NOT TO BE CITED WITHOUT PRIOR
REFERENCE TO THE AUTHOR(S)

International Commission for the Northwest Atlantic Fisheries

Serial No. 5475

ICNAF Res. Doc. 79/VI/110

ANNUAL MEETING - JUNE 1979

On Estimating the Variance of the Estimate of Catch Per Unit Effort

by

S. J. Smith
Newfoundland Environment Center
St. John's, Newfoundland, Canada

Introduction

Catch per unit effort where the effort is defined here with respect to days or hours spent fishing for a specific species is one index of abundance used in fisheries work. In order for this statistic to be used in a comparative sense, say month to month or year to year comparisons, an estimate of the variance of the statistic is needed.

Two methods of estimating this variance are investigated empirically here. The first method, a regression technique, requires specific assumptions to be met in order to be valid. The second method is an application of the so called 'Jackknife' estimator and is more general in its use.

Both methods are applied to commercial data on Yellowtail founder (Limanda ferruginea) and American plaice (Hippoglossoides platessoides) catches by Newfoundland vessels for the Grand Bank area (ICNAF Div. 3LNO) in the years 1976, 1977 and 1978.

The estimators

Define,

K_i = the number of trip components reported for the ith month,

L = the number of trip components reported for the year,

f_{ij} = the number of days (or hours) in the jth component, for the ith month, where $j = 1, 2, \ldots, K_i$ and $i = 1, 2, \ldots, 12$,

$F_{i.}$ = the total number of days (or hours) recorded for the ith month, i.e., $F_{i.} = \sum_{j=1}^{K_i} f_{ij}$,

$F_{\cdot\cdot}$ = the total number of days (or hours) recorded for the year,

$$F_{\cdot\cdot} = \sum_{i=1}^{12} \sum_{j=1}^{K_i} f_{ij} = \sum_{i=1}^{12} F_{i\cdot} ,$$

C_{ij} = the catch (in kgs) of the jth trip component in the ith month.

At this point it is appropriate to define the term 'trip component' first introduced in the section above. The data, summarized from the original fishing logs, is categorized according to the ICNAF Division being fished, unit area being fished, depth, and main species. It is conceivable that in anyone trip a vessel may fish in many different divisions, unit areas, etc., with each day's fishing results entered in the log for the trip. When summarizing the results of this trip for data processing, all fishing entries in the same division, unit area and depth zone for the same main species are combined as one entry; hence these records are terms 'trip components' (expressed as tenths of trips). Catch per unit effort is calculated by,

(1) $$C_{i\cdot}/F_i = \frac{\sum_{j=1}^{K_i} C_{ij}}{F_{i\cdot}} ,$$

for the month and

(2) $$C/F = \frac{\sum_{i=1}^{12} \sum_{j=1}^{K_i} C_{ij}}{F_{\cdot\cdot}} ,$$

for the year.

The estimators as defined in (1) and (2) are in the class of ratio estimators since both the C_{ij} and the f_{ij} are random variables. In general estimators of this class are biased estimators with bias of order $^1/_n$, where n is the sample size (Cochran 1977). The techniques explored in this paper provide means by which the bias can be reduced or eliminated altogether (given certain conditions are met).

I. Regression Technique

A well-known result from regression states that if a linear relationship of the form,

(3) $$C = \beta F + \varepsilon ,$$

holds then $\frac{\Sigma\, W\, CF}{\Sigma\, W\, F^2}$ is an unbiased estimator of the population ratio β, where W is defined as a weight in the sense of a weighted least squares

(Snedecor and Cochrane 1967, Cochrane 1977). The weight, W, would be equal to 1, $^1/F$, or $^1/F^2$ according to whether the variance of C was constant with respect to F, proportional to F or proportional to F^2 respectively. Given one of these three conditions then, the estimator of β would be one of,

(4) $$T_1 = \frac{\Sigma CF}{\Sigma F^2},$$

(5) $$T_2 = \frac{\Sigma C}{\Sigma F},$$

or,

(6) $$T_3 = \frac{\Sigma C/F}{n},$$

where n in this case would be either K_i or L depending upon whether it was the monthly or yearly estimates we are interested in. The estimator T_1 is the usual least squares estimators, T_2 is identical to the estimators in (1) and (2) and T_3 is a form of a grand mean of individual trip component ratios.

The first step in using this technique is to fit a model of the form

(7) $$C = \alpha + \beta F + \varepsilon,$$

and then determine if the line passes through the origin by testing the null hypothesis, H_0: $\alpha = 0$ versus the alternative, H_a: $\alpha \neq 0$ by the usual procedures. If the null hypothesis of zero intercept can not be rejected, that is (3) is the appropriate model, then a plot of the residuals $(C-\hat{C})$ against the values of F would indicate the appropriate value of W to use.

