NOAA Technical Memorandum NMFS F/NWC-95

Studies of the Distribution and Abundance of Juvenile Groundfish in the Northwestern Gulf of Alaska, 1980-82: Part III, Estimation of Sample Size Requirements

By
Charles D. Knechtel

January 1986

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service

This TM series Is used for documentation and timely communication of preliminary results, Interim reports, or special purpose Information, and has not received complete formal review, editorial control, or detailed editing.

STUDIES OF THE DISTRIBUTION AND ABUNDANCE OF JUVENILE GROUNDFISH IN THE NORTHWESTERN GULF OF ALASKA, 1980-82: PART III, ESTIMATION OF SAMPLE SIZE REQUIREMENTS

by

Charles D. Knechtel

Northwest and Alaska Fisheries Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way Northeast
Bin Cl5700
Seattle, Washington 98115

January 1986

References to trade names do not imply
endorser&t by the National Marine
Fisheries Service, NOAA.

ABSTRACT

Sample sizes necessary to detect changes between 2 yr in the total annual' population abundance of juvenile groundfish in the northwestern Gulf of Alaska were estimated using Monte Carlo simulation for a statistical test described by Brown and Forsythe.-- Parameters for simulations were estimated from actual bottom trawl survey data, covering 12 geographic regions and 3 yr, by the method of maximum likelihood; these data were described in detail in earlier reports. The catch per unit effort (CPUE) data from these surveys were generally heteroscedastic and skewed. Fishing effort was measured by calculating an estimate of area swept during tows of the survey trawl net. Total abundance in each region was estimated by multiplying the sample mean of CPUE in that region by the area of the region. Fish catches were usually simulated using a negative binomial distribution. The two parameters of the negative binomial distribution were made into functions of fishing effort and mean fish abundance in such a manner that mean catch was proportional to effort and to mean abundance, and CPUE variance was a power function of mean CPUE.

Simulated survey data covering 2 yr were analyzed using analysis of variance (ANOVA), but the usual constraints on ANOVA coefficients were modified because the areas of the geographic regions were unequal. In addition, the usual F-test of the significance of differences between years was modified to account for heteroscedasticity using the method of Brown and Forsythe. In simulations using sample sizes of 180 hauls/year (about 0.12 hauls per square kilometer-for each of the 12 geographic regions),the predicted probability was greater than 87% of correctly detecting the direction of a change by a factor of 3.16 in annual abundance of young-of-the-year walleye pollock, *Theragra chalcogramma*. The corresponding probabilities

Preceding Page Blank

of correctly detecting the direction of a change by a factor of 10 in annual abundance of 1-yr-old walleye pollock? 1-yr-old sablefish, Anoplopoma fimbria; and young-of-the-year Pacific cod, Gadus macrocephalue, were greater than 93%, 94%, and 48% respectively.

Simulated surveys were also used to test the validity of some statistical methods commonly used with this type of data. The method of Brown and Forsythe worked reasonably well for 2 yr of data despite heteroscedasticity and non-normality as long as at least 2 hauls/year were allocated to each geographic region, but some other method may be necessary for more than 2 yr of data. The validity of the method of Brown and Forsythe suggests that the heteroscedasticity of this data may be more crucial than non-normality in selecting a valid statistical test of the significance of differences in total abundance between years. In contrast, application of ordinary ANOVA to CPUE data transformed using log(CPUE+l) , rank, or power transformations worked poorly, especially when geographic region by year interactions in abundance were present. When the null hypothesis of no difference in total abundance between years was true, the expected probability of rejecting the null hypothesis was 5%, and geographic region by year interactions in abundance were present, then estimated probabilities of rejecting the null hypothesis were greater than 60% in some cases in simulations using the log(CPUE+l) transformation. There are theoretical reasons to expect that similar problems will occur in general if total abundance is estimated as in this paper, ordinary ANOVA is applied to CPUE data transformed with any nonlinear transformation, and geographic region by year interactions are present.

Limitations of this study and suggestions for further research are also discussed.

CONTENTS

INTRODUCTION

This study, the final paper in a series of three reports, presents the results of an analysis of sample size requirements necessary to detect changes in year-class strengths of one or two juvenile age groups of three commercially important species of groundfish in the northwestern Gulf of Alaska. Each year-class strength was estimated using an area swept method described in Smith and Bakkala (1982: equations 13-14), except that number of individuals instead of biomass was estimated. The age groups analyzed were young-of-the-year and 1-yr-old walleye pollock, Theragra chalcogramma; 1-yr-old sablefish, Anoplopoma fimbria; and young-of-the-year Pacific cod, Gadus macrocephalus. The data used for this analysis were collected in 12 major inlets of Kodiak Island and along the central Alaska Peninsula during the months August-September 1980-82 as part of trawl surveys to assess the abundance of shrimp populations. The surveys were conducted by the Alaska Department of Fish and Game (ADF&G), but biologists from the National Marine Fisheries Service (NMFS) also participated to make possible increased sampling of juvenile fish. Detailed descriptions and initial analyses of these surveys are given in the first two reports of this series, Smith et al. (1984) and Walters et al. (1985). These reports described the geographical distribution and abundance of juvenile age groups of major fish species, and their annual variations: provided initial evaluations of the feasibility of measuring year-class strengths? related results to other research in the region; and provided recommendations for further work.

Although the primary goal of this final study was to estimate sample size requirements, a number of problems had to be solved before this was possible. First, statistical characteristics of the survey data had to be further evaluated. Then an appropriate statistical test had to be found to detect differences in year-class strength between years. And, finally, a method had to be chosen and implemented to estimate the sample size requirements for this test.

The statistical and mathematical methods relevant to the above three problems are discussed in this paper, and important statistical characteristics of the survey data are presented. The statistical models and parameter estimates used to describe the survey data are listed, and the statistical test which was selected to detect differences in annual population size between 2 yr (Brown and Forsythe 1974a) is presented. It is shown why common statistical tests based on nonlinear variance-stabilizing transformations were inappropriate. The use of Monte Carlo simulation to-estimate sample size requirements for the selected statistical test is described and the resulting estimates presented. Limitations of this study, conclusions, and suggestions for further research complete the paper.

Eventually it may be possible to use juvenile year-class strengths to provide warning a year or more in advance of possible strong or weak-year-class recruitment to commercial fisheries. The reliable detection of significant changes in juvenile year-class strengths by means of adequate sampling and appropriate statistical tests appears to be a prerequisite for such a method.

MATERIALS AND METHODS

Background

Analysis of variance (ANOVA) has often been used to determine which changes in a continuous dependent variable can be attributed to one or more categorical variables. In this study, for example, one wishes to determine what changes in year-class population density can be attributed to changes in the categorical variables "year" and "geographical location." It is frequently assumed in ANOVA that each measurement is statistically independent, that effects due to the independent variables are fixed, and that any-variation (random error) not caused by the independent variables has a normal (Gaussian) distribution with constant variance. The assumption that' the variance is constant implies that it is independent of any changes in the independent variables. An ordinary F-test is appropriate when these assumptions are essentially met. Published charts of the power of the F-test can be used to determine approximate sample sizes needed to detect specified changes using ANOVA in conjunction with an ordinary F-test under these assumptions (Scheffe 1959: section 2.8). However, in this study, catch per unit effort (CPUE) variances were not constant with changes in year and stratum, but were heteroscedastic ("heteroscedastic" is a statistical term which means that the variances were not constant). In addition, the CPUE data were highly skewed and therefore non-normal (e.g., Smith et al. 1984: fig. 9). For these reasons, sample size requirements were not determined using charts of the power of the F-test. Instead, the performance of different sample sizes was evaluated using Monte Carlo simulation.

Simulation may be defined as a numerical computational technique for conducting experiments,. It utilizes mathematical and logical models that describe the behavior of a system or system component. Monte Carlo simulation may be defined as simulation which includes stochastic sampling from a probability distribution or distributions (Rubinstein 1981: 6,11).

An important consideration in the design of the simulation model described in this paper was that variance of CPUE in a given year and stratum appeared to be a power function of mean CPUE. A weighted negative binomial model described by Bissell (1972) and used by Zweifel and Smith (1981) to model fish catches was modified to account for this power function relationship,, and used to model fish catch in each simulated survey trawl haul. Each haul was considered to be one "sample." The assumptions in the modified model were: 1) the estimated count of fish in the catch of each haul had a negative binomial, Poisson, or binomial probability distribution; 2) catch counts were stochastically independent of counts in other hauls; 3) mean catch was a linear function of fishing effort and fish population density in the environment; 4) fishing effort was measured essentially without error; 5) variance of CPUE was essentially a power function of fish density: and, 6) mean fish density was constant in a given year and geographic location (stratum). In this paper, CPUE is defined to equal catch divided by effort;' consequently assumptions 3 and 4 imply that the catch variance was proportional to the square of the fishing effort. The parameters in this model are the fish densities in each year and stratum; and a multiplier and exponent used to calculate CPUE variance as a power function of fish density. Both the model of Bissell (1972) and the model used in this paper are examples of models which are linear except for nonlinear, non-normal, heteroscedastic error terms. The method of maximum likelihood (ML) is widely applicable to estimating parameters of distribution functions (Rao 1973: 354), and has certain optimal properties (Mood et al. 1974: 284-286, 358-360). For instance, in many cases ML estimates are asymptotically unbiased and asymptotically of minimum variance. Utilizing the ML method has the disadvantage that estimates often must be calculated iteratively, which may require large amounts of computer time. In addition, for some distribution functions, poor starting points for the iterations may cause a failure to converge to the true ML estimates of the parameters. However, ML was considered the method of choice to estimate parameters for the negative binomial model (Bliss and Fisher 1953) and the weighted negative binomial model of Bissell (1972). Since the model used in this study usually assumes an underlying negative binomial distribution, ML was also used to estimate its parameters.

The data used in this study were collected in a total of 366 trawl hauls made in 12 different geographic regions in the Gulf of Alaska (Fig. 1, Table 1) in August-September of 1980-82; Smith et al. (1984) describe and present results of initial analyses of this data. The allocation of sampling effort (number of hauls per region per year) was not proportional to the areas of the regions given in Table 1; further information regarding the allocation of sampling effort is given in Smith et al. (1984: 10-12).

Estimated counts of fish at age and estimates of fishing effort were analyzed in this study. Estimated counts were determined as follows. The entire catch was weighed, and a sample (sample A) of the catch was selected as described by Hughes (1976). Sample A was sorted to the lowest feasible taxonomic group (usually species), and each taxonomic group was weighed and counted. For species of interest, a sample (sample B) was selected from sample A, and the ages of fish in sample B were estimated using methods

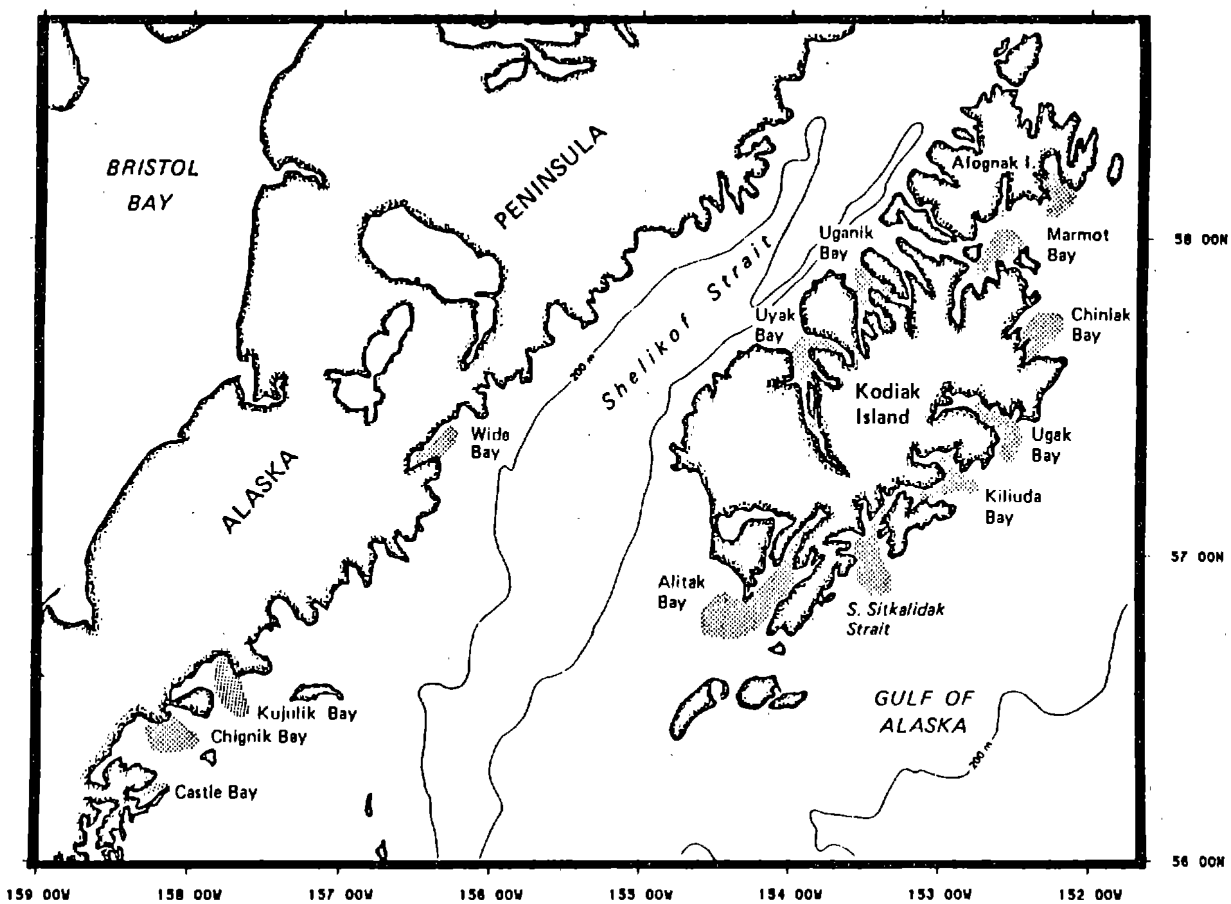

Figure 1.--Map of the study area in the northwestern Gulf of Alaska showing the 12 strata used for estimating juvenile fish abundance, 1980-82 (Smith et al. 1984).

Table 1.--Survey areas used in the northwestern Gulf of Alaska for estimating juvenile fish abundance, geographic areas, and sampling effort, 1980-82 (after Smith et al. 1984: table 1).

Stratum*	Geographic area (km^2)	Number of tows			
		1980	1981	1982	Total
Castle Bay	28.5	4	4	4	12
Chignik Bay	114.0	9	8	7	24
Kujulik Bay	74.0	10	11	10	31
Wide Bay	25.8	6	6	6	18
Alitak Bay	426.6	18	29	21	68
S. Sitkalidak Strait	235.8	12	12	16	40
Kiliuda Bay	154.2	10	10	9	29
Ugak Bay	91.9	11	11	10	32
Chiniak Bay	58.2	6	10	8	24
Marmot Bay	195.9	11	12	16	39
Uganik Bay	69.6	8	8	7	23
Uyak Bay	88.2	9	9	8	26
Total	1562.7	114	130	122	366

*See Figure 1.

described in Smith et al. (1984: 13-15). Counts of a given species i at age j in a given trawl haul were then estimated by

$$(W_T/W_A) \times (N_{Ai}/N_{Bi}) \times (N_{Bij}), \tag{1}$$

where WT is the total weight of the catch, WA is the total weight of sample A, N_{Ai} is the number of fish in sample A of species i, N_{Bi}. is the number of fish in Sample B of species i, and N_{Bij} is the estimated number of fish in sample B of species i, and age j. Consequently, estimated counts of fish at age were not necessarily integers.

