Special Publication No. 98-1

The Mechanics of Onsite Creel Surveys in Alaska

by

David R. Bernard,

Allen E. Bingham,

and

Marianna Alexandersdottir

February 1998

Alaska Department of Fish and Game

Division of Sport Fish

Symbols and Abbreviations

The following symbols and abbreviations, and others approved for the Système International d'Unités (SI), are used in Division of Sport Fish Fishery Manuscripts, Fishery Data Series Reports, Fishery Management Reports, and Special Publications without definition. All others must be defined in the text at first mention, as well as in the titles or footnotes of tables and in figures or figure captions.

Weights and measures (metric)

centimeter	cm
deciliter	dL
gram	g
hectare	ha
kilogram	kg
kilometer	km
liter	L
meter	m
metric ton	mt
milliliter	ml
millimeter	mm

Weights and measures (English)

cubic feet per second	ft^3/s
foot	ft
gallon	gal
inch	in
mile	mi
ounce	oz
pound	lb
quart	qt
yard	yd

Spell out acre and ton.

Time and temperature

day	d
degrees Celsius	°C
degrees Fahrenheit	°F
hour (spell out for 24-hour clock)	h
minute	min
second	s

Spell out year, month, and week.

Physics and chemistry

all atomic symbols	
alternating current	AC
ampere	A
calorie	cal
direct current	DC
hertz	Hz
horsepower	hp
hydrogen ion activity	pH
parts per million	ppm
parts per thousand	ppt, ‰
volts	V
watts	W

General

All commonly accepted abbreviations.	e.g., Mr., Mrs., a.m., p.m., etc.
All commonly accepted professional titles.	e.g., Dr., Ph.D., R.N., etc.
and	&
at	@
Compass directions:	
east	E
north	N
south	S
west	W
Copyright	©
Corporate suffixes:	
Company	Co.
Corporation	Corp.
Incorporated	Inc.
Limited	Ltd.
et alii (and other people)	et al.
et cetera (and so forth)	etc.
exempli gratia (for example)	e.g.,
id est (that is)	i.e.,
latitude or longitude	lat. or long.
monetary symbols (U.S.)	$, ¢
months (tables and figures): first three letters	Jan,...,Dec
number (before a number)	# (e.g., #10)
pounds (after a number)	# (e.g., 10#)
registered trademark	®
trademark	™
United States (adjective)	U.S.
United States of America (noun)	USA
U.S. state and District of Columbia abbreviations	use two-letter abbreviations (e.g., AK, DC)

Mathematics, statistics, fisheries

alternate hypothesis	H_A
base of natural logarithm	e
catch per unit effort	CPUE
coefficient of variation	CV
common test statistics	F, t, χ^2, etc.
confidence interval	C.I.
correlation coefficient	R (multiple)
correlation coefficient	r (simple)
covariance	cov
degree (angular or temperature)	°
degrees of freedom	df
divided by	÷ or / (in equations)
equals	=
expected value	E
fork length	FL
greater than	>
greater than or equal to	≥
harvest per unit effort	HPUE
less than	<
less than or equal to	≤
logarithm (natural)	ln
logarithm (base 10)	log
logarithm (specify base)	log_2, etc.
mideye-to-fork	MEF
minute (angular)	'
multiplied by	x
not significant	NS
null hypothesis	H_O
percent	%
probability	P
probability of a type I error (rejection of the null hypothesis when true)	α
probability of a type II error (acceptance of the null hypothesis when false)	β
second (angular)	"
standard deviation	SD
standard error	SE
standard length	SL
total length	TL
variance	Var

SPECIAL PUBLICATION NO. 98-1

THE MECHANICS OF ONSITE CREEL SURVEYS IN ALASKA

by

David R. Bernard
Allen E. Bingham,
and
Marianna Alexandersdottir
Division of Sport Fish, Anchorage

Alaska Department of Fish and Game
Division of Sport Fish, Research and Technical Services
333 Raspberry Road, Anchorage, Alaska 99518-1599

February 1998

This investigation was partially financed by the Federal Aid in Sport Fish Restoration Act (16 U.S.C. 777-777K) under project F-10-13, Job No. RT-4

The Special Publication series was established in 1991 for the publication of techniques and procedures manuals, informational pamphlets, special subject reports to decision-making bodies, symposia and workshop proceedings, application software documentation, in-house lectures, and other documents that do not fit in another publication series of the Division of Sport Fish. Special Publications are intended for fishery and other technical professionals and lay persons. Distribution is to state and local publication distribution centers, libraries and individuals and, on request, to other libraries, agencies, and individuals. This publication has undergone editorial and peer review.

David R. Bernard, and Allen E. Bingham
Alaska Department of Fish and Game, Division of Sport Fish, Research and Technical Services
333 Raspberry Road, Anchorage, AK 99518-1599, USA
Marianna Alexandersdottir
Northwest Indian Fisheries Commission, 6730 Martin Way East, Olympia, WA 98516-5540, USA

This document should be cited as:
Bernard, D.R., A.E. Bingham, and M. Alexandersdottir. 1998. The mechanics of onsite creel surveys in Alaska. Alaska Department of Fish and Game, Special Publication No. 98-1, Anchorage.

TABLE OF CONTENTS

Page

PREFACE v
INTRODUCTION 1
CHAPTERS 7

1. Sampling Designs 7

    1.1 Sampling Periods 7
    1.2 Stratification 8

        1.2.1 Full Stratification 9
        1.2.2 Partial Stratification 9
        1.2.3 Post-Stratification 9

    1.3 Sampling Stages 10
    1.4 Scheduling for Stratified Multistage Designs 13

        1.4.1 Order of Selection 13
        1.4.2 Random Selection 13
        1.4.3 Systematic Selection 14
        1.4.4 What if it doesn't fit? 14

    1.5 Specific Types of Onsite Surveys 15
    1.6 Remaining Potential for Bias 16

2. Harvest, Catch, and Effort 18

    2.1 Access-Point Surveys 18

        2.1.1 Harvest, Catch, and Effort 18
        2.1.2 Estimating Variance 22
        2.1.3 Missing Data 29

    2.2 Roving Surveys 30

        2.2.1 Fishing Effort 31
        2.2.2 Mean Harvest Rate 31

            2.2.2.1 Avoiding LOS Bias 32
            2.2.2.2 Avoiding Opportunity Bias 35

        2.2.3 Harvest, Catch, Effort, and Angler-Trips 35
        2.2.4 Estimating Variance 36
        2.2.5 Post-stratification and Covariance 39
        2.2.6 Missing Data 42

    2.3 References for Examples 47

TABLE OF CONTENTS (Continued)

Page

3. Harvest Composition 48

3.1 Simultaneous Estimation 49

3.1.1 All Inspected Harvest Sampled 49
3.1.2 Inspected Harvest Subsampled 49

3.2 Separate Estimation 53
3.3 Incompletely Marked Cohorts 58

3.3.1 Stratified "Random" Sampling 58
3.3.2 Self-Weighted Sampling 61

3.4 Opportunity Bias 63

3.5 References for Examples 65

4. Angler Success 66

4.1 The Basics 66
4.2 Self-Weighted Designs 68
4.3 Stratified "Random" Designs 71
4.4 Avoiding Bias 73

4.4.1 LOS Bias 73
4.4.2 Opportunity Bias 73

4.5 Reference for Examples 74

5. CPUE as an Index of Abundance 75

5.1 The CPUE Survey 75

5.1.1 Planning and Calculation 75
5.1.2 Graphics and Confidence Intervals 77

5.2 Avoiding Bias 79

5.2.1 LOS Bias 79
5.2.2 Opportunity Bias 80

5.3 Reference for Examples 80

TABLE OF CONTENTS (Continued)

Page

6. One-Technician Surveys 81

6.1 Systematic Scheduling........ 81

6.1.1 Access-point Surveys 81
6.1.2 Other Surveys........ 82

6.2 Random Scheduling 84

6.2.1 The Basic Schedule 84
6.2.2 Access-point Surveys 85
6.2.3 Roving Surveys 85

6.3 Comparison, Compromise, and Bias 86

7. Counts, Interviews, and Creels........ 87

7.1 Counting Anglers 87

7.1.1 Identifying Anglers 87
7.1.2 Duration of Counts........ 88
7.1.3 Geography of Counts 88

7.2 Interviewing Anglers........ 88

7.2.1 Avoiding Bias........ 88
7.2.2 Fishing Effort 89
7.2.3 Catch and Harvest 90

7.3 Sampling the Creel 90

7.4 Avoiding Errors........ 90

7.5 References for Examples........ 91

LITERATURE CITED........ 92

APPENDICES........ 95

A: Estimating Mean Length of an Angler-Trip........ 97

B: Selecting Sampling Units in Proportion to their Size 101

TABLE OF CONTENTS (Continued)

Page

C: Variance Equations for Totals: Unequal Sampling Fractions 105

D: Jackknife Estimator for Mean Harvest Rate in Roving Surveys 107

E: Covariance among Post-Stratified Estimates of Harvest in Roving Surveys 111

F: Roving Surveys with Missing Data on Harvest Rates 115

G: Relationships among Equations to Estimate Relative Harvest Composition 119

H: Source Code for the QuickBasic™ Program "DELTA.BAS" 121

GLOSSARY 123

PREFACE

This report was promised as a logical conclusion to a short course about onsite creel surveys given in Anchorage in 1990. Eight years have past (flown by) since then; many of the ideas presented in 1990 have been updated, and some have been rejected. More importantly, many more of the ideas presented during the course and many developed since have been put into practice. The result is a paradigm of how to plan and conduct onsite creel surveys in Alaska.

Methods described in this report are based on standard statistical procedures for stratified, multistage, and self-weighted sampling with new procedures developed specifically to address the peculiarities of onsite creel surveys. Emphasis is on estimating totals: harvest or catch of fish, fishing effort by anglers, or composition of harvest. Access-point and roving surveys are described along with surveys to estimate harvest composition and angler success. How to avoid bias in statistics and what to do if data are missing figure prominently in this report. There are many examples of surveys on fisheries for migratory and non-migratory species taken from the year-to-year operations of our research program.

Our original intent was to include a detailed discussion of planning onsite creel surveys in this report. Scattered throughout the report are details and suggestions on how to design creel surveys to avoid bias and simplify scheduling of sampling. Earlier drafts of the report also contained a chapter (a large chapter) on how to design surveys to minimize variance through optimal definition of strata and optimal allocation of sampling effort. Although these planning tools are currently used in Alaska, we decided that these methods needed more statistical development, so that chapter was dropped from this report. A second report is anticipated that will cover these planning methods in detail.

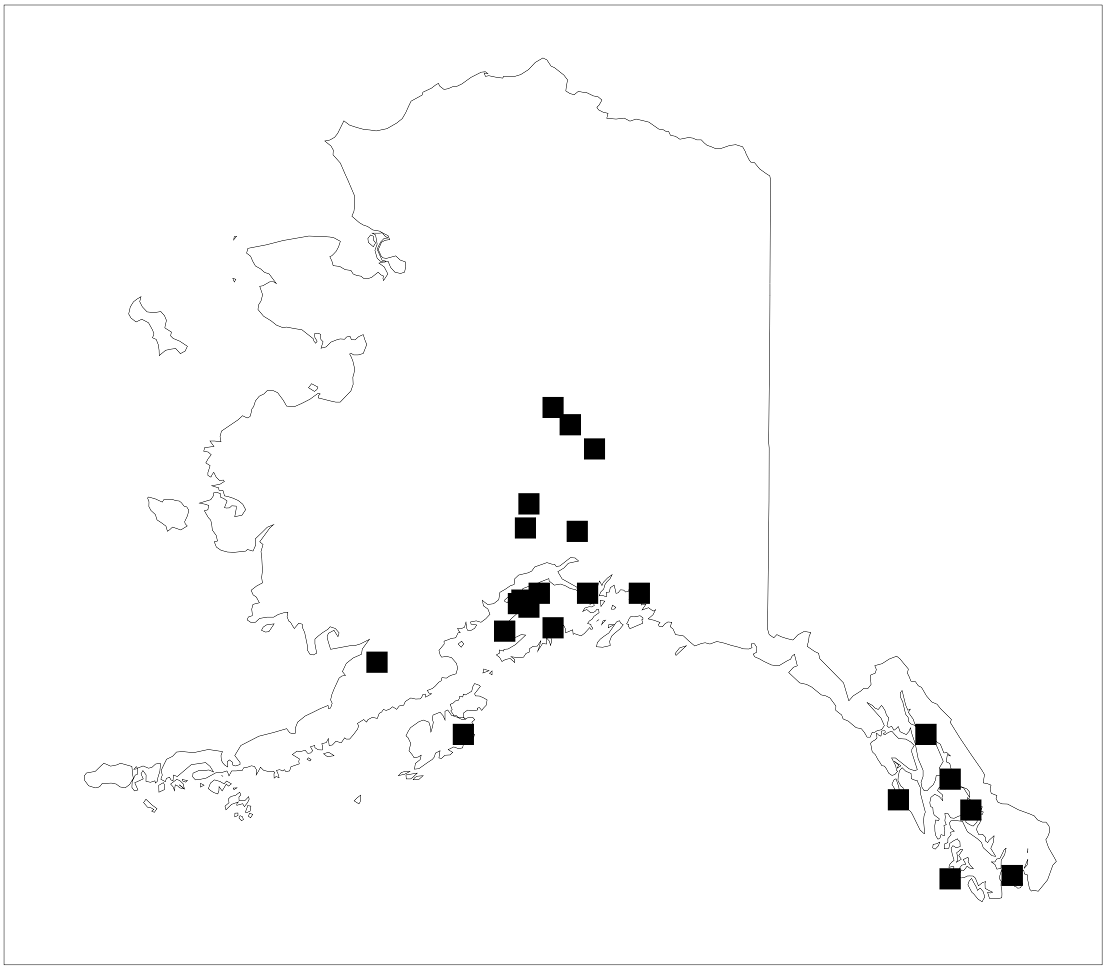
Another chapter covering methods of estimating opinions and demographics of anglers with onsite surveys was included in earlier drafts, but was dropped for this report. Our basic conclusion on estimating opinions and demographics in this way is that it is a bad idea because of avidity bias. Since that's the same conclusion drawn in Pollock et al. (1994:159, 185-6), we saw no further need to burden our readers by repeating this message.

Efforts by many people have made this work what it is. Many biologists over the years have given us impetus, insight, and inspiration, without which this document would not be possible. We specifically thank Sandy Sonnichsen, Bob Marshall, Jim Hasbrouck, Keith Webster, Paul Suchanek, and Steve Fleischman who over the years implemented and tested the designs in this report. We are also grateful to Sandy for her editing and formatting of the data used in examples. Sandy and Bob Marshall reviewed earlier drafts of this report. We are grateful to Dave Bowden of Colorado State University for his suggestion that roving creel surveys, like access-point surveys, should be designed around sampling periods. We also wish to acknowledge the Federal Aid to Sport Fish Restoration Act and the anglers of Alaska for financially supporting this report and for supporting the creel surveys upon which it is based. And finally, we are grateful to Bob Conrad who pioneered our efforts to understand and better implement onsite creel surveys in Alaska.

Dave Bernard
Allen Bingham
Anchorage

Marianna Alexandersdottir
Olympia

This work is dedicated to the memory of Keith A. Webster, our friend and colleague.



