# Chapter 4

# Statistical Modeling

In this chapter statistical modeling was performed to identify the strength of association between outcome and determinants. Multiple linear regressions with natural logarithm transformation were fitted to the low value fish in the Songkhla Lake. The objective in this chapter is to develop an appropriate statistical model for describing the low value fish catches, categorised by gear and group.

## 4.1 Models Fitting

The models fitting were selected by Akaike Information Criterion (AIC). The full model has four determinants (month-year, gear, species and fish group). After selection, the group variable was omitted from the full model. The reduced model has two or three determinants, which it was based on the model.

*The low value fish model*

In this section we fit linear regression models to the transformed the total low value fish catch weight. Linear regression of low value fish, trap was treated as the baseline for gear, 2003 January was treated as the baseline for month-year and *Macrobrachium equidens* species was treated as the baseline for species. The determinants, month-year, gear and species were statistically significant (p-value<0.0001) and trend was statistically significant also (p-value<0.0001).

Figure 4.1 shows the scatter plot of observed values and fitted value (in the left panel) and the residuals plot of the total value fish (in the right panel). The scatter is along the diagonal line which means that the model fitted with the data acceptable. The r-squared is 0.66. The scatter of residual is along the diagonal middle line. It can be seen that linear models are, as expected, acceptable. Thus, we could fit the linear regression models by gear in the next section.

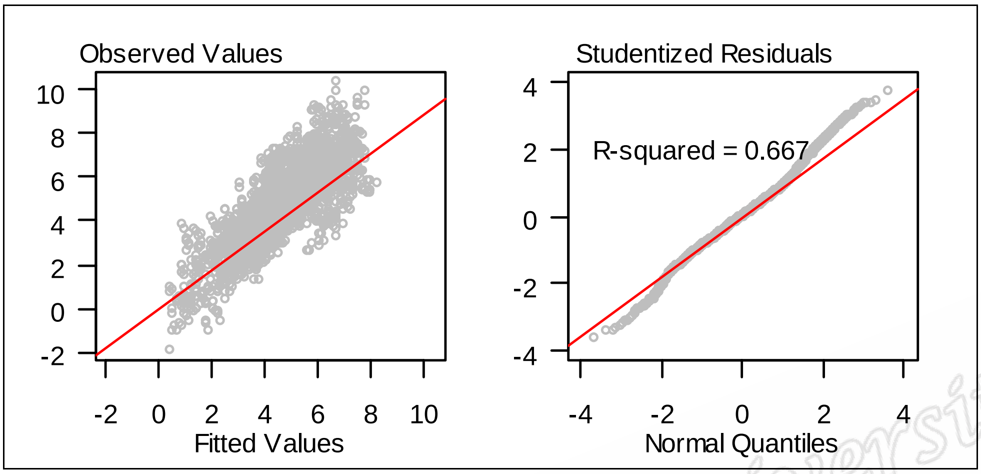


*Figure 4.1: Scatter plot of observed values and fitted value, residuals plot of the low value fish*

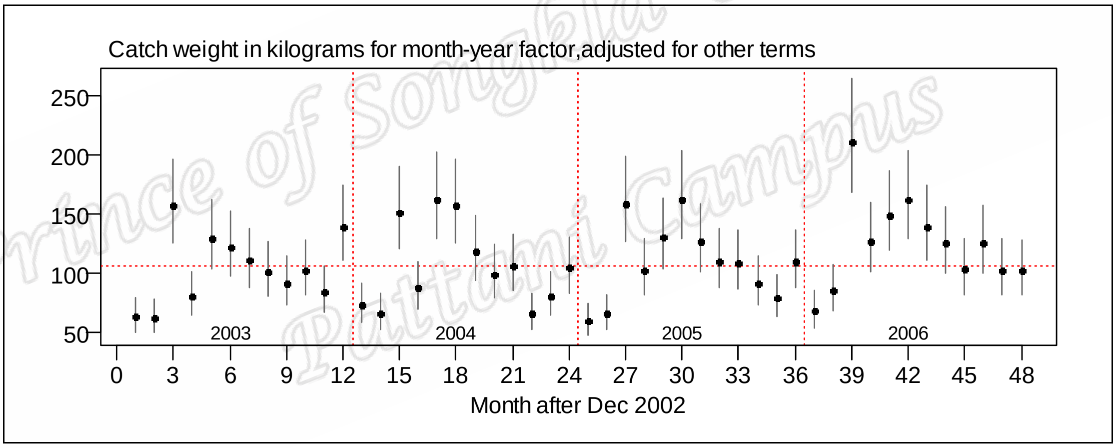


*Figure 4.2: The catch weight in kilograms of low value fish, adjusted for all other factors*

Sum contrasts were used to obtain confidence intervals. The 95% confidence intervals of the low value fish model for each factor; gear, species and month-year, were identified, after adjusting for the effects of the other factors in the model. The overall mean catch weight was 105.6 kilograms.