The estimate of the variance of β would then be of the form (Draper and Smith 1966),

(8) $$\widehat{Var(\beta)} = \frac{\dfrac{\Sigma\, W\,(C-\hat{C})^2}{n-1}}{\Sigma\, W\, F^2}.$$

One limitation of this method, given all assumptions are valid, is that a minimum of three observations are required in order that there will be at least one degree of freedom available for the testing of hypothesises.

II. <u>The Jackknife Technique</u>

The jackknife technique was originally developed by Quenouille (1949) as a means of reducing bias of a serial correlation estimator. The name 'Jackknife' was coined by John Tukey in an unpublished work. A recent review of technique has been made by Rupert (1974).

Let $Y_1, \ldots, Y_n$ be a sample of independent and identically distributed random variables. Also let $\hat{\theta}$ be an estimator of the population parameter θ based on the sample of n observations. Further define $\hat{\theta}_{-i}$ to be the corresponding estimator based on all the sample points, excluding the ith point.

The estimator

(9) $$\hat{\theta}_J = n\,\hat{\theta} - (n-1)\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i},$$

known generally as the first order Jackknife estimator has the property of eliminating bias of the order $\frac{1}{n}$ (Rupert 1974).

Durbin (1959) presented results in a note that demonstrated that, for the class of ratio estimators where there existed a linear relationship between the numerator and the denominator, not only did the use of the Jackknife reduce the bias but it also reduced the variance. These results were specific for cases when the denominator had a normal distribution.

The Jackknife estimator used here, denoted by R_J is the average of the n quantities (K_i for the month, L quantities for year)

(10) $$R'_{-j} = n\,{}^{C}/_{F} - (n-1)\,R_{-j},$$

where $R_{-j} = \frac{\Sigma C_{ij}}{\Sigma f_{ij}}$, with the jth observation removed. The estimate of the variance of R_J, $\hat{V}(R_J)$ would be

(11) $$\widehat{V(R_J)} = \frac{1}{n(n-1)}\sum^{n}(R_{-j} - R_J)^2 .$$

If the R_{-j} can be considered to be independent estimates of the population ratio R, then (11) is an unbiased estimator (Cochran 1977).

Further it has been postulated by Tukey (1958) that if the n values (10) can be treated as independent and identically distributed random variables, the statistic

(12) $$\frac{R_J - R}{(V(R_J))^{\frac{1}{2}}}$$

should have an approximate Students'-t distribution with n-1 degrees of freedom. This postulate, if valid, would provide the means for interval estimation and hypothesis testing.

A limitation of the Jackknife technique, is that for calculation, at least two observations must be available.

Results

The methods discussed in the previous section were applied to commercial fishing data collected from Newfoundland vessels. The first technique was only used on the American plaice and Yellowtail flounder fishery for 1978. The Jackknife was applied to both plaice and flounder for 1976, 1977 and 1978 data.

The results for method one are presented in Table 1, Table 2 and 3 are the results for the Jackknife method for plaice and flounder respectively.

The regression method presented itself to be an extremely lengthly form of analysis. The model in (7) was fitted to data for each month and the year as a whole. Hypothesis tests were performed and residual plots were examined for each regression. Based on the patterns of the residuals an appropriate weight was chosen and the model in (3) was then fitted to the data. Given the time available for this study, the analysis was restricted to one year's data.

The results for the regression technique were not satisfactory on the whole. For both the plaice and the flounder data the intercepts were significantly different from zero for the data over the whole year ($p > 0.0001$ for both cases). This was also the case for some of the months. The residual plots for the plaice data suggested a weight of $^1/_{f_{ij}}$, hence the use of the estimator T_2. The results using this estimator are presented in Table 1.

The small sample sizes for the Yellowtail flounder data reduced the interpretative value of the residual plots. The residual plot for the year as a whole indicated that the variance of the catch was constant with respect to the value of f_{ij}. The estimator T_1 was deemed suitable therefore, on the basis of the years' residual plots.

With respect to computation and analysis time the Jackknife technique was more efficient than the regression method. In general the coefficients of variation (CV(j)) for American plaice were much lower than those for the Yellowtail flounder. This may be in part a function of the larger number of trip components for the plaice data in comparison to the flounder data. Table 4 presents the correlation between the coefficients of variation and the number of trip components. There is evidence that for five out of the six cases, a strong inverse relationship exists between the coefficient of variation and the number of trip components.