Locations of onsite access-point and roving-access creel surveys in Alaska to estimate harvest, fishing effort, catch, angler success, and species, sex, age, and cohort composition of harvest in 1992. Surveyed fisheries were located in streams, lakes, and on the ocean. Populations of whitefish *Coregonus* sp., salmonids *Oncorhynchus* sp., rockfish *Sebastes* sp., Pacific halibut *Hippoglossus stenolepis*, and lingcod *Ophiodon elongatus* were exploited in these fisheries.

INTRODUCTION

The Alaska Department of Fish and Game spent in excess of $1.5 million in fiscal year 1997 to survey, and sometimes census, recreational fisheries. On the Kenai River, an onsite survey was used as part of a management plan to meet escapement goals for chinook salmon *Oncorhynchus tshawytscha*. At Resurrection Bay, an onsite survey was used to estimate relative species composition of harvested rockfish *Sebastes* sp. Onsite creel surveys based in Sitka, Ketchikan, Petersburg, Wrangell, and Juneau were used to estimate marine harvests of chinook salmon to attain allocations set by the Alaska Board of Fisheries. Some of our onsite creel surveys, such as those on the Kenai River, have many technicians (10 on the Kenai River in 1992); others such as those near the towns of Haines, Yakutat, and Sitka, and those on the Delta Clearwater River and Piledriver Slough have had but one technician each. Our onsite surveys have covered recreational fisheries on migrating sockeye salmon *O. nerka* in the Russian River, feeding coho salmon *O. kisutch* in the ocean near Ketchikan, spawning whitefish *Coregonus* sp. harvested in the Chatanika River, Arctic grayling *Thymallus arcticus* in rivers such as the Delta Clearwater, stocked fish such as rainbow trout *O. mykiss* in Piledriver Slough, groundfish such as lingcod *Ophiodon elongatus* from Resurrection Bay, and Dolly Varden *Salvelinus malma* of Prince William Sound following the grounding of the *Exxon Valdez*.

Information from surveying or censusing a fishery is distilled into a battery of statistics, the pattern of which is used to regulate the fishery in the current or in future years. These statistics can be divided into three basic groups: ***totals***, ***proportions***, and ***means*** with totals the most popular statistics for managing recreational fisheries. Harvest measured as the number of fish caught and kept is a total, as is fishing effort measured in hours spent fishing, days spent fishing, or in angler trips. Composition of harvest by age group, by sex, or by species is a series of totals. Angler success is a series of proportions describing the relative distribution of anglers by their daily success. Catch per unit effort (CPUE) as an index of abundance is averaged over anglers in a sample to produce a mean. These and other statistics commonly sought through onsite creel surveys and their usual uses are:

TOTALS:

Harvest (number caught and kept) - regulate fishery and estimate sustained yields
Catch (number caught) - estimate size of fishery
Fishing Effort (hours, days, trips) - estimate size of fishery
Number of Angler-Trips - estimate size of fishery
Composition of Harvest by:
- *Age Group* - estimate sustained yields
- *Sex* - estimate sustained yields
- *Species* - allocate harvest
- *Length Group* - evaluate length limits
- *Marked Cohort* - allocate harvest

PROPORTIONS:

Angler Success (Relative distribution of anglers according to their success at harvesting 0, 1, 2, 3, etc., fish per day) - evaluate daily bag limits

---

MEANS:

Mean Catch Per Unit of Effort (CPUE) - index fish abundance
Mean Length of Angler-trip - planning surveys

Information is usually obtained from a creel survey, because a creel census is too expensive. In a census, ***all anglers or all harvest*** are interviewed or inspected; information is known with certainty. In a creel survey, anglers or harvest are ***sampled*** to produce estimates (statistics) of totals, means, or proportions.

The tradeoff with creel surveys is that sampling produces some uncertainty in statistics and some risk that by chance decisions on regulations could be misinformed. For each statistic there is a measure of its precision, a ***variance***, which can be estimated. For each statistic there is also a measure of its inaccuracy, a ***bias***, which is unknown, but can be minimized. Design of creel surveys is the most influential factor affecting bias in statistics, while size and allocation of sampling effort largely affect precision of estimates. Both design and sampling effort affect the cost of a survey. A "good" survey is one that produces the right statistics for fishery management with minimal risk of error and at an acceptable cost.

Most creel surveys in Alaska are onsite surveys designed to produce timely harvest estimates used to manage fisheries on migrating populations. These onsite surveys are of two general types: ***access-point*** and ***roving*** surveys. Access-point surveys can be implemented in fisheries where anglers exit the fishing grounds at only a few locations. Average harvest among anglers exiting at a location during a span of time is estimated by interviewing some anglers, then the average is expanded for the anglers not interviewed and the times and places not sampled. In access-point surveys, all exiting anglers must be counted and some, if not all, interviewed as they exit the fishery. A roving survey is used when access to a fishery is geographically continuous (or nearly so), making counting all anglers as they exit the fishery prohibitively expensive. In what are called ***roving-roving*** surveys (Pollock et al. 1994:242-251), technicians pass (rove) through the fishery to both count and interview actively fishing anglers. Harvest rate estimated from interviews is multiplied by fishing effort estimated from counts to produce an estimate of harvest for a fishing day, then estimated harvest is expanded for the days not sampled. In ***roving-access*** surveys (Pollock et al. 1994:251-253; Bernard et al. 1998), fishing anglers are still counted by roving technicians, but anglers are interviewed as they exit the fishery at an access point. Statistics from access-point surveys are generally more precise and less subject to bias than those from roving surveys. For these reasons, access-point surveys are preferred over roving surveys when the former can be implemented. If not, roving-access surveys are used because they are less prone to bias than roving-roving surveys (Pollock et al. 1994:250-1). When harvest estimates are not needed for inseason decision-making, an onsite survey is usually not conducted at all, and estimates are taken from an annual mail survey, the Statewide Harvest Survey (see Howe et al. 1996).

Sampling in onsite creel surveys to estimate harvest is designed around the nature of the fishery to be surveyed. Anglers access and exit the recreational for sockeye salmon on the Russian River through five campgrounds, four along the Russian River and the fifth across the Kenai River at its confluence with the Russian River. A single technician can count all exiting anglers and interview most of them at each campground. Anglers can walk to the river from the first four campgrounds, but must take a ferry across the Kenai River at the fifth. Most anglers using the ferry fish only at the confluence, while those accessing the fishery from the most upstream campground fish almost exclusively on the Russian River. Anglers using the other campgrounds can easily fish either at the confluence or upstream in the Russian River. Anglers fish during daylight hours from late May through early August; there is little variation within the week in the number of anglers participating in the fishery. Harvest is restricted with a daily bag limit. Two populations of sockeye salmon enter the river from late May through late June and from late June through early August. Each population has its own escapement goal, and harvest is regulated inseason to attain each goal. Also, estimated age composition of harvest from each population is coupled with estimated age composition of escapement to determine escapement goals that produce sustainable yields from these populations. ***Because of limited access to this fishery, the ability to count all exiting anglers, and the need for inseason harvest estimates, an access-point survey is the logical onsite survey for this fishery.*** Harvest from the Russian River proper would be isolated from that at its mouth and estimated for both populations of sockeye salmon to provide information with which to determine sustainable yields for both. ***Subsampling during the fishing day is appropriate here because harvest estimates from access-point surveys are free of bias from such subsamling when bag limits are restrictive***.

A roving-access survey is the logical choice for the fishery for coho salmon on the Kenai River because there are too many possible access-

points at which to inexpensively count all exiting anglers. The fishery begins in early August and ends in late September with almost all anglers fishing from boats. Two populations are fished, the first passing through the fishing grounds in August, the second in September. About a third of the fishing effort is from guided anglers who have a higher rate of success and a lower chance of being interviewed than their unguided peers. More anglers tend to fish during mornings and weekends. Harvest is restricted with a daily bag limit and possibly premature closure of the fishery if harvests are too large. Given these characteristics, harvest should be estimated separately for guided and unguided anglers to reduce possible bias in estimates, for weekends versus weekdays to increase their precision, and August versus September to provide detailed information on the two populations of coho salmon. ***Harvest should not be estimated separately for mornings and afternoons in this situation because doing so would involve subsampling the fishing day, which often biases estimated harvest rates when harvest is restricted with a bag limit***.

After harvest, its composition is the most often sought after statistic from onsite creel surveys. Estimates of harvest by age and sex are used to investigate sustainability of harvests by "reconstructing" recruitment and spawning biomass in individual year classes. Harvest composition can be estimated simultaneously with harvest in access-point or roving-access surveys, or estimated as the product of statistics from two independent surveys; ***relative harvest composition*** is estimated with an onsite creel survey, and harvest is estimated with the Statewide Harvest Survey. Products of statistics from these two surveys are estimates of harvest composition. When timely estimates of harvest are not needed, estimating harvest composition from two surveys is attractive because its cheaper. The Statewide Harvest Survey provides annual harvest estimates for all Alaska's fisheries, and sampling to estimate relative harvest composition is less complex than estimating harvest. Composition of a fish population does not appreciably change from morning to night, from week day to week end, and sometimes for non-migratory species, change across a fishing season can be negligible. If there is no temporal or spatial variation (change) in harvest composition, when, where, or how often harvest is sampled is not important. Under this hypothetical situation, an onsite survey becomes very simple and very cheap. However, there is always some variation, especially in harvest from a migratory population.

The International Pacific Halibut Commission can use estimated age composition of Pacific halibut *Hippoglossus stenolepis* harvested in the recreational fishery in Homer to "reconstruct" the Pacific halibut population with catch-at-age analysis. Estimates of relative age composition from an onsite creel are multiplied by harvest estimates from the Statewide Harvest Survey to produce estimated harvest composition. The Pacific halibut population is represented by many age groups producing a relatively stable age composition over a short span of time (several months). Pacific halibut farther from shore tend to be larger and older. The recreational fishery in Homer begins in May and runs through late September with anglers fishing from private and from chartered boats. Chartered boats are usually larger than private vessels and travel far from shore. Most anglers access the fishing grounds from the single harbor on the Homer Spit in the morning and return in late afternoon. The boat harbor is too large to be covered by a single technician. Because many participants in the fishery are tourists, there is little week day versus weekend variation in fishing effort. Small Pacific halibut are cleaned at sea on board chartered vessels with the carcasses thrown overboard; large Pacific halibut are brought ashore intact to be photographed. In contrast, private anglers usually clean their harvest ashore at one of the many cleaning stations on the dock. Under these circumstances, harvest by chartered and private anglers should be treated separately, and some cooperation solicited from charter operators to save carcasses of fish cleaned at sea. Harvest from these two groups should also be (and are) estimated separately in the Statewide Harvest Survey. Money can be saved in the onsite survey by taking advantage of the stable age composition of the population and of the subsequent harvest. Sampling can be scheduled during the late afternoon when most anglers return to the Spit and need only occur on a few days throughout each month. The onsite survey would need employ only a single technician.

An onsite creel survey strictly to estimate relative harvest composition can be designed as a ***self-weighted survey*** (see Cochran 1977:91). Time and geography (if needed) are divided, and a sample of the harvest is taken from each division in such a way that each sample represents the same fraction of harvest as in the others. If successful, data from self-weighted sampling can be pooled across divisions to produce an unbiased estimate of relative harvest composition. In reality, sampling in self-weighted surveys is only somewhat proportional with samples representing similar, but not the same fraction across divisions. Because temporal or spatial variation in relative harvest composition is usually subdued, differences in sampling fractions fortunately produce negligible bias in estimates of relative harvest composition from self-weighted surveys. Self-weighted surveys are also used to estimate angler success.

The remainder of this report is a technical explanation of how to design and execute "good" onsite creel surveys in Alaska. Whenever possible we followed the notation and terminology in Cochran (1977), Thompson (1992), and Pollock et al. (1994). Notation has been standardized across chapters with definitions placed in hopefully convenient locations. Notation is also defined at the end of this introduction. Terms are defined as the text unfolds and in a closing glossary. Topics covered in the following text are:

Chapter 1: Basic definitions and procedures of sampling designs, types and scheduling of sampling, and solutions to several common problems in setting up onsite creel surveys.

Chapter 2: Specific designs for access-point and roving creel surveys to estimate harvest, catch, and fishing effort are described.

Chapter 3: Stratified and self-weighted sampling designs to estimate composition of harvest when all groups can be identified (usually by age, length, weight, sex, etc.) or only when some of each group can be identified (usually by carrying coded wire tags).

Chapter 4: Self-weighted surveys to estimate angler success are described along with how to evaluate bag limits with these estimates.

Chapter 5. Surveys to estimate catch per unit of effort in a recreational fishery as an index of fish abundance is covered in this chapter.

Chapter 6: Special sampling designs and schedules that can be used for surveys with only one technician are listed.

Chapter 7: Tips on how to identify, count, and interview anglers; correctly label fishing effort; and properly sample harvest to avoid bias in estimates from onsite creel surveys is discussed in this chapter.

Appendices: Topics in appendices include analysis of unequal probability sampling, derivations, and methods for unusual situations.

What to do with missing data and potential biases in estimates for specific types of onsite surveys are discussed in each chapter. The report contains many examples, all of which are based on actual data from onsite surveys in Alaska.

* * * *

Notation and terms in this report are defined according to a few general rules. Because the same equations can be used to calculate several statistics, Y or y is used as a general variable in these equations. Under specific circumstances, like in examples, Y is replaced by H (harvest), E or e (fishing effort), C or c (catch), or M (number of angler-trips). With one exception, lower case letters refer to data from individual anglers while upper case letters to totals of statistics across units in a sampling stage; the exception is measured harvest for an angler-trip which is denoted with H. When H denotes estimated total harvest for a sampling unit, it will be flagged as an estimate. All estimated totals are noted with "^" and all means with "−" in the standard fashion. Variance of estimates is implied by the prefix *Var* and estimated variance by *var*; the same convention holds for covariance. When there is multistage sampling in onsite creel surveys, estimated variance (of estimates) is composed of two or more summed components, one for each sampling stage. At the core of each component is a sample variance noted as s^2 for data from anglers or S^2 for estimated totals for units within that sampling stage. Specific definitions for notation are:

B Proposed daily bag (harvest) limit for an individual angler.

B_o Current daily bag (harvest) limit for an individual angler.

c Catch during an individual angler-trip. This statistic is always a datum.

cov[...]
Estimated covariance between statistics in the brackets.

CPUE
Acronym for Catch Per Unit of Effort which denotes the number of fish caught for each hour of fishing in an angler-trip; its estimated mean $\overline{cpue}$ can be used as an index of fish abundance.

D Maximum number of finite calendar days in a survey in which sampling can occur. In access-point surveys, *D* is either the maximum number of first or second-stage units. In roving surveys, *D* is always the maximum number of first-stage units.

d Number of calendar days in a survey in which sampling occurs. In all surveys, *d* is always a subset of *D*.

Δ (Delta) Percent reduction in harvest from reducing a daily bag limit for an angler from B_o to *B*.

E Fishing effort in hours for any strata or stage in a stratified, multistage sampling design.

e Length of an angler-trip in hours. This statistic is usually a datum, but can be an estimate.

η (eta) Number of fish in a sample of harvest.

f Fraction of harvest sampled.

f_1 Fraction of first-stage units sampled.

f_2 Fraction of second-stage units sampled.

f_3 Fraction of third-stage units sampled.

f_4 Fraction of fourth-stage units sampled.

G[...]
Squared coefficient of variation for the term in the brackets.

g 1) Subscript that denotes a specific category in the composition of harvest or 2) number of fish in an angler's daily bag.

H Harvest for elements of any strata or stage in a stratified, multistage sampling design. This statistic is usually an estimate, but can be a datum.

h Subscript that denotes a specific stratum.

