

An analysis of the homogeneity of salmon tag recoveries in ocean troll fishery catches

Ken Newman, U.S. Fish and Wildlife Service

August 28, 1991

1 Introduction

Experiments with loading densities, feeding regimes, and time of release, to name a few factors, have been conducted for years at anadromous fish hatcheries. The effects of these treatments on responses such as ocean survival rates and contribution rates to particular regional fisheries need to be assessed using appropriate statistical methodology. Most classical experimental designs and analyses begin with an assumption of statistical independence, i.e., the probability of an experimental unit taking on some value is not affected by the response of a different experimental unit.

The purpose of this report is to describe a procedure for detecting statistical dependencies between fish and to present the results of an application of the procedure to real data. In particular attention is focused on the distribution of recoveries of coded-wire-tagged salmon in troll catches off the Washington coast. It has been noted by others [3] that it appears that recoveries of a given tag code tend to cluster by catches. E.g., if there were 50 catches of roughly equal size during a week in some area and 20 total recoveries of a particular tag code, then the majority of the recoveries are found in a few catches. Such a phenomenon would be consistent with a theory that fish schooling is a function of kinship, a particular theory being one of inherited pheromonal attraction[2].

The basic idea behind the procedure described in this report is that if the fish do not cluster or school on the basis of kinship, or say hatchery of origin, then the distribution of recoveries should be relatively evenly distributed

across the catches. To say this in conventional statistical terms, the distribution of recoveries should be homogenous with respect to catch and the test applied is generally referred to as a test of homogeneity of proportions.

2 Summary

There appears to be little evidence that recoveries of fish in the Washington ocean troll fishery cluster by stock of origin on a weekly-WDF¹ catch area basis. Statistical tests of the hypothesis of homogeneity of recoveries by stock type were conducted on 1988 and 1989 data and significant deviations from homogeneity occurred only sporadically.

This does not prove that fish do not school with fish of like origin in the ocean, however. There are a couple of possible explanations for the apparent independence of recoveries and stock origin. One is that even if there is clustering of fish stocks in particular catches, the percentage of the catch that is identifiable by stock is so low, generally less than 5%, that it is extremely difficult to detect the clustering. Another possibility is that the troll fishery operates in such a manner that an individual boat's catch is more of a random sample of the fish in an area than a random sample of clusters or schools of fish. For purse seine catches of chinook salmon in Alaska, for instance, it has been observed that most of the recoveries of tagged fish were found in a few catches [3], but I have not analysed this data.

Again, however, based on the current level of information for the Washington ocean troll fishery, I have found little evidence for statistical dependency in the catches with respect to stock.

3 Methods

The analysis methods are aimed at testing the hypothesis that for some spatial-temporal level of resolution the fish are evenly mixed with regard to their region of origin. There still might be clustering occurring, but the basis for clustering is not a function of kinship. Assuming that the total population of fish is large relative to the harvest taken, each catch can be viewed as a draw from a multinomial population where the attributes are the

¹Washington Department of Fisheries.

region of origin. A two way classification of the harvested fish by catch and region of origin (which includes unknown region, or untagged fish) yields a contingency table.

A standard statistical procedure for the analysis of a contingency table for the homogeneity of proportions in each sample, in this case each catch, is a chi-square test of homogeneity². The accuracy of the chi-square test depends upon how closely the calculated test statistic follows the null distribution. The null distribution is chi-square with degrees of freedom equal to $(C-1) \times (R-1)$, where C is the number of catches and R is the number of regions of origin. The calculated test statistic, unfortunately, is very poorly approximated by a chi-square distribution when the table is sparse, i.e., when there are cells with 5 or fewer counts.

The tables generated for the Washington ocean troll fishery were very sparse. An example is presented on the next page.

²An asymptotically equivalent procedure is a likelihood ratio test based on an underlying multinomial distribution.

An example of a typical troll catch by recovery type (stock or tag code) contingency table³.