Figure 4.2 show the 95% confidence intervals of the catch weight in kilograms for month-year factor. The peaks are high annually in March; the maximum peak was

210.3 kilograms in March 2006. In 2005 and 2006, the peak resembles clearly the parabola, except the catch weight was higher in March than in other months. For gear and species factor, the highest catch weight of low value fish was in set bag net and *Rasbora lateristriata*, respectively.

*The low value fish model by gear*

In this section we fit linear regression models to the transformed the low value fish catch weight and show the scatter plot of observed values and fitted value, and residuals plot of the low value fish by gear; trap, set bag net and gill net. Linear regression of low value fish, 2003 January was treated as the baseline for month-year and *Macrobrachium equidens* species was treated as the baseline for species. The determinants in each gear model, month-year and species were statistically significant (p-value<0.0001) and trend was statistically significant also; trap (p-value<0.01), set bag net (p-value <0.01) and gill net (p-value<0.002).

Figure 4.3 show the scatter plot of observed values and fitted value (in the left panel) and the residuals plot of the total value fish (in the right panel), with top panel for trap, middle panel for set bag net and bottom panel for gill net. The scatter is along the diagonal line which means that the model fitted with the data acceptable, but the scatter is rather underestimate for set bag net and gill net. The r-squared values are 0.76, 0.68 and 0.77 for trap, set bag net and gill net, respectively. The scatter of residual is along the diagonal middle line, it is straight line for trap. It can be seen that linear models are, as expected, acceptable more than the total low value fish model; however, we could also fit the linear regression model for groups.

Sum contrasts were used to obtain confidence intervals. The 95% confidence intervals of the low value fish model for each factor; species and month-year were identified after each factor had been adjusted for the effects of the other factors in the model. The overall mean catch weights were 44.7, 376.6 and 69.9 kilograms in trap, set bag net and gill net model, respectively.