Fishing effort was measured by estimating the area swept by a 61-foot (19-meter) shrimp trawl net during a tow: i.e., the estimated area swept equaled the distance traveled during the trawl haul multiplied by an assumed path width of 9.75 m. The net was towed using 1.7 x 2.7 m V-doors (otterboards), weighing approximately 540-590 kg, to spread the net. The distance traveled during the haul was measured using Loran-C or radar fixes; it was intended that the standard tow be either 15 or 30 min in duration (Smith et al. 1984: 12). The resulting estimates of effort had a range of 0.52×10^4 m^2 to 3.54×10^4 m^2. Wathne (1977: 21-22) described three tows using 370 kg, 1.5 x 2.1 m V-doors and suggested that between-tow variability in path width may be of some significance with this gear; from this data an assumed mean path width of 9.75 m appears reasonable for tows using 1.5 x 2.1 m V-doors. In NWAFC (1981), 11 tows are described which used 363 kg, 1.5 x 2.1 m V-doors and resulted in mean path widths in a range of about 4.6-13.1 m, and 11 tows are described which used 526 kg, 1.5 x 2.1 m V-doors and resulted in mean path widths in the range 9.1-12.2 m. Despite the different V-door size in the present study, it was assumed that mean path widths equaled 9.75 m.

The heteroscedasticity and non-normality of the CPUE data were important characteristics considered when selecting an appropriate statistical test of population size changes. It is expected that these characteristics would not be greatly changed if better estimates of path width were available. Relative population size estimates also might not be greatly affected. However, the estimates of absolute population size used in this study are inversely proportional to the path width assumed, so that estimates of the probability of detecting specified changes in absolute population size may be affected to some degree by this assumption.

The CPUE estimates (catch divided by effort) used in this study have units of number/10,000 m^2 and are summarized in Tables 2-5. The sample coefficient of skewness is defined as the third sample moment about the mean divided by the 1.5th power of the 'second sample moment about the mean (Snedecor and Cochran 1980: section 5.13). It is a measure of the degree of symmetry of the data; a sample coefficient of skewness close to zero indicates that the data appears symmetric. If the CPUE data were symmetrically distributed, the sample coefficient of skewness (g1) shown in Tables 2-5 would be expected to be less than zero with the same frequency as it was greater than zero. However, it was less than zero only 4 times, but greater than zero 100 times. For each of the four species or age groups in Tables 2-5, a two-sided statistical test based on tables of the binomial distribution showed that the count of negative values of the sample coefficient of skewness was significantly different at the 99.6% level from the count of positive values of the sample coefficient of skewness, under the null hypothesis that the data were symmetrically distributed. In each case the count of positive sample coefficients of skewness was greater than the count of negative sample coefficients of skewness.

Table 2.--Summary statistics describing catch per unit effort (CPUE) of young-of-the-year walleye pollock, by year and stratum, in the northwestern Gulf of Alaska..

Year	Stratum	Number of hauls	Sample mean (no./10,000 m^2)	Sample variance ([no./10,000 m^2]^2)	Sample coefficient of skewness*
1980	Castle	4	6.64×10^1	6.24×10^3	1.597
	Chignik	9	9.77×10^1	9.94×10^3	1.203
	Kujulik	10	2.71×10^2	1.99×10^5	2.759
	Wide	6	6.80×10^2	1.55×10^5	-0.932
	Alitak	18	2.01×10^2	1.39×10^5	2.881
	Sitkalidak	12	8.66×10^1	7.93×10^4	3.443
	Kiliuda	10	5.00×10^1	2.01×10^4	3.094
	Ugak	11	4.34×10^2	3.08×10^5	1.297
	Chiniak	6	3.19×10^3	4.34×10^7	2.430
	Marmot	11	1.84×10^2	1.20×10^5	2.271
	Uganik	8	0.0	0.0	0.000
	Uyak	9	2.26×10^1	4.06×10^3	2.981
1981	Castle	4	2.45×10^2	1.14×10^5	1.958
	Chignik	8	3.54×10^2	1.32×10^5	1.003
	Kujulik	11	5.55×10^2	1.34×10^6	3.166
	Wide	6	1.00×10^3	6.55×10^5	1.104
	Alitak	29	3.96×10^3	8.16×10^7	4.110
	Sitkalidak	12	1.13×10^2	2.16×10^4	1.575
	Kiliuda	10	9.72×10^2	3.93×10^6	2.451
	Ugak	11	3.24×10^3	1.78×10^7	2.073
	Chiniak	10	1.71×10^3	3.56×10^6	0.856
	Marmot	12	6.11×10^2	1.05×10^6	2.650
	Uganik	8	1.80×10^2	2.14×10^5	2.820
	Uyak	9	5.40×10^1	5.56×10^3	1.319
1982	Castle	4	3.26×10^1	5.40×10^2	-1.247
	Chignik	7	9.04	1.42×10^2	1.691
	Kujulik	10	1.19×10^2	6.18×10^4	3.080
	Wide	6	3.62×10^2	1.15×10^5	1.066
	Alitak	21	2.33×10^2	8.93×10^4	1.319
	Sitkalidak	16	2.01×10^1	1.07×10^3	2.261
	Kiliuda	9	8.12×10^1	1.05×10^4	0.965
	Ugak	10	1.25×10^2	1.50×10^5	3.162
	Chiniak	8	3.60×10^2	2.82×10^5	1.355
	Marmot	16	4.79×10^2	1.23×10^6	2.605
	Uganik	7	9.66	6.53×10^2	2.646
	Uyak	8	1.39×10^1	8.52×10^2	2.261

*The sample coefficient of skewness is defined as the third sample moment about the mean divided by the 1.5th power of the second sample moment about the mean (Snedecor and Cochran 1980: section 5.13).

Table 3.--Summary statistics describing catch per unit effort (CPUE) of 1-yr-old walleye pollock, by year and stratum, in the northwestern Gulf of Alaska.

Year	Stratum	Number of hauls	Sample mean (no./10,000 m^2)	Sample variance ([no./10,000 m^2]^2)	Sample coefficient of skewness*
1980	Castle	4	7.18×10^3	3.10×10^7	0.415
	Chignik	9	2.54×10^3	1.96×10^7	1.991
	Kujulik	10	8.24×10^2	1.68×10^6	2.973
	Wide	6	1.35×10^2	1.32×10^4	0.035
	Alitak	18	3.96×10^2	2.99×10^5	1.554
	Sitkalidak	12	8.04×10^{-1}	6.49	3.417
	Kiliuda	10	0.0	0.0	0.000
	Ugak	11	3.57	1.41×10^2	3.317
	Chiniak	6	0.0	0.0	0.000
	Marmot	11	1.36×10^1	1.64×10^3	3.289
	Uganik	8	1.44×10^2	4.22×10^4	1.341
	Uyak	9	1.81×10^2	3.74×10^4	0.746
1981	Castle	4	7.85×10^1	2.17×10^3	0.211
	Chignik	8	5.17×10^2	1.75×10^5	0.310
	Kujulik	11	5.88×10^1	5.23×10^3	2.644
	Wide	6	3.96×10^2	4.58×10^5	2.288
	Alitak	29	4.60	1.97×10^2	4.472
	Sitkalidak	12	2.25×10^1	1.92×10^3	2.937
	Kiliuda	10	3.12×10^1	4.86×10^3	2.881
	Ugak	11	2.57	7.26×10^1	3.317
	Chiniak	10	4.61×10^1	4.39×10^3	2.062
	Marmot	12	2.87×10^2	3.96×10^5	3.104
	Uganik	8	1.15×10^2	5.86×10^4	2.260
	Uyak	9	1.84×10^2	7.01×10^4	1.651
1982	Castle	4	4.07×10^2	1.51×10^5	0.583
	Chignik	7	6.90×10^1	6.79×10^3	1.534
	Kujulik	10	1.43×10^2	1.16×10^4	0.361
	Wide	6	8.49×10^1	8.18×10^3	1.437
	Alitak	21	7.37×10^2	3.99×10^6	4.058
	Sitkalidak	16	3.45×10^1	1.49×10^4	3.970
	Kiliuda	9	2.96	2.13×10^1	1.127
	Ugak	10	6.78×10^1	1.94×10^4	1.816
	Chiniak	8	3.17×10^{-1}	8.04×10^{-1}	2.828
	Marmot	16	1.09×10^1	2.86×10^2	1.452
	Uganik	7	2.15×10^2	1.87×10^5	2.271
	Uyak	8	5.34×10^2	2.02×10^6	2.815

*The sample coefficient of skewness is defined as the third sample moment about the mean divided by the 1.5th power of the second sample moment about the mean (Snedecor and Cochran 1980: section 5.13).

Table 4.--Summary statistics describing catch per unit effort (CPUE) of 1-yr-old sablefish by year and stratum, in the northwestern Gulf of Alaska.

Year	Stratum	Number of hauls	Sample mean (no./10,000 m^2)	Sample variance ([no./10,000 $m^2]^2$)	Sample coefficient of skewness*
1980	Castle	4	6.31×10^1	7.97×10^3	1.414
	Chignik	9	5.71×10^1	4.72×10^3	1.090
	Kujulik	10	7.56×10^1	8.97×10^3	2.573
	Wide	6	1.85×10^{-1}	2.04×10^{-1}	2.449
	Alitak	18	2.63×10^{-1}	5.88×10^{-1}	2.730
	Sitkalidak	12	2.49	6.87×10^1	3.455
	Kiliuda	10	0.0	0.0	0.000
	Ugak	11	0.0	0.0	0.000
	Chiniak	6	4.09×10^{-1}	2.31×10^{-1}	0.527
	Marmot	11	2.89×10^{-1}	4.62×10^{-1}	2.352
	Uganik	8	3.95	7.44×10^1	2.427
	Uyak	9	0.0	0.0	0.000
1981	Castle	4	5.53	2.17×10^1	-0.166
	Chignik	8	2.86	1.92×10^1	2.195
	Kujulik	11	1.18×10^1	5.43×10^1	-0.046
	Wide	6	0.0	0.0	0.000
	Alitak	29	0.0	0.0	0.000
	Sitkalidak	12	1.90×10^{-1}	4.35×10^{-1}	3.464
	Kiliuda	10	6.09×10^{-1}	1.71	1.936
	Ugak	11	0.0	0.0	0.000
	Chiniak	10	5.99×10^{-2}	3.58×10^{-2}	3.162
	Marmot	12	0.0	0.0	0.000
	Uganik	8	1.20	4.83	2.422
	Uyak	9	8.55×10^{-2}	6.57×10^{-2}	3.000
1982	Castle	4	6.57×10^{-1}	9.14×10^{-1}	1.526
	Chignik	7	3.10×10^{-1}	6.74×10^{-1}	2.646
	Kujulik	10	1.32×10^1	3.50×10^2	2.149
	Wide	6	1.74	1.82×10^1	2.449
	Alitak	21	0.0	0.0	0.000
	Sitkalidak	16	2.56	1.05×10^2	4.000
	Kiliuda	9	2.94×10^{-1}	3.52×10^{-1}	1.761
	Ugak	10	3.03×10^{-1}	2.89×10^{-1}	1.620
	Chiniak	8	7.28	1.70×10^2	2.134
	Marmot	16	1.31×10^1	5.34×10^2	2.202
	Uganik	7	0.0	0.0	0.000
	Uyak	8	0.0	0.0	0.000

*The sample coefficient of skewness is defined as the third sample moment' about the mean divided by the 1.5th power of the second sample moment about the mean (Snedecor and Cochran 1980: section 5.13).

Table 5.--Summary statistics describing catch per unit effort (CPUE) of young-of-the-year Pacific cod, by year and , stratum, in the northwestern Gulf of Alaska.

Year	Stratum	Number of hauls	Sample mean (no./10,000 m^2)	Sample variance ([no./10,000 $m^2]^2$)	Sample coefficient of skewness*
1980	Castle	4	0.0	0.0	0.000
	Chignik	9	0.0	0.0	0.000
	Kujulik	10	0.0	0.0	0.000
	Wide	6	2.22	2.94×10^1	2.449
	Alitak	18	0.0	0.0	0.000
	Sitkalidak	12	0.0	0.0	0.000
	Kiliuda	10	0.0	0.0	0.000
	Ugak	11	0.0	0.0	0.000
	Chiniak	6	0.0	0.0	0.000
	Marmot	11	0.0	0.0	0.000
	Uganik	8	0.0	0.0	0.000
	Uyak	9	0.0	0.0	0.000
1981	Castle	4	0.0	0.0	0.000
	Chignik	8	0.0	0.0	0.000
	Kujulik	11	1.99×10^{-1}	4.37×10^{-1}	3.317
	Wide	6	1.00×10^2	2.08×10^4	2.054
	Alitak	29	1.97×10^1	2.86×10^3	4.891
	Sitkalidak	12	0.0	0.0	0.000
	Kiliuda	10	0.0	0.0	0.000
	Ugak	11	0.0	0.0	0.000
	Chiniak	10	3.32	1.10×10^2	3.162
	Marmot	12	0.0	0.0	0.000
	Uganik	8	0.0	0.0	0.000
	Uyak	9	0.0	0.0	0.000
1982	Castle	4	0.0	0.0	0.000
	Chignik	7	0.0	0.0	0.000
	Kujulik	10	7.20×10^{-1}	5.18	3.162
	Wide	6	2.33×10^1	1.18×10^3	1.044
	Alitak	21	1.99	9.83	1.303
	Sitkalidak	16	0.0	0.0	0.000
	Kiliuda	9	0.0	0.0	0.000
	Ugak	10	0.0	0.0	0.000
	Chiniak	8	0.0	0.0	0.000
	Marmot	16	0.0	0.0	0.000
	Uganik	7	0.0	0.0	0.000
	Uyak	8	5.57	1.55×10^2	2.504

*The sample coefficient of skewness is defined as the third sample moment about the mean divided by the 1.5th power of the second sample moment about the mean (Snedecor and Cochran 1980: section 5.13).

This indicates that the data are not symmetrically distributed but generally have positive skewness, and that this asymmetry is not alleviated by stratifying by year and geographical region. Since the normal (Gaussian) distribution is symmetric, this also indicates that the data are non-normal. Figure 2 shows a histogram of a subset of this data with a typical sample coefficient of skewness. The non-normality and asymmetry of the subset are evident: in fact, the underlying probability density function may well be strictly monotone decreasing.

Another characteristic of the data was the frequent occurrence of zero catches. Catch per unit effort values cannot be less than zero, but values equal to zero frequently occurred. No young-of-the-year pollock were caught in 109 hauls, no 1-yr-old pollock in 171 hauls, no 1-yr-old sablefish in 283 hauls, and no young-of-the-year Pacific cod in 328 hauls.

From Tables 2-5, it is evident that the data are heteroscedastic. For a given species or age group, the maximum sample variance was more than 400 times the minimum nonzero sample variance. Harris (1975: section 8.2) recommended that an ordinary F-test not be used in ANOVA if the ratio of the maximum sample variance to the minimum sample variance is greater than about 20. Sample variances equal to zero occurred for each of the species or age groups. The true underlying variance being measured can plausibly equal zero for this type of data, because a species or age group may be completely absent from a region during the time of the survey, and the resultant population density estimate of zero would then also have a variance of zero. For this reason, it seems unreasonable to assume that the variance-covariance matrix of the observations is positive definite; rather it may be only non-negative definite. This can affect the computational methods used- (Searle 1983: equation 16).