HPUE
Acronym for Harvest Per Unit of Effort which denotes the number of fish caught and kept for each hour of fishing in an angler-trip. The estimated mean $\overline{hpue}$ is used in the relationship between fishing effort and harvest in roving surveys; it should not be used as an index of abundance.

i Subscript that denotes a specific first-stage unit.

j Subscript that denotes a specific second-stage unit.

k Subscript that denotes a specific third-stage unit.

L Number of strata in the sampling design.

l Subscript that denotes a specific fourth-stage unit.

λ (lambda) Compound fraction of sampled salmon with missing adipose fins whose heads reach a lab and the fraction of coded wire tags detected in these heads that are subsequently decoded.

M Number of anglers (angler-trips) that could be interviewed in a sampled period. In access-point surveys, *M* is the number of second, third, or fourth-stage units, and is the number of anglers ***exiting*** the fishery during a sampled period. In roving surveys, M is the number of second-stage units, and is the number of anglers that ***participate*** in the fishery during a sampled period.

m Number of anglers interviewed during a sampled period. In all surveys, m is always a subset of M.

N Number of access points that anglers have to a fishery. In access-point surveys, N is usually the number of first-stage units.

n Number of access points where anglers are interviewed as they exit a fishery. In access-point surveys, n is always a subset of N.

ω (omega) Anticipated ratio of angler-days of fishing effort under a new daily bag limit for an individual angler versus angler-days under the old limit.

p 1) Proportion of harvest composed of particular category of fish or 2) fraction of daily bags of a certain size.

ϕ (phi) Scaling factor for missing data that equals 1 if data for a unit are available, 0 if not.

' (prime) Denotes a different unit of the same kind, for instance, h and h' denote two strata. The prime is used to avoid confusion.

" (double prime) Denotes a different unit of the same kind, for instance, h' and h'' denote two strata. The double prime is used to avoid confusion.

Q Number of sampling periods in a typical fishing day in an access-point survey; Q represents either second or third-stage units.

q 1) Number of sampled periods in a typical fishing day; this q is always a subset of Q. 2) Catchability coefficient (the fraction of a population captured instantaneously with one unit of fishing effort).

r Number of times anglers are counted in a sampled period in a roving survey. Each count corresponds to passing through the entire fishery once.

S_1^2 First-stage sample variance for harvest, catch, effort, or number of angler-trips.

S_2^2 Second-stage sample variance for harvest, catch, effort, or number of angler-trips.

S_3^2 Third-stage sample variance for harvest, catch, effort, or number of angler-trips.

s_2^2 Sample variance for mean harvest, mean catch, or mean fishing effort across angler-trips in a two-stage sampling design.

s_3^2 Sample variance for mean harvest, mean catch, or mean fishing effort across angler-trips in a three-stage sampling design.

s_4^2 Sample variance for mean harvest, mean catch, or mean fishing effort across angler-trips in a four-stage sampling design.

T Length of a sampling period in hours.

t Subscript that denotes a specific count during a sampled period in a roving-access survey.

θ (theta) Fraction of a cohort of fish that is marked.

u Superscript that denotes the statistic was calculated as a ratio of means.

var[...]
Estimated variance for the statistic in the brackets.

x Number of anglers counted through a single pass through the fishery by a technician in a roving survey.

Y General statistic that could be total harvest, total catch, total fishing effort, or total number of angler-trips.

y General statistic that could be length of an angler-trip (hrs), fish caught or harvested during that trip, number of anglers in a sample with a particular success, or number of fish in a sample sharing a particular attribute. This statistic is always a datum.

CHAPTER 1: Sampling Designs

The ***basic sampling unit*** in an onsite creel survey, the unit for which data are recorded in the field, is the fishing trip or the harvested fish. Collectively, these basic units comprise the ***target*** population, the population about which information is wanted. Because access to anglers or harvested fish is limited in an onsite creel survey, basic units can not be randomly selected for a sample. Instead, samplers (technicians) encounter anglers and harvested fish opportunistically or haphazardly at access-points to a fishery. Since random sampling of a population is a prerequisite for an unbiased estimate in any sampling project, opportunistic sampling represents a real source of problems for onsite creel surveys.

These problems can be solved to a large extent, and sometimes completely, by grouping basic sampling units according to a sampling design composed of ***strata***, and within strata, a hierarchy of ***sampling stages***. All access-point, roving, and self-weighted surveys follow sampling designs with strata and sampling stages. Organization of strata and sampling stages for each survey represents a specific ***sampling frame*** that permits a practical, yet random scheduling of work for technicians. Some divisions in the sampling frame are obvious; some are arbitrary. Since our society has a diurnal pattern of activity, days would represent an obvious sampling stage for many of our recreational fisheries. In contrast, how to divide days to schedule subsampling would be an arbitrary choice. Access points can be another obvious division in a sampling frame. Behavior, distribution, and migration of fish, the needs of fishery management, and union rules are also used to establish sampling frames.

The random scheduling of work for technicians according to a ***sampling design*** produces random sampling on all levels of the sampling frame except one: the encounter. Although sampling designs effectively reduce the target population to those basic units associated with when or where a technician is actually working, the technician still can only sample these units opportunistically. This final failure to sample randomly still has the potential to produce significantly biased estimates from onsite creel surveys unless specific remedies are part of the sampling design.

This chapter continues with explanations of the basic concepts behind sampling designs for onsite creel surveys, such as sampling periods and fishing days, stratification and multi-stage sampling. Methods of scheduling work for technicians are described along with common problems and their acceptable and unacceptable solutions. The chapter concludes with an overview of potential for bias in statistics arising from not randomly encountering and interviewing anglers during a sampled period in access-point, roving, and self-weighted surveys.

1.1 SAMPLING PERIODS

A ***sampling period*** is the working unit used in scheduling a creel survey. A sampling period is a specific interval of time at a specific location during which sampling could occur. Length of a sampling period is generally a function of the consecutive hours in a day that anglers as a group spend fishing (the ***fishing day***), and of the length of a working day for technicians (7.5 hr in Alaska). Exact length of a sampling period is generally set according to convenience, but is kept consistent throughout the survey or at least throughout each stratum (see the next section for a definition of stratum). In Alaska, the union rule that a technician must work continually for at least 4 hr of each day in which they work at all, effectively bounds the sampling period to be at least 4-hr long. The only hard rule is that length of the sampling period must divide evenly into the length of a fishing day. If a sampling period is selected for sampling, it becomes a ***sampled***

period; during a sampled period, one or more technicians count and interview anglers and inspect their harvest while anglers are fishing or as they leave the fishery.

1.2 STRATIFICATION

A ***stratum*** is a collection of basic sampling units that are associated with the same location(s), the same time of day, the same week, part of a week, month, or that share a common characteristic. Because all anglers associated with the same sampling period are in the same time and location-defined stratum, such a stratum is also a collection of sampling periods. Statistics for each stratum are calculated separately, then combined to produce a statistic for the entire fishery or for any part of the fishery. Calculations within such strata are based on an independently selected set of sampled periods which make the resulting statistics independent among strata. If statistics are totals, say harvest, fishing effort, or harvest composition, estimates are added across strata to get a statistic for the whole or for part of a fishery:

$$\hat{Y} = \sum_{h=1}^{L} \hat{Y}_h \tag{1.1a}$$

$$var(\hat{Y}) = \sum_{h=1}^{L} var(\hat{Y}_h) \tag{1.1b}$$

where L is the number of strata, and $\hat{Y}_h$ is the estimated total for stratum h.

Sampling designs are stratified in onsite creel surveys for four reasons (Pollock et al. 1994:33; Bernard et al. 1998) :

- for convenience in staffing;
- to tailor statistics for use in managing a fishery,
- to improve accuracy of statistics; and
- to improve precision of statistics.

Onsite creel surveys to estimate harvest of whitefish from the Chatanika River where stratified by ***location*** to simplify staffing. Fishing occurred only at night (1800 to 2400 hours) by the light of hand-held lanterns. Anglers accessed the fishery near Olnes Pond and at two other distant locations. A sampling period covered an entire fishing day (in this case, fishing night) with all sampling periods (every night) at Olnes Pond comprising a single stratum. Sampling periods at the other two locations comprised two other strata. Since one technician could not service all three locations at the same time, stratifying the survey by location kept the staff to a single technician occasionally augmented by permanent staff. The most common reason for ***weekly stratification*** in surveys is to conform with the established work week for technicians.

Onsite creel surveys to estimate harvest of coho salmon on the lower Kenai River had ***seasonal stratification***, in this case monthly stratification, to improve accuracy and to produce statistics useful for management. As daylight waxes and wanes during the year, the length of a fishing day changes dramatically for prolonged fisheries in Alaska. Since length of a fishing day and length of sampling periods can be redefined in each stratum, monthly stratification minimizes bias in statistics from passing seasons. Sampling periods were 1.5 hr longer in August than in September. Also, two separate populations of coho salmon pass through the fishing grounds, one in August and the other in September. Harvest estimates for each month are harvest estimates for each population.

Grouping sampling periods into strata will almost always improve precision of estimates (Cochran 1977:99). However, as sampling periods are grouped into ever more strata, the marginal increase in precision declines and difficulty in scheduling sampling by staff intensifies. Because more anglers exit the marine fishery near Juneau later in the day, surveys of that fishery have had “time-of-day” (***TOD) stratification*** to improve precision of estimated harvest of chinook salmon. The 15-hr fishing day has been divided into two, 7.5-hr sampling periods: “mornings” and afternoons”. If all "mornings" comprise one stratum and all "afternoons" the other, one technician working 37.5 hr per week can theoretically meet the minimal requirements of the design (at least two sampled periods for each of two strata, four sampled periods in all). Past surveys of this fishery also have used ***DOW*** (day-of-week) *stratification.* More fishing effort and harvest in this fishery had occurred during weekends than during the rest of the week, so stratifying sampling periods into weekends (Saturday and Sunday) and week days (Monday through Friday) would probably improve precision of estimated harvest in upcoming surveys. Together, TOD and DOW stratification define four strata: weekend mornings, weekend afternoons, week-day mornings, and week-day afternoons. One technician can not meet the minimal requirements of this more complex design (at least two sampled periods for each of four strata each week, eight sampled periods in all). The gain in precision promised by the second design comes at the cost of a second technician.

1.2.1 Full Stratification. The maximum number of strata in a sampling design is a function of the dimensions of stratification. Week, month, location, TOD, and DOW are examples of dimensions in stratification. If there are K dimensions of stratification with k_1, k_2 .... k_K elements in each dimension, the population of sampling periods is divided into at most $L = \prod_{i=1}^{K} k_i$ strata. Weeks, months, locations, morning and afternoons, and week days and weekends are the respective elements in the dimensions listed above. Besides DOW and TOW stratification, surveys of the 22-week marine fishery near Juneau have been stratified by location and by fortnights (biweeks). These four dimensions of stratification were used because early surveys showed that: 1) more anglers completed their fishing trips during afternoons than during mornings, 2) more during weekends than during week days, 3) more at some harbors than others, and because 4) statistics had to be reported to the Pacific Salmon Commission once every two weeks. For the sake of this example, only two harbors near Juneau are considered with one harbor considerably more popular than the other (a high-use and a low-use harbor). Full stratification with this design produces 88 strata (=L=2x2x2x11) (Fig. 1.1). The minimum sample size for a stratum in these surveys is two sampled periods, and 176 sampled periods (=2x88) minimally cover the sampling design. Since a sampling period has the same length as a work day (7.5 h) for these surveys, cost of full stratification is 176 technician-days.

1.2.2 Partial Stratification. In full stratification, every combination of every element in every dimension of stratification is represented in the sampling design; in partial stratification, elements in one dimension have been "collapsed" within an element of another dimension. In the example in previous sections involving marine surveys near Juneau, full stratification produced 88 strata with 8 strata occurring every two weeks during the 22-week fishing season (Figure 1.1). If contrary to the scenario in Section 1.2.1, daily variation in harvest exiting from low-use harbors is nil, but significant at the high-use harbor, DOW stratification at the low-use harbor will not meaningfully improve precision of estimated harvest for the season. The partial stratification proposed in Fig.1.1 would eliminate two strata every two weeks, decreasing strata from 88 to 66 for the season. This 25% reduction in the number of strata corresponds to a 25% reduction in work days of sampling effort and a 25% reduction in the cost of the survey at no appreciable loss of precision in estimated harvest.

1.2.3 Post Stratification. For some onsite creel surveys, not all stratification can be determined during planning and design. Assignment of fishing trips and harvest to the appropriate stratum is pre-destined if strata are spatially or temporally defined, as is the case with seasonal, location, TOD, and DOW stratification. However, if strata are defined by characteristics of anglers or how they fished, data can not be assigned to the appropriate stratum until anglers have been encountered and/or interviewed. Most common types of post-stratification are based on:

- the type of fishing gear used (i.e., bait vs. artificial lures);
- residence of the angler (i.e., resident vs. visitor vs. military);
- mode of fishing (i.e., boat vs. shore); and
- expertise of anglers (i.e., guided vs. unguided).

In surveys to estimate harvest of coho salmon in the lower Kenai River, angler-trips were post-stratified into those by guided and unguided anglers. Both unguided and guided anglers were counted and interviewed during each

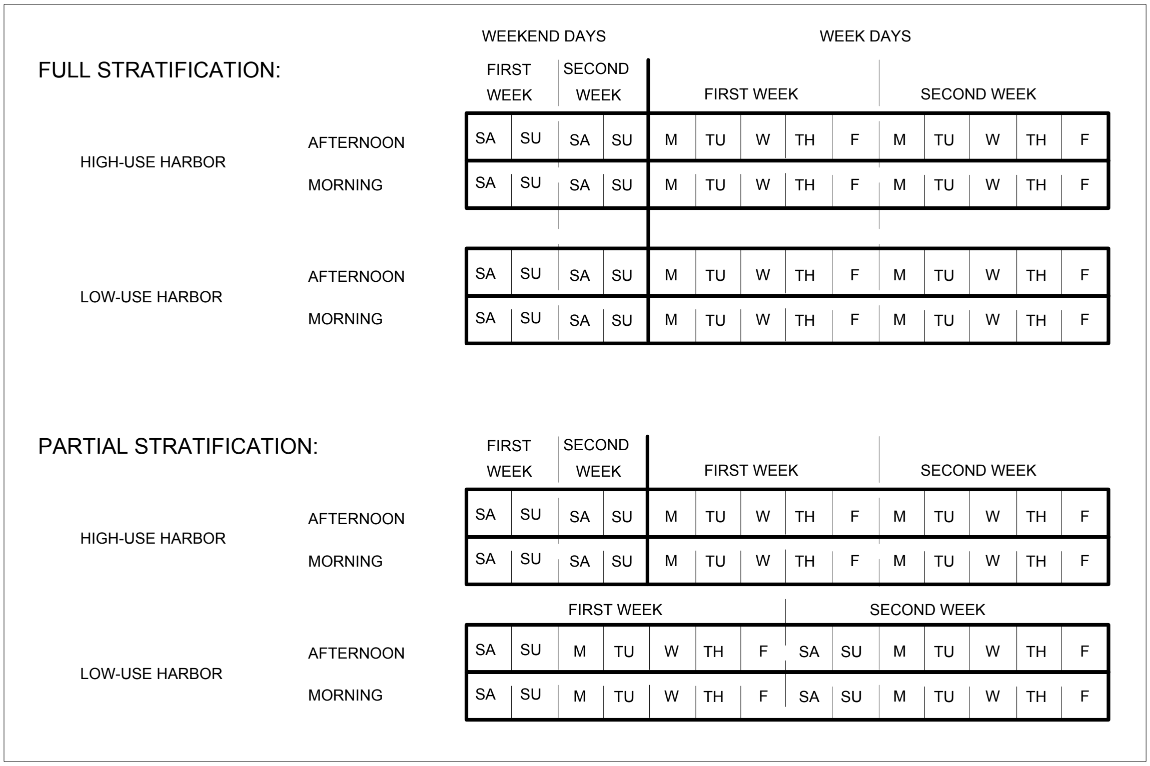


Figure 1.1 - Schematic showing full and partial stratification for a typical two-week period in surveys of the marine creel near Juneau. Sampling periods in the same stratum are bound with broad, solid lines and are separated from one another within the stratum by dashed lines.

sampled period. Post-stratification doubled the number of strata in these surveys, but obviously without changing the cost of the survey. Post-stratification is used to produce more useful estimates and in some situations to avoid bias in statistics (Section 2.2.2.2). Post-stratification can also be a part of accurately estimating angler success and using catch per unit of fishing effort to index abundance (see Chapters 4 and 5).

