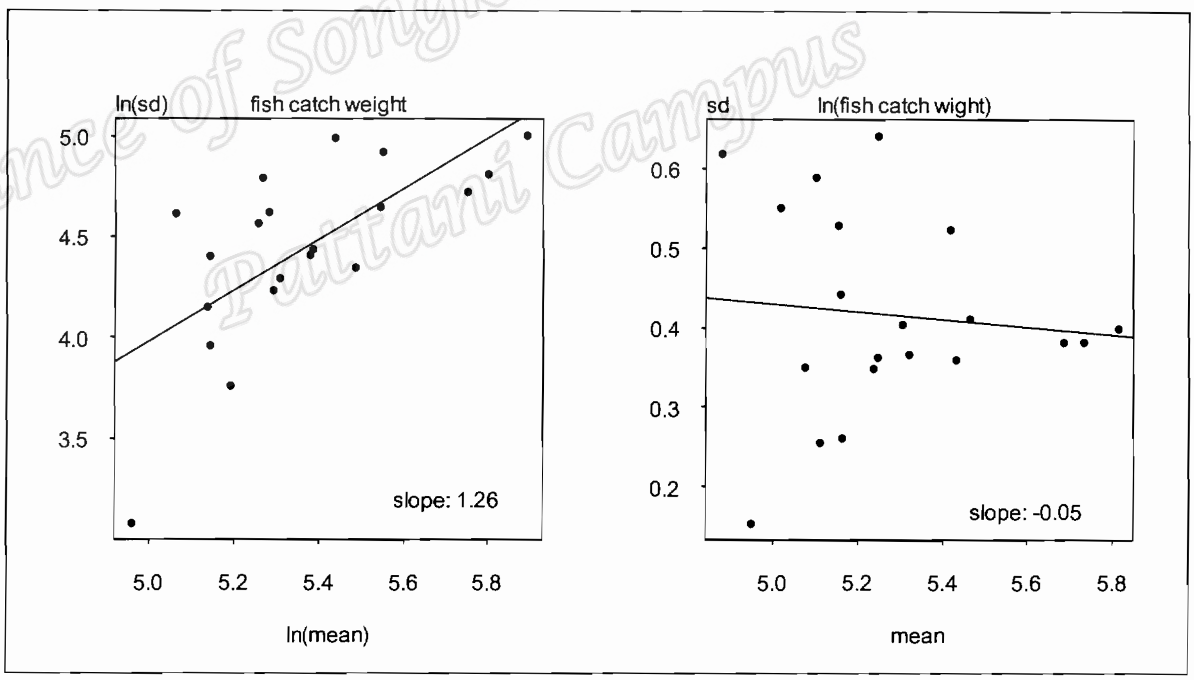

Figure 3.3 Relationship between mean and standard deviation

The left panel of Figure 3.3 shows the relationship between the mean and standard deviation in the 20 samples. It clearly shows a positive relationship, with a slope of

Prince of Songkla University Pattani Campus

1.26, confirming the need to transform the data. The right panel shows the relationship between the mean and standard deviation after taking natural logarithms. The slope of the fitted line is close to zero and we conclude that a logarithmic transformation is appropriate.

3.2 Linear regression model

The linear regression model is fitted to the natural logarithms of the fish catch weights. The predictor variables are the month of the year, the linear trend, the seasonal effect, and the catches in the proceeding two months as described in Chapter 2, equation (2.4), the result shown in Table 3.2, which gives the estimated values of the parameters.