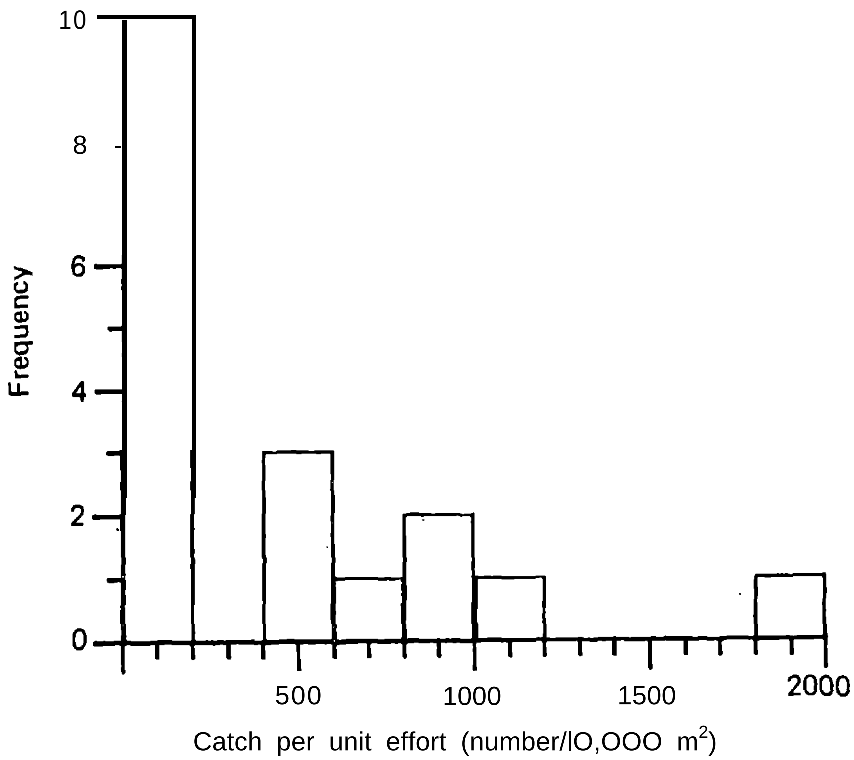

Figure 2.--Histogram of catch per unit effort (CPUE) of 1-yr-old walleye pollock in Alitak Bay in 1980. The sample coefficient of skewness is 1.554.

In all but four cases in Tables 2-5, if the sample mean was unequal to zero, then the sample variance exceeded the mean. All four exceptions occurred for sablefish.

It was found that representing the sample variances as a power function of the sample means provided a good fit to the data, since the log-log plots of the nonzero sample means and variances were close to a straight line (Figs. 3-6). The unweighted correlation of the logarithms of the nonzero sample variances with the logarithms of the nonzero sample means was more than 96% for each of the species or age groups shown.

Statistical Models and Parameter Estimation

A three-parameter Weibull distribution (Bury 1975: section 12.11) was used to model the statistical distribution of effort. The Weibull distribution is often used to model lifetime data; in this case the amount of effort expended until the haul was terminated was considered analogous to "lifetime." The three-parameter Weibull cumulative distribution function used was

$$1 - \exp(-[(E - P_3)/P_2]^{P_1}), \qquad (2)$$

where E is fishing effort (10,000 m^2), and P_1, P_2, and P_3 are parameters. The probability that E < P3 is zero. It was assumed that the distribution of fishing effort was independent of geographical region, year, species, and species density. The probability density function of the three-parameter Weibull distribution was fitted using the method of maximum likelihood with a FORTRAN computer program which used the International Mathematical and Statistical Libraries (IMSL 1982) subroutine ZXMIN. (However, Kappenman

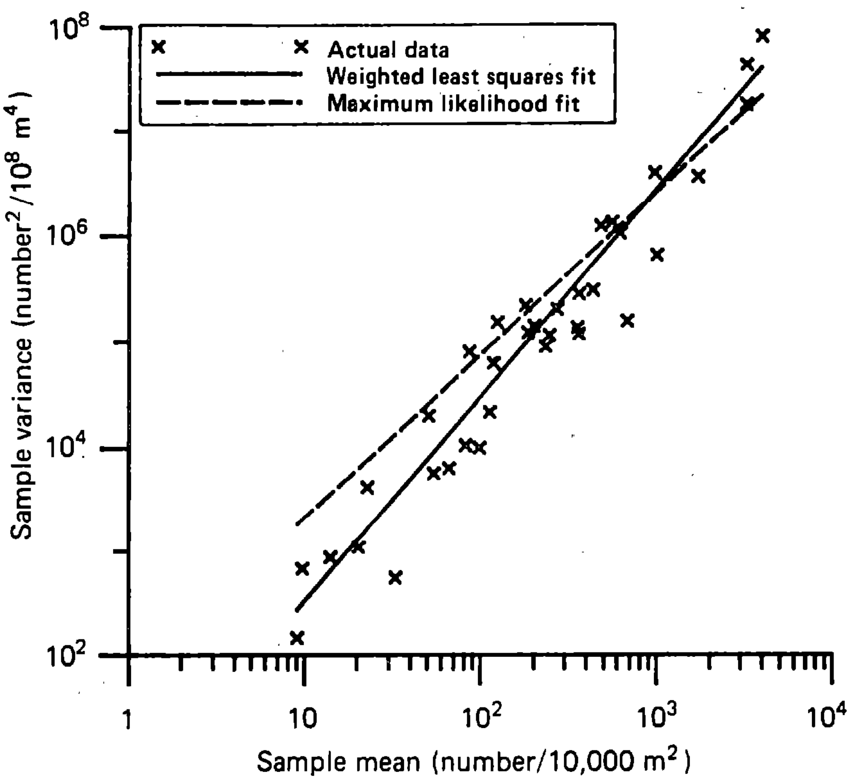

Figure 3.--Log-log plot of sample variance versus sample mean of CPUE of young-of-the-year walleye pollock, with weighted least squares and maximum likelihood fitted lines.

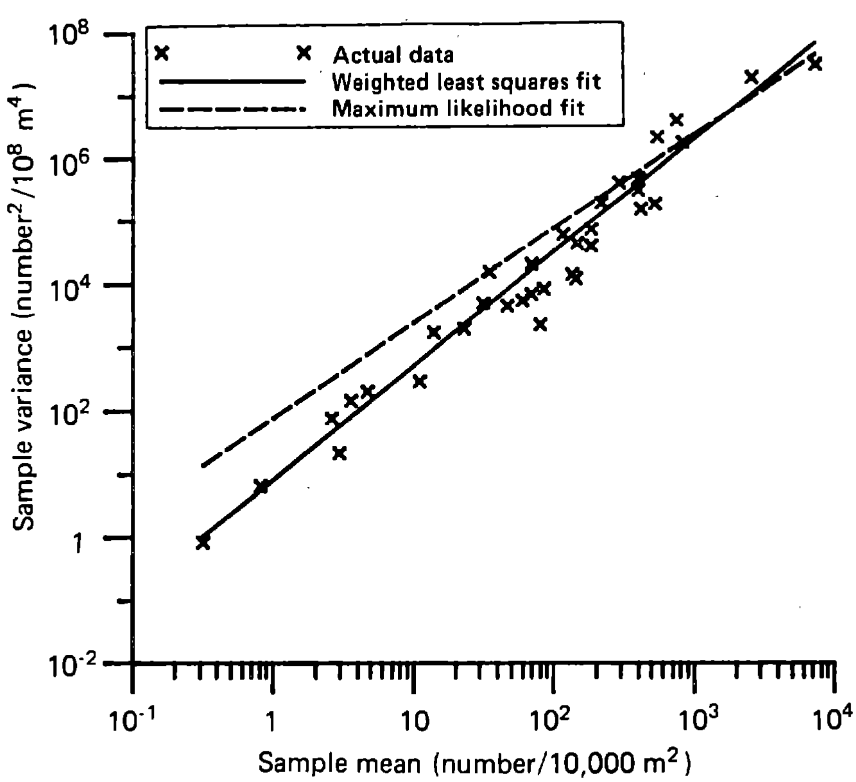

Figure 4.--Log-log plot of sample variance versus sample mean of CPUE of 1-yr-old walleye pollock, with weighted least squares and maximum likelihood fitted lines.

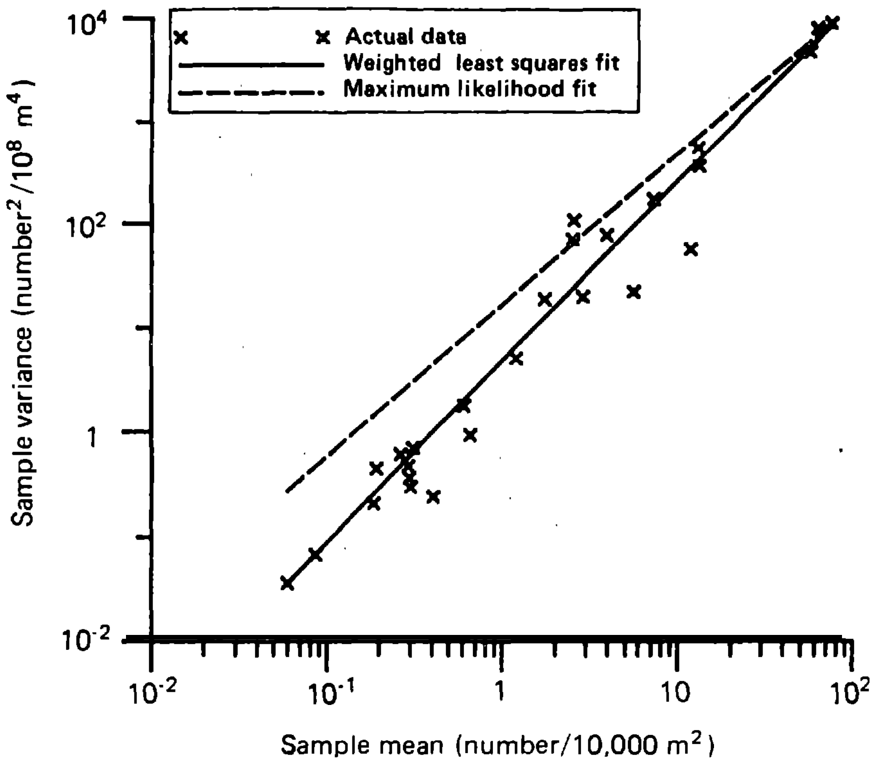

Figure 5.--Log-log plot of sample variance versus sample mean of CPUE of 1-yr-old sablefish, with weighted least squares and maximum likelihood fitted lines.

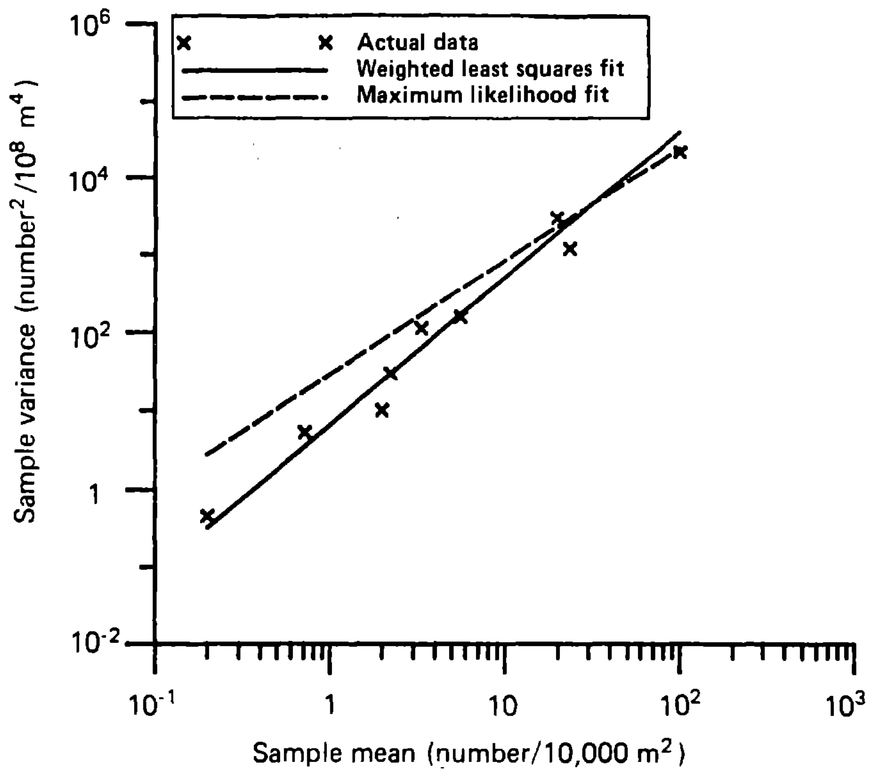

Figure 6.--Log-log plot of sample variance versus sample mean of CPUE of young-of-the-year Pacific cod, with weighted least squares and maximum likelihood fitted lines.

(1985) has described a method of parameter estimation for the three-parameter Weibull distribution which appears preferable to the method of maximum likelihood; this new method may be used in the future.)

The resulting parameter estimates were P1 = 3.806, P2 = 1.683, and P3 = 0.2826. Parameter P1 is unitless; parameters P2 and P3 have units of 10,000 m^2. Parameter P_3 was significantly different from zero at the 96.0% level using a generalized likelihood ratio test (Mood et al. 1974: 440-441).

Figure 7 gives an indication of the goodness of fit of the fitted Weibull distribution to the effort data. It appears most of the lack of fit was caused by nonrandom termination of 117 hauls at 0.50 or 1.00 nautical miles (0.90×10^4 m^2 or 1.81×10^4 m^2); the average duration of these hauls was 17 and 30 min. Extremely small ($<0.55 \times 10^4$ m^2) and extremely large ($>3.0 \times 10^4$ m^2) values of effort also occurred more frequently than expected from the fitted Weibull distribution, but the resulting lack of fit appears minor. In future work, a double exponential (Laplace) or related distribution may be used instead of a Weibull distribution in order to improve the fit. Additional improvement to fit may also be obtained by separating the hauls into two populations according to whether a duration of 15 or 30 min was originally intended, and fitting a different statistical distribution to each population.

The statistical distribution of catch was modeled as follows. Let C_{ijuv} and, E_{ijv} be the catch (no.) and fishing effort (10,000 m^2) for geographic region i, year j, species or age group u, and haul v. It is assumed that

$$\mathrm{var}(C_{ijuv}/E_{ijv}) = P_{4u} \times [\mathrm{mean}(C_{ijuv}/E_{ijv})]^{P_{5u}}, \tag{3}$$

where "var" and "mean" are functions giving the expected value of the mean and variance, and P_{4u} and P_{5u} are parameters. When it is apparent from the

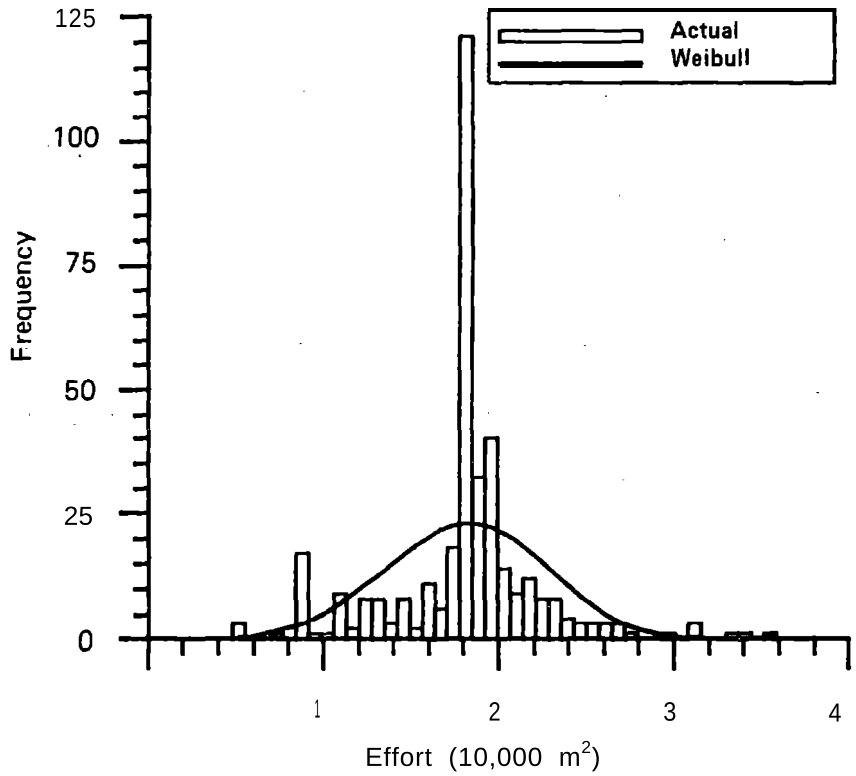

Figure 7.--Histogram of fishing effort during 1980-82 surveys in the northwestern Gulf of Alaska, with graph of expected frequencies calculated using a fitted Weibull distribution.

context which species or age group is being referred to; P_{4u} and P_{5u} may be- referred to as P_4 and P_5. The above equation implies that the variance of CPUE is a power function of its mean. It is assumed that

$$\text{mean}\ (C_{ijuv}) = E_{ijv} \times M_{iju}, \tag{4}$$

where M_{ijv} is the apparent underlying population density (no./10,000 m^2). Although E_{ijv} has a Weibull distribution, it is assumed to be measured without significant error, so that if E_{ijv} is given, it may be treated as a constant. This implies that given E_{ijv},

$$\begin{aligned} \text{var}(C_{ijuv}) &= \text{var}(E_{ijv} \times C_{ijuv} / E_{ijv}) \\ &= (E_{ijv})^2 \times \text{var}(C_{ijuv}/E_{ijv}) \\ &= (E_{ijv})^2 \times P_{4u} \times [\text{mean}(C_{ijuv}/E_{ijv})]^{P5u} \end{aligned} \tag{5}$$

and

$$\begin{aligned} \text{mean}(C_{ijuv}/E_{ijv}) &= \text{mean}(C_{ijuv})/E_{ijv} \\ &= E_{ijv} \times M_{iju} / E_{ijv} \\ &= M_{iju}\ . \end{aligned} \tag{6}$$

Combining the above two equations gives

$$\text{var}(C_{ijuv}) = (E_{ijv})^2 \times P_{4u} \times [M_{iju}]^{P5u}\ . \tag{7}$$

It was assumed that C_{ijuv} is an integer. Equations 4 and 7 imply that $\text{var}(C_{ijuv})$ could be greater than, equal to, or less than mean (C_{ijuv}), depending on the value of E_{ijv}, P_{4u}, P_{5u}, and M_{iju}. It was almost always the case in this study that var $(C_{ijuv}) > 1.01 \times$ mean (C_{ijuv}).