1.3 SAMPLING STAGES

By design a dimension of stratification in a sampling design can be "demoted" to a sampling stage with subsampling. For stratification, all elements in each dimension are sampled; if some are not sampled by design, the "dimension" becomes a sampling stage, and its "elements" sampling units. For example, anglers access the marine fishery near Juneau through 13 harbors, not just two harbors as described earlier. In this instance, popularity of harbor is a dimension of stratification with two elements (high-use and low-use), and individual harbors are ***first-stage sampling units***. Harbors are treated as a sampling stage instead of a dimension of stratification in the sampling design because sampling all 13 harbors within a fortnight is prohibitively expensive. Only a subset of these harbors are sampled during each fortnight, time of day, or time of week (the other dimensions of stratification).

Possible sampling stages for all onsite surveys are location, fishing day, sampling period, and angler-trip. The hierarchy in Fig. 1.2 corresponds to a four-stage sampling design with location as the first stage, day as the ***second stage***, sampling period as the ***third stage***, and angler-trip as the

fourth stage. In the sampling schedule for an onsite survey, a technician is sent to interview and/or count anglers during a specific sampling period on a specific day at a specific location. Sampling these stages is "nested" such that technicians will interview anglers during at least two sampling periods on each of at least two days at each location sampled. In this design, locations are first-stage sampling units, fishing days second-stage sampling units, sampling periods third-stage units, and anglers-trips the fourth-stage units. At least eight sampling periods must be sampled to fulfill the requirements of the design. ***At least two sampling units must be sampled in each stage; if only one unit is sampled, there is no means of accurately calculating variance for statistics***. More about this rule is covered in Section 1.4.

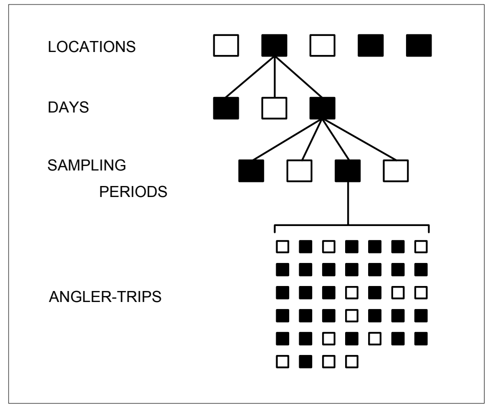


Figure 1.2 - Schematic of a four-stage sampling design within a single stratum. Filled boxes correspond to sampled units. Lines connect examples in the hierarchy.

Note that basic sampling units in Fig. 1.2 are angler-trips and not anglers. When harvest, fishing effort, catch, angler success, or mean catch per unit of effort (CPUE) as an index of abundance are to be estimated, angler-trips are the basic units. If composition of the harvest is to be estimated, individual fish in the creel are the basic sampling units.

Hierarchy and number of sampling stages in a specific design are determined through convenience and the number of access points (Table 1.1). The number of stages in the design is often a function of the number of locations by which anglers access the fishery. Whenever anglers have but one point of access to the fishery, location is no longer a sampling stage, and the sampling design falls a notch from four stages to three or from three to two. If there are two access points, the four- or three-stage sampling design above is the same as the three or two-stage designs stratified by location. If there are three or more access points, location can be kept as a first stage or changed into a dimension of stratification as is convenient. When access points are numerous, the sampling burden becomes onerous if every access point is to be a stratum. In this situation, location can be better treated as a sampling stage or ignored by abandoning the access-point survey in favor of a roving survey.

Sampling period can often be treated as a sampling stage or can be transformed into a dimension of stratification (remember TOD stratification). Since the length of a sampling period is generally set as a matter of convenience, a fishing day can be conveniently divided. For instance, if a fishing day is 16- hr long and the sampling period is set at 8 hr, the sampling design could be stratified into "mornings" and "afternoons". If the fishing day is 18 hr long, sampling periods could be reduced to 6 hr, and the design stratified into "mornings", "afternoons", and "evenings". If there is a diurnal cycle to statistics, TOD stratification will simplify scheduling of technicians (Section 1.4.4). If there is no such diurnal cycle, TOD stratification could be kept to simplify the scheduling of sampling, or sampling periods could become sampling units in a sampling stage. In the 18-hr fishing day with three sampling periods and no TOD stratification, two must be sampled for each day that sampling takes place.

In all sampling designs for our onsite creel surveys with locations and days as sampling stages, location is the preferred first stage while day is the second. Although there is no theoretical reason these roles could not be reversed, there is a practical one: scheduling of

Table 1.1 - Types of sampling designs used in onsite creel surveys.

DESIGN:

- **First/Second/Third/Fourth Stages**
  - Conditions

FOUR-STAGE :

- **Location/Day/Sampling Period/Angler-trip**

THREE-STAGE:

- **Location/Day/Angler-trip**
  - Length of a sampling period is the length of the fishing day; **OR**
  - Sampling period is a dimension of stratification (TOD stratification).
- **Day/Sampling Period/Angler-trip**
  - There is only one point of access to the fishery; **OR**
  - Access points are a dimension of stratification (location stratification).

TWO-STAGE:

- **Day/Angler-trip**
  - There is only one point of access to the fishery; **OR**
  - Access points are a dimension of stratification (location stratification);

AND

  - Length of a sampling period is the length of the fishing day; **OR**
  - Sampling period is a dimension of stratification (TOD stratification).

---

sampling. If day is the first sampling stage and location the second, anglers must be interviewed at two or more locations on the same day, requiring two or more technicians. If location is the first sampling stage and days the second, one technician can interview anglers on two or more days.

Strata and stages have "***size***". The size of a sampling unit from a particular stage is the number of sampling units it contains from the next lower stage (see Cochran 1977:299). In our basic four-stage design, there are (usually) the same number of days at any possible sampling location, therefore all of the first-stage units are the same size. Dropping lower in the multistage hierarchy, each day has the same number of hours and hence the same number of sampling periods, therefore all the second-stage units are the same size. However, sampling periods are not the same size because each period represents different numbers of angler-trips, anglers, or harvest. This heterogeneity of size among sampled periods occurs in all multistage onsite creel surveys. Because strata are a collection of sampling periods, strata also have different sizes (unequal numbers of basic units).

Only for the rarely calculated mean length of an angler-trip are sizes of sampled periods and strata considered when calculating statistics (see Appendix A). Sizes of sampling periods or strata are irrelevant when calculating totals like harvest, catch, or fishing effort from access-point or roving creel surveys (see Chapter 2). Size is also irrelevant in self-weighted surveys to estimate angler success and relative harvest composition because size of the sample is proportional to the size of the sampling period (see Chapters 3 and 4). Size is also irrelevant in onsite surveys to calculate $\overline{cpue}$ as an index of abundance because this index is never combined over sampling periods (see Chapter 5).

Multistage designs in our onsite creel surveys can change from stratum to stratum for our convenience. Because sampling periods are chosen independently among strata (see the section below), each stratum represents a different "fishery", and statistics for each stratum are calculated independently. A minor disadvantage in having different multistage designs in different strata is that more than one set of equations must be used.

1.4 SCHEDULING IN STRATIFIED MULTI-STAGE DESIGNS

After setting the length of sampling periods and deciding how many and what kind of strata and sampling stages are in a survey design, a sampling schedule is built that commits technicians to sampling at specific places at specific times. The first step is to divide the available technician time into sampling periods. For instance, if sampling periods are 7.5-hr long (the same length as a work day) and the fishery continues for 6 weeks, one technician represents 30 sampled periods (=5x6x7.5/7.5); if sampling periods are 4-hr long, one technician represents 56 sampled periods (=5x6x7.5/4). The product of the number of available technicians and the number of sampling periods each technician can sample is the sample size. The next step in this process is to allocate sampled periods among the *L* strata in the design. The minimal number of sampling periods that will be sampled in each stratum depends on the number of sampling stages in that stratum. ***At least two sampling units must be sampled from each sampling stage to estimate variance (precision) of statistics***. If a stratum has three stages (say location, day, and angler-trips), technicians must be instructed to interview anglers during at least four sampling periods in that stratum (two days at each of two locations). If a stratum has four stages, then at least eight sampling periods must be sampled (two sampled periods at each of two days at each of two locations). If a stratum has but two stages, then there must be at least two sampled periods. Since all strata in surveys of the marine fishery near Juneau have three stages, at least eight sampling periods have to be sampled during each stratum. Obviously, more than the minimum can be sampled in a stratum if technicians are available.

1.4.1 Order of Selection. Once the sampling design, sample sizes, and an allocation of those samples have been determined, specific periods can be selected for sampling, one stratum at a time. ***Start with the "largest" stratum, the stratum that will have the most angler-trips. Select sampling periods from it first. Then progress from stratum to stratum in descending order according to their "sizes".*** If conflicts with scheduling technicians occur (see Section 1.4.4), they will do so during the least important strata if this sampling rule is followed. Larger strata generally contribute more to the variance of the final statistic than do smaller strata (Cochran 1977:98). Any "bending" of schedules for "small" strata will have minimal impact on accuracy and precision of overall estimates.

Within each stratum, sampling units are scheduled for sampling stage-by-stage beginning with first-stage units. Two or more first-stage units are selected first, then if need be, two or more second-stage units are selected for each first-stage unit chosen, and two or more third-stage units are selected for each second-stage unit chosen. For instance, if a stratum is a week of seven days, a sampling period is 6-hr long, the fishing day is 18-hr, and four sampling periods have been assigned to this stratum, two days are selected for sampling, say Tuesday and Saturday, out of the seven possible. On Tuesday, the morning and afternoon sampling periods are picked for sampling out of the three possible periods. ***The rule for selecting sampling units in order by stages must be strictly followed***. If this rule is not followed, the implemented survey will not follow the original design. Calculations for the original design will be inconsistent with how data were collected, and ***resulting statistics will be biased***.

1.4.2 Random Selection. Within a sampling stage, sampling units are almost always selected with ***equal probability and without replacement***. The most common means of selecting sampling units is randomly choosing numbers that correspond to each unit. A random number is selected, usually from a table or from a calculator or computer.

Since our society is based on the 7-d week, our tendency is to build a sampling schedule one week at a time. Such a construction is appropriate only if weeks are a dimension of stratification. If weeks are not a dimension of stratification, yet sampling schedules are built by week, sampling periods are not randomly scheduled within the stratum, and stratification of the survey will not match the calculation of statistics. If fortnights are a dimension of stratification, then all 14 days in a stratum have an equal chance of being the first sampling unit selected, all remaining 13 days have an equal chance of being selected as the second sampling unit, all remaining 12 etc. If 10-d periods are a dimension of stratification, then all 10 days in a stratum have an equal chance of being selected as the first sampling unit, all remaining nine have equal chance of being selected as the second unit, etc.

Sampling units can also be randomly selected with probability proportional to their expected size and with replacement (see Pollock et al. 1994:43,231-53). Expected size of a sampling period is the expected number of angler-trips or expected harvest during the period. The larger the sampling period, the greater the probability that it will be scheduled. Advantages and disadvantages of this alternative approach were investigated with the recommendation that it not be used in creel surveys in Alaska (Bernard et al. 1998; Appendix B).

1.4.3 Systematic Selection. Instead of randomly selecting sampling units within a stage, sampling units can be chosen systematically (Pollock et al. 1994:39-40). When sampling units have a natural order, such as hours in a day, days in a week, days in a month, and harbors along a coast, units can be selected for sampling according to this order. If n units are to be sampled from N units in a stage ($n<N$), only one out of every k ($=N/n$) members are sampled. To draw a systematic sample, the first member sampled is randomly selected from the first k members in the order; thereafter, every kth member in the order beyond the previous member sampled will be selected. The number of units N in the sampling stage should be an even multiple of k. For instance, if three sampling periods are to be systematically selected (k=3) each day, there must be 6, 9, 12, 15, etc. (=N) sampling periods in a day, and the length of the sampling period must be adjusted accordingly. Systematic selection assures that the spatial or temporal range of sampling units within a stage will be covered even for small sample sizes.

There are some caveats when sampling units are chosen systematically. Variances in onsite surveys even partially based on systematically drawn samples must be approximated. Under systematic sampling, there is no true estimate of variance since there are only k possible samples of which only one is chosen (Cochran 1977:207). Because of the method chosen to approximate variance of a systematic sample (Chapter 2), ***at least three units in each stage must be sampled.*** This is in contrast to the minimal requirement that at least two units in each stage be sampled when selecting units randomly. Another concern with systematic sampling is "frequency bias" when the frequency of sampling is in synchrony with cyclical patterns in the parameter to be estimated (see Cochran 1977:218). For instance, sampling one day a week would most likely produce biased estimates of harvest or fishing effort for a fishery in which most participation occurs during weekends. Under these circumstances, systematic sampling should be avoided or modified to break the synchrony. These considerations aside, systematic sampling can be especially useful in designing surveys, especially one-technician surveys (see Chapter 6).

1.4.4 What if it doesn't fit? Once the periods that will be sampled have been selected, specific technicians are assigned to specific periods at specific times and locations. Sometimes during these assignments, some problem arises in getting available technician time to "fit" the design. Problems that have been encountered along these lines are:

- Not enough technician time is available to meet minimal requirements;
- More technicians must be used to sample in a given day than are available; and
- The sampling schedule would violate union rules for one or more technicians.

If there is a shortage of technician time, the correct action is to either add more technician time or change the design. Because adding technician time can be expensive, especially for fisheries in remote locations, changing the survey design has been the more popular alternative. Two correct changes are to 1) reduce the length of sampling periods and 2) use partial stratification. If stratification is collapsed over dimensions that represent small amounts of variation in parameters being measured, partial stratification can be used with little loss in the precision of statistics. However, if dimensions represent significant variation, simplifying the schedule will significantly degrade precision.

One union rule in Alaska that has complicated scheduling in small onsite surveys is that ***each technician should have two consecutive days off per week***. In the past, the common means of avoiding this problem in small surveys was to randomly assign two consecutive days off during the week, then randomly schedule sampled periods on the remaining days. ***This remedy should be avoided because the first and last days in each week will have twice the probability of other days of being sampled.*** If statistics are cyclical on a weekly basis (as are fishing effort and subsequently harvest in many fisheries), statistics will be biased with this remedy. Acceptable solutions to this problem are to 1) extend sampling effort through the judicious use of overtime pay, 2) assign some sampling to permanent staff, and 3) redesign the survey. Chapter 6 contains several sampling designs for small surveys that avoid violation of this union rule. In large onsite surveys with several technicians, violation of this rule is easily avoided.