	Catch\ "Stock"											
	1	2	3	4	5	6	7	8	9	10	11	12
1 :	36	1	1	0	0	0	0	0	0	0	0	0
2 :	2	0	0	0	0	0	0	0	0	0	0	0
3 :	22	0	0	1	2	1	0	0	0	0	0	0
4 :	22	0	0	0	0	0	1	0	0	0	0	0
5 :	2	0	0	0	0	0	0	0	0	0	0	0
6 :	2	0	0	0	0	0	0	0	0	0	0	0
7 :	4	0	0	0	0	0	0	0	0	0	0	0
8 :	2	0	0	0	0	0	0	0	0	0	0	0
9 :	5	0	0	0	0	0	0	1	0	0	0	0
10:	1	0	0	0	0	0	0	0	0	0	0	0
11:	8	0	0	0	0	0	0	0	0	0	0	0
12:	2	0	0	0	0	0	0	0	0	0	0	0
13:	42	0	0	0	0	0	0	0	1	0	0	0
14:	4	0	0	0	0	0	0	0	0	0	0	0
15:	1	0	0	0	0	0	0	0	0	0	0	0
16:	9	0	0	0	0	0	0	0	0	0	0	0
17:	5	0	0	0	0	0	0	0	0	1	0	0
18:	11	0	0	0	0	0	0	0	0	0	1	0
19:	3	0	0	0	0	0	0	0	0	0	0	0
20:	5	0	0	0	0	0	0	0	0	0	0	0
21:	20	0	0	0	0	0	0	0	0	0	0	1
22:	7	0	0	0	0	0	0	0	0	0	0	0
23:	2	0	0	0	0	0	0	0	0	0	0	0
24:	2	0	0	0	0	0	0	0	0	0	0	0
25:	1	0	0	0	0	0	0	0	0	0	0	0
26:	3	0	0	0	0	0	0	0	0	0	0	0
27:	3	0	0	0	0	0	0	0	0	0	0	0
28:	11	0	0	0	0	0	0	0	0	0	0	0

³The first column is the number of fish of unknown origin. Details of table construction are given below. This particular table is for the 1988, statistical week 19, catch area 74 troll fishery with "stocks" being tag codes pooled by common hatchery and release sites.

The methods described in this report are procedures for analysing sparse contingency tables. Several methods have been developed to deal with this problem, but I have chosen to use just two of the procedures which appear to perform well based on simulated test examples. The two procedures are the iterated bootstrap test and a monte carlo version of an exact test. Before detailing the procedures, the way the tables were constructed and the pooling of tag codes are discussed.

3.1 Table creation and pooling of tag codes

There were several steps to the table preparation. Three types of data files containing catch, fish head labels, and CWT code information, respectively, were provided by WDF. Information about the hatchery and release sites corresponding to each code were provided by the U.S. Fish and Wildlife Service. Each record of a catch file contains information about total catch size and the number of chinook and coho fish heads taken. Each record of the head label file corresponds to a particular head and provides information about the catch it came from. Finally each record of the CWT code file corresponds to a fish head and reports the tag number if available (i.e., if a tag was found in the head and it was readable).

To create a two-way table, a computer program first read a single record of the catch file, and retrieved the time-area-gear information and the number of chinook and coho heads taken. The head label file was then read, searching for the records corresponding to each chinook and coho head in the catch. There were many instances where not all the heads could be accounted for—such heads were then lumped into the 'unknown' stock category. Then for each head in the head label file, the CWT code file was read and the CWT code found. There were several instances of unknown CWT codes, either no tag was found or the tag was unreadable. These heads were also tallied in the 'unknown' stock category.

Three important digressions:

- Should the unknown category be included in the table? Under the null hypothesis of homogeneity, the probability of a stock being unknown is simply the sum of the probabilities for all the untagged stocks. If the untagged stocks were not homogeneous but their proportions summed to a constant value, then the tests will fail to detect the non-

homogeneity because of the aggregation. If the proportions do not sum to a constant value, however, then given sufficiently large deviations in the unknown proportion between catches, the test should detect the non-homogeneity.