The negative binomial distribution has often been used to model the count of fish in catches (Zweifel and Smith 1981, Taylor 1953). No test was made of the goodness of fit of the negative binomial to the raw catch data in this study, and it is possible that there were significant deviations from a negative binomial distribution, perhaps because some counts were estimated using subsamples or because of some other reason. However, examination of the data indicated that the negative binomial appeared plausible. It also seems reasonable that inferences drawn in this study assuming a negative binomial distribution will not be severely affected by plausible deviations from this assumption. Nonetheless, it would be useful to develop some test of the validity of an underlying negative binomial distribution.

In addition to the negative binomial, the Poisson and binomial distributions can be used to model counts of fish in catches (Elliott 1977). The variances of the negative binomial, Poisson, and binomial distributions are respectively greater than, equal to, or less than the mean of the distribution. These distributions were used to model the catch of fish in this study. Although the below parameterizations may appear complex, all three distributions satisfy Equation 4 exactly, so that expected catch is a linear function of effort. In addition, the negative binomial distribution satisfies Equation 3 exactly, and the Poisson and binomial distributions satisfy Equation 3 approximately, so that the expected variance of CPUE is essentially a power function of its expected mean.

If var $(C_{ijuv}) > 1.01 \times$ mean (C_{ijuv}), then C_{ijuv} was assumed to have a negative binomial distribution with the probability density function

$$(P_{ijuv})^{k_{ijuv}} \times (1 - P_{ijuv})^{C_{ijuv}} \qquad (8)$$

$$\times \ \mathrm{gamma}(k_{ijuv} + C_{ijuv}) \ / \ [\mathrm{gamma}(C_{ijuv} + 1) \times \mathrm{gamma}(k_{ijuv})],$$

where

$$k_{ijuv} = [\mathrm{mean}(C_{ijuv})]^2/[\mathrm{var}(C_{ijuv}) - \mathrm{mean}(C_{ijuv})], \tag{9}$$

$$p_{ijuv} = k_{ijuv}/[\mathrm{mean}(C_{ijuv}) + k_{ijuv}], \tag{10}$$

gamma is the gamma function, and $\mathrm{mean}(C_{ijuv})$ and $\mathrm{var}(C_{ijuv})$ are given by Equations 4 and 7.

If $0.99 \times \mathrm{mean}(C_{ijuv}) \leq \mathrm{var}(C_{ijuv}) \leq 1.01 \times \mathrm{mean}(C_{ijuv})$, then C_{ijuv} was assumed to have a Poisson distribution with the probability density function

$$[\mathrm{mean}(C_{ijuv})]^{C_{ijuv}} \times \exp(-\,\mathrm{mean}(C_{ijuv}))/\mathrm{gamma}(C_{ijuv}+1), \tag{11}$$

where $\mathrm{mean}(C_{ijuv})$ is defined by Equation 4.

If $\mathrm{var}(C_{ijuv}) < 0.99 \times \mathrm{mean}(C_{ijuv})$, then C_{ijuv} was assumed to have a binomial distribution with probability density function

$$[r_{ijuv}]^{C_{ijuv}} \times [1 - r_{ijuv}]^{N_{ijuv}-C_{ijuv}} \times \binom{N_{ijuv}}{C_{ijuv}}, \tag{12}$$

where

$$N_{ijuv} = \begin{cases} \mathrm{round}(-k_{ijuv}), \text{ if } \mathrm{round}(-k_{ijuv}) > \mathrm{mean}(C_{ijuv}) \\ 1 + \mathrm{round}(-k_{ijuv}), \text{ if } \mathrm{round}(-k_{ijuv}) \leq \mathrm{mean}(C_{ijuv}), \end{cases} \tag{13}$$

$$r_{ijuv} = \mathrm{mean}(C_{ijuv})/N_{ijuv}, \tag{14}$$

"round" is the function which rounds values to the nearest integer and rounds values ending in 0.5 to the next largest integer, and $\mathrm{mean}(C_{ijuv})$ and k_{ijuv} are defined by Equations 4 and 9, respectively.

The parameters $\log_{10}M_{iju}$, $\log_{10}P_{4u}$, and P_{5u} were estimated by the method of maximum likelihood using a log likelihood function based on the negative binomial (Equation 8), whenever the estimate of var (C_{ijuv}) was greater than the estimate of mean(C_{ijuv}). The corresponding estimates of M_{iju} and P_{4u} were calculated using antilogarithms. It was not found necessary to include the Poisson and binomial distributions (Equations 11 and 12) in the log likelihood functions. An iterative process was used to find the maximum likelihood estimates. First, preliminary estimates of $\log_{10}M_{iju}$ were found by taking logarithms of the nonzero sample means of the CPUE values for each given year and geographical region; these sample means and associated sample variances are listed in Tables 2-5. If a sample mean equaled zero, then the corresponding estimate of M_{iju} was assumed to also equal zero. All CPUE values corresponding to sample means which equaled zero were deleted from the data sets used for maximum likelihood estimation. Preliminary estimates of $\log_{10}P_{4u}$ and P_{5u} were estimated by weighted least squares linear regression of logarithms of sample variances on logarithms of sample means (Perry 1981: equation 2); each case weight used in the regression was set equal to one less than the number of hauls used in the associated sample mean. The regression lines calculated using weighted least squares are shown in Figures 3-6. After calculating preliminary parameter estimates, but before beginning the maximum likelihood estimation process, the data values C_{ijuv} were set to equal round(C_{ijuv}) . In the case of sablefish, 1.1 was used as the initial estimate of $\log_{10}P_{4u}$ instead of the weighted least squares estimate of 0.674, so that additional computer programming was not necessary in order to include Poisson or binomial terms in the log likelihood function. However, it was later found that the maximum likelihood estimation process converged to essentially the same final parameter estimates for sablefish, even if the initial estimate of

$\log_{10}P_{4u}$ was not increased and Poisson and binomial terms were initially included.

It is known that the log likelihood function of a negative binomial is unimodal (Levin and Reeds 1977: theorem 2) However, it is not certain that the log likelihood function is unimodal when the variance of the negative binomial is forced to be a power function of the mean, as in Equation 8. Perhaps a mathematical proof of this could be found. If the log likelihood function is not unimodal, it is possible that the iterative maximum likelihood estimation process did not converge to the global maximum of the log likelihood function, but rather to a local maximum. As a partial check of this possibility, 41 sets of initial parameter estimates were tried for both young-of-the-year and 1-yr-old pollock, and 3 different sets were tried for young-of-the-year Pacific cod. Only one set of initial estimates was used for 1-yr-old sablefish. Each set of initial estimates was improved ,by using a FORTRAN computer program which used IMSL (1982) subroutine ZXMIN to maximize the log likelihood function. In the cases where more than one initial set of parameters was used, the resulting sets of final parameter estimates for each species or age group did not appear to differ significantly from each other; the differences which did occur appeared to be a result of rounding error in calculation of the log likelihood function coupled with flatness of the log likelihood function near its maximum. This provided evidence that the estimates were indeed maximum likelihood estimates.

Final maximum likelihood estimates of the M_{iju}, P_{4u}, and P_{5u} are shown in Tables 6-7 for each species or age group. Regression lines corresponding to the final maximum likelihood estimates of P_{4u} and P_{5u} are also plotted in Figures 3-6. These figures indicate that the slopes of the maximum likelihood estimates of log variance as a function of log mean appear biased toward zero

Table 6.--Maximum likelihood estimates (no./10,000 m^2) of apparent fish density, M_{iju}, as a function of year, stratum, and species or age group.

Year	Stratum	Walleye pollock young-of-the-year	Walleye pollock 1-yr-old	Sablefish 1-yr-old	Pacific cod young-of-the-year
1980	Castle	171.7	7687	30.79	0.0
	Chignik	208.3	1341	38.81	0.0
	Kujulik	571.0	1237	91.64	0.0
	Wide	1187	297.8	0.5659	1.822
	Alitak	123.6	202.2	0.3639	0.0
	Sitkalidak	38.53	4.345	1.369	0.0
	Kiliuda	20.55	0.0	0.0	0.0
	Ugak	459.2	2.430	0.0	0.0
	Chiniak	2534	0.0	3.301	0.0
	Marmot	156.7	23.09	0.7041	0.0
	Uganik	0.0	211.2	2.522	0.0
	Uyak	13.43	278.5	0.0	0.0
1981	Castle	596.4	400.3	13.17	0.0
	Chignik	764.5	963.2	6.999	0.0
	Kujulik	842.8	312.6	27.55	0.4667
	Wide	1692	437.9	0.0	96.02
	Alitak	2842	7.870	0.0	20.52
	Sitkalidak	195.8	66.82	0.2469	0.0
	Kiliuda	767.2	36.87	0.9640	0.0
	Ugak	3525	2.106	0.0	0.0
	Chiniak	1876	93.18	0.2757	1.440
	Marmot	613.9	320.8	0.0	0.0
	Uganik	139.3	38.12	3.577	0.0
	Uyak	78.61	175.1	0.2742	0.0
1982	Castle	139.0	824.5	3.309	0.0
	Chignik	49.70	152.0	0.5105	0.0
	Kujulik	288.9	379.3	13.20	0.6788
	Wide	797.4	334.0	1.154	18.42
	Alitak	275.5	500.8	0.0	4.378
	Sitkalidak	35.13	18.42	0.7889	0.0
	Kiliuda	183.2	16.84	0.9407	0.0
	Ugak	56.17	45.02	1.367	0.0
	Chiniak	182.6	2.401	5.244	0.0
	Marmot	472.5	25.66	8.361	0.0
	Uganik	5.333	175.1	0.0	0.0
	Uyak	12.60	144.7	0.0	4.071

Table 7.--Maximum likelihood estimates of parameter values used to calculate variance as a power function of the mean.

Parameter name (units)	Walleye pollock young-of-the-year	Walleye pollock' 1-yr-old	Sablefish 1-yr-old	Pacific cod young-of-the-year
multiplier P_4 ($[\text{no.}/10{,}000\ m^2]^{2-P_5}$)	59.68	75.09	16.45	29.02
exponent P_5 (unitless)	1.545	1.497	1.459	1.456

relative to the initial weighted least squares regression estimates. The slopes estimated using weighted least squares may also be biased toward zero; Bloch (1978) discusses this problem for unweighted least squares. Consequently, the maximum likelihood estimates are probably more biased than the weighted least squares estimates. However, it is possible that the mean-squared errors of the maximum likelihood estimates are lower, so that despite bias the maximum likelihood estimates might be better from the point of view of minimizing mean-squared error.

Selection of a Statistical Test

When estimating sample sizes, one must specify the null and alternative hypotheses as well as a statistical test of whether to accept or reject the null hypothesis. In order to be appropriate for this study, the statistical test used had to take into account the heteroscedasticity and non-normality of the CPUE data. The null hypothesis specified that the apparent total population size (number) of a given species or age group in the entire survey area did not vary from year to year. The alternative hypothesis was that the total number in the entire survey area did vary from year to year.

An ANOVA method used in conjunction with an F-test based on the method of Brown and Forsythe (1974a) was found to be appropriate for a comparison between 2 yr. The ANOVA method and F-test differed from the usual ANOVA and F-test. It was necessary to modify the usual equations for ANOVA because of the particular null hypothesis used, and a modified F-test was necessary because of heteroscedasticity.

Mathematically,

$$\sum_{i=1}^{I} (A_i \times M_{iju}) \qquad (15)$$

equals the total population in the entire survey area in year j of species or age group u, where I is the total number of strata, and A_i is the area of the ith stratum. The null hypothesis, H_o, that the total population size did not vary between years 1,2,...,J can be expressed as

$$\sum_{i=1}^{I} (A_i \times M_{iqu}) = \sum_{i=1}^{I} (A_i \times M_{iru}) , \tag{16}$$

where q and r are any distinct elements of the set 1,2,...,J. Indices 1,2,...,J may represent any years; the years need not be consecutive or in order. In simulations in this study, I=12 and J=2. It is assumed that the sum of the Ai is greater than zero.

The parameters M_{iju} are assumed to follow a general linear model; i.e.,

$$M_{iju} = a_u + b_{iu} + c_{ju} + d_{iju} , \tag{17}$$

where a, is the overall mean population density (no./10,000 m^2) of species *or* age group u, b_{iu} is the change in density (no./10,000 m^2) attributed to stratum, c_{ju} is the change in density (no./10,000 m^2) attributed to year, and d_{iju} is the change in density (no./10,000 m^2) attributed to stratum by year interaction. As is common in ANOVA, the bi_u and Cju are defined so that

$$0 = \sum_{i=1}^{I} b_{iu} = \sum_{j=1}^{J} c_{ju} . \tag{18}$$

In this study the parameters a_u, bi_u, Cju, and diju are also defined in such a manner that Equation 16 is true if and only if

$$0 = c_{qu} \tag{19}$$

for each q. This is convenient because a test of the null hypothesis that

Equation 16 is true becomes equivalent to a test of the null hypothesis that Equation 19 is true. It can be algebraically proved that the condition

$$0 = \sum_{i=1}^{I} (A_i \times d_{iqu}) \qquad (20)$$

for each q is sufficient to insure that Equation 16 is true if and only if Equation 19 is true. Consequently the diju are defined in this study so that Equation 20 is true, which insures that a test of the null hypothesis that Equation 16 is true is equivalent to a test of the null hypothesis that Equation 19 is true. In addition the d_{iju} are defined so that

$$0 = \sum_{j=1}^{J} d_{1ju} = \sum_{j=1}^{J} d_{2ju} = \cdots = \sum_{j=1}^{J} d_{Iju} . \qquad (21)$$

The parameter restrictions expressed by Equations 18 and 21 are quite common in ANOVA, but the parameter restrictions expressed by Equation 20 are relatively unusual. When performing an ANOVA of this type, care must be taken to choose an algorithm or statistical package program that permits specification of coefficient restrictions such as in Equation 20.

The parameters d_{iju} represent interannual shifts in the geographic distribution of the population. Even if the total population does not change from year to year, so that each Cju = 0 and the null hypothesis is true, the geographical distribution of the population may change, which implies that some of the d_{iju} are nonzero. Consequently, even if the interaction effect is significant, it is still meaningful in this situation to test whether the main effect due to years is also significant, which would indicate that the total population size has changed between years.