Prince of Songkla University Pattani Campus

Table 3.2 Linear regression model for logarithms of monthly catches

Determinant	Coefficient	St Error	z-value	p-value
Constant α	1.9303	0.2805	6.88	< 0.0001
Trend β	0.0001	0.0002	0.51	0.6109
Lag1 δ_1	0.4955	0.0536	9.25	< 0.0001
Lag2 δ_2	0.1511	0.0536	2.82	0.0051
Season:				< 0.0001
Jan γ_1	0	-	-	-
Feb γ_2	-0.4147	0.0984	-4.21	< 0.0001
Mar γ_3	0.0871	0.1020	0.85	0.3938
Apr γ_4	-0.3907	0.0925	-4.23	< 0.0001
May γ_5	-0.0111	0.1027	-0.11	0.9137
Jun γ_6	0.0715	0.0943	0.76	0.4491
Jul γ_7	-0.2274	0.0937	-2.43	0.0157
Aug γ_8	-0.4081	0.0985	-4.14	< 0.0001
Sep γ_9	-0.0890	0.1018	-0.87	0.3825
Oct γ_{10}	0.1869	0.0968	1.93	0.0545
Nov γ_{11}	-0.0861	0.0929	-0.93	0.3549
Dec γ_{12}	0.4330	0.0963	4.50	< 0.0001
r-sq: 0.5119 df: 344 Residual Sum of Squares: 41.7690 σ: 0.3485				

Both the time-lagged variables were statistically significant but the trend was not. The seasonal effect was statistically significant with overall p-value < 0.0001. However, the seasonal effects vary from large negative values in February, April, and August to a large positive value in December. This model provided an r-squared of 0.51, implying that it accounted for 51% of the variation in the data. The correlation between observed values and the values fitted by the model is 0.714.

Figure 3.4 shows a scatter plot of the observed total fish catch with the natural log transformed values plotted against the fitted values in the left panel and the residuals plot in the right panel. The normality assumption for the errors is plausible because the points in the plot follow the line corresponding to normality with no extreme outliers. We use this model to forecast the monthly fish catch weight in Songkhla Lake during the twelve months in 2007 based on the equation (2.6) in Chapter 2.

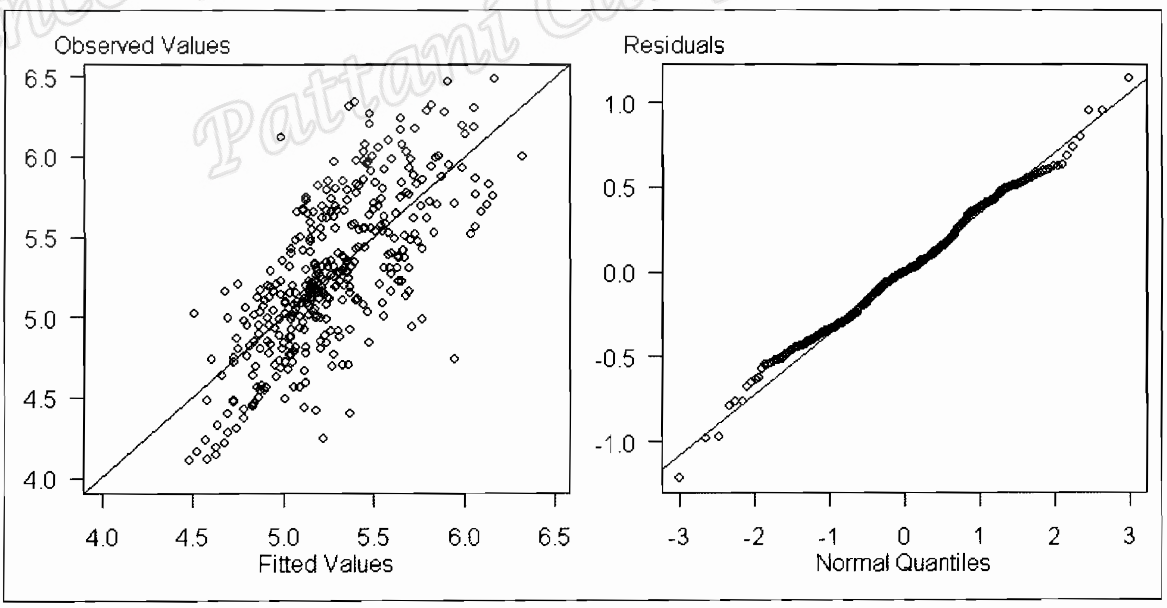

Figure 3.4 Plot of observed values and fitted values and residuals plot

Prince of Songkla University Pattani Campus

Figure 3.5 shows a plot of the time series of the data with the forecasts of 2007. It shows that the monthly forecast of fish catch weights were 167.87, 127.29, 191.26, 139.19, 184.83, 220.21, 185.92, 146.52, 174.64, 242.16, 222.54 and 376.81 tonnes.

