DETECTING CHANGES IN PINK SALMON (*Onchorhynchus gorbuscha*) FRY DENSITY FOR HUMPY CREEK, ALASKA, AND DETERMINATION OF SAMPLE SIZE REQUIREMENTS

By:

B. Alan Johnson

Regional Information Report[1] No. 4K90-3

Alaska Department of Fish and Game
Division of Commercial Fisheries
211 Mission Road
Kodiak, Alaska 99615

February 1990

[1]The Regional Information Report Series was established in 1987 to provide an information access system for all unpublished division reports. These reports frequently serve diverse ad hoc informational purposes or archive basic uninterpreted data. To accommodate timely reporting of recently collected information, reports in this series undergo only limited internal review and may contain preliminary data; this information may be subsequently finalized and published in the formal literature. Consequently, these reports should not be cited without prior approval of the author or the Division of Commercial Fisheries.

Abstract

The counts of pink salmon *(Onchorhynchus gorbuscha)* fry in Humpy Creek, Alaska, for the period 1977–1988 were analyzed to determine if differences in the number of live fry per dig could by related to escapement levels. The Wilcoxon rank sum test was used to test the differences in live fry per dig between years. When the escapement was more than 270,000 it was found that statistical differences (α=0.1) existed between adjacent years. Sample sizes necessary to detect a difference (or lack of difference) were developed using two sets of years representing equal escapement and unequal escapement. A sample size of 30 digs with live fry will provide a stable Type I and Type II error for testing the hypothesis of equal fry counts.

Introduction

This paper provides a preliminary analysis of the question of necessary sample sizes to detect changes in pick salmon *(Onchorhynchus gorbuscha)* fry densities from year to year due to changes in the level of escapement. It is based on the 1977–1988 fry dig data for Humpy Creek along with the escapement estimates. The data from a fry dig is the information obtained at a single site sampled along the creek. The gravel at the site is flushed with water to force eggs and fry from the gravel into a net. The number of eggs, dead fry, and live fry are counted. This analysis is limited to the number of live fry per dig, hereafter referred to as fry per dig. The number of digs varied from 40 to 90 for the period analyzed.

Methods

This analysis used exploratory data analysis techniques, both graphical and mathematical, to determine if any relation existed between escapement and the number of fry per dig. Also looked at was the relationship between the count distributions of fry per dig, both between and across the years sampled.

The t-test, the median test, the Wilcoxon rank sum test, the t-test on rank data, QQ-plots, and contingency table analysis were used in the exploratory phase. Final analysis used the Wilcoxon due to its robustness for non-Gaussian distributions.

The two sets of years 1986/1988 and 1977/1978 were used as a basis for a Monte Carlo simulation of reduced sample sizes. One hundred samples were used to determine the number in the test statistic critical region along with an overall mean test statistic.

All analysis was undertaken using S (an AT&T registered trademark).

Results and Discussion

The data was highly variable and most of the exploratory analysis was fruitless. The following results are based on zero-truncated fry counts (positive counts) as it was the only informative cut of the data at this preliminary stage.

Table 1 presents the result of testing adjacent years for statistical differences at the $\alpha = 0.1$ level. The is a weak pattern associated with the escapement level. What would be thought to be affected or unaffected years based on the magnitude of the escapement are evident with the small P-values. A pattern does exist for the differences when the escapement is greater than 270,000.

Table 1.— Differences in the distributions of live fry per dig counts in adjacent years for the period 1977–1988 for Humpy Creek

Year	Escapement (x1000)	Number of digs	Wilcoxon T_1	P-value
1977	92	44		
1978	290	73	2.271	0.024
1979	306	33	4.181	<0.001
1980	260	52	-4.519	<0.001
1981	240	44	0.452	0.652
1982	160	68	0.630	0.528
1983	142	68	0.910	0.362
1984	95	40	0.501	0.616
1985	273	37	1.507	0.132
1986	118	79	-4.071	<0.001
1987	208	63	0.522	0.602
1988	120	62	0.583	0.560

The Monte Carlo simulations are developed by simulating the full sample size followed by simulating sample sizes that reduce each original sample by the same integer value. Each of the simulations is repeated 100 times by sampling the original data with replacement at the reduced sample size. Each simulation took approximately 2 hours.

The 1986/1988 set of data has a test statistic of $T_1 = 0.998$ ($P = 0.841$). The two years are not statistically different and have similar escapement of 118,000 and 120,000 respectively. The first line in Table 2 shows what the simulation with original samples sizes (79,62) gives for number of samples rejected. There are 24 of the samples with a test statistics greater than the critical value (1.645) and the average of the 100 test statistics is $T_1 = 1.075$. The question is how well do reduced sample sizes compare with the full sample.

Table 2.— Tail frequencies for a Monte Carlo simulation of reduced sample sizes for 1986 and 1988.

Integer reduction	Sample size 1986	Sample size 1988	Number greater than 1.645	Average Wilcoxon T_1
0	79	62	24	1.075
5	74	57	28	0.962
10	69	52	30	1.042
15	64	47	22	0.849
20	59	42	19	0.726
25	52	37	27	0.899
32	47	30	28	0.845

From the results in Table 1, we can conclude that we would fail to reject the null hypothesis of equality of distributions just as likely with a sample size of 30 as with the original sample size of 62.

The data set for 1977/1978 gives us a picture of rejecting the null hypothesis of equality under a reduced sample size scenario. Table 3 provides the same information as extracted for Table 2. From this we can conclude that the null hypothesis would just as likely be rejected with a sample size of 30 as with the original sample size of 44.

Table 3.— Tail frequencies for a Monte Carlo simulation of reduced sample sizes for 1977 and 1978

Integer reduction	Sample size 1977	Sample size 1978	Number greater than 1.645	Average Wilcoxon T_1
0	44	73	76	2.370
5	39	68	74	2.326
10	34	63	65	2.104
14	30	59	64	1.940

Taking into account the decisions of rejection and failure of rejection in the two sets of data, combined together we would expect a reasonable Type I and Type II error.

Conclusions

We have shown a pattern (although not explicit) of differences in counts of live fry per dig between adjacent years from the period 1977–1988 in Humpy Creek. The test under a reduced sample size would yield the same results using sample sizes of 30. The Type I error and Type II error appear to hold under this reduced sampling scheme. This is to be expected with the Wilcoxon test, which has an asymptotic relative efficiency (compared to the t-test) of never less than 0.86 (Conover 1980). A sample design of sampling until 30 fry digs have at least one live fry would provide the necessary sample size to detect a difference in fry counts. The difference could possibly be due to a major change in the previous year escapement. This is preliminary work on an extremely variable data set and may or may not apply to other systems. However, the robustness of the test statistic will detect differences and the only further work that might be needed is to address that the differences may not be due to escapement levels.

Literature Cited

Conover, W. J. 1980. Practical nonparametric statistics. John Wiley, New York. 493 pp.