For each species or age group u, the estimates of au, b_{iu}, C_{ju}, and d_{iju} were simultaneously calculated using. an ordinary least squares multiple linear regression approach which was programmed in FORTRAN. The algorithm was mathematically equivalent to the method of Lagrange multipliers used by IMSL (1982) subroutine AGLMOD, but certain changes were made to reduce the computer central processor time necessary to perform replicate ANOVA's. The estimates and tests of significance of the parameters in Equation 17 were made assuming a fixed effects model. This was felt justified on the following basis: An effect may be considered random because it is not presently possible to predict its magnitude very far in advance. Nonetheless, at the time the effect is measured, the only observable random variation may be due to measurement and sampling variation, which can both be grouped into a "measurement error" term. The result is a fixed effects model for describing the magnitude of the effect at the time of measurement.

An ordinary F-test can be used to test the significance of an ANOVA effect, but this test can be adversely affected by heteroscedasticity. However, Brown and Forsythe (1974a: sections 4,7) reexpressed this test as an F-test of the significance of a set of orthonormal contrasts on the ANOVA cell means, and compensated for the effects of heteroscedasticity on this test by using an approximation originated by Satterthwaite (1946: equation 7) to make an adjustment to the denominator degrees of freedom (df) of the F statistic. The equations of Brown and Forsythe (1974a) can be reexpressed in matrix form as follows. Let n_{ij} be the number of hauls in stratum i and year j, $MCPUE_{iju}$ be the sample mean of CPUE in stratum i and year j of species or age group u, and $[s_{iju}]^2/n_{ij}$ be the sample variance of $MCPUE_{iju}$. Let m be the (I x J) by 1 dimensional vector of sample means in each ANOVA cell: i.e., element [I x (j-1) + i] of m is set equal to $MCPUE_{iju}$. Let V be the corresponding (I x J) by

1 dimensional vector of the $[s_{iju}]^2/n_{ij}$, and let W be the corresponding (I x J) by 1 dimensional vector of the quantities $1/(n_{ij} - 1)$. If $n_{ij} = 1$, then $[s_{iju}]^2/n_{ij}$ is set to zero, and the corresponding element of, W is arbitrarily set equal to one. Let T be an (I x J) by (J - 1) dimensional matrix of orthonormal contrast coefficients appropriate- for testing whether Equation 19 is true: i.e., whether there is an overall difference between years: Let T' be the transpose of T. Define the function "diag" such that diag(TT') is the 1 by (I x J) dimensional vector equal to the diagonal of TT'. Let f_{ij} equal element [I x (j-1) + i] of diag(TT'). Define the function "DIAG" such that DIAG(V) is a square matrix with its diagonal equal to V and zeroes elsewhere, and let $\underline{V}$ = DIAG(V), $\underline{W}$ = DIAG(W), and $\underline{g}$ = DIAG([diag(TT')]'). The F statistic used in this study to test whether there is an overall difference between years was mathematically equivalent to

$$F = \frac{m'TT'm}{\mathrm{trace}(TT'\underline{V})} \tag{22}$$

$$= \frac{m'TT'm}{\sum_{i=1}^{I} \sum_{j=1}^{J} f_{ij} \times [s_{iju}]^2/n_{ij}} ,$$

where trace (TT'V) is the sum of the diagonal elements of TT'V. Brown and Forsythe (1974a) assumed this F statistic approximately followed an F distribution with the df of the numerator equal to J-1 and the df of the denominator approximated by a formula mathematically equivalent to

$$DFD = \frac{[\mathrm{trace}(TT'\underline{V})]^2}{\mathrm{trace}([\underline{DV}]^2\underline{W})} \tag{23}$$

$$= \frac{\left(\sum_{i=1}^{I}\sum_{j=1}^{J} f_{ij} \times [s_{iju}]^2/n_{ij}\right)^2}{\sum_{i=1}^{I}\sum_{j=1}^{J} [f_{ij} \times [s_{iju}]^2/n_{ij}]^2/(n_{ij}-1)} .$$

It is unnecessary to calculate T in Equations 22-23, because TT' can be calculated as in Appendix B (Equation 54).

If $n_{ij} = 1$, then it was assumed in Equation 22 that $[s_{iju}]^2 = 0$. In Equation 23 terms involving $n_{ij} = 1$ were not included in summations.

If the denominator of Equation 22 was ≤ 0, the F statistic was set equal to the numerator multiplied by 10^{18}. If the denominator of Equation 23 was ≤ 0, the estimated df was arbitrarily set equal to 0.5. However, such situations rarely occurred.

Since the denominator df was generally not an integer, IMSL (1982) subroutine MDFDRE was used to evaluate the significance of the F statistic in Equation 22.

Advantages of the method described by Equations 22-23 which were important to this study were 1) no transformation of the data is necessary, 2) no computational difficulties arise if some of the S_{iju} equal zero, and 3) the method is based on asymptotic normality of ANOVA cell means, so from the central limit theorem (Snedecor and Cochran 1980: section 4.5; Cochran 1977: section 2.15) it is expected to be robust whenever the nij are sufficiently large, even if actual distributions of catch and fishing effort differ from those used in this study. Since the method does not assume that variance of CPUE is a power function of its mean, the method is still applicable even if the variance and mean do not follow this relationship.

Other examples of the approach of Brown and Forsythe (1974a) are given in Brown and Forsythe (1974b), Iman and Davenport (1976), and Tamhane (1979). Unless sample sizes were quite small, the use of the method of Brown and Forsythe (1974a) to calculate an F statistic was found in the cited examples to be reasonably powerful with approximately correct type I error rates and to be robust for one kind of non-normality. (A type I error is defined as the rejection of the null hypothesis, when in fact the null hypothesis is true. In this study, a type I error rate is defined as the percentage of a set of replicates for which a type I error occurred.) If $J = 2$, then the method of Brown and Forsythe (1974a) is a special case of the method of Rubin (1982: section 4.2(iv)) discussed in Appendix C; Rubin applied an approximation described by Box (1954: equation 6.1). Improvements to the method of Brown and Forsythe (1974a) have been found important in some cases (Rubin 1982, 1983; Tan 1982a,b; Kaiser and Bowden 1983). One of these approximations may be used in future studies; an improved approximation often is particularly important if $J > 2$ (Rubin 1982: section 4.4).

Other approaches besides that of Brown and Forsythe (1974a) do exist for dealing with heteroscedasticity. A common method of dealing with heteroscedasticity is to apply a nonlinear variance-stabilizing transformation to the data, such as a power, logarithmic, or rank transformation, and then perform an ANOVA on the transformed data (Green 1979: section 2.3.9). However, nonlinear transformation is generally inappropriate if interaction effects are present and one wishes to combine density data from several different regions into a weighted-estimate of total density or population Size, such as in Equation 16. Appendix A gives examples of uncontrolled type I error rates which result from employing the 'commonly used nonlinear transformation log(CPUE+l); such uncontrolled error rates may be typical for other nonlinear transformations as well;

These uncontrolled type I error rates are in large part caused by the presence of interaction effects. If any nonlinear transformation is used and interactions are present, then Equations 16 and 19 may no longer be equivalent, and a test of whether Equation 19 is true will no longer be equivalent in general to a test of whether Equation 16 is true. This causes apparent type I errors, and leaves one without a convenient way to test the null hypothesis. Accordingly, use of a nonlinear transformation in conjunction with a test of a null hypothesis of the type expressed by Equations 16 and 19 should be avoided for multi-way ANOVA unless it is appropriate to assume that interaction effects are not present. Even then the transformation may not be appropriate (Appendix A). Nonlinear transformations may sometimes be useful for one-way ANOVA or for multi-way ANOVA tests of hypotheses that can be conveniently expressed in a form compatible with the nonlinear transformation.

Monte Carlo Simulation

Because of heteroscedasticity and non-normality of the original data, no attempt was made to analytically estimate the sample' sizes needed to detect departures from the null hypothesis with given probability. Instead, the method of Brown and Forsythe (1974a) described in the previous section was applied to replicate surveys generated using Monte Carlo simulation, and the ratio of the number of times the null hypothesis was rejected to the total number of replicates was used to estimate the probability of detecting departures from the null hypothesis. The resulting probabilities can be used to indicate what sample sizes are deemed necessary. Monte Carlo simulation is not used in the method of Brown and Forsythe (1974a).

The model used for Monte Carlo simulation was implemented in FORTRAN IV on the Burroughs B7800 computer at the Northwest and Alaska Fisheries Center.

The simulated surveys were assumed to take place in the 12 geographic regions listed in Table 1, and during 2 different years; therefore I=12 and J=2. Because variances of CPUE in the simulations are a function of the underlying population densities, the predicted probabilities of detecting departures from the null hypothesis are also a function of the underlying population densities. Consequently, it is necessary to specify the mean overall population densities in each year as input parameters. Any departures from these means from one geographic region to another must also be specified or calculated within the simulation program. Preliminary analyses indicated that some significant systematic differences in population density between regions appeared to be repeated from 1980 to 1982 for the species or age groups surveyed. In addition, interaction effects between regions and years also appeared significant. However, precise mathematical or statistical descriptions of these effects were not formulated. Application of an ANOVA method and an approximate F-test which takes heteroscedasticity into account may make improved descriptions possible. An empirical formulation of possible region (stratum) by year interaction effects based on the survey estimates of apparent population density was used in simulations. This formulation is related to the bootstrap method described by Efron and Gong (1983). It was assumed that stratum by year interactions were essentially random between years, because it is not known how to predict them in advance. The population densities in each stratum and year, Miju, were used in simulations to specify stratum by year interaction effects. Let MESTij, be a survey estimate of

M_{iju}; the values of $MEST_{iju}$ were used as simulation input parameters and are given in Table 6. Values of M_{iju} calculated during simulations were forced to be proportional to some value of $MEST_{iju}$, where j was picked randomly. This was done as follows.

Define $MEST_{.ju}$ by

$$MEST_{.ju} = \sum_{i=1}^{I} A_i \times MEST_{iju} \, , \tag{24}$$

where index j equals 1, 2, or 3, since there were 3 yr of survey data. Define A. by

$$A_{.} = \sum_{i=1}^{I} A_i \quad . \tag{25}$$

Then $MEST_{.ju}/A.$ equals the average estimated density weighted by area in year j of population u. Define Z_{iju} by

$$Z_{iju} = MEST_{iju}/(MEST_{.ju}/A_{.}) \quad . \tag{26}$$

Before simulating each replicate survey, J integers were picked from a random permutation of the indices of the years of original survey data. For example, in this study, since there were 3 yr of survey data and J=2 in simulations, the integers g and h were chosen to equal the first two elements of a random permutation of the set {1,2,3}. Consequently, the set,{g,h].equaled {1,2}, {2,3), or 11,31 in any given replicate survey with equal probability. The first J integers of the random permutation were generated by a FORTRAN subroutine which used an algorithm similar to that of IMSL (1982) subroutine GGPER. Let PDY_{1u} and PDY_{2u} equal the input parameters of the simulation which give the true mean overall density (no./10,000 m^2) in years 1 and 2 of species or age group u. Then for each replicate simulated survey, the Miju were calculated using the formulas

$$M_{i1u} = PDY_{1u} \times z_{igu} \tag{27}$$

and

$$M_{i2u} = PDY_{2u} \times z_{ihu} \quad . \tag{28}$$

Consequently, each Miju was proportional to either MESTigu or MESTihu.

Equations 27 and 28 make it possible to specify both stratum by year interaction effects and annual mean overall population densities in simulations, because in any given simulated year j, the average of the Miju weighted by stratum areas equals PDYju; i.e.,

$$PDYju = (\sum_{I=1}^{I} Ai \times Miju)/A, \; l \tag{29}$$

These equations preserve any systematic effects attributed to strata which may exist, because the index I referring to strata was not randomized. But since index j was randomized, the equations do introduce a random component to the stratum 'by year interactions. It is uncertain to what extent this method may underestimate or overestimate variability. For instance, g is always unequal to h, which may tend to overestimate variability due to stratum by year interactions. In contrast, whenever the survey estimate of population density, MESTiju, equals zero, then the corresponding simulated density parameter Miju Always equals zero also. Since MESTiju may have equaled zero by coincidence, even though the actual population density in that geographic region and year was nonzero, this could cause an underestimate in simulations of actual variability.

Total sample sizes (number of simulated trawl hauls) of 60, 120, and 180 hauls/year were allocated to the geographic regions in proportion to the areas listed in Table 11 slight adjustments were made so that the sample sizes allocated to each region would sum to the total specified for each year despite rounding. The resulting values of n_{ij} are listed in Table 8. Depending on the objectives of a given survey, some other method of allocation may be more optimal, but this method was chosen for illustration.

Fishing effort in each trawl haul was simulated using the three-parameter Weibull distribution specified by Equation 2. Samples were generated from this distribution by the method of inversion (Bury 1975: 542-543). Unless noted otherwise, the Weibull distribution parameters were P1=3.866, P2=1.683, and P_3=0.2826; P, is unitless and P_2 and P_3 have units of 10,000 m^2.

Catches in each trawl haul were simulated using the negative binomial, Poisson, or binomial distributions as specified by Equations 8, 11, or 12. The values of parameters P4u and $P5_u$ are given in Table 7.

As a result of the parameters used in these simulations, only the negative binomial distribution was used in the simulation of catches of young-of-the-year and 1-yr-old walleye pollock and young-of-the-year Pacific cod; the Poisson and binomial distributions (Equations 11-12) were never used. The probability of the use of Poisson or binomial distributions (Equations 11-12) to simulate catches of 1-yr-old sablefish was less than 0.04% in any given trawl haul and occurred only for very small values of fishing effort; in other cases the negative binomial was used.

A variate from the negative binomial distribution was generated by first generating a value from a gamma distribution; this value was then used as the parameter of a Poisson distribution, which in turn was used to generate one

Table 8.--Allocation of sampling effort (trawl hauls/year) to geographic regions in simulated surveys as a function of total sampling density.

Stratum	Total sampling density (hauls/km^2) 0.0384	0.0768	0.1152
Castle Bay	1	2	3
Chignik Bay	4	9	13
Kujulik Bay	3	6	9
Wide Bay	1	2	3
Alitak Bay	16	33	48
S. Sitkalidak Strait	9	18	27
Kiliuda Bay	6	12	18
Ugak Bay	4	7	11
Chiniak Bay	2	4	7
Marmot Bay	8	15	23
Uganik Bay	3	5	8
Uyak Bay	3	7	10
Total	60	120	180

variate. The resulting variate can be considered to come from a negative binomial distribution (Bratley et al. 1983: 174).

In simulations, the IMSL (1982) subroutines GGAMR, GGEON, and GGBN were used to generate values from the gamma, Poisson, and binomial distributions. When uniformly distributed random numbers were required, a pseudorandom multiplicative congruential generator was used which was based on the algorithm of the IMSL (1982) subroutine GGUBS, except that instead of a multiplier of 7^5-16,807, a multiplier of 764,261,123 was used. The latter. multiplier performed well in tests by Fishman and Moore (1982), and was among those recommended by Maindonald (1984: 280-281). The algorithm using the latter multiplier was also substituted whenever the IMSL subroutines made calls to the IMSL uniform variate generators GGUBFS or GGUBS. Goodness of fit tests using the G statistic for data classified into categories (Sokal and Rohlf 1969: sections 16.1-16.2) indicated that the permutation, Weibull, negative binomial, Poisson, and binomial generators used in this study performed as expected.

RESULTS

Tables 9-12 list simulation estimates of the probability of a type I error as a function of population density when using a statistical test with a nominal significance level of 95%. The population densities used as parameters in these tables were estimated as follows. First, typical apparent annual population densities were estimated by taking averages of the values of MESTiju (Table 6), weighted by stratum area (Table 1). The resulting estimates for young-of-the-year and 1-yr-old walleye pollock, 1-yr-old sablefish, and young-of-the-year Pacific cod were, respectively, 610, 256, 4.34, and 3.02 fish/10,000 m^2. Populations sparser and denser than these typical densities were then simulated in half order of magnitude increments. (A half order of magnitude increment is defined as a change by a factor of approximately 3.16.) This was done to indicate how simulated type I error rates change as a function of population density. Since the null hypothesis was assumed to be true in Tables 9-12, in each case $PDY1_u = PDY2u$.