Discussion and Conclusions

If we define the better estimator as the one which has the smallest coefficient of variation, then the regression estimators were consistently better for the 1978 data, albeit the differences in many cases were minimal. The fact that for some months, and the year overall the intercept was found to be significantly different from zero must be taken into consideration before the regression method can be unequivically excepted as the best of the two methods. For these cases of non zero intercepts, the estimates are biased and more seriously the estimates of the variance are also biased having a component due to lack of fit added to them. For the Jackknife technique the estimates of the variance are unbiased if independence can be assumed.

Therefore, considering the small differences between the values provided by the methods both for C/F estimates and variance estimates and the possibility of violation of the assumptions required for the regression method, it is suggested here that the Jackknife estimator be preferred over the regression method. The general utility of the method and comparative ease of application further support this suggestion.

It should also be added that the data sets analyzed in this paper are ideal with respect to the detailed information available. Such complete information is usually not available. For any other data sets, if information on species catch and directed effort is available within months or years, the Jackknife method will retain its utility.

Finally further work is required in order to determine if Tukey's postulate of a students'-t distribution is valid for the data of the type considered here. If this can be established then equation (12) could be used as a pivotal value for interval estimation.

References

Cochran, W. G. 1977. Sampling techniques. John Wiley and Sons, Toronto, 428 pp.

Draper, N. R., and H. Smith. 1966. Applied regression analysis. John Wiley and Sons, New York, 407 pp.

Durbin, J. 1959. A note on the application of Quenouille's method of bias reduction to the estimation of ratios. Biometrika, 46: 447-480.

Quenouille, M. H. 1949. Approximate tests of correlations in time-series. J. R. Statist. Soc. B 11: 68-84.

Rupert, G. M. 1974. The Jackknife - a review. Bioemtrika, 61: 1-15.

Snedecor, G. W. and W. G. Cochran. 1967. Statistical methods. IOWA State University Press. Ames. IOWA. (Sixth printing), 593 pp.

Tukey, J. W. 1958. Bias and confidence in not quite large samples (Abstract). Ann. Math. Statist. 29: 614.

Table 1. Regression technique.

Month	T_2(Kg)	SE(T_2)	CV(T_2)	No. of trip components
American plaice 1978				
Jan.*	6150.78	339.259	0.055	65
Feb.	6175.59	515.449	0.083	44
March	6418.45	543.937	0.085	45
April	4298.22	299.92	0.0698	71
May	6221.85	264.88	0.043	123
June	6914.27	195.87	0.028	186
July*	7666.10	206.93	0.027	185
August*	8372.73	238.93	0.029	195
Sept.	7896.75	272.78	0.035	156
Oct.	8292.34	275.83	0.033	161
Nov.	7128.69	295.75	0.041	118
Dec.*	9353.04	602.99	0.064	92
Year*	7453.76	93.048	0.0125	1441

Month	T_1(Kg)	SE(T_1)	CV(T_1)	No. of trip components
Yellowtail flounder 1978				
Jan.	8948.461	1372.783	0.153	6
Feb.+	4786.656	981.320	0.205	3
March	7881.992	918.312	0.117	6
April	5863.395	513.669	0.088	38
May	6166.504	474.649	0.077	33
June	5118.836	229.340	0.045	32
July	6448.164	651.116	0.101	19
August*	6646.086	412.137	0.062	28
Sept.	7604.953	675.269	0.089	41
Oct.*	9969.473	553.544	0.056	73
Nov.*	7857.199	375.341	0.048	53
Dec.	8291.691	867.205	0.105	14
Year*	7714.484	195.078	0.025	346

Note (1) an asterisk in the first column indicates that the intercept was significantly different from zero (α = 0.05).

(2) a cross (+) in the first column indicates that the slope was not significantly different from zero (α = 0.05).