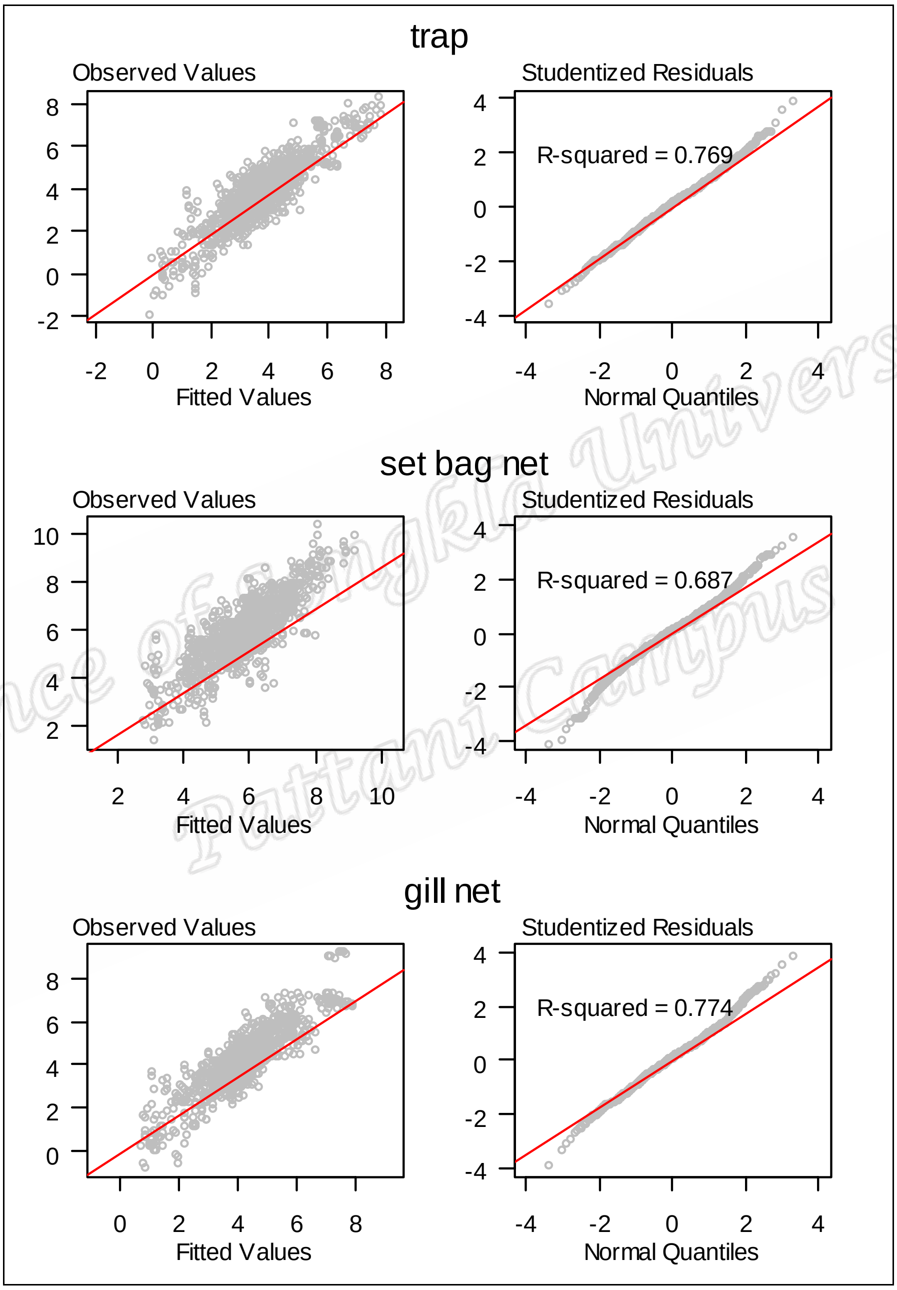


*Figure 4.3: Scatter plot of observed values and fitted value, residuals plot of the low value fish by gear*