Similarly, Tables 13-16 list simulation estimates of the probability of correctly detecting specified changes in population levels between years. A change was counted as correctly detected if 1) the approximate F-statistic corresponding to a change between years was significant, and 2) the estimate of c_{1u} was greater than or essentially equal to the estimate of c_{2u}. This latter condition was used because $PDY_{1u} > PDY_{2u}$.

In Tables 9-16, each probability was estimated by dividing the number of times a type I error or correct detection occurred by the total number of replicates; 400 replicates were used for each table entry. This large number of replicates was used so that multiple comparisons could be made with confidence. A constant number of replicates was used for simplicity, although

Table 9.--Predicted probabilities of a type I error? as a function of total population density of young-of-the-year walleye pollock and annual sample size.

Population density (no./10,000 m^2)	Sample size (hauls/year)	Type I error probability (%)
61	60	4.75
	120	2.50
	180	4.50
193	60	7.00
	120	5.25
	180	5.00
610	60	4.25
	120	4.50
	180	3.25
1930	60	5.75
	120	3.75
	180	4.50
6100	60	6.25
	120	4.75
	180	5.50

*A type I error is defined as rejecting the null hypothesis when the null hypothesis is true. The null hypothesis was that there was no change in total annual population size in the survey area. An F-test based on 'the method of Brown and Forsythe (1974a) with a nominal significance level of 95% was used in conjunction with a modified form of analysis of variance applied to catch per unit effort data. Each probability was estimated using 400 replicates.

Table 10.--Predicted probabilities of a type I error* as a function of total population density of 1-yr-old walleye pollock and annual sample size.

Population density (no./10,000 m^2)	Sample size (hauls/year)	Type I error probability (%)
25.6	60	9.00
	120	2.25
	180	4.25
81.0	60	9.25
	120	4.75
	180	5.75
256	60	9.25
	120	6.00
	180	5.25
810	60	10.25
	120	5.00
	180	4.75
2560	60	10.50
	120	5.50
	180	7.00

*A type I error is defined as rejecting the null hypothesis when the null hypothesis is true. The null hypothesis was that there was no change in total annual population size in the survey area. An F-test based on the method of Brown and Forsythe (1974a) with a nominal significance level of 95% was used in conjunction with a modified form of analysis of variance applied to catch per unit effort data. Each probability was estimated using 400 replicates.

Table 11.--Predicted probabilities of a type I error* as a function of total population density of 1-yr-old sablefish and annual sample size.

Population density (no./10,000 m^2)	Sample size (hauls/year)	Type I error probability (%)
0.434	60	2.25
	120	0.75
	180	2.25
1.37	60	3.00
	120	2.25
	180	4.75
4.34	60	3.25
	120	3.25
	180	3.50
13.7	60	2.25
	120	3.50
	180	5.00
43.4	60	4.50
	120	4.00
	180	4.50

*A type I error is defined as rejecting the null hypothesis when the null hypothesis is true. The null hypothesis was that there was no change in total annual population size in the survey area. An F-test based on the method of Brown and Forsythe (1974a) with a nominal significance level of 95% was used in conjunction with a modified form of analysis of variance applied to catch per unit effort data. 'Each probability was estimated using 400 replicates.

Table 12.--Predicted probabilities of a type I error* as a function of total population density of, young-"of-the-year Pacific cod and annual sample size.

Population density (no./10,000 m^2)	Sample size (hauls/year)	Type I error probability (%)
0.302	60	26.00
	120	2.75
	180	6.75
0.955	60	28.25
	120	4.50
	180	4.25
3.02	60	27.25
	120	8.75
	180	7.25
9.55	60	32.25
	120	8.00
	180	9.25
30.2	60	31.00
	120	6.50
	180	6.25

*A type I error is defined as rejecting the null hypothesis when the null hypothesis is true. The null hypothesis was that there was no change in total annual population size in the survey area. An F-test based on the method of Brown and Forsythe (1974a) with a nominal significance level of 95% was used in conjunction with a modified form of "analysis of variance applied to catch per unit effort data. Each probability was estimated using 400 replicates.

Table 13.--Predicted probabilities* (%) of the correct detection of the direction of change in total annual population size of young-of-the-year walleye pollock, as a function of annual population densities and annual sample size.

Population density year 2 (no./10,000 m^2)	Sample size (hauls/year)	Population density year 1 (no./10,000 m^2)			
		61	193	610	1930
193	60	40.00			
	120	66.50			
	180	87.75			
610	60	85.00	58.75		
	120	97.25	88.50		
	180	100.00	97.25		
1930	60	89.00	86.00	78.50	
	120	99.75	98.25	94.75	
	180	100.00	100.00	100.00	
6100	60	90.00	88.75	90.25	86.50
	120	100.00	100.00	100.00	99.75
	180	100.00	100.00	100.00	100.00

*For an F-test based on the method of Brown and Forsythe (1974a) with a nominal significance level of 95% in conjunction with a modified form of analysis of variance applied to catch per unit effort data. The null hypothesis was that there was no change in total annual population size in the survey area. Each probability was estimated using 400 replicates..

Table 14.--Predicted probabilities* (%) of the correct detection of the direction of change in total annual population size of 1-yr-old walleye pollock, as a function of annual population densities and annual sample size.

Population density year 2 (no./10,000 m^2)	Sample size (hauls/year)	Population density year 1 (no./10,000 m^2)			
		25.6	81.0	256	810
81.0	60	31.50			
	120	44.50			
	180	62.00			
256	60	79.00	49.00		
	120	85.75	67.50		
	180	93.75	77.50		
810	60	96.75	94.25	72.50	
	120	88.75	93.25	80.75	
	180	97.75	99.00	91.50	
2560	60	99.50	99.50	99.00	88.75
	120	92.50	92.00	90.50	89.00
	180	100.00	99.75	99.25	99.25

*For an F-test based on the method of Brown and Forsythe (1974a) with a nominal significance level of 95% in conjunction with a modified form of analysis of variance applied to catch per unit effort data.. The null hypothesis was that there was no change in total annual population size in the survey area. -Each probability was estimated using 400 replicates.

Table 15.--Predicted probabilities* (%) of the correct detection of the direction of change in total annual population size of 1-yr-old sablefish, as a function of annual population densities and annual sample size.

Population density year 2 (no./10,000 m^2)	Sample size (hauls/year)	Population density year 1 (no./10,000 m^2)			
		0.434	1.37	4.34	13.7
1.37	60	5.25			
	120	18.75			
	180	37.00			
4.34	60	35.00	15.00		
	120	74.50	37.50		
	180	94.50	64.75		
13.7	60	63.00	49.25	25.00	
	120	98.00	94.75	65.00	
	180	99.50	99.75	85.00	
43.4	60	82.00	82.50	76.25	53.75
	120	99.75	99.50	99.00	92.25
	180	100.00	100.00	100.00	99.25

*For an F-test based on the method of Brown and Forsythe (1974a) with a nominal significance level of 95% in conjunction with a modified form of analysis of variance applied to catch per unit effort data. The null hypothesis was that there was no change in total annual population size in the survey area. Each probability was estimated using 400 replicates.

Table 16.--Predicted probabilities*, (%) of- the correct detection of the direction of change in total annual population size of young-of-the-year Pacific cod, as a function of annual population densities and annual sample size.

Population density year 2 (no./10,000 m^2)	Sample size (hauls/year)	Population density year 1 (no./10,000 m^2)			
		0.302	0.955	3.02	9.55
0.955	60	25.00			
	120	10.00			
	180	12.50			
3.02	60	47.00	32.00		
	120	35.75	16.25		
	180	48.75	23.75		
9.55	60	77.75	67.50	43.50	
	120	58.50	48.50	29.00	
	180	64.25	62.00	39.50	
30.2	60	96.00	96.25	90.50	63.25
	120	63.25	61.00	61.75	42.50
	180	73.75	77.00	72.25	52.25

*For an F-test based on the method of Brown and Forsythe (1974a) with a nominal significance level of 95% in conjunction with a modified form of analysis of variance applied to catch per unit effort data.. The null hypothesis was that there was no change in total annual population size in the survey area. Each probability was-estimated using 400 replicates.

multistage sampling and other methods discussed by Angers (1984) could presumably be used to reduce the number of replicates needed to achieve high levels of precision. Each replicate used for Tables 9-16 and Appendix Table A-l was calculated using a sequence of pseudorandom numbers which did not overlap the sequences of pseudorandom numbers used to calculate the other replicates. Each replicate can therefore be regarded as statistically independent of the other replicates. Since each probability being estimated can be regarded as fixed for a given table entry, and since each replicate is independent of the other replicates, the counts used to estimate the probabilities can be regarded as coming from a binomial distribution. Any of the usual methods for confidence intervals for the parameter of a binomial distribution can be applied to these probabilities, such as those reviewed by Matuszetiski and Sotres (19851. For multiple comparisons, the usual methods of contingency or frequency table analysis can be used to analyze the original count data -(for instance, see Dixon 1983: chapter 11). Suppose that p represents any of the 180 estimated Probabilities in Tables 9-16, and we wish to determine appropriate- multiple confidence intervals for these probabilities. Then using a normal approximation (Snedecor and Cochran 1980: section 7.8) for the most imprecise case of p= 50% it can be shown that the 180 confidence intervals given by (p-9.05%, p+9.05%) have a greater than 950 chance of including all of the true underlying probabilities. Although this illustrates the precision resulting from use of 400 replicates, more precise confidence intervals are possible for p unequal to 50% by taking the actual estimates of p into consideration instead of using the worst case p=50%.

Table 17 indicates that estimated type I error rates in Tables 9-12 were not different at individual significance levels of 95% from the expected type I error rate of 5% in the following cases: sample densities of 60, 120,

Table 17.--Significance of differences between actual counts of type I errors in Tables 9-12 and the expected 5% type I error rate, as a function of species, age group, and survey sample size.

Species	Age group (year)	Survey sample size (hauls/year)	G statistic*	df	Probability of a smaller G-statistic (%)
Walleye pollock	0	60	5.24	5	61.3
		120	8.16	5	85.2
		180	3.57	5	38.7
Walleye pollock	1	60	73.26	5	>99.9
		120	8.99	5	89.1
		180	4.07	5	46.1
Sablefish	1	60	22.94	5	>99.9
		120	37.25	5	>99.9
		180	10.32	5	93.3
Pacific cod	0	60	1213.29	5	>99.9
		120	23.25	5	>99.9
		180	20.12	5	99.9

*The G-statistic is defined in Sokal and Rohlf (1969: section 16.1).

and 180 hauls/year for young-of-the-year pollock; 120 and 180 hauls/year for 1-yr-old pollock) and 180 hauls/year for 1-yr-old sablefish. For young-of-the-year Pacific cod, estimated type I error-ratesdiffered at individual significance levels of 95% from the expected type I error rate at all three sampling densities. Nonetheless, for many uses the departures from the expected type I error rate may not be of practical significance except for the sample density of 60 hauls/year for young-of-the-year Pacific cod; estimated type I error rates were 26% or more for this case, in contrast to the maximum estimate of 10.5% for the other cases in Tables 9-12.

The excessive type I error rates for young-of-the-year Pacific cod using the sample density of 60 hauls/year (Table 12) apparently were the result of only allocating 1 haul/year in the simulations to the Wide Bay stratum (Table 8). When 2 hauls/year were allocated to the Wide Bay stratum and 8 hauls/year were allocated to the south Sitkalidak Strait stratum, with all other stratum allocations and parameters the same as for Table, 12, then the type I error rates were greatly reduced (Table 18). However, the rates were still significantly different from the expected 5% rate at a significance level of >99.9% (G=35.872, df=5), mainly due to the low 0.250 estimated type I error rate which occurred at the population density 0.302 fish/lO,OOO m^2. In 1980, young-of-the-year Pacific cod were only caught in Wide Bay (Tables 5 and 6). As a result, whenever the 1980 data were used to determine simulated population density in some year, which occurred with probability 2/3 for each replicate survey, all the cod in that year were concentrated in Wide Bay. When only 1 haul/year was allocated to Wide Bay, the sample variance of CPUE in Wide Bay for each year was always assumed to equal zero, when in fact the true variance was larger. This apparently was a principal cause of excessive type I errors for this case, and increasing the allocation to 2 hauls/year

Table 18.--Predicted probabilities of a type I error* with annual sample sizes of 60 hauls/year (0.0384 hauls/km^{2}1 when 2 hauls/year are allocated to the Wide Bay stratum and 8 haulslyear are Allocated to the south Sitkalidak Strait stratum, as a function of total population density of young-of-the-year Pacific cod.

Population density (no./10,000 m^2)	Type I error probability (%)
0.302	0.25
0.955	3.75
3.02	4.50
9.55	4.75
30.2	6.25

*A type I error is defined as rejecting the null hypothesis when the null hypothesis is true. The null hypothesis was that there was no change in total annual population size in the survey area. An F-test based on the method of Brown and Forsythe (1974a) with a nominal significance level of 95% was used in conjunction with a modified form of analysis of variance applied to catch per unit effort data. Each probability was estimated using 400 replicates.

made possible a more accurate variance estimate. Since the geographic distribution of fish is presumably not well known in advance of a survey, it would seem prudent to require that a minimum of at least two hauls be allocated to each stratum in each year to insure such problems do not occur. Although the possibility was not investigated in this study, requiring a minimum of two or more hauls per stratum in each year might result in improved control of type I error rates for other species or age groups as well. Improvements to Satterthwaite's approximation might also result in improved control of type I error rates.

Examination of Tables 8 and 13-16 yields the following results regarding the power of the F-test based on the method of Brown and Forsythe (1974a).

The sampling rate of 0.08 hauls/km^2 appeared sufficient to correctly detect half order of magnitude changes in young-of-the-year pollock with >66% probability, and order of magnitude changes in young-of-the-year and 1-yr-old pollock and 1-yr-old sablefish with >74% probability.

The sampling rate of 0.12 hauls/km^2 appeared sufficient to correctly detect half order of magnitude changes in young-of-the-year and 1-yr-old pollock with >62% probability, and order of magnitude changes in young-of-the-year and 1-yr-old pollock and 1-yr-old sablefish with >94% probability.

These simulations provide evidence that young-of-the-year Pacific cod were less well sampled in this survey. Ignoring the 0.04 hauls/km^2 sampling rate because of the problems with type I error rates, it appears that only 1.5 order of magnitude changes or greater were correctly detected in simulations with probability >64%, and then only for the most dense sampling rate of 0.12 haul/km^2 (Table 16).

DISCUSSION

Limitations

The ANOVA method and F-test of Brown and Forsythe (1974a) based on the Satterthwaite approximation appear valid for 2 yr of data (i.e., for J=2) as long as enough samples are taken in each ANOVA cell so that the Satterthwaite approximation is appropriate and the cell means are approximately normal; at least two or more hauls per cell appears prudent. Simulations were performed only for 2 yr of data, which meant that the ANOVA method and approximate F-test were mathematically equivalent to a Student's t-test with df estimated by Satterthwaite's approximation. For more than 2 yr of data, an improved approximation often is necessary (Bubin 1982: section 4.4).

The ANOVA method and approximate F-test used in this study are expected to be fairly robust if applied to other statistical distributions besides those tested in this study., However, some other method may be more powerful in specific cases. For instance, a maximum likelihood method based on a specific distribution may be more powerful so long as it is applied to data from that distribution or similar distributions.

No use is made in this study of correlations which may exist between the CPUE of a given species or age group and additional explanatory variables besides year and geographic region. For instance, if CPUE of one species or age group is correlated with CPUE of another species or age group, it may be possible to use this correlation to increase precision and reduce sample sizes.

The estimate of trawl net path width used to calculate the original CPUE data was apparently based on measurements made for nets towed using different size V-doors than the V-doors used in this study. This may have caused some degree of bias in population size estimates in this study and in estimates of the probability of detecting specified changes in population size.