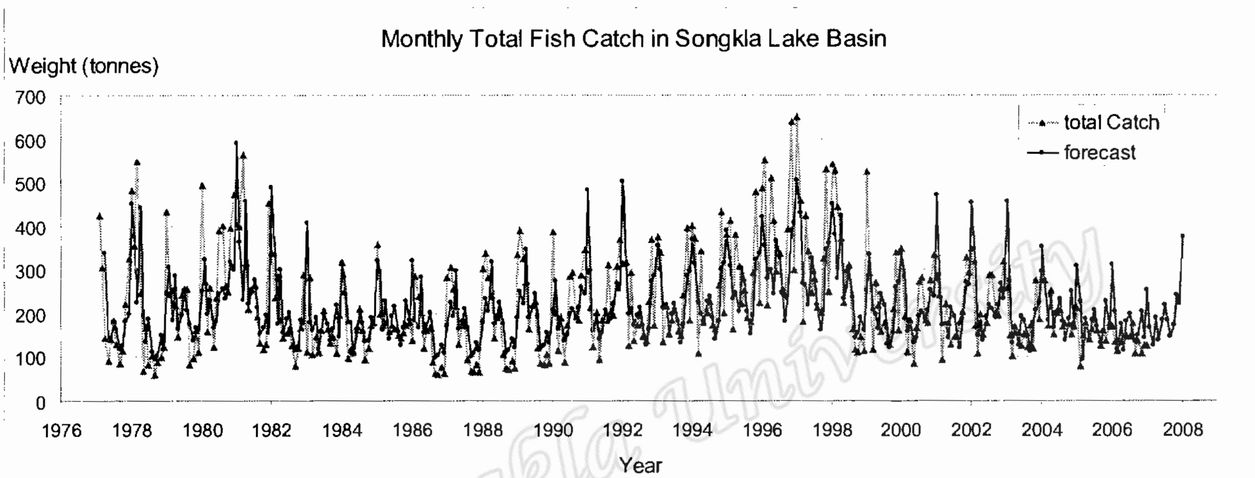

Figure 3.5 Monthly total fish catch in Songkhla Lake with model-based forecasts

Prince of Songkla University Pattani Campus

These forecasts of fish catch weights were compared to the actual values in 2006 with the percentage errors as shown in Table 3.3.

Table 3.3 Forecast and actual values of fish catches (tonnes) in 2006

Month	Forecast (95% CI)	Actual	% |error|
January	204.3 (180.9 - 227.7)	135.2	51.2
February	119.5 (96.1 - 142.9)	115.6	3.4
March	176.5 (153.1 - 199.9)	137.0	28.9
April	116.3 (92.8 - 139.7)	147.9	21.4
May	181.1 (157.7 - 204.5)	150.6	20.3
June	200.8 (177.4 - 224.2)	146.4	37.2
July	147.2 (123.8 - 170.7)	178.6	17.5
August	135.0 (111.6 - 158.4)	109.4	23.4
September	150.2 (126.8 - 173.6)	138.0	8.8
October	206.2 (182.8 - 229.6)	109.1	89.1
November	144.7 (121.2 - 168.1)	127.9	13.1
December	253.8 (230.4 - 277.3)	126.7	100.4
Average	169.6	135.2	34.6

Table 3.3 shows that the average percentage error of predicted fish caches was moderate.

3.3 Summary

The time series of monthly total fish catch weights from January 1977 to December 2006 revealed that the weights varied between 1,622 and 4,817 tonnes with an average of 2,639 tonnes. The catch peaked at 3,639 tonnes in 1980 and gradually declined to 1,817 tonnes in 1986, then increased to a second peak of 4,817 tonnes in 1996 before steadily declining to 1,622 tonnes in 2006. Fish catch weights were transformed using the natural logarithm, since the variation was not constant over the whole period. A linear regression model fit reasonably well to the data. The model showed a seasonal effect and the two time-lagged variables were statistically significant. The r–squared was 0.51, indicating that the model accounted for 51% of the variation in the data.

We consider the fish species composition and environmental parameters for the fish catch in the next chapter.

Prince of Songkla University Pattani Campus