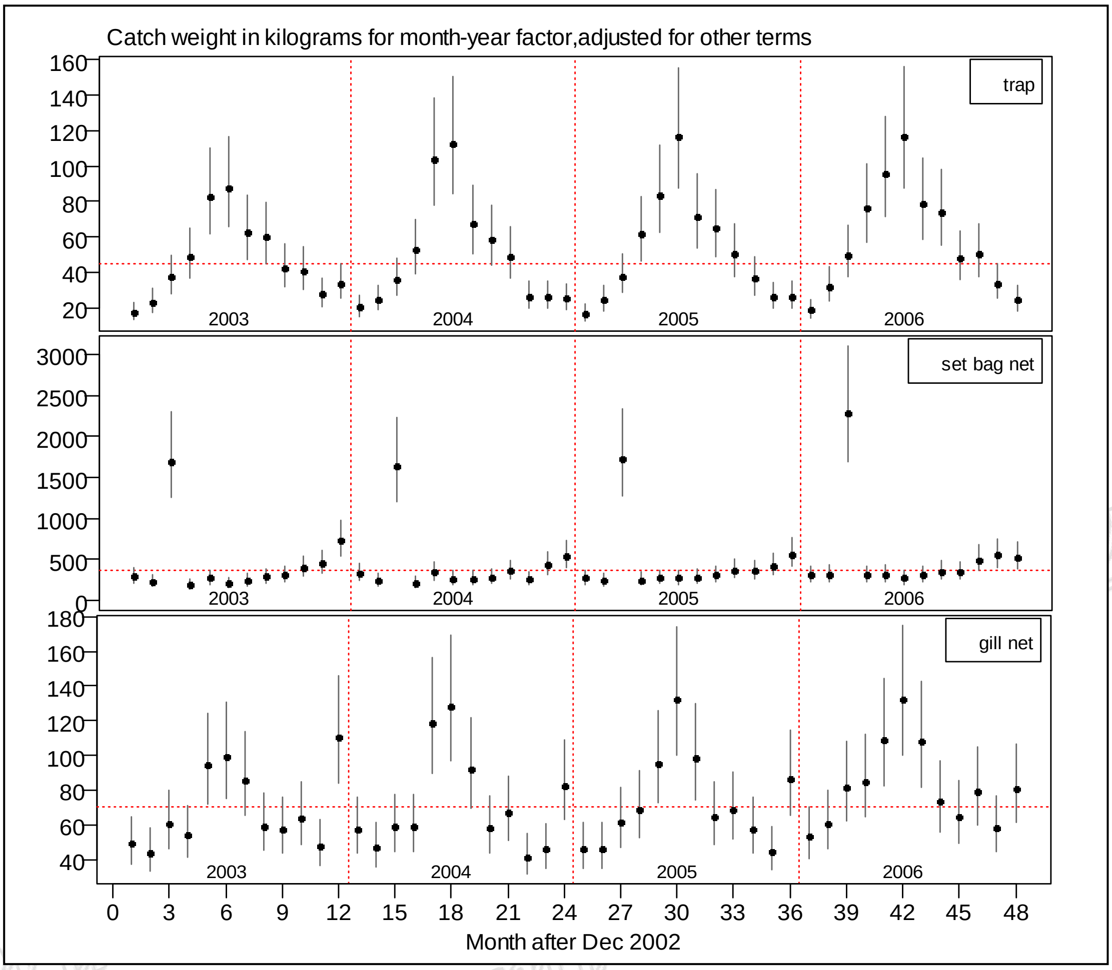


*Figure 4.4: The catch weight in kilograms of low value fish in each gear, adjusted for all other factors*

Figure 4.4 show the 95% confidence intervals of the catch weight in kilograms for month-year factor. The peak resembles the parabola in trap and maximum peaks in catch weights in June. The highest peak in catch weight was in March for set bag net but in the other months this gear had catch weights near the overall mean. The peak quite resembles the parabola in gill net, but with a high catch weight again in December, especially in December 2003. However the highest peaks for gill net catch in 2004, 2005 and 2006, were in June of those years. For species, the highest catch weight *Macrobrachium equidens* and *Rasbora lateristriata* in trap and gill net, *Ambassis marianus* and *Ambassis gymnocephalus* in set bag net.

*The low value fish model by group*

In this section we fit linear regression models to the transformed the low value fish catch weight and show the scatter plot of observed values and fitted value, residuals plot of the low value fish by group; estuarine and marine invertebrates, freshwater, estuarine and marine vertebrates. Linear regression of low value fish, trap was treated as the baseline for gear, 2003 January was treated as the baseline for month-year and species factor was treated as the various baseline based on amount species in each group. There was only one species in estuarine invertebrates group. *Sphaerozius nitidus*, *Trichogaster trichopterus*, *Tylosurus crocodilus* and *Tetraodon nigroviridis* was treated as the baseline for marine invertebrates, freshwater, estuarine and marine vertebrates respectively. The determinants in each gear model, month-year, gear and species were statistically significant (p-value<0.0001).

Trend was statistically significant for marine invertebrates (p-value<0.02), estuarine vertebrates (p-value <0.003) and marine vertebrates (p-value<0.0001), but it was not statistically significant for estuarine invertebrates and freshwater vertebrates.

Figure 4.5 show the scatter plot of observed values and fitted value (in the left panel) and the residuals plot of the total value fish (in the right panel), estuarine and marine invertebrates, freshwater, estuarine and marine vertebrates. The scatter is along the diagonal line which means that the model fitted with the data well, but the scatter is rather underestimate for estuarine invertebrates and freshwater vertebrates. The r-squared is 0.90, 0.82, 0.78, 0.85 and 0.74 respectively, obviously seen at the r-squares close to 1. The scatter of residual is along the diagonal line, it is rather straight line for in each group. Therefore linear models were fitted data well.