- Is the treatment of the missing or unreadable tags appropriate? If the missing or unreadable tags are missing at random, i.e., independently of catch and stock of origin, then the inclusion of those fish in the unknown group should not affect the probability of a fish being in the unknown category.
- Should catches that are quite small in total number be included in the tables? In some cases the troll boat catch was only 1. I do not know how much such small catches affect the power of the test to detect deviations from homogeneity. I did conduct the tests on several reduced tables where only catches greater than 20 were included, but found little difference from the results for the full tables. The results presented later include all the catches in a given time-area cell. The issue of testing power was not explicitly addressed, however.

Two levels of aggregation of tag codes were analyzed. The first level was at the finest stock identification resolution available. All recoveries with identifiable tag codes were tallied in a category unique to the tag code; all unidentifiable recoveries, due to lost tags or unreadable tags, were aggregated in the unmarked category.

The other level of resolution was to pool all tag codes that came from the same hatchery, were released from the same site in the same year, and were the same species. This approach is not an ideal means of aggregating tag codes since different tag codes might still be associated with different hatchery treatments. However, given the available data, it seems a reasonable means of pooling tag codes.

3.2 The iterated bootstrap test

The iterated, or two stage, bootstrap test is a computer intensive, non-parametric general purpose hypothesis testing procedure proposed by Beran [1]. It is a refinement of the notion of single stage bootstrap hypothesis test-

ing that reduces the magnitude of the potential bias in that procedure. The mechanics of the procedure are as follows.

1. Choose a test statistic such that large values of the test statistic suggest a deviation from the null hypothesis.
2. Generate $n-1$ bootstrap samples and calculate $n-1$ test statistics for each sample.
3. Calculate the relative rank of the original sample's test statistic. This is the first stage p-value.
4. From each of the $n-1$ bootstrap samples, generate $m-1$ 'second stage' bootstrap samples, treating the 'first stage' samples as if they were real samples. Calculate $m-1$ test statistics in each case and calculate p-values for each 'first stage' test statistic relative to the 'second stage' test statistics.
5. Calculate the relative rank of the first stage p-value to the $n-1$ p-values. This is the second stage p-value and the basis for 'rejecting' or 'accepting' the null hypothesis.

One might visualize the process as one of creating an $n-1$ by $m-1$ matrix of test statistics, where in the leftmost column sit the $n-1$ first stage bootstrap sample test statistics. For each row r all the columns to the right of the first column contain the second stage bootstrap sample test statistics based on the r th (bootstrap) sample. $n-1$ p-values are then calculated on the basis of rankings of the first column's test statistic value relative to the rest of the row entries.

Symbolically, the first stage p-value for the observed test statistic T ,

$$p_1 = \frac{\#T^* \geq T}{n},$$

and the second stage p-value,

$$p_2 = \frac{\#p^* \leq p_1}{n},$$

where $\#T^*$ and $\#p^*$ refer to the bootstrapped test statistics and p-values.

3.3 A monte carlo exact test

The monte carlo exact test is an extension of Fisher's exact test (FET), a testing procedure for analyzing 2 by 2 contingency tables. I'll briefly explain the logic behind the FET for tests of homogeneity. The 2 by 2 table is a classification of entities from 2 samples (of fixed sizes) into one of two categories, say successes and failures. Suppose that the samples are simply a random division of the $n_1 + n_2$ objects and the number of successes, $m_.$, was fixed beforehand. Then the arrangement of m_1 successes in sample 1 and $m_2 = m_1 - m_1$ successes in sample 2 is due to chance alone. The FET p-value is simply the calculation of the probability of getting a table with successes as 'extremely unbalanced' as was observed. E.g., if the samples were of equal size and there were 10 successes in sample 1 and 1 success in sample 2, the p-value is the probability of 10 or 11 successes in sample 1. Incidentally, this is just the calculation of hypergeometric probabilities where $n_.$ is the population size, $m_.$ is the number of 'successes' and n_1 is the sample size.