Other "fixes" to avoid when fitting a sampling schedule to a sampling design are:

- ignoring a stratum;
- ignoring a sampling stage;
- single samples;
- subsampling periods; and
- rerandomizing.

Not sampling in a stratum will bias estimated totals such as harvest and fishing effort downward. Because each stratum represents some angler-trips, ignoring a stratum is the same as ignoring the harvest, catch, or fishing effort associated with those trips. Although bias may or may not result, ***ignoring a sampling stage will underestimate variances for statistics and promote a false sense of confidence.*** Only when parameters do not vary across dimensions of stratification or across units in a sampling stage can strata and sampling stages be ignored with little consequence. Relative harvest composition and to a lesser degree angler success epitomize this exception (see Chapters 3 and 4). Unfortunately, diurnal, weekly, and seasonal variations in fishing effort, harvest, and catch is expected. For these totals, ignoring strata or stages will reduce the credibility of estimates.

Another "fix" when fitting schedules to designs is to deliberately sample only one sampling period in a stage. ***One sample per stage guarantees that there will be "missing data" in the survey***. Statistics for a stage will be based on a sample size of one, and estimated variances can be salvaged only with imputed data. Solving the problem of missing data in onsite creel surveys is bad enough (see Sections 2.1.3, 2.2.6) without producing the problem by design.

Subsampling periods, usually to save technician time, both understates variance and produces single samples. An additional sampling stage is created (subsamples within sampled periods), which is usually ignored when estimating variance. Even if this additional stage is included when estimating variance, there remains the problem of missing data with only one subsample per sampled period.

On occasion a randomly chosen schedule proves inconvenient; a single technician must be in two locations at the same time, a technician will not get their two consecutive days off during a particular week, etc. The temptation is to ignore the first schedule and randomly select another. If the second is still inconvenient, it is discarded, and other schedules are drawn until one is generated that has no such problems. ***Rerandomizing a sampling schedule in this way to make it convenient will bias estimates.*** Such a schedule is not really random, but is purposely selected. Some sampled periods will have higher probability of being in the schedule when schedules are rerandomized in this way.

When changing the survey design is no longer an option, yet no additional sampling effort is available, the survey should be canceled. If a survey is of such low priority that additional resources can not be found, or if there is so little concern for bias in its statistics, the original justification for doing the survey is suspect. Expenditures that would have supported this survey would obviously be better spent elsewhere.

1.5 SPECIFIC TYPES OF ONSITE SURVEYS

Rules of stratification, definition of sampling stages, and scheduling are the same for both access-point and roving surveys to estimate harvest, catch, or fishing effort (Pollock et al. 1994; Bernard et al. 1998). Calculations above the level of sampling periods are also the same. Statistics are averaged across sampled periods, then expanded for those periods not sampled within a stage or stratum. The same procedure is followed up the hierarchy of sampling stages to obtain an estimate for each stratum. Where these designs differ is in sampling procedures and calculations within each sampled period (see Chapter 2).

Surveys to estimate relative harvest composition, angler success, or mean CPUE as an index of abundance are actually two-stage stratified surveys with strata defined by an auxiliary variable, either time, location, or both (see Cohcran 1977:132-4). Because of proportional sampling in self-weighted surveys, strata can be ignored during calculations, and data simply pooled to estimate statistics for the season. In surveys to index abundance, seasonal statistics are not wanted, so statistics from each stratum stand alone.

1.6 REMAINING POTENTIAL FOR BIAS

Regardless of the sampling design chosen or how sampling is scheduled, there still remains the potential for bias in estimates from onsite creel surveys by not randomly selecting anglers to interview during sampled periods. ***Opportunity bias*** occurs when anglers not encountered represent different statistics than those that are. The related ***length-of-stay (LOS) bias*** happens when the probability of encountering an angler is related to his or her harvest rate (Pollock et al. 1994:71). Estimates of fishing effort are biased from ***sampling shadows*** when interviewing some anglers interferes with counting others (see Pollock et al. 1994:176 for an example).

In ***access-point surveys*** to estimate harvest, catch, or fishing effort, the sampling design reduces the target population for technicians to *be all anglers or harvested fish* ***exiting*** *the fishery during the sampled period*. If technicians interview all members of their target population, the angler or harvested fish ceases to be the basic sampling unit. The sampling period becomes the basic unit, which has been selected (scheduled) randomly, and so estimates of harvest, catch, fishing effort, or harvest composition will be unbiased. If not all anglers are interviewed during a sampled period, the estimated average statistic per angler is multiplied by the number of exiting anglers to estimate the total harvest, catch, or fishing effort for that sampled period. If there is a relationship between catch or harvest and the probability of an angler not being interviewed, the statistic for that sampled period will be biased in relation to the fraction of exiting anglers not interviewed. If relatively few anglers are not interviewed, bias should be relatively small.

In ***roving surveys***, the target population for technicians are *all anglers that* ***fished*** *during a sampled period.* Because some of these anglers exit the fishery unobserved by technicians during a sampled period, some can not be interviewed, and achieving a random sample of interviews becomes impossible. If there is a relationship between harvest rate or catch rate and the probability of an angler not being interviewed, estimated harvest rate or catch rate for that sampled period, and subsequently the estimate of harvest or catch, will be biased, often grossly so. There are two main strategies used to minimize the potential for this bias:

- set the sampling period to cover the fishing day; and

- stratify the target population by area with harvest rates of anglers more similar within strata than between strata.

The first strategy counteracts LOS bias in estimated harvest rates while the second counteracts opportunity bias. If technicians work throughout a fishing day with equal sampling effort, any temporal trend in harvest rates among exiting anglers within that day will be covered. If there is no spatial trend in harvest rates within a stratum, unequal probabilities of encountering a particular angler are of no concern. More on these strategies is given in Section 2.2.2.

Self-weighted surveys to estimate relative harvest composition or angler success represent the widest divergence between sampled and target populations of any type of onsite creel survey, and therefore the farthest deviation from random sampling. Most harvest or angler-trips will be unavailable to technicians, so theoretically the potential is great for bias in estimates from self-weighted surveys. In practice, however, bias is often negligible. In surveys to estimate relative harvest composition, the probability of angler being interviewed is generally unrelated to the relative composition of his/her harvest because there are little or no temporal or spatial trends in these statistics. Self-weighted surveys to estimate angler success are prone to many of the same biases as roving surveys because angler success is related to harvest rate. For this reason, the remedies are the same (see above). More on potential bias in self-weighted surveys and how to avoid them can be found in Sections 3.4 and 4.4.

CHAPTER TWO: Harvest, Catch, and Effort

Estimates of harvest, catch, and fishing effort from onsite surveys are used to regulate some of the biggest, most important sport fisheries in Alaska. Four annual, onsite surveys along the Kenai River have been used to estimate harvest of sockeye salmon, chinook salmon, and coho salmon. Fishing effort in these fisheries totaled an estimated 297 thousand angler-trips in 1995. The access-point survey of the fishery for sockeye salmon and the roving-access survey of the fishery for chinook salmon are used to attain escapement goals according to management plans. The roving-access survey on the fishery for coho salmon is an important element in a comprehensive approach to manage this stock to attain proper escapements to the Kenai River. Fishing effort in marine recreational fisheries for chinook salmon and coho salmon in Southeast Alaska annually tops 300 thousand angler-trips. Access-point surveys based in Juneau, Ketchikan, and Sitka are used to provide weekly and daily estimates that are critical in meeting guidelines for harvest of chinook salmon set by the Alaska Board of Fisheries and the governing body of the U.S.-Canada Salmon Treaty. Both access-point and roving-access surveys have been used to estimate harvests during the four-week spear fishery for whitefish in the Chatanika River. Annual harvests in this fishery have topped 25 thousand whitefish.

Access-point and roving-access surveys to estimate harvest, catch, and fishing effort in recreational fisheries are the topics of this chapter. Access-point surveys based on stratified two-, three-, and four-stage sampling designs are described, including methods to accommodate missing data. Roving-access surveys based on stratified, two-stage sampling designs are described for situations when access-point surveys are impractical. Recipes for handling missing data and covariances involved with some post-stratified data are given for roving-access surveys along with discussions on how to avoid and detect opportunity and length-of-stay (LOS) bias in roving surveys. Designs and methods are demonstrated with examples based on past onsite surveys in Alaska, including fisheries for whitefish on the Chatanika River, for sockeye salmon on the Russian River, for coho salmon on the Kenai River, and for chinook salmon on the Ninilchik River.

2.1 ACCESS-POINT SURVEYS

2.1.1 Harvest, Catch, and Effort. Estimating harvest, fishing effort, or catch in an access-point survey is described as a single procedure. Anglers are interviewed at the completion of their fishing trip when they exit the fishery. They are asked how many fish they caught (catch), how many they caught and kept (harvest), and how many hours or days they fished (effort). Their answers become the basic datum *y* for any stratified, two-, three-, or four-stage sampling design to estimate harvest, catch, or fishing effort:

	Two-Stage	Three-Stage	Four-Stage
	y_{hij}	y_{hijk}	y_{hijkl}
Harvest	H_{hij}	H_{hijk}	H_{hijkl}
Effort (hr)	e_{hij}	e_{hijk}	e_{hijkl}
Catch	c_{hij}	c_{hijk}	c_{hijkl}

h = stratum
i = first-stage unit
j = second-stage unit
k = third-stage unit
l = fourth-stage unit

In an access-point survey, harvest (or catch or fishing effort) for a sampled period is estimated by finding the average harvest (or catch or fishing effort) for sampled anglers, then expanding that average to ***all*** anglers exiting during the sampled period. If anglers are not post-stratified, these averages and expansions for a four-stage design are:

$$\hat{Y}_{hijk} = M_{hijk}\,\bar{y}_{hijk}$$

where $\bar{y}_{hijk}$ is the average for anglers interviewed during sampled period *hijk* and M_{hijk} is the count during that period of all exiting anglers. If anglers are post-stratified and exiting anglers can be assigned to a post-stratified group while being counted:

$$\hat{Y}_{h'ijk} = M_{h'ijk}\,\bar{y}_{h'ijk}$$

where h' represents the post-stratified group. If an exiting angler can not be assigned to a post-stratified group without an interview:

$$\hat{Y}_{h'ijk} = \hat{M}_{h'ijk}\,\bar{y}_{h'ijk}$$

$$\hat{M}_{h'ijk} = M_{hijk}\frac{m_{h'ijk}}{m_{hijk}}$$

where m_{hijk} is the number of anglers interviewed during a sampled period, and $m_{h'ijk}$ is the subset representing post-stratified group *h'*. Such prorated estimates of angler-trips are unbiased if ***probability of interviewing all types of exiting anglers is the same within each sampled period***. This condition must also hold if estimates of harvest, catch, and fishing effort are to be unbiased, regardless of post-stratification. The same equations listed above for sampled periods in four-stage sampling hold for sampled periods in simpler three and two-stage designs, only the number of subscripts are reduced accordingly.

Once harvest (or catch or fishing effort) has been estimated for each sampled period, calculations to estimate harvest for each stratum follow the equations in Table 2.1. These formulations are algebraically equivalent to those for stratified multistage sampling designs listed in Cochran (1977) when the penultimate sampling stage is the only stage with units of different sizes. For instance, in a four-stage sampling design with locations as first-stage units, fishing days as second-stage units, sampling periods within days as third-stage units, and anglers within sampling periods as fourth-stage units, only sampling periods are of different sizes in terms of the number of units in the next lower stage. Mathematically this means that all days $D_{hi} = D_h$ and all sampling periods $Q_{hij} = Q_h$ in stratum *h.* These constraints can be relaxed when designing surveys, however, such added complexity is hardly if ever needed, and when needed, better handled through stratification. The intent in sampling designs is that sample sizes within a stratum be equivalent as well; all $d_{hi} = d_h$ and all $q_{hij} = q_h$. However, sampling designs are not always followed (see Section 2.1.3), so equations have been written to incorporate differences in sample sizes (see Appendix C for equations).

Finally, estimated harvest, effort, or catch for the entire fishery, for the fishery up to a specific date, or for any part of the fishery are sums of statistics across appropriate strata as needed (see Equation 1.1). If post-stratification is involved in the survey but an estimate of totals across strata is needed, samples within each sampled period should be pooled to remove post-stratification, then combined estimates recalculated on the pooled data. The reason for this special treatment is that post-stratified estimates are not independently calculated when multistage sampling is involved. While this dependence is not critical to calculating unbiased estimates of harvest, catch, or fishing effort, it is to calculating unbiased estimates of their variances (see Example 2.4). Since the same sampled periods are used to obtain post-stratified estimates, the same sampling periods are involved in all calculations. Adding post-stratified variances would overestimate the true variance of the sum of post-stratified estimates of harvest, catch, or fishing effort. No such problem arises if post-stratified samples within sampled periods are pooled and estimates recalculated.

Example 2.1. In 1989, the spear fishery for whitefish on the Chatanika River started in early September and lasted to mid October. Harvest

Table 2.1 - Equations to estimate harvest, catch, or fishing effort with creel surveys based on stratified multistage sampling designs used in Alaska. Equations were adapted from those in Cochran (1977:Chapter 10) and Thompson (1992:Chapter 13) for multistage, equal probability sampling without replacement. All sampling units except sampling periods are of equal size within each sampling stage.

Two Stages (Day/Trip): **(2.1)**

$$\hat{Y}_h = D_h \hat{\bar{Y}}_h \; ; \qquad \hat{\bar{Y}}_h = \frac{\sum_{i=1}^{d_h} \hat{Y}_{hi}}{d_h}$$

Three Stages (Day/Period/Trip): **(2.2)**

$$\hat{Y}_h = D_h \hat{\bar{Y}}_h \; ; \qquad \hat{\bar{Y}}_h = \frac{\sum_{i=1}^{d_h} \hat{Y}_{hi}}{d_h} \; ; \qquad \hat{Y}_{hi} = Q_h \hat{\bar{Y}}_{hi} \; ; \qquad \hat{\bar{Y}}_{hi} = \frac{\sum_{j=1}^{q_{hi}} \hat{Y}_{hij}}{q_{hi}}$$

Three Stages (Location/Day/Trip): **(2.3)**

$$\hat{Y}_h = N_h \hat{\bar{Y}}_h ; \qquad \hat{\bar{Y}}_h = \frac{\sum_{i=1}^{n_h} \hat{Y}_{hi}}{n_h} ; \qquad \hat{Y}_{hi} = D_h \hat{\bar{Y}}_{hi} ; \qquad \hat{\bar{Y}}_{hi} = \frac{\sum_{j=1}^{d_{hi}} \hat{Y}_{hij}}{d_{hi}}$$

Four Stages (Location/Day/Period/Trip): **(2.4)**

$$\hat{Y}_h = N_h \hat{\bar{Y}}_h \; ; \quad \hat{\bar{Y}}_h = \frac{\sum_{i=1}^{n_h} \hat{Y}_{hi}}{n_h} ; \quad \hat{Y}_{hi} = D_h \hat{\bar{Y}}_{hi} \; ; \quad \hat{\bar{Y}}_{hi} = \frac{\sum_{j=1}^{d_{hi}} \hat{Y}_{hij}}{d_{hi}} ; \quad \hat{Y}_{hij} = Q_h \hat{\bar{Y}}_{hij} ; \quad \hat{\bar{Y}}_{hij} = \frac{\sum_{k=1}^{q_{hij}} \hat{Y}_{hijk}}{q_{hij}}$$

---

h = stratum	N = number of locations	n = number of locations sampled
i = first-stage unit	D = number of days	d = number of days sampled
j = second-stage unit	Q = number of sampling periods	q = number of periods sampled
k = third-stage unit	Y = statistic	
l = fourth-stage unit		

---

waxed and waned as the migratory whitefish entered, then left the fishing grounds. Fishing occurred nightly from about 1800 hours with almost all fishing completed by 2400 hours. The fishery was really three discrete fisheries with anglers reaching the river by a single road past Olnes Pond (downstream), along both banks near the Elliot Highway Bridge (midstream), and continuously along the Steese Highway (upstream).