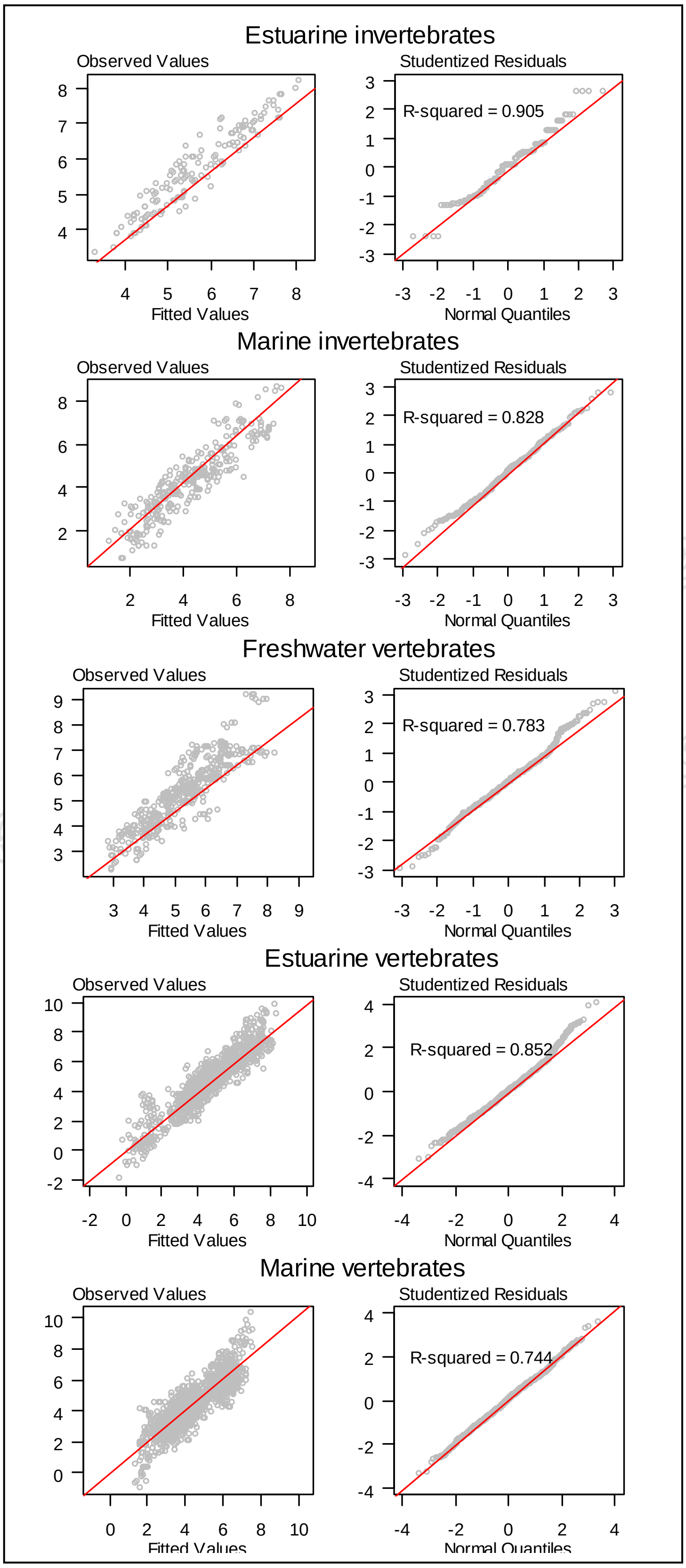


*Figure 4.5: Scatter plot of observed values and fitted value, residuals plot of the low value fish by groups*

Sum contrasts were used to obtain confidence intervals. The 95% confidence intervals of the low value fish model in each factor; gear, species and month-year were plotted, after each factor was adjusted for the effects of the other factors in the model. The overall mean catch weights were 297.7, 74.0, 202.1, 104.2 and 83.0 kilograms in estuarine and marine invertebrates, freshwater, estuarine and marine vertebrates' model, respectively.

Figure 4.7 show the 95% confidence intervals of the catch weight in kilograms for month-year factor. The peak of estuarine invertebrates varied in each year. In 2003 and 2004, the lowest peak for mean catch was in February, then the catch increased slowly until the highest peak in June. Then it reduced from July until November and peaked again in December. In 2005, the lowest peak was in February and highest peak was in March. In 2006, the peak was different to other years, with the highest peak in November, the second highest in March and lowest peak in January.

For marine invertebrates, the catch weight peak is near the overall mean, the peak is rather high in March to June in 2004. In 2005, the highest peak was in March, then there was a decrease until near overall mean in April and May and a high peak again in June. The peak resembles the parabola and the catch weight is near the overall mean.

The peak of freshwater vertebrates was similar in each year. The peak slightly decreased in January until June and increased in July until December. The catch weights in January and February were near the overall mean, lower than in March to June and higher than the overall mean in July to December.

The peaks of estuarine vertebrates were rather similar, and slightly resembled the parabola. The catch weight was lowest in January, except that in 2004 it was similar during January and February and highest in June. In 2006 it was highest in March, and in 2003 highest in May.

Also, the peak of marine vertebrates slightly resembles the parabola. The peaks are very much higher than the overall mean for March, especially in 2003 and 2006. The peaks are quite higher than the overall mean in May and June and for the other months are near the overall mean.

For gear, the highest catch weight by trap in estuarine invertebrates, by gill net in freshwater vertebrates and by set bag net in 'the rest' group. For species, the highest catch weights were for *Sphaerozius nitidus*, *Rasbora lateristriata*, *Ambassis Marianus* and *Thryssa dussumieri* in estuarine and marine invertebrates, freshwater, estuarine and marine vertebrates.