The extension to two way tables with more than 2 categories is to fix the row and column totals and calculate the probabilities for all tables with probabilities of occurrence that are less than or equal to the probability of the observed table. The monte carlo test modification is simply to generate tables with the same row and column totals and calculate the probabilities empirically. The algorithm used for this project is one developed by Patefield [5].

4 Results

The choice of a spatial-temporal resolution at which stocks are hypothesized to be randomly mixed is somewhat arbitrary. At one extreme one might begin by assuming that the entire Washington coast fish population is randomly mixed throughout an entire fishing season which may span 4 or 5 months. I chose to begin at a much finer level, however, intending to aggregate levels if homogeneity was not disproved. The temporal level chosen was a statistical week and the spatial level was a Washington Department of Fisheries catch area.

The data came from 1988 and 1989 catch samples of the Washington coastal fishery collected by WDF port samplers at Ilwaco, Westport, La

Push, and Neah Bay. The particular weeks chosen were weeks 18 through 30. Although there were some relatively large catches for some weeks after week 30, they were not included in the analysis because of the potentially confusing influence of freezer boat catches⁴. The primary areas covered were 1, 2, 3, and 4, corresponding to the four ports, although some of the catch came from areas 5, 74 and 84. I further restricted attention to troll catches only. Sport catches are rarely large enough to yield one recovery, let alone multiple recoveries.

For the exact test 999 random samples of tables with fixed margins were generated. For the iterated bootstrap 99 iterations of the first stage were combined with 99 iterations at the second stage (for each of the first stage samples). The tables on the following pages present the estimated p-values for both tests for areas 1, 2, 74, and 84. The first set of tables are based on contingency tables where each unique tag code is treated as a distinct outcome and results are presented for 1988 and 1989. The second set of results are based on contingency tables with tag codes pooled by common hatchery-release site combinations, as was discussed earlier, but includes only 1988 data.

⁴'Freezer boats ... may fish for weeks and accumulate large numbers of fish that aren't accounted for on fish tickets until the season closes.[4]'

4.1 Unpooled tag codes

Area 1 results- 1988

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
19	153	13	6	0.85	0.72
20	274	11	13	0.93	0.91
21	209	16	8	0.76	0.76
22	19	5	2	0.58	0.49
23	33	7	1	1.00	0.49
24	420	35	12	0.31	0.41
25-30:	0 catch (mostly)				

Area 1 results- 1989

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
19	96	2	4	0.32	0.30
20	83	5	0	NA	NA
21	5	1	0	NA	NA
22-30:	0 catch				

Area 2 results- 1988.

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
18	51	15	1	0.23	0.15
19	1868	166	44	0.01*	0.04*
20	1872	147	55	0.12	0.11
21	2493	194	70	0.46	0.26
22	2376	161	72	0.59	0.58
23	1118	139	31	0.96	0.96
24	2889	214	73	0.43	0.47
25	730	56	31	0.21	0.26
26	851	2	39	0.46	0.58
27	202	6	10	0.82	0.83
28	178	7	10	1.00	0.96
29	196	3	7	0.45	0.36
30	286	8	16	0.54	0.60

Area 2 results- 1989.

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
19	3244	159	80	0.45	0.31
20	2321	151	70	0.73	0.71
21	1875	94	56	0.03	0.05
22	2028	112	61	0.24	0.13
23	3429	118	86	0.01*	0.01*
24	2582	101	74	0.05	0.15
25	1157	70	42	0.21	0.27
26	14	4	0	NA	NA

Area 74 results- 1988.