An access-point, two-stage survey was conducted at Olnes Pond with fishing days (nights) as the first sampling stage and angler-trips as the second. Each night comprised a single sampling period. Nights were scheduled systematically for sampling across the duration of the fishery. All anglers were counted as they exited the fishery and a subset were opportunistically interviewed during each night of sampling (each sampled period). Eighteen first-stage units (nights) were sampled in all (d_h = 18) out of the 35 days in the fishery (D_h = 35). Sampling on two nights (3 and 11 October) was canceled due to uncontrollable circumstances and was rescheduled for adjacent evenings. The example begins with data collected on 17 September with interviews of seven anglers (m_{hi} = 7) with harvests of 0, 0, 4, 4, 0, 15, and 15 fish. Estimated mean harvest across angler-trips that evening is:

$$\hat{\bar{H}}_{hi} = 5.43 = \frac{0+0+4+4+0+15+15}{7}$$

Since 18 anglers were counted exiting the fishery that night near Olnes Pond (M_{hi} = 18), estimated harvest on 17 September ($\hat{H}_{hi}$) is 98 whitefish [=18(5.43)]. Statistics for other sampled periods are in Table 2.2. From Equation 2.1 in Table 2.1, estimated mean harvest across periods is:

$$\hat{\bar{H}}_{h} = 310 = \frac{10+16+30+98+\ldots+160+46}{18}$$

and the estimated harvest $\hat{H}_h$ for the entire stratum (Olnes Pond) is 10,850 whitefish [=35(310)]. An estimate of harvest for the entire fishery would be obtained by adding the estimates from this stratum to those from the other two strata (access at the Elliot Highway Bridge and along the Steese Highway) according to Equation 1.1. Calculating estimates of fishing effort for this fishery would follow the same pattern as that for harvest, only $y_{hij} \leftarrow e_{hij}$ instead of $y_{hij} \leftarrow H_{hij}$. Because this fishery is a spear fishery, $c_{hij} = H_{hij}$, and catch is harvest.

Table 2.2 - Average harvest $\hat{\bar{H}}_{hi}$ among interviewed anglers, number of anglers counted M_{hi}, and estimated harvest by period $\hat{H}_{hi}$ for all sampled periods during the creel survey of the fishery for whitefish near Olnes Pond on the Chatanika River in 1989.

	$\hat{\bar{H}}_{hi}$	M_{hi}	$\hat{H}_{hi}$		$\hat{\bar{H}}_{hi}$	M_{hi}	$\hat{H}_{hi}$
September				October			
11	1.00	10	10	1	8.73	53	463
13	16.00	1	16	2	8.13	27	220
15	0.96	31	30	5	10.72	38	407
17	5.43	18	98	7	8.44	82	692
19	4.44	23	102	9	6.50	21	137
21	8.48	49	416	12	10.73	28	300
23	6.36	62	394	13	4.58	35	160
25	10.58	57	603	15	3.29	14	46
27	10.47	59	618				
29	9.42	92	867				

Example 2.2 In 1991, an access-point survey was used to estimate the harvest of sockeye salmon at the Russian River. The survey design had two dimensions of stratification: access point (five elements) and location of fishing (two elements). Anglers were interviewed as they left the area through the (1) Ferry Parking Lot, (2) Grayling Parking Lot, (3) Rainbow Campground, (4) Pink Salmon Parking Lot, or (5) Red Salmon Campground. Also, anglers leaving were asked if they had fished on the Russian River proper (A) or at the confluence of the Russian River with the Kenai River (B). This post-stratification was used to segregate estimated harvest of stocks reproducing in the Russian River from harvests of a mixture of several stocks at the confluence. Harvests were estimated for each of these 10 strata (A1-A5 and B1-B5). Fishing days were first-stage

sampling units, sampling periods second-stage units, and angler-trips the third. Days were systematically selected for sampling in all five standard strata. Because a work-day was 7.5 hr long in 1991, the day was divided into six, 3-hr periods of which two were randomly selected for sampling (the remainder of the day was spent in transport to and from the fishery and in other duties).

Although the fishery and survey ran from July through August during passage of two runs of sockeye salmon, this example pertains only to the fishery on the later run which migrates through the fishery from late July to mid August. Two periods were sampled on 16 August at the Grayling Parking Lot. During the first sampled period (0900 - 1200 hours), 24 anglers exited the fishery of which 21 were interviewed; 16 of these fished on the river proper and 5 around its confluence with the Kenai River. During the second sampled period (1200-1500 hours), 47 anglers exited and 42 were interviewed of which 12 had fished on the river and 30 at its confluence. Prorated estimates $\hat{M}_{h'ij}$ are 18 [=24(16/21)] and 6 [=24(5/21) for anglers fishing at the river and at the confluence, respectively, during the first period. During the second period, prorated estimates are 13 [=47(12/42)] and 34 [=47(30/42)]. Average harvest among interviewed anglers exiting during these two sampled periods was then expanded to estimate the harvest by sampled period:

	River			Confluence		
	$\hat{\bar{H}}_{h'ij}$	$\hat{M}_{h'ij}$	$\hat{H}_{h'ij}$	$\hat{\bar{H}}_{h'ij}$	$\hat{M}_{h'ij}$	$\hat{H}_{h'ij}$
First Period:	(0.83)	18	= 15	(2.50)	6	= 15
Second Period:	(1.31)	13	= 17	(1.26)	34	= 43

Calculations proceeded according to Equations 2.2 in Table 2.1. Estimated average harvest by anglers exiting the Grayling Parking Lot in a sampling period and estimated harvest for 16 August are:

	$\hat{H}_{h'ij}$ Period 1st	2nd	q_{hi}	$\hat{\bar{H}}_{h'i}$	Q_h	$\hat{H}_{h'i}$
River	{[(15	+ 17)	/ 2 =	16] }	6	= 96
Confluence	{[(15	+ 43)	/ 2 =	29] }	6	= 174

When $\hat{H}_{h'i}$ for the other six days (sampled first-stage units) had been calculated, they were averaged across days sampled (d_h) to estimate mean harvest per day, which was then expanded by the number of days in each stratum (D_h) to estimate harvest by stratum:

$\hat{H}_{h'i}$	d_h	$\hat{\bar{H}}_{h'i}$	D_h	$\hat{H}_{h'}$

River:
{[(51+ 33+ 51+
42+ 39+ 81+ 96) / 7 = 56.14] } 22 = 1,235

Confluence :
{[(69+360+651+ 390+
306+117+174) / 7= 295.29] } 22 = 6,496

Similar calculations for other strata produced estimates of harvests for all strata (Table 2.3).

2.1.2 Estimating Variance. Equations to estimate variances for estimated harvest, catch, or effort for a stratum in an access-point survey are located in Table 2.4. Each variance equation contain a sampling fraction (*f*) for each sampling stage, correction factors for finite numbers of sampling units in each sampling stage (1- f), numbers of sampling units in each stage (*N, D, Q*, or *M*), numbers of units sampled per stage (*n, d, q*, or *m*), and sample variances s^2 or S^2 for each stage of sampling. The quantity 1- f is a finite population correction factor, an *fpc*.

Table 2.3 - Stratified estimates of harvest of sockeye salmon in 1991 on the Russian River and at its confluence with the Kenai River.

	River	Confluence
Ferry Parking Lot	0	19,451
Grayling Parking Lot	1,235	6,496
Rainbow Campground	394	132
Pink Salmon Parking Lot	3,878	314
Red Salmon Campground	1,518	0
Sums	7,025	26,393

When all sampling units in a stage are sampled, $f = 1$, the *fpc* $= 1 - f = 0$, and that component of estimated variance contributed by that sampling stage is zero. If one or less units in a stage are sampled, neither sample variance nor the corresponding component of estimated variance can be calculated. Two different formulations for the S_h^2 in Table 2.4 correspond to random and systematic selection of sampling units. If units were selected randomly without replacement, the sum of squared deviations from the mean is used to calculate the S_h^2 (left-hand choice in Table 2.4). If units were selected systematically, the square of differences in the sequence of samples is used (the right-hand choice). Because an unbiased estimate of variance can not be calculated from systematically drawn sample, variance for the estimate must be approximated. Wolter (1985) recommends the formulation in Table 2.4 as the best approximation in most instances, especially when data are autocorrelated or follow trends, as is often the case with migrating fish, such as salmon or whitefish. Sample variance s_h^2 for harvest, catch, or effort across angler-trips is calculated as if anglers had been randomly selected for sampling even though this never happens. Anglers are interviewed in sequence as they leave the fishery; some are not interviewed when more leave than technician(s) can interview at one time. Although this lack of random sampling of final-stage units (angler-trips) can impart some opportunity bias in estimated variance, that bias becomes negligible as more anglers are interviewed and the *fpc* approaches 0.

Example 2.3. This example is Example 2.1 revisited, only now to estimate variance for estimated harvest of whitefish near Olnes Pond. Nights were first-stage units (sampling periods) and angler-trips the second-stage units in this access-point survey. Eighteen nights (d_h=18) were chosen systematically for sampling without replacement in the 35-d fishery (D_h=35). Sampling on two nights (3 and 11 October) was canceled due to unforeseen circumstances and was rescheduled for adjacent evenings. During a sampled period, all anglers were counted (M_{hi}) as they exited the fishery while a subset (m_{hi}) was interviewed.

As in Example 2.1, calculations in this example begin with data collected on 17 September. On that night, seven anglers were interviewed and had harvested 0, 0, 4, 4, 0, 15, and 15 whitefish for an average harvest of 5.43 fish. By Equation 2.5 in Table 2.4:

$$s_{hi}^2 = 45.95 =$$

$$= \frac{\left[\begin{array}{c}(0-5.43)^2+(0-5.43)^2+(4-5.43)^2+(4-5.43)^2\\ +(0-5.43)^2+(15-5.43)^2+(15-5.43)^2\end{array}\right]}{7-1}$$

for 17 September. The sample fraction (f_{2h}) is 0.39 ($= m_{hi}/M_{hi}$=7/18) for that night, and the *fpc* is 0.61 ($=1-f_{2h}=$ 1–0.39). Sampling fractions, *fpc*, and s_{hi}^2 for all days are listed in Table 2.5. For 17 September, the bracketed term in Table 2.5 (the bracketed term in Equation 2.5) is 1,297 [$=18^2(1-0.39)(45.95)/7$]. The sum of the bracketed terms across all sampled days is 16,907. Since 18 ($=d_h$) of a possible 35 days ($=D_h$) were sampled, the sampling fraction for first-stage units is 0.51 ($=f_{1h}=d_h/D_h$). The product of the reciprocal of this sampling fraction and the sum of the bracketed terms is 33,151 [=(1/0.51)16,907], which is the ***second-stage component of estimated variance.***

Table 2.4 - Equations to estimate variance of estimated harvest, catch, or fishing effort with access-point creel surveys based on stratified multistage sampling designs used in Alaska. Formulations for sample variances S_h^2 corresponds to sampling units having been selected randomly or systematically.

Two Stages (Day/Trip): **(2.5)**

$$var(\hat{Y}_h) = (1 - f_{1h})D_h^2 \frac{S_{1h}^2}{d_h} + f_{1h}^{-1} \sum_{i=1}^{d_h} \left[M_{hi}^2 (1 - f_{2hi}) \frac{s_{2hi}^2}{m_{hi}} \right] \qquad f_{1h} = \frac{d_h}{D_h} \qquad f_{2hi} = \frac{m_{hi}}{M_{hi}}$$

$$s_{2hi}^2 = \frac{\sum_{j=1}^{m_{hi}} (y_{hij} - \bar{y}_{hi})^2}{m_{hi} - 1} \qquad S_{1h}^2 = \frac{\sum_{i=1}^{d_h} (\hat{Y}_{hi} - \bar{\hat{Y}}_h)^2}{d_h - 1} \text{ or } \frac{\sum_{i=2}^{d_h} (\hat{Y}_{hi} - \hat{Y}_{h(i-1)})^2}{2(d_h - 1)}$$

Three Stages (Day/Period/Trip): **(2.6)**

$$var(\hat{Y}_h) =$$

$$(1 - f_{1h})D_h^2 \frac{S_{1h}^2}{d_h} + f_{1h}^{-1} Q_h^2 \sum_{i=1}^{d_h} \left[(1 - f_{2hi}) \frac{S_{2hi}^2}{q_{hi}} \right] + f_{1h}^{-1} \sum_{i=1}^{d_h} \left[f_{2hi}^{-1} \sum_{j=1}^{q_{hi}} \left[M_{hij}^2 (1 - f_{3hij}) \frac{s_{3hij}^2}{m_{hij}} \right] \right]$$

$$s_{3hij}^2 = \frac{\sum_{k=1}^{m_{hij}} (y_{hijk} - \bar{y}_{hij})^2}{m_{hij} - 1} \qquad S_{2hi}^2 = \frac{\sum_{j=1}^{q_{hi}} (\hat{Y}_{hij} - \bar{\hat{Y}}_{hi})^2}{q_{hi} - 1} \text{ or } \frac{\sum_{j=2}^{q_{hi}} (\hat{Y}_{hij} - \hat{Y}_{hi(j-1)})^2}{2(q_{hi} - 1)}$$

$$S_{1h}^2 = \frac{\sum_{i=1}^{d_h} (\hat{Y}_{hi} - \bar{\hat{Y}}_h)^2}{d_h - 1} \text{ or } \frac{\sum_{i=2}^{d_h} (\hat{Y}_{hi} - \hat{Y}_{h(i-1)})^2}{2(d_h - 1)} \qquad f_{1h} = \frac{d_h}{D_h} \qquad f_{2hi} = \frac{q_{hi}}{Q_h} \qquad f_{3hij} = \frac{m_{hij}}{M_{hij}}$$

---

f_{1h} = fraction of first-stage units sampled

f_{2hi} = fraction of second-stage units sampled

f_{3hij} = fraction of third-stage units sampled

f_{4hijk} = fraction of fourth-stage units sampled

s_h^2 = last-stage sample variance for measurements

S_h^2 = first, second, or third-stage sample variance for totals

---

Table 2.4 - Equations to estimate variance of estimated harvest, catch, or fishing effort with access-point creel surveys based on stratified multistage sampling designs used in Alaska (continued).

Three Stages (Location/Day/Trip): **(2.7)**

$$var(\hat{Y}_h) =$$

$$(1-f_{1h})N_h^2\frac{S_{1h}^2}{n_h} + f_{1h}^{-1}D_h^2\sum_{i=1}^{n_h}\left[(1-f_{2hi})\frac{S_{2hi}^2}{d_{hi}}\right] + f_{1h}^{-1}\sum_{i=1}^{n_h}\left[f_{2hi}^{-1}\sum_{j=1}^{d_{hi}}\left[M_{hij}^2(1-f_{3hij})\frac{s_{3hij}^2}{m_{hij}}\right]\right]$$

$$s_{3hij}^2 = \frac{\sum_{k=1}^{m_{hij}}(y_{hijk}-\bar{y}_{hij})^2}{m_{hij}-1} \qquad S_{2hi}^2 = \frac{\sum_{j=1}^{d_{hi}}(\hat{Y}_{hij}-\bar{\hat{Y}}_{hi})^2}{d_{hi}-1} \text{ or } \frac{\sum_{j=2}^{d_{hi}}(\hat{Y}_{hij}-\hat{Y}_{hi(j-1)})^2}{2(d_{hi}-1)}$$

$$S_{1h}^2 = \frac{\sum_{i=1}^{n_h}(\hat{Y}_{hi}-\bar{\hat{Y}}_h)^2}{n_h-1} \text{ or } \frac{\sum_{i=2}^{n_h}(\hat{Y}_{hi}-\hat{Y}_{h(i-1)})^2}{2(n_h-1)} \qquad f_{1h} = \frac{n_h}{N_h} \quad f_{2hi} = \frac{d_{hi}}{D_h} \quad f_{3hij} = \frac{m_{hij}}{M_{hij}}$$

Four Stages (Location/Day/Period/Trip): **(2.8)**

$$var(\hat{Y}_h) = (1-f_{1h})N_h^2\frac{S_{1h}^2}{n_h} + f_{1h}^{-1}D_h^2\sum_{i=1}^{n_h}\left[(1-f_{2hi})\frac{S_{2hi}^2}{d_{hi}}\right] +$$

$$f_{1h}^{-1}Q_h^2\sum_{i=1}^{n_h}\left[f_{2hi}^{-1}\sum_{j=1}^{d_{hi}}\left[(1-f_{3hij})\frac{S_{3hij}^2}{q_{hij}}\right]\right] + f_{1h}^{-1}\sum_{i=1}^{n_h}\left[f_{2hi}^{-1}\sum_{j=1}^{d_{hi}}\left[f_{3hij}^{-1}\sum_{k=1}^{q_{hij}}\left[M_{hijk}^2(1-f_{4hijk})\frac{s_{4hijk}^2}{m_{hijk}}\right]\right]\right]$$

$$s_{4hijk}^2 = \frac{\sum_{l=1}^{m_{hijk}}(y_{hijkl}-\bar{y}_{hijk})^2}{m_{hijk}-1} \qquad S_{3hij}^2 = \frac{\sum_{k=1}^{q_{hij}}(\hat{Y}_{hijk}-\bar{\hat{Y}}_{hij})^2}{q_{hij}-1} \text{ or } \frac{\sum_{k=2}^{q_{hij}}(\hat{Y}_{hijk}-\hat{Y}_{hij(k-1)})^2}{2(q_{hij}-1)}$$

$$S_{1h}^2 = \frac{\sum_{i=1}^{n_h}(\hat{Y}_{hi}-\bar{\hat{Y}}_h)^2}{n_h-1} \text{ or } \frac{\sum_{i=2}^{n_h}(\hat{Y}_{hi}-\hat{Y}_{h(i-1)})^2}{2(n_h-1)} \quad S_{2hi}^2 = \frac{\sum_{j=1}^{d_{hi}}(\hat{Y}_{hij}-\bar{\hat{Y}}_{hi})^2}{d_{hi}-1} \text{ or } \frac{\sum_{j=2}^{d_{hi}}(\hat{Y}_{hij}-\hat{Y}_{hi(j-1)})^2}{2(d_{hi}-1)}$$

$$f_{1h} = \frac{n_h}{N_h} \qquad f_{2hi} = \frac{d_{hi}}{D_h} \qquad f_{3hij} = \frac{q_{hij}}{Q_h} \qquad f_{4hijk} = \frac{m_{hijk}}{M_{hijk}}$$

Calculating the ***first-stage component of estimated variance*** begins with a sample variance for estimated harvest by night (sampling periods). Estimated harvests by sampled periods ($\hat{H}_{hi}$) at Olnes Pond are listed in Table 2.2. Because nights were scheduled for sampling systematically, sums of the squared differences of consecutive estimates using Equation 2.3 approximated first-stage sample variance according to procedures in Wolter (1985):

$$S_{1h}^2 = \frac{(16-10)^2 + (30-16)^2 + (98-30)^2 + \ldots + (300-137)^2 + (160-300)^2 + (46-160)^2}{2(18-1)}$$

$$= 26{,}927$$

There were two aberrations from a strictly systematically drawn sample: sampling on 3 and 11 October were rescheduled to 2 and 12 October. For the sake of this demonstration, calculating first-stage sample variance proceeded as if there had been no rescheduling. For an unbiased approximation of sample variance, the appropriate approach would be to treat the lack of data on 3 and 11 October as missing data (and ignore data collected on 2 and 12 October). This approach is discussed in Section 2.1.3.

Expansion of first-stage sample variance S_{1h}^2 to become the first-stage component of estimated variance followed Equation 2.5:

$$(1-f_{1h})D_h^2 \frac{S_{1h}^2}{d_h} = (1-0.51)(35)^2 \frac{26{,}927}{18} = 897{,}941$$

The sum of first and second-stage components of estimated variance (897,941 and 33,151) is the estimated variance for the stratum, 931,072 which makes the SE = 965 whitefish for the estimated harvest near Olnes Pond of 10,850.

* * * *

When post-stratification is involved in a survey, variance is somewhat larger than given in Table 2.4. When angler-trips are post-stratified, sample sizes are no longer considered fixed, but become variables. Variance is increased by a component that is a function of the reciprocal of the square of the sample size across a strata (Thompson 1992:109). Cochran (1977:134) showed that this component of variance is negligible whenever sample sizes are >20 over all strata. Because more than 20 anglers are interviewed in our creel surveys, we suggest equations in Table 2.4 can be used regardless of the type of stratification in the survey.

Table 2.5 - Sampling statistics for the access-point survey of the fishery for whitefish near Olnes Pond on the Chatanika River for 1989. The term [......] corresponds to the bracketed operations in Equation 2.5.

	m_{hi}	M_{hi}	f_{2hi}	$1-f_{2hi}$	s_{2hi}^2	[......]
September						
11	8	10	0.80	0.20	1.71	4.28
13	1	1	1.00	0.00	0.00	0.00
15	25	31	0.81	0.19	2.20	16.07
17	7	18	0.39	0.61	45.95	1,297.37
19	16	23	0.70	0.30	22.00	218.21
21	33	49	0.67	0.33	30.13	723.42
23	11	62	0.18	0.82	24.25	6,948.90
25	33	57	0.58	0.42	39.06	1,615.17
27	49	59	0.83	0.17	21.42	258.69
29	48	92	0.52	0.48	27.83	2,355.53
October						
1	22	53	0.42	0.58	27.73	2,053.56
2	23	27	0.85	0.15	20.75	98.65
5	25	38	0.66	0.34	19.96	391.98
7	59	82	0.72	0.28	22.01	702.35
9	16	21	0.76	0.24	8.13	53.78
12	26	28	0.93	0.07	31.16	65.77
13	31	35	0.89	0.11	23.78	103.37
15	14	14	1.00	0.00	10.84	0.00
Sum						16,907.09

Example 2.4. This example is Example 2.2 revisited, only now to estimate variance for estimated harvest of sockeye salmon from the late-run to the Russian River. Because data were post-stratified according to where anglers had fished,

calculations are presented for two strata: anglers who fished on the Russian River proper and exited through Grayling Campground, and anglers who exited through this same campground, but who had fished at the confluence of the Russian and Kenai Rivers. Equation 2.6 in Table 2.4 is used to estimate variance because this survey followed a stratified three-stage design with days as first-stage units, sampling periods as second-stage units, and angler-trips as third-stage units. To simplify the example, Equation 2.6 has been decomposed into three components, one for each stage in the design. Calculations begin with the two sampled periods on 16 August:

River			Confluence		
$\hat{M}_{h'ij}$	$m_{h'ij}$	$s^2_{3h'ij}$	$\hat{M}_{h'ij}$	$m_{h'ij}$	$s^2_{3h'ij}$
18	16	1.496	6	5	0.800
13	12	1.477	34	30	1.444

where the number of anglers are considered estimates because of post-stratification (see Example 2.2). Because the desire here is to describe method and not how all statistics were calculated, statistics from days other than 16 August are included in the calculations as needed without description. Our demonstration continues with those parts of the ***third-stage component of estimated variance*** due to sampled periods on 16 August:

$$\sum_{j=1}^{q_{hi}}\left[\hat{M}^2_{h'ij}(1-\hat{f}_{3h'ij})\frac{s^2_{3h'ij}}{m_{h'ij}}\right] =$$

$$18^2\left[1-\frac{16}{18}\right]\frac{1.496}{16}+13^2\left[1-\frac{12}{13}\right]\frac{1.477}{12} = 4.966$$

$$6^2\left[1-\frac{5}{6}\right]\frac{0.800}{5}+34^2\left[1-\frac{30}{34}\right]\frac{1.444}{30} = 7.506$$

The top set of calculations corresponds to the stratum associated with harvest by anglers who had fished in the Russian River proper while the bottom corresponds to harvest by anglers who had fished at the confluence of the two rivers. This dual format is used to describe all calculations for both strata below. The complete formulation to calculate the third-stage component of estimated variance is an expansion over all days in the stratum:

$$f_{1h}^{-1}f_{2h}^{-1}\sum_{i=1}^{d_h}\left[\sum_{j=1}^{q_h}\left[\hat{M}^2_{h'ij}(1-\hat{f}_{3h'ij})\frac{s^2_{3h'ij}}{m_{h'ij}}\right]\right] =$$

$$\frac{22}{7}\frac{6}{2}\left[\begin{matrix}3.544+15.960+1.906+17.264+\\2.632+7.804+4.966\end{matrix}\right] = 510$$

$$\frac{22}{7}\frac{6}{2}\left[\begin{matrix}10.347+100.416+11.747+\\91.271+24.091+12.669+7.506\end{matrix}\right] = 2{,}433$$

Note that Equation 2.6 was simplified for this demonstration by taking advantage of $q_{hi} = q_h = 2$ for all days, making $f_{2hi} = f_{2h} = 6/2$.

Because periods were randomly scheduled for sampling without replacement, second-stage sample variance is based on sums of squared deviations from the mean of harvest across sampled periods:

$$S^2_{2h'i} = \frac{\sum_{j=1}^{q_h}(\hat{H}_{h'ij}-\hat{\bar{H}}_{h'i})^2}{q_h-1} =$$

$$\frac{(15-16)^2+(17-16)^2}{2-1} = 2.0$$

$$\frac{(15-29)^2+(43-29)^2}{2-1} = 392.0$$

Second-stage component of estimated variance is a function of the second-stage sample variances corrected for the finite number of sampling periods in a day, and expanded:

$$f_{1h}^{-1}Q_h^2(1-f_{2h})\sum_{i=1}^{d_h}\left[\frac{S^2_{2h'i}}{q_h}\right] =$$

$$\frac{22}{7}6^2\left[1-\frac{2}{6}\right]\left[\frac{0.5}{2}+\frac{0.5}{2}+\frac{0.5}{2}+\frac{72.0}{2}+\right.$$

$$\left.+\frac{60.5}{2}+\frac{220.5}{2}+\frac{2.0}{2}\right] = 13{,}445$$

$$\frac{22}{7}6^2\left[1-\frac{2}{6}\right]\left[\frac{40.5}{2}+\frac{162.0}{2}+\frac{264.5}{2}+\frac{512.0}{2}+\right.$$

$$\left.+\frac{648.0}{2}+\frac{364.5}{2}+\frac{392.0}{2}\right] = 89{,}892$$

Because days were systematically sampled, first-stage sample variance is based on sums of squared differences between consecutive estimates:

$$S_{1h'}^2 = \frac{\sum_{i=2}^{d_h}(\hat{H}_{h'i} - \hat{H}_{h'(i-1)})^2}{2(d_h - 1)} =$$

$$\frac{\left[\begin{array}{c}(33-51)^2+(51-33)^2+(42-51)^2+\\(39-42)^2+(81-39)^2+(96-81)^2\end{array}\right]}{2(7-1)} = 227$$

$$\frac{\left[\begin{array}{c}(360-69)^2+(651-360)^2+\\(390-651)^2+(306-390)^2+\\(117-306)^2+(174-117)^2\end{array}\right]}{2(7-1)} = 23{,}626$$

The resulting sample variance was then expanded to produce the first-stage components of estimated variance:

$$(1-f_{1h})D_h^2\frac{S_{1h'}^2}{d_h} =$$

$$\left[1-\frac{7}{22}\right]22^2\frac{227}{7} = 10{,}701$$

$$\left[1-\frac{7}{22}\right]22^2\frac{23{,}626}{7} = 1{,}113{,}797$$

Putting all three components of estimated variance together in a single equation produces Equation 2.6. The corresponding calculations of estimated variance for the estimated harvest by anglers exiting the fishery at Grayling Campground are:

$$var(\hat{H}_{h'}) =$$

$$= 10{,}701 + 13{,}445 + 510 = 24{,}656$$

$$= 1{,}113{,}797 + 89{,}892 + 2{,}433 = 1{,}206{,}122$$

Estimated variances for estimated harvest by anglers exiting the fishery through the other four access points were calculated with the same procedures:

	River	Confluence
Ferry Parking Lot	0	11,675,394
Grayling Parking Lot	24,656	1,206,122
Rainbow Campground	53,651	232
Pink Salmon Parking Lot	2,897,219	29,423
Red Salmon Campground	1,556,204	0
Total	4,531,730	12,911,171
SE	2,129	3,593

Estimated standard error (SE) is the square root of an estimated variance, and coefficient of variation (CV) is the ratio of a SE to its statistic in percent. Since estimated harvests from the river and from the confluence are 7,025 and 26,843 sockeye salmon, respectively, the CVs for the two estimates are 30% and 13%.

A harvest estimate for the entire fishery would not be obtained by summing estimates for post-stratified groups fishing at the confluence and the river (as per Equation 1.1), but would be obtained by pooling information from interviews within each sampled period and recalculating statistics. In short, information on where exiting anglers had fished on the Russian River in 1991 would be ignored. Data can be pooled without fear of bias because anglers that fished at the confluence in this example would have the same probability of being encountered in an access-point survey as would those that fished in the river. Data collected on 16 August would now be:

First Period			Second Period		
M_{hij}	m_{hij}	s_{3hij}^2	M_{hij}	m_{hij}	s_{3hij}^2
24	21	1.890	47	42	1.418

Calculations would proceed from this point according to Equation 2.6 in Table 2.4.

2.1.3. Missing Data. Occasionally a technician is ill, sleeps through the alarm clock, has car trouble, forgets, or is instructed incorrectly with the result that one or more sampling periods scheduled for sampling are not. Because this kind of "failure" is usually independent of fishing effort in the fishery, remaining sampled periods can still produce unbiased estimates so long as failures are few relative to the planned sample size and are spread throughout the fishery. Under these circumstances, missed periods can be ignored as if they had never been scheduled for sampling. Equations in Tables 2.1 and 2.4 are still appropriate just so long as sampling levels (the n_h, $d_{h(i)}$, and $q_{h(ij)}$) are adjusted downward for "missing data" at the appropriate stages. When samples are drawn systematically, calculation of sample variance will have *two* less squared differences between consecutive data for each missing sample. An adjusted calculation of sample variance would be:

$$S^2 \cong \frac{\sum_{i=2}^{\tilde{n}} \phi_i \phi_{i-1} (\hat{Y}_i - \hat{Y}_{i-1})^2}{2\sum_{i=2}^{\tilde{n}} \phi_i \phi_{i-1}}$$

where $\tilde{n}$ is the number of periods scheduled to be sampled, $\phi_i = 1$ if scheduled period i was sampled and 0 if not, and the actual sample size $n = \sum_{i=1}^{\tilde{n}} \phi_i$. Regardless of how sampled periods were selected, the entire stage should be treated as “missing” if one or fewer periods are sampled. ***A datum should not be treated as a mean.*** The same rule should apply to higher stages in the design.

Example 2.5. Two systematically scheduled sampled periods were missed in 1989 during the onsite creel survey to estimate the harvest of whitefish from the Chatanika River near Olnes Pond (see Examples 2.1 and 2.3). Although sampling was rescheduled, the pattern of systematic sampling was broken. For reasons provided in Example 2.3, this departure from systematic sampling was ignored. However, the "by-the-book" response is to ignore data from rescheduled nights (sampled periods) and treat the aberration as resulting from missing data. With data "missing" from scheduled sampling periods on 3 and 11 October, estimated harvests by sampled period ($\hat{H}_{hi}$) are:

Sep	$\hat{H}_{hi}$	Sep	$\hat{H}_{hi}$	Oct	$\hat{H}_{hi}$	Oct	$\hat{H}_{hi}$
11	10	21	416	1	463	11	miss
13	16	23	394	3	miss	13	160
15	30	25	603	5	407	15	46
17	98	27	618	7	692		
19	102	29	867	9	137		

This loss of two nights of sampling reduced d_h from 18 to 16 and changed the sampling fraction f_{1h} from 0.51 (=18/35) to 0.46 (=16/35). The new estimate of mean harvest across all sampled periods is:

$$\hat{\bar{H}}_h =$$

$$\frac{\begin{bmatrix} 10+16+30+\ldots+618+867+463+ \\ 407+692+137+160+46 \end{bmatrix}}{16} = 316$$

and estimated harvest near Olnes Pond 11,060 whitefish [= $D_h \hat{\bar{H}}_h$ =35(316)] (see Equation 2.1).

Calculations for estimating both components of variance were adjusted for "missing" data. From Table 2.5, the sum of the values from operations in the bracketed terms is 16,743 without data from the "missing" periods on 2 and 12 October. The second-stage component of estimated variance becomes 36,397 [=(1/0.46)16,743]. First-stage sample variance was calculated without four squared differences that correspond to two missing data:

$$S_{1h}^2 \cong 29{,}820 =$$

$$\frac{\left[\begin{array}{c} 1(1)(16-10)^2 + 1(1)(30-16)^2 + \ldots + 1(1)(137-692)^2 \\ +0(1)(?-137)^2 + 1(0)(160-?)^2 + 1(1)(46-160)^2 \end{array}\right]}{2[1(1)+1(1)+\ldots+1(1)+0(1)+1(0)+1(1)]}$$

Final calculation of the first-stage component of estimated variance is:

$$(1-f_{1h})D_h^2 \frac{S_{1h}^2}{d_h} =$$

$$(1-0.46)35^2 \frac{29{,}820}{16} = 1{,}232{,}871$$

The sum of its two components produces the estimated variance of harvest for the stratum, which is 1,269,268 (= 1,232,871 + 36,397). The SE = 1,127 whitefish for an estimated harvest near Olnes Pond of 11,060. When aberrations in systematic sampling were conveniently ignored (Examples 2.1 and 2.3), estimated harvest of whitefish and its SE are 10,850 and 965, respectively.

* * * *

Another kind of missing data is not independent of fishing effort. Occasionally fishing is so sporadic that no angler-trips are sampled during a sampled period. When M_h = m_h = 0 for a sampled period, calculations of harvest, catch, effort, and number of angler-trips are unaffected. In calculating variances for these statistics, the *fpc* for the sampled period is 0, thereby avoiding the embarrassment of trying to divide by m_h = 0. However, when $M_h > m_h = 0$ for a sampled period, there is no information on catch, harvest, or effort from this period as a result of a sampling failure. Fortunately, such instances are extremely rare, and only occur when M_h is extremely small and by rare chance the few anglers encountered are uncooperative. In this instance, arbitrarily setting catch, harvest, and fishing effort to 0 for the sampled period will bias estimates downward for the fishery, but only negligibly. Again, the *fpc* for this sampled period is set to 0.

Sampling "failures" should be an infrequent occurrence in any onsite survey. In small surveys, effects of a few missed sampling periods can cascade through the sampling design. As sample sizes drop, precision degrades, and potential bias increases. Calculations become more subjective and complex as sampling failures become more frequent. If sampling failures have been common in past onsite surveys of a fishery, the survey is in need of redesign, the fishery is too small to warrant an onsite creel survey, or performance of technicians needs better evaluation.

2.2 ROVING-ACCESS SURVEYS

Roving surveys are used to estimate harvest in recreational fisheries when access points are too numerous to practically conduct an access-point survey. Fishing effort is estimated for each sampled period as the product of counted anglers and the length of the period in hours. Estimated fishing effort is then multiplied by an estimate of harvest rate in fish per hour for that sampled period to produce an estimate of harvest for the period. Counting all *fishing* anglers and expanding these counts by a harvest rate in a roving survey compensates for a technician's inability to count all *exiting* anglers during a sampled period.

Roving-access surveys are preferred over roving-roving surveys because the former are less prone to bias. In "traditional" roving-roving creel surveys, anglers are interviewed during counts to estimate fishing effort. Whenever the pace of counting is slowed by taking "incompleted-trip" interviews, subsequent estimates of fishing effort will be biased low (see Pollock et al. 1994:176). This "shadow" in counts due to interviewing can be avoided in roving-roving surveys if one technician counts unabated while others pass through the fishery interviewing anglers. Early roving-roving surveys to estimate harvest of chinook salmon on the Kenai River used this technique. In roving-access surveys, no "shadow bias" in estimates of fishing effort occurs because counting and interviewing anglers are separate activities.

Both roving-access and roving-roving surveys are prone to LOS bias (Pollock et al. 1994:251; Bernard et al. 1998), especially when harvest in a recreational fishery is restricted with a daily bag limit. Estimated harvest will tend to be biased low in roving-roving surveys and could be biased either way in roving-access surveys. No remedy is known for LOS bias in roving-roving surveys other than discarding this type of survey in favor of a roving-access survey (Pollock et al. 1994:251). The remedy for LOS bias in roving-access surveys is to equate the length of the sampling period to that of the fishing day (Bernard et al. 1998).

2.2.1 Fishing Effort for a Sampled Period. In a roving survey, equations to estimate fishing effort expended during each sampled period and its estimated variance are:

$$\hat{E} = T\,\bar{x} \qquad var(\hat{E}) = T^2 var(\bar{x}) \qquad (2.9)$$

where $\hat{E}$ = estimated fishing effort in hours, $\bar{x}$ = average number of anglers counted fishing, and T is the number of hours in a sampling period. Average numbers ($\bar{x}$) are usually based on three or more systematically scheduled passes through the fishery during a sampled period. The sampling period is divided into segments based on how many counts (passes through the fishery) are desired. For instance, if a sampling period is 6 hr long, four passes are desired, and each pass takes 30 min to complete, the first count (pass) can begin anytime within the first 60 min of the sampled period, but the next three will begin at exactly 90-min intervals after the beginning of the previous count. If three counts are desired, the first begins anytime within the first 90 min and the other two at exact intervals of 2 hr. If five counts are desired, the first count begins anytime within the first 42 min and the other four passes occur at 72-min intervals. The systematic pattern of times for counting spreads counts across sampled periods, thereby minimizing potential for bias from small sample sizes when size of the angling population changes during a sampled period. At least three counts per sampled period are needed to approximate sample variance for systematically drawn samples. The maximum number of counts has no theoretical limit, although a practical one is obvious.

How to calculate sample variance for estimated fishing effort is described as part of Equation 2.9 as the product of the sample variance for the mean number of anglers counted in a sampled period and the square of the length of the period in hours. Estimating variance for the mean number of anglers follows the formulation of Wolter (1985) for approximating a variance from a systematically drawn sample of r counts:

$$var(\bar{x}) = \frac{\sum_{t=2}^{r} (x_t - x_{(t-1)})^2}{2r(r-1)} \qquad (2.10)$$

Example 2.6. In 1989, a roving-access, two-stage survey near the Elliot Highway Bridge across the Chatanika River was used to estimate harvest of whitefish in a spear fishery from early September through mid October (see Example 2.1 for more details on the fishery and the creel survey in 1989). Fishing days (nights) were first-stage sampling units and angler-trips and counts were the second. Each night comprised a single sampling period. Nights were scheduled systematically for sampling across the duration of the fishery. There was no stratification in this survey. Anglers were counted six times at hourly intervals that evening for an average number of anglers fishing during the period. For 10 October:

$$\bar{x} \;=\; \frac{4+5+6+7+3+0}{6} \;=\; 4.167$$

From Equation 2.10:

$$var(\bar{x}) = 0.4667 =$$

$$= \frac{\begin{bmatrix}(5-4)^2 + (6-5)^2 + (7-6)^2 + \\ (3-7)^2 + (0-3)^2\end{bmatrix}}{2(6)(6-1)}$$

Since sampling periods (nights) were 6-hr long, estimated fishing effort for 10 October is 25 hr [=6(4.167)] according to Equation 2.9 and its estimated variance 16.8 [$=6^2(0.4667)$].

2.2.2 Mean Harvest Rate. Estimated mean harvest rate ($\overline{hpue}$) for a sampled period is obtained by asking anglers "How long have you fished today?" and "How many fish have you caught and kept during that time?" during an interview. Answers to these questions produce two measurements for each angler: harvest in fish and the length of the angler-trip in hours. In a "traditional" roving-roving survey, anglers are contacted while they are engaged in fishing to produce an "incompleted-trip" interview. In a roving-access survey, anglers are interviewed as they exit the fishery at the end of their fishing trip to produce "completed-trip" interviews.

When information is gathered through "completed-trip" interviews, the best estimate of mean harvest rate is the ratio of means for harvest and trip length (Jones et al. 1995; Hoenig et al. 1997). For each sampled period:

$$\overline{hpue} = \frac{\hat{\bar{H}}}{\bar{e}} = \frac{\sum_{j=1}^{m} H_j}{\sum_{j=1}^{m} e_j} \qquad (2.11a)$$

where H_j is harvest by interviewed angler j, e_j is the length of his or her trip, and m is the number of anglers interviewed in the sampled period. From Thompson (1992:61-2):

$$var(\overline{hpue}) = \frac{\sum_{j=1}^{m} (H_j - e_j \overline{hpue})^2}{\bar{e}^2 m(m-1)} \qquad (2.11b)$$

Estimators based on a ratio of means as above have a sampling-induced bias of magnitude m^{-1} when harvest and effort are perfectly correlated (Cochran 1977:162). Experience has shown for onsite creels surveys that this bias is negligible (Appendix D).

Example 2.7. Only data from fishery near the Elliot Highway Bridge (see Example 2.6) on 10 October were used to demonstrate estimating mean harvest rates. Sixteen persons were interviewed as they left the river at the campground off the Elliot Highway that night (Table 2.6). By Equation 2.11, estimated mean harvest rate and its estimated variance are:

$$\overline{hpue} = \frac{165}{32} = 5.1625$$

Table 2.6 - Harvest, fishing effort, and interim calculations for anglers interviewed while exiting the fishery for whitefish near the Elliot Highway Bridge across the Chatanika River on the night of 10 October, 1989.

H_j	e_j	$(H_j - e_j \overline{hpue})^2$
15	3.0	0.22
22	3.0	42.66
11	2.0	0.47
8	2.0	5.35
13	2.0	7.22
11	2.0	0.47
11	2.0	0.47
11	2.0	0.47
14	2.0	13.60
3	1.5	22.41
3	1.5	22.41
7	1.0	3.40
4	1.0	1.34
5	1.0	0.02
12	3.0	12.03
15	3.0	0.22
Sums:		
165	32	132.78

$$var(\overline{hpue}) = \frac{132.78}{\left(\frac{32}{16}\right)^2 16\,(16-1)} = 0.1383$$

2.2.2.1 Avoiding LOS Bias. Conditions for accurately estimating mean harvest rate from onsite interviews is that ***every angler fishing during a sampled period has an equal chance of being interviewed*** or ***that mean harvest rate is the same for sampled and unsampled anglers***. In roving-roving surveys, the chance of interviewing an angler that has completed his or her fishing trip is considerably less than interviewing an actively fishing angler. If mean harvest rates are different for these two groups of anglers, estimates of mean harvest rate and subsequently estimated harvest will be biased. One way of detecting this

Table 2.7 - Statistics calculated from information gathered during "incompleted-trip" and "completed-trip" interviews of anglers in the spear fishery for whitefish near the Elliot Highway Bridge across the Chatanika River in 1989.

	Interview	m	$\overline{hpue}$	$\bar{e}$		Interview	m	$\overline{hpue}$	$\bar{e}$
16-Sep	Incompleted	4	0.00	0.58	5-Oct	Incompleted	6	2.24	2.13
	Completed	20	0.05	1.11		Completed	13	3.07	1.88
22-Sep	Incompleted	7	0.00	1.64	6-Oct	Incompleted	11	1.02	2.05
	Completed	30	0.17	1.91		Completed	41	2.11	3.07
30-Sep	Incompleted	8	2.12	2.13					
	Completed	26	2.86	2.89					

problem is to take both "incompleted-trip" and "completed-trip" interviews and compare estimates of mean harvest rate (Malvestuto 1983:405; Pollock et al. 1994:179).

Example 2.8. During the roving survey in 1989 near the Elliot Highway Bridge, both "incompleted-" and "completed-trip" interviews were taken during five sampled periods (nights). Mean harvest rate was estimated according to Equation 2.11 for each type of interview (Table 2.7). In all five nights, $\overline{hpue}$ for "completed-trip" interviews was larger than for statistics based on information from "incompleted-trip" interviews (sign test, P = 0.03 for one-sided alternative). On average, estimated mean harvest rate was 35% lower as measured from fishing anglers (1.08 vs. 1.65).

* * * *

A daily bag limit is the reason for the discrepancy between estimates of mean harvest in Example 2.8. The more successful anglers are limited, have shorter trips, and are less likely to be interviewed while still fishing (incompleted-trip interviews). The daily bag limit at the Elliot Highway Bridge in 1989 was 15 whitefish. The most egregious problem will occur when the daily bag limit is one fish; the mean harvest rate of fishing anglers will be near zero. Estimated harvests for roving-roving surveys of fisheries regulated with bag limits will be biased low. Because all of Alaska's fisheries are so regulated, roving-roving creel surveys are no longer conducted in the state.

Roving-access surveys based on "completed-trip" interviews have been suggested as a remedy for LOS bias (Robson 1961; Pollock et. al. 1994:251; Bernard et al. 1998; Pollack et al. 1997). ***While this remedy is intuitively reasonable, possibility of LOS bias remains in roving-access surveys***. When harvest is regulated with a daily bag limit, more successful anglers will exit the fishery early while the less successful will tend to exit at the end of the day. If a sampling period covers only the early part of a fishing day, mean harvest rate among exiting anglers will be ***higher*** than the harvest rate for anglers fishing during that period (Bernard et al. 1998). In contrast, mean harvest rate for anglers exiting late in the fishing day will be lower than the actual harvest rate (Bernard et al. 1998).