Note that, the linear models are as expected acceptable, it is high by group, moderate by gear and the last, total low value fish. It obviously shows at the r-squares close to 1 and the residual line quite straight line. The month-year factor is show clearly different of pattern in the catch weight by gear. For group, it shows different the month-year pattern in some group, contrast some group is not clear.

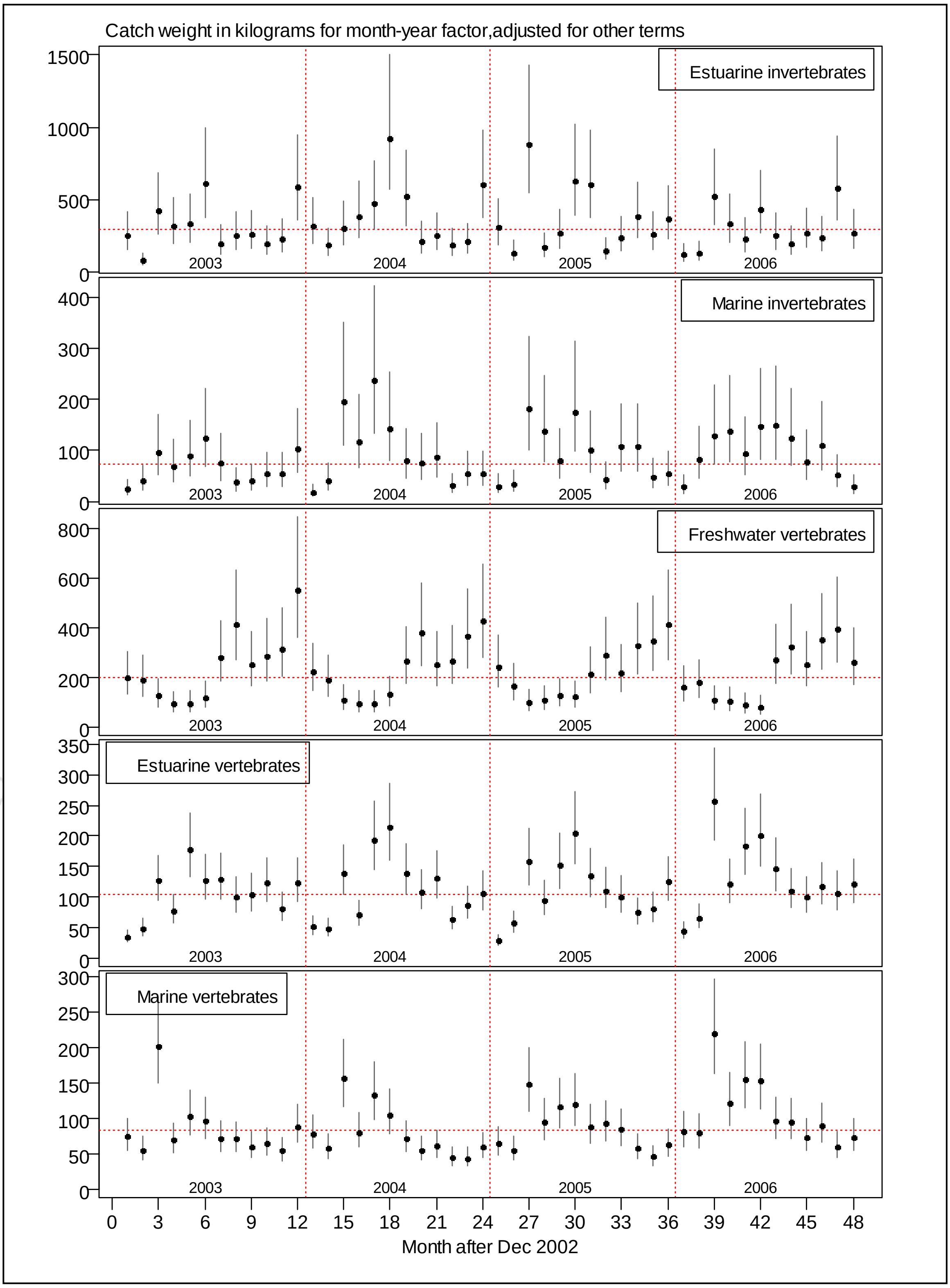


*Figure 4.6: The catch weight in kilograms of low value fish in each group, adjusted for all other factors*