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
18	9	2	0	NA	NA
19	249	28	12	0.40	0.50
20	648	17	30	0.01*	0.03*
21	585	21	33	0.48	0.56
22	555	10	34	0.37	0.48
23	456	12	17	0.34	0.37
24	1735	26	48	0.11	0.21
25	2794	32	82	0.29	0.40
26	210	8	4	1.00	0.77
27	434	19	22	0.09	0.22
28	433	24	4	0.20	0.20
29	52	8	0	NA	NA
30	133	10	2	0.69	0.50

Area 74 results- 1989.

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
18	0	NA	NA	NA	NA
19	371	5	28	0.32	0.48
20	44	2	2	1.00	0.57
21	67	2	5	0.63	0.44
22	262	4	11	0.00*	0.01*
23	264	13	15	0.82	0.66
24	386	11	19	0.75	0.86
25	46	9	0	NA	NA
26	100	8	3	0.44	0.55
27	11	2	0	NA	NA
28	0	NA	NA	NA	NA

29	129	3	6	0.75	0.80
30	371	6	8	0.83	0.86

Area 84 results- 1988.

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
18	285	54	4	0.45	0.40
19	394	44	9	0.12	0.08
20	369	43	13	0.37	0.30
21	312	40	12	0.06	0.09
22	455	56	14	0.48	0.46
23	407	47	11	0.44	0.33
24	350	50	11	0.15	0.10
25	93	24	0	NA	NA
26	133	26	5	0.11	0.26
27	1	1	0	NA	NA
28	13	2	0	NA	NA
29	4	3	0	NA	NA
30	62	4	3	0.60	0.35

Area 84 results- 1989.

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
18	280	40	8	0.35	0.45
19	124	28	1	0.55	0.54
20	96	23	4	0.79	0.82
21	218	36	11	0.33	0.28
22	1604	37	15	0.00*	0.04*
23	152	15	9	0.19	0.24
24	117	7	3	0.06	0.04*
25	180	32	6	0.71	0.80
26	179	24	9	0.56	0.43
27	65	14	3	0.91	0.70
28	0	NA	NA	NA	NA
29	0	NA	NA	NA	NA
30	211	7	3	0.84	0.86

4.2 Pooled tag codes: 1988 only

Area 1 results- 1988

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
19	153	13	5	0.70	0.64
20	274	11	11	0.86	0.85
21	209	16	6	0.77	0.77
22	19	5	2	0.58	0.49
23	33	7	1	1.00	0.49
24	420	35	10	0.31	0.37
25-30: 0 catch (mostly)					

Area 2 results- 1988.

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
18	51	15	1	0.23	0.15
19	1868	166	27	0.02*	0.03*
20	1872	147	33	0.12	0.15
21	2493	194	40	0.32	0.14
22	2376	161	43	0.40	0.41
23	1118	139	22	0.72	0.97
24	2889	214	40	0.43	0.63
25	730	56	24	0.18	0.22
26	851	2	25	0.31	0.45
27	202	6	7	0.93	0.94
28	178	7	8	0.99	0.99
29	196	3	7	0.45	0.36
30	286	8	14	0.59	0.61

Area 74 results- 1988.

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
18	9	2	0	NA	NA
19	249	28	11	0.28	0.36
20	648	17	25	0.01*	0.03*
21	585	21	25	0.46	0.55
22	555	10	21	0.45	0.57
23	456	12	13	0.39	0.54
24	1735	26	32	0.01*	0.17
25	2794	32	47	0.22	0.33
26	210	8	4	1.00	0.77
27	434	19	15	0.13	0.26

28	433	24	4	0.20	0.20
29	52	8	0	NA	NA
30	133	10	2	0.69	0.50

Area 84 results- 1988.