Even if the estimate of average trawl net path width is essentially correct, between-haul variability in path width may be significant (Wathne 1977; NWAFC 1981). This also could cause bias in CPUE data, since effort calculated from average trawl width is used as a divisor. Catchability coefficients unequal to 1.0 may also cause bias when CPUE is used as an estimate of fish density.

If the CPUE values are unbiased estimators of fish per unit area in a given geographical region, then the ANOVA method used here is essentially unbiased in the following sense. Let the function "est" denote the ANOVA estimate of a given parameter, and let MCPUE denote the sample mean of CPUE. It can. be shown that

$$\mathrm{est}(a_u) + \mathrm{est}(b_{iu}) + \mathrm{est}(c_{ju}) + \mathrm{est}(d_{iju}) = MCPUE_{iju} \; . \tag{30}$$

The quantity MCPUEiju is an unbiased estimator of fish per unit area if individual values of CPUJ3 are unbiased estimators of fish per unit area. Furthermore

$$A_i \times MCPUE_{iju} \tag{31}$$

closely approximates an unbiased estimator of the total population given by Royall (1970: equation 11) as long as the total area swept by the trawl hauls in region i is small compared to Ai. Therefore

$$\sum_{i=1}^{I} A_i \times MCPUE_{iju} \tag{32}$$

$$= \sum_{i=1}^{I} A_i \times [est(a_u) + est(b_{iu}) + est(c_{ju}) + est(d_{iju})]$$

provides an essentially unbiased estimate of the total population size in the. entire survey area in year j, if individual CPUE values are unbiased. However, if the CPUE values are biased estimators of the number of fish per unit area, the ANOVA method used here does nothing to correct such biases.

The probabilities in Tables 9-16 can be considered uncertain estimates of binomial or multinomial parameters , and have associated variances which measure this uncertainty. These probabilities, especially the probabilities of a "correct detection" in Tables 13-16, are also conditional on the correctness of the mathematical models used as well as on the correctness of the maximum likelihood parameter estimates. The maximum likelihood parameter estimates undoubtedly include a certain amount of random error. The maximum likelihood method is only asymptotically unbiased in general, and therefore may have added some additional bias to the estimates, but such bias may be minor compared to the random error and biases already in the data.

In Tables 13-16 the actual population levels in both years must be specified in order to determine the probability of detecting the difference between those levels. However, it is expected that this limitation will apply to any method applied to similar data, since the variance of CPUE appears dependent on mean CPUE. In these tables a qualitative definition of a correct detection was used; namely, that the direction of change was correctly estimated. Different sampling densities may be necessary if another definition of a correct detection is used. Different sampling densities may also be necessary if the number of years is unequal to 2, the number of strata is unequal to 12, or the amount of area surveyed is changed.' For example, a survey covering a larger area may require fewer samples per unit area to achieve a given probability of correctly detecting a change in total population levels between years; and a survey covering a smaller area may require more samples per unit area- (e.g., Cochran 1977: section 4.9).

The simulations described in this study were expensive in terms of the central processor time needed on the Burroughs B7800 computer. The simulations used to generate Tables 9-16 took 3.3, 2.9, 2.0, and 0.9 h off processor time, respectively, for young-of-the-year and 1-yr-old walleye pollock, 1-yr-old sablefish, and young-of-the-year Pacific cod. For young-of-' the-year pollock about 770 of the central processor time was spent generating variates from' the negative binomial distribution.

Conclusions

Despite heteroscedasticity and non-normality, if sample sizes 'are sufficiently large, then analysis of variance with appropriate constraints on estimated interaction coefficients coupled with a modified F-test using Satterthwaite's approximation provides a valid method to measure the changes in estimated total population between 2 yr. This indicates that the heteroscedasticity of the data used in this study was a more crucial factor than non-normality in selecting a valid statistical test. At least 60 to 120 hauls/year (0.04 and 0.08 hauls/km^2) were necessary to control type I errors; a minimum of 2 hauls/year in each stratum also appeared necessary in some cases. The direction of annual population changes of young-of-the-yeal walleye pollock differing by a factor of 3.16 and direction of annual population changes of 1-yr-old walleye pollock and sablefish differing by a factor of 10 can be detected with better than, 66% probability using 120 haul/year (0.08 hauls/h^2). Smaller changes in population size can be detected and type I error rates are better controlled if sample sizes (number of trawl hauls) are increased.

It is suggested that this or related methods have wider applicability than just to the data used in this study because: 1) it is frequently of interest to combine density data from several different regions to make estimates of total population in a combined region; 2) sample variances calculated from density data frequently appear to be power functions of sample means (Taylor et al, 1978), which generally implies that the data are heteroscedastic; and 3) geographic region by year interactions may exist in population density data for other regions, species, or age groups. For example, see Pereyra et al. (1976: figs. VIII-11, VIII-21) regarding walleye pollock and Pacific cod in the Bering Sea , and Francis and Hollowed (1984) regarding Pacific whiting, *Merluccius productus,* in U.S. and Canadian waters of the Pacific Ocean.

Suggestions for Further Research

Total population estimates in these surveys were inversely proportional to the estimate of average trawl net path width. Either a normal or a beta distribution could be fitted to- the path width data given in Wathne (1977) and NWAFC (1981). A simple modification of the, present simulation model could then be used to predict the effect of between-tow variability in trawl net path width on population estimates; this may be helpful in assessing the significance of path width variability.

Improvements to Brown and Forsythe's (1974a) use of Satterthwaite's approximation could be tested for robustness to non-normality; this may be especially important for more than 2 yr of data.- Possible methods to test are the method of Rubin (1982: section 4.2(iv)) discussed in Appendix C, and the methods of Tan (1982a: 45) and Kaiser and Bowden (1983). Because both the numerator and denominator of the F statistic in Equation 22 can be expressed as quadratic forms, this F statistic can be reexpressed as a single quadratic form (e.g., Harrison and. McCabe 1979: 498). Consequently, it may be possible to evaluate its exact significance assuming the ANOVA cell means are approximately normal, perhaps by using the algorithm of Davies (1980). This method might also prove to be robust for non-normality.

The effects of different experimental designs for allocation of sample sizes (number of hauls) could be investigated with the present simulation model. An example is further investigation of the effect on type I error rates of requiring sample sizes of two or more hauls per year to-be allocated to each stratum.

A more rigorous definition of a "correct detection" could be implemented in simulations. Instead of merely considering the correct direction of change, confidence intervals could be calculated for the estimated total annual population sizes in a simulated replicate survey, perhaps by using methods based on those of Rubin (1982: section 4.2(iv)) or Raiser and Bowden (1983). A correct detection could then be defined as the inclusion of all the actual population levels used as simulation parameters within the estimated confidence intervals.

Improved models of the statistical distribution of fishing effort could be implemented.

A faster method of generating negative binomial variates may substantially speed up simulations.

Multistage sampling and- other methods (Angers 1984) could reduce the number of replicate simulated surveys needed to estimate probabilities of type I error and probabilities of a correct detection of a change in population size: this would also speed up simulations.

Sensitivity analyses using the present model or simple modifications would be useful. The effect of random errors in parameter estimates and in measurement of fishing effort could be predicted. Predicted consequences of different statistical distributions for fishing effort could be explored.

Alternative statistical methods and tests could be implemented and compared to the method in this study. The significance of the F statistic in Equation 22 could be evaluated using a bootstrap or jackknife method (Efron and Gong 1983) instead of Satterthwaite's approximation. Maximum likelihood or iteratively reweighted least squares methods (Stirling 1984) could be used to estimate parameters: statistical tests commonly used with these methods could be used in simulations instead of an F-test using Satterthwaite's approximation. Methods for estimating the regression of logarithms of sample variances of CPUE on logarithms of sample means of CPUE could be implemented which take into consideration that sample means have random variation not primarily due to measurement error (ticker 1973; Dolby 1976).

Satterthwaite's approximation does not make use of the assumption that variance of CPUE was a power function of the mean. When this assumption is valid, it may be possible to use it to improve the estimate of df in Equation 23 by substituting estimates of variance calculated using the power function relationship in place of the sample variances. It may also be possible to extend the robust weighted method of Carroll and Ruppert (1982) to take into account that the true underlying variance of CPUE may plausibly equal zero and that the sample variance of CPUE may frequently equal zero.

Either of these methods may be more powerful than the method used in this paper.

Alternative tests and methods could be compared in simulations using criteria such as power, robustness , control of type I error rates, bias, variance, mean-squared error, and computational speed. The bootstrap, maximum likelihood, and iteratively reweighted least squares methods may be computationally slower than the method used in this study, so that comparisons may need to be done using fewer geographic regions (strata) than were used in this study.

Despite non-normality, it may be possible to approximate the sample sizes needed to detect population changes without using simulation (Tan 1982a: 54).

A test of the assumption that catches have an underlying negative binomial, Poisson, or binomial distribution would be useful. Perhaps the method of Bol'shev and Mirvaliev (1978), which increases the power of a chi-square goodness-of-fit test by appropriately grouping the data, could be extended so that it is applicable despite the variability of fishing effort from haul to haul and despite possible changes in fish density between strata and years.

Further investigation of appropriate ways to quantitatively model the density dependence of geographic and statistical distributions of fish species and age groups may suggest improvements to the simulation model used in this study.

THIS PAGE INTENTIONALLY LEFT BLANK

REFERENCES

Angers, C.
1984. Large sample sizes for the estimation of multinomial frequencies from simulation studies. Simulation 43(4):175-178.

Bissell, A. F.
1972. A negative binomial model with varying element sizes. Biometrika 59(2):435-441.

Bliss, C. I., and R. A. Fisher.
1953. Fitting the negative binomial distribution to biological data. Biometrics 9(2):176-200.

Bloch, F. E.
1978. Measurement error and statistical significance of an independent variable. Am. Stat. 32(1):26-27.

Bol'shev, L. N., and M. Mirvaliev.
1978. Chi square goodness-of-fit test for the Poisson, binomial and negative binomial distributions. Theory Probab. Appl. 23(3):461-474.

Box, G. E. P.
1954. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. Ann. Math. Stat. 25(2):290-302.

Bratley, P., B. L. Fox, and L. E. Schrage.
1983. A guide to simulation. Springer-Verlag, New York, 383 p.

Brown, M. B., and A. B. Forsythe.
1974a. The ANOVA and multiple comparisons for data with heterogeneous variances. Biometrics 30(4):719-724.

1974b. The small sample behavior of some statistics which test the equality of several means. Technometrics 16(1):129-132.

Bury, K. V.
1975. Statistical models in applied science. John Wiley & Sons, New York, 625 p.

Carroll, R. J., and D. Ruppert.
1982. A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. J. Am. Stat. Assoc. 77(380):878-882.

Cochran, W. G.
1977. Sampling techniques. John Wiley & Sons, New York, 428 p.

Preceding Page Blank

Davies, R. B.
1980. Algorithm AS 155. The distribution of a linear combination of chi-square random variables. Appl. Stat. 29(3):323-333.

Dixon, W. J. (chief editor).
1983. BMDP statistical software. Univ. Calif. Press, Berkeley, 735 p.

Dolby, G. R.,
1976. The ultrastructural relation: a synthesis of the functional and structural relations. Biometrika 63(1):39-50.

Efron, B., and G. Gong.
1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am. Stat. 37(1):36-48.

Elliott, J. M.
1977. Some methods for the statistical analysis of samples of benthic invertebrates. Sci. Publ. 25. Freshwater Biol. Assoc., Ferry House, U.K., 156 p.

Fishman, G. S., and L. R. Moore.
1982. A statistical evaluation of multiplicative congruential random number generators with modulus $2^{31} - 1$. J. Am. Stat. Assoc. 77(377):129-136.

Francis, R. C., and A. Hollowed.
1984. Status of the Pacific hake resource and recommendations for management in 1985. In Status of Pacific Coast groundfish fishery and recommendations for management in 1985, Appendix 2. Unpubl. rep., 17 p. Pac. Fish. Manage. Count., 526 S.W. Mill St., Portland, Ore. 97201.

Green, R. H.
1979. Sampling design and statistical methods for environmental biologists. John Wiley & Sons, New York, 257 p.

Harris, R. J.
1975. A primer of multivariate statistics. Academic Press, New York, 332 pa

Harrison, M.J., and B. P. M. McCabe.
1979. A test for heteroscedasticity based on ordinary least squares residuals. J. Am. Stat. Assoc. 74(366):494-499.

Hughes, S. E.
1976. System for sampling large trawl catches of research vessels. J. Fish. Res. Board Can. 33(4, Part 1):833-839.

Iman, R. L., and J. M. Davenport.
1976. New approximations to the exact distribution of the Kruskal-Wallis test statistic. Commun. Stat. Theory Methods 5(14):1335-1348.

IMSL (International Mathematical & Statistical Libraries).
1982. IMSL library reference manual, edition 9. IMSL, Inc., 2500 ParkWest Tower One, 2500 CityWest Blvd., Houston, TX 77042.

Kaiser, L. D., and D. C. Bowden.
1983. Simultaneous confidence intervals for all linear contrasts of means with heterogenous variances. Commun. Stat. Theory Methods 12(1):73-88.

Kappenman, R. F.
1985. Estimation for the three-parameter Weibull, lognormal, and gamma distributions. Comput. Stat. Data Anal. 3(1):11-23.

Levin, B., and J. Reeds.
1977. Compound multinomial likelihood functions are unimodal: proof of a conjecture of I. J. Good. Ann. Stat. 5(1):79-87.

Maindonald, J. H.
1984. Statistical computation. John Wiley & Sons, New York, 370 p.

Matuszewski, A., and D. Sotres.
1985. A basic statistical problem: confidence interval for the Bernoulli parameter. Comput. Stat. Data Anal. 3(2):103-114.

Mood, A. M., F. A. Graybill, and D. C. Boes.
1974. Introduction to the theory of statistics. McGraw-Hill, New York, 564 p.

NWAFC (Northwest and Alaska Fisheries Center).
[1981.] Results of gear observations on the NOAAS Chapman 1981 Gulf of Alaska shrimp survey, cruise 81-4. Unpubl. manuscr., 7 p. Northwest and Alaska Fish. Cent., Natl. Mar. Fish. Serv., 7600 Sand Point Way N.E., Bin C15700, Seattle, WA 98115.

Pereyra, W. T., J. E. Reeves, and R. G. Bakkala.
1976. Demersal fish and shellfish resources of the eastern Bering Sea in the baseline year 1975. Processed Rep., 619 p. Northwest Fish. Cent., Natl. Mar. Fish. Serv., NOAA, 2725 Montlake Blvd. E., Seattle, WA 98112.

Perry, J. N.
1981. Taylor's power law for dependence of variance on mean in animal populations. Appl. Stat. 30(3):254-263.

Rao, C. R.
1973. Linear statistical inference and its applications. -John Wiley & Sons, New York, 625 p.

Ricker, W. E.
1973. Linear regressions in fishery research. J. Fish. Res. Board Can. 30(3):409-434.

Royall, R. M.
1970. On finite population sampling,-theory under certain linear regression models. Biometrika 57(2):377-387.

Rubin, A. S.
1982. The use of weighted contrasts in the analysis of models with heterogeneity of variance. Ph.D. Thesis, Univ. Rochester, Rochester, New York, 226 p.

1983. The use of weighted contrasts in analysis of models with heterogeneity of variance. In American Statistical Association. 1983. Proceedings of the business and economic statistics section, p. 347-352. Am. Stat. Assoc., 806 15th St. N.W., Washington, D.C. 20005.

Rubinstein, R. Y.
1981. Simulation and the Monte Carlo method. John Wiley & Sons, New York, 278 p.

Satterthwaite, F. E.
1946. An approximate distribution of estimates of variance components. Biometrics 2(6):110-114.

Scheffe, H.
1959. The analysis of variance. John Wiley & Sons, New York, 477 p.

Searle, S. R.
1971. Linear models. John Wiley & Sons, New York, 532 p.

1982. Matrix algebra useful for statistics. John Wiley & Sons, New York, 438 p.

1983. General linear model. In S. Kqtz, N. L. Johnson, and C. B. Read (editors), Encyclopedia of statistical sciences, Volume 3., p. 357-372. John Wiley & Sons, New York.