Table 2. Jackknife technique

Month	Catch per day	Jackknife estimates	$SE(R_J)$	$CV(R_J)$	No. of trip components
American Plaice					
1976					
Jan.	6695.77	6700.32	259.38	0.0387	158
Feb.	6436.45	6441.10	342.38	0.0532	107
March	6585.74	6609.85	751.00	0.1136	95
April	6864.20	6888.74	1014.64	0.1473	61
May	5066.14	5069.08	241.23	0.0476	100
June	6215.57	6218.49	201.47	0.0324	192
July	6345.64	6349.52	192.04	0.0302	189
August	6945.10	6948.81	183.88	0.0265	148
Sept.	7319.11	7323.75	266.81	0.0310	266
Oct.	6405.16	6412.03	288.35	0.0450	186
Nov.	5427.62	5430.08	161.98	0.0298	136
Dec.	6198.29	6203.33	251.60	0.0406	76
Year	6393.63	6396.12	81.05	0.0127	1714
1977					
Jan.	-	-	-	-	-
Feb.	7145.87	7153.31	535.51	0.0749	107
March	4397.75	4414.39	437.90	0.0992	37
April	4876.79	4898.87	542.43	0.1107	28
May	6708.60	6715.02	271.29	0.0404	84
April	6950.88	6954.30	193.63	0.0278	173
July	6638.65	6641.87	224.27	0.0338	168
August	6493.39	6494.05	179.53	0.0276	202
Sept.	6289.41	6292.21	193.60	0.0308	226
Oct.	5253.83	5255.62	148.67	0.0283	258
Nov.	6734.71	6738.42	220.34	0.0327	171
Dec.	6827.21	6832.44	306.00	0.0448	150
Year	6416.47	6417.51	76.19	0.119	1604
1978					
Jan.	6150.66	6155.85	362.83	0.0589	65
Feb.	6175.45	6204.53	627.98	0.1012	44
March	6418.01	6414.88	574.81	0.0896	45
April	4298.05	4308.04	331.48	0.0769	71
May	6221.68	6220.51	328.09	0.0527	123
June	6913.98	6915.34	211.50	0.0306	186
July	7665.77	7668.52	228.35	0.0298	185
August	8372.25	8375.37	266.20	0.0318	195
Sept.	7896.34	7900.49	297.25	0.0376	156
Oct.	8291.95	8292.73	294.09	0.0355	161
Nov.	7128.50	7131.35	437.16	0.0487	118
Dec.	9352.96	9369.32	743.21	0.0793	92
Year	7451.99	7454.02	109.92	0.0147	1441

Table 3. Jackknife technique

Month	Catch per day	Jackknife estimates	$SE(R_J)$	$CV(R_J)$	No. of trip components
Yellowtail flounder					
1976					
Jan.	6473.71	6610.21	1978.03	0.2992	7
Feb.	5703.94	5756.38	864.54	0.1502	5
March	6140.23	6217.57	820.84	0.1320	21
April	6159.30	6161.75	307.01	0.0498	50
May	4737.41	4745.04	278.68	0.0587	70
June	5109.05	5817.94	1771.35	0.3045	7
July	2763.47	2785.89	573.72	0.2059	8
August	4531.25	4880.24	1914.06	0.3922	5
Sept.	4277.31	4322.08	566.76	0.1311	10
Oct.	3948.69	3884.36	620.90	0.1598	16
Nov.	3213.38	3353.24	1041.37	0.3106	6
Dec.	6836.27	6962.43	248.33	0.0357	2
Year	5342.08	5344.45	189.41	0.0354	207
1977					
Jan.	-	-	-	-	-
Feb.	10269.48	10269.48	-	-	1
March	6870.96	5714.37	3469.77	0.6072	2
April	6535.06	6562.00	1230.36	0.1875	5
May	5934.68	5928.96	295.17	0.0498	44
June	5350.22	5346.52	335.52	0.0627	34
July	6563.53	6543.45	474.00	0.0724	41
August	5838.38	5847.93	394.28	0.0674	45
Sept.	5400.01	5430.38	866.67	0.1596	15
Oct.	4916.27	4648.72	1621.17	0.3487	5
Nov.	8968.03	8942.65	673.68	0.0753	52
Dec.	8388.27	8384.92	965.07	0.1151	18
Year	6628.96	6626.64	200.21	0.0302	262
1978					
Jan.	9192.81	9131.66	1422.67	0.1558	6
Feb.	4488.92	4634.01	969.77	0.2093	3
March	7596.98	7656.30	925.55	0.1209	6
April	6431.97	6411.53	595.73	0.0929	38
May	7073.23	7035.55	580.61	0.0825	33
June	5163.22	5161.33	255.07	0.0494	32
July	6720.50	6700.00	748.82	0.1118	19
August	7110.58	7089.23	466.72	0.0658	28
Sept.	8354.62	8330.11	780.23	0.0937	41
Oct.	11463.36	11436.06	688.14	0.0584	73
Nov.	8474.87	8460.16	429.26	0.0507	53
Dec.	7780.14	7831.03	1047.15	0.1337	14
Year	8267.59	8266.04	229.46	0.0278	346

Table 4. Correlations between coefficients of variation (CV) and number of trip components (t). (Asterisk denotes p significantly different from zero at α = 0.05, one tailed test).

Species	1976	1977	1978
American plaice	$r_{CV,t}$	$r_{CV,t}$	$r_{CV,t}$
	-0.606*	-0.866*	-0.935*
Yellowtail flounder	$r_{CV,t}$	$r_{CV,t}$	$r_{CV,t}$
	-0.543	-0.754*	-0.780*