Week	#Fish	#Catches	#Tagcodes	Exact(999)	Boot(99*99)
18	285	54	4	0.45	0.40
19	394	44	8	0.05*	0.02*
20	369	43	11	0.38	0.33
21	312	40	10	0.07	0.08
22	455	56	12	0.53	0.48
23	407	47	11	0.44	0.33
24	350	50	9	0.18	0.11
25	93	24	0	NA	NA
26	133	26	4	0.18	0.30
27	1	1	0	NA	NA
28	13	2	0	NA	NA
29	4	3	0	NA	NA
30	62	4	2	0.06	0.35

5 Discussion

First, the two tests yielded generally similar results throughout all the contingency tables. I had made some preliminary comparisons between the procedures on non-sparse matrices, ones for which the ordinary chi-square test of homogeneity would perform well. The results of these comparisons indicated that the exact, the iterated bootstrap, and the usual chi-square tests yielded essentially identical results. The consistency with sparse matrices between the monte carlo exact test and the iterated bootstrap tests is reassuring.

The results of the tests for tables in which tag codes are separated at the finest resolution did not differ that greatly from tests of tables based on pooled tag codes. The number of tag groups often greatly decreased after pooling, e.g., 1988, area 2, week 21 tag groups decreases from 70 to 40. But the decrease in sparseness really did not alter the results 'significantly'.

As a sample table indicated earlier these were extremely sparse tables to analyse. With the exception of the unknown category the observed proportions for particular categories were generally less than 1%. How sensitive the

tests are to deviations in the 'rare' categories was not examined. Inspection of some of the cases of significant non-homogeneity suggests that instances of 3 or more recoveries of one rare type in one catch with 1 or fewer of the same category in other catches could be the reason for a small p-value. In other cases it was more likely variation in the percentage in the unknown category- wide variation here is indicative of non-homogeneity, too, as was explained in the Methods subsection on table creation.

When a large number of tests are conducted one expects to find instances of small p-values even if the 'true' proportions are homogenous because of sampling variation alone. 106 tests were carried out with 6 instances in which both the exact test and the iterated bootstrap both yielded p-values less than or equal to 0.05. The expected number of falsely significant results for 106 tests when testing at the 5% level is 5.3 occurrences. There is little evidence for consistent non-homogeneity of stock proportions in the Washington ocean troll fishery on a weekly-catch area basis. This could be because the troll fishery itself catches fish at random throughout an area. I.e., the troll fishery is not intensively fishing schools of fish at random, where the pools may be segregated by stock type or 'kin'. On the other hand it could be that because the samples are largely unidentifiable in terms of stock type, roughly 95% in the proportions are too difficult to detect. Given the current information, however, there is insufficient evidence for clustering of salmon by kin in the Washington ocean troll fishery.

6 Acknowledgements

I want to thank two people for their very helpful assistance with this project, Richard Comstock of the US Fish and Wildlife Service and Dick O'Connor of the Washington Department of Fisheries. Mr. Comstock posed the original question for this research and provided several insightful perspectives with respect to the analysis and interpretation of results. Mr. O'Connor provided the data along with helpful explanations of various data anomalies.

References

- [1] Beran, R. [1988.] 'Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements'. *J. Am. Stat. Soc.* **83**. pp687-697.
- [2] McBride, I. and T. P. Quinn. [1988.] 'Evidence for hereditary component in the homing behavior of chinook'. *Can. Jr. Fish. and Aquat. Sci.* **45**. pp2201-2205.
- [3] Geiger, Hal. [1988.] In discussion of 'Comparative marine survivals of chinook salmon fed different diets' by Kenneth Leon in 'Report of the 1987 Alaska chinook salmon workshop'. NWAFC processed report 88-06, National Marine Fisheries Service.
- [4] Moore, Rick, Wendy Beeghley-White, and Martin Peoples. [1991.] '1989 Ocean Sampling Program Annual Report'. Washington Department of Fisheries Progress Report 291.
- [5] Patefield, W.M. [1981.] 'Algorithm AS 159. An efficient method of generating random $R \times C$ tables with given row and column totals.' *Applied Statistics*, 30 91-97.