Smith, G. B., and R. G. Bakkala.
1982. Demersal fish resources of the eastern Bering Sea: spring 1976. U.S. Dep. Commer., NOAA Tech. Rep. NMFS SSRF-754, 129 p.

Smith, G. B., G. E. Walters, P. A. Raymore, Jr., and W. A. Hirschberger.
1984. Studies of the distribution and abundance of juvenile groundfish in the northwestern Gulf of Alaska, 1980-82: part I, three-year comparisons. U.S. Dep. Commer., NOAA Tech. Memo. NMFS F/NWC-59, 100 p. (Available from U.S. Dep. Commer., Natl. Tech. Inf. Serb, Springfield, VA 22161.)

Snedecor, G. W., and W. G. Cochran.
1980. Statistical methods. Iowa State Univ. Press, Ames, Iowa, 507 p.

Sokal, R. R., and F. J. Rohlf.
1969. Biometry. W. H. Freeman, San Francisco, Calif., 776 p.

Stirling, W. D.
1984. Iteratively reweighted least squares for models with a linear part. Appl. Stat. 33(1):7-17.

Tamhane, A. C.
1979. A comparison of procedures for multiple comparisons of means with unequal variances. J. Am. Stat. Assoc. 74(366):471-480.

Tan, W. Y.
1982a. On approximating probability distributions of a statistic of Brown-Forsythe from normal and non-normal universes. J. Stat. Comput. Simulation 16(1):35-55.

1982b. On comparing several straight lines under heteroscedasticity and robustness with respect to departure from normality. Commun. Stat. Theory Methods 11(7):731-750.

Taylor, C. C.
1953. Nature of variability in trawl catches. U.S. Fish. Wildl. Serv., Fish. Bull. 54:145-166.

Taylor, L. R., I. P. Woiwod, and J. N. Perry.
1978. The density-dependence of spatial behaviour and the rarity of randomness. J. Anim. Ecol. 47(2):383-406.

Walters, G. E., G. B. Smith, P. A. Raymore, Jr., and W. Hirschberger.
1985. Studies of the distribution and abundance of juvenile groundfish in the northwestern Gulf of Alaska, 1980-82: part II, biological characteristics in the extended region. U.S. Dep. Commer., NOAA Tech. Memo. NMFS F/NWC-77, 95 p. (Available from U.S. Dep. Commer., Natl. Tech. Inf. Serv., Springfield, VA 22161.)

Wathne, F.
1977. Performance of trawls used in resource assessment. U.S. Natl. Mar. Fish. Serv., Mar. Fish. Rev. 39(6):16-23.

Zweifel, J. R., and P. E. Smith.
1981. Estimates of abundance and mortality of larval anchovies (1951-75): application of a new method. Rapp. P.-V. Reun. Cons. Int. Explor. Mer 178:248-259.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A

UNCONTROLLABLE TYPE I ERRORS DUE TO A NONLINEAR TRANSFORMATION

Table A-l illustrates the uncontrollable type I errors which can result from the use of the popular nonlinear variance-stabilizing transformation log(CPUE+l), especially when interaction terms are present; Use of this transformation is reviewed in Green (1979: section 2.3.9). Although not reported here, excessive type I error rates of similar magnitude occurred when interaction terms were present and a rank or power transformation was used in place of log(CPUE+l).

The null hypothesis (Equation 161, parameters, and methods used to generate Table 10 were also used to generate Table A-l, except the values loglO(CPUE+l) were substituted for the CPUE values in the ANOVA calculations, an ordinary F-test was used, and Equation 20 was changed so that

$$0 = \sum_{i=1}^{I} d_{iqu} \qquad (33)$$

for each q. In the case without the presence of geographic region by year interactions, the values of MESTiju for 1-yr-old walleye pollock calculated from 1980 data (Table 6) were substituted for the values of MESTiju calculated from 1981-82 data.

For the case with interactions, the actual counts of type I errors used to calculate the probabilities in Table A-l were significantly different from the expected 5% error rate at greater than the 99.99% level (G=4449, df-15). Although the counts of type I errors for the case without interactions also were significantly different from the expected 5% error rate at greater than

Preceding Page Blank

Table A-l .--Predicted probabilities of a type I error* for an ordinary F-test applied to catch per unit effort (CPUE) values of 1-yr-old walleye pollock transformed using the function log(CPUE+l), as a function of the presence or absence of geographic region by year interactions, the total population density of 1-yr-old walleye pollock, and annual sample size.

Population density (no./10,000 m^2)	Sample size (hauls/year)	Type I error probability (%)	
		With interactions	Without interactions
25.6	60	14.75	20.50
	120	16.25	17.00
	180	15.00	15.00
81.0	60	12.50	10.25
	120	13.75	5.50
	180	14.75	7.00
256	60	15.75	4.50
	120	17.75	6.00
	180	21.75	6.25
810	60	21.50	2.00
	120	33.25	2.00
	180	46.75	2.25
2560	60	38.75	1.50
	120	64.25	0.50
	180	75.25	1.25

*A type I error is defined as rejecting the null hypothesis when the null: hypothesis is true. The null hypothesis was that there was no change in total annual population size in the survey area. An F-test with a nominal significance level of 95% was used in conjunction with ordinary two-way analysis of variance. Each probability was estimated using 400 replicates.

the 99.99% level (G=360.4, df=l5), the smaller G statistic indicates that the departures from the expected rate were less drastic. This shows that the presence of two-way interactions were one major cause of uncontrollable type I errors when this nonlinear transformation was used. Nonetheless, the transformation log(CPUE+l) is not recommended for this data, because of the problems with incorrect type I error rates even when interactions were not present.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B

COMPUTATIONAL FORMULAS FOR THE METHOD OF BROWN AND FORSYTHE

Brown and Forsythe (1974a) presented their method using orthonormal contrast coefficients. However, in this Appendix it is shown that it is not necessary to actually calculate the orthonormal contrast coefficients. It is assumed that the null hypothesis being tested is that no changes in total population size occurred between years (Equations 16 and 19). The formulas derived, however, can be easily generalized to test other null hypotheses: for instance, 'that there are no changes in total population size between geographic regions, or that there are no geographic region by year interactions in total population size. The formulas derived can also be used in the method of Rubin (1982: section 4.2(iv)) described in Appendix C.

Let n be the total number of samples i.e., in this study

$$n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \quad . \tag{34}$$

The general linear model can be expressed in the form (Searle 1971: 164)

$$Y = X b + e , \tag{35}$$

where Y is an n by 1 matrix of observed values of the dependent variable, X is an n by p dimensional design matrix, b is a p by 1 dimensional matrix of parameters, and e is an n by 1 dimensional matrix of random "errors." Let r equal the rank of X; it is assumed that r < p. Letb be the least squares estimate of b. In this study b is the vector of parameters in Equation 17; i.e.,

Preceding Page Blank

$$b' = (a_u, b_{1u}, \ldots, b_{Iu}, c_{1u}, \ldots, c_{Ju}, d_{11u}, d_{21u}, \ldots, d_{IJu}) \tag{36}$$

and

$$p = I \times J + I + J + 1 = (I+1) \times (J+1) . \tag{37}$$

Let K be a matrix such that the elements of K'b are a linearly independent subset of the elements of b. For example, if

$$K = (E_{I+3}, E_{I+4}, \ldots, E_{I+J+1}) , \tag{38}$$

where E_i is a p by 1 dimensional vector, with one in its i^{th} entry and zeros elsewhere, then

$$K'b = (c_{2u}, \ldots, c_{Ju})' \tag{39}$$

is a vector of linearly independent elements of b. In addition, K'b is estimable because of the constraints expressed by Equations 18, 20, and 21 (Scheffe 1959: section 1.4, especially theorem 4). The constraints expressed by Equations 18, 20, and 21 can be reexpressed using a matrix equation of the form

$$P'b = (0 .m. 0)' , \tag{40}$$

where p is a p by (I+J+2) dimensional matrix of appropriate constants. A p by (p-r) dimensional matrix P is formed by choosing p-r linearly independent columns of P in such a manner (Searle 1971: 21-22) that all rows-of the matrix

$$S = \begin{bmatrix} X'X & P \\ P' & 0 \end{bmatrix} \tag{41}$$

are linearly independent (a convenient computational formula for X'X will be given in Equation 48). Matrix P satisfies an equation similar to Equation 40; i.e.,

$$P'b = (0 \ . \ . \ . \ 0)' \ . \tag{42}$$

The matrix P'b is not estimable (Searle 1971: section 5.7a). If I=4 and J=3, then an example of a P' which could be used in this study is

$$P' = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_1 & A_2 & A_3 & A_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_1 & A_2 & A_3 & A_4 & 0 & 0 & 0 & 0 \end{bmatrix} . \tag{43}$$

In this Appendix, let G be a p by p dimensional matrix equal to the upper left submatrix of S-l ; matrix G is a particular generalized inverse of X'X (Searle 1971: section 1.5). Since S is symmetric, so is S-l, which implies that G is' symmetric and G = G'. It is suggested that the method used to calculate S^{-1} include a test of whether S is algorithmically singular, as did the IMSL (1982) subroutine LEQlS used in this study. If the test indicates S is algorithmically singular, then the rows of S are effectively linearly dependent, which may indicate that the matrix P was incorrectly chosen. If a suitable matrix P does not exist which causes the rows of S to be linearly independent, then the method of Searle (1971: section 5.6a) may be appropriate, but such a problem did not arise in this study and this alternative was not investigated.

Once G has been calculated, the least squares estimate of b satisfying the constraints imposed by Equations 18, 20, and 21 is given by (Searle 1971:

table 5.13 and section 5.7a)

$$\underline{b} = GX'Y \ . \tag{44}$$

The null hypothesis can be expressed in the form

$$K'b = (c_{2u},\ldots,c_{Ju})' = (0 \ \ldots \ 0)'. \tag{45}$$

If Equation 45 is true, then Equation 18 implies also that $c1_u = 0$. One possible numerator sum of squares for an F-test of the null hypothesis is (Searle 1971: 192)

$$Q = Y'XG'K(K'GK)^{-1}K'GX'Y \ . \tag{46}$$

For ordinary analysis of variance, it can be proven that there exists an (I x J) by p dimensional matrix Z such that

$$m'NZ = Y'X \ , \tag{47}$$

where m is the (I x J) by 1 dimensional vector of the sample means in each ANOVA cell, and N is the (I x J) by (I x J) dimensional matrix with the nij along its diagonal and zeros elsewhere. The matrix Z consists of all the unique rows of X; i.e., Z can be determined from X by eliminating all rows of X which are duplicates of any other row. When calculating S using Equation 41, it is convenient to note that

$$X'X = Z'NZ \ . \tag{48}$$

If I = 4 and J = 3, then an example of Z of the type used in this study is

$$Z = \begin{bmatrix}
1&1&0&0&0&1&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
1&0&1&0&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
1&0&0&1&0&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0\\
1&0&0&0&1&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
1&1&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0\\
1&0&1&0&0&0&1&0&0&0&0&0&0&1&0&0&0&0&0&0\\
1&0&0&1&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&0\\
1&0&0&0&1&0&1&0&0&0&0&0&0&0&0&1&0&0&0&0\\
1&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&1&0&0&0\\
1&0&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0&1&0&0\\
1&0&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&1&0\\
1&0&0&0&1&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1
\end{bmatrix} \quad (49)$$

Equations 46 and 47 imply that

$$Q = m'NZG'K(K'GK)^{-1}K'GZ'N'm \ . \quad (50)$$

If (K'GK)" is positive definite, which was the case in this study, then there exists a nonsingular square matrix R (Searle 1982: 206). such that

$$RR' = (K'GK)^{-1} \ . \quad (51)$$

It can be proved that

$$(R')^{-1}R^{-1} = K'GK. \quad (52)$$

Let

$$T = NZG'KR \ . \quad (53)$$

Then Equations 51 and 53 imply that

$$TT' = NZG'K(K'GK)\text{-'}K'GZ'N' \ . \quad (54)$$

Equation 54 is a convenient computational formula for TT', because it is unnecessary to calculate T and only the operations of matrix multiplication, transposition, and inversion are used. Although Brown and Forsythe (1974a: section 4) used a different notation, in effect they required that the eknents of T satisfy the conditions for orthonormal contrast coefficients; T does satisfy these conditions, as is proved below.

First, it is proved that the elements of T-satisfy the conditions for contrast coefficients. In the remainder of this paragraph, it is assumed that m is an (I x J) by 1 dimensional vector with each element equal to one and that Y is an n by 1 dimensional vector with each element equal to one. The elements of T satisfy the condition for contrast coefficients if

$$m'T = (0 \ . \ . \ . \ 0) \ . \tag{55}$$

Equations 47 and 53 imply

$$m'T = m'NZG'KR = Y'XG'KR = (GX'Y)'KR \ . \tag{56}$$

From Searle (1971: 80, equation 20) , it is known that the least squares solution b corresponding to Y must satisfy the equation

$$X'Xb = X'Y \ . \tag{57}$$

Now Equations 17, 35, and 36 imply that each element of the first column of X is equal to one, which means that

$$XE1 = Y \ , \tag{58}$$

which in turn implies that

$$X'XE1 = X'Y \ ; \tag{59}$$

l.e., that E_l satisfies the condition for 2 given by Equation 57. In addition, setting b = El satisfies the constraints imposed by Equations 18, 20, and 21. This implies that

$$b = E_l \tag{60}$$

is the unique solution satisfying these constraints (Scheffe 1959: section 1.4). Consequently, Equation 44 implies that

$$E_1 = GX'Y \ . \tag{61}$$

Equation 38 implies that

$$(Ej)^*K = (0 \ . \ . \ . \ 0) \ . \tag{62}$$

Then combining Equations 56, 61, and 62 we have

$$m'T = (GX'Y)'KR = (E_1)'KR = (0 \ ... \ 0)R = (0 \ ... \ 0), \tag{63}$$

which shows that the condition for contrast coefficients given by Equation 55 is satisfied.

It is next proved that the elements of T satisfy the condition for orthonormality; i.e., that

$$T'N^{-1}T = I \ , \tag{64}$$

where, in this paragraph, I represents the identity matrix of appropriate dimension. It is known that the G used in this study satisfies the second Penrose condition (Searle 1971: 2317 i.e.,

$$GX'XG = G \ . \tag{65}$$

Making use of the fact that this G = G' and that N = N', and using Equations 48, 52, 53, and 65, we have

$$T'N^{-1}T = (R'K'GZ'N')N^{-1}(NZG'KR) = R'K'GZ'(N'I)ZG'KR \tag{66}$$

$$= R'K'GZ'(N)Z(G)KR = R'K'G(Z'NZ)GKR$$

$$= R'K'G(X'X)GKR = R'K'(GX'XG)KR = R'K'(G)KR$$

$$= R'(K'GK)R = R'(R')^{-1}R^{-1}R = II = I ,$$

which shows that the condition for orthonormality is satisfied.

APPENDIX C

COMPUTATIONAL FORMULAS FOR THE METHOD OF RUBIN

Equations 22, 23, and 54 can be used in the method of Rubin (1982: section 4.2(iv)) and Rubin (1983: equation 5). However, instead of assuming that the df of the numerator of the F statistic in Equation 22 equals J-l as in the method of Brown and Forsythe (1974a), the df of the numerator is assumed to equal

$$DFN = \frac{[\mathrm{trace}(TT'v)1^2}{\mathrm{trace}(\ [TT'I;')} \qquad (67)$$

$$= \frac{\left(\sum_{i=1}^{I} \sum_{j=1}^{J} f_{ij} \times [s_{iju}]^2/n_{ij} \right)^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} [R_{ij}] \times [C_{ij}]},$$

where Rij equals row [I x (j-l) + i] of TT'V, and Cij equals column [I x (j-1) + i] of TT'V. If J = 2, then it can be proved that DFN = 1 (Rubin 1982: section 4.2, equation 24). Since in general DFN is not an integer, it is appropriate to use IMSL (1982) subroutine MDFDRE to evaluate the significance of the F statistic in Equation 22. If the denominator of Equation 67 is < 0, then DFN can be set equal to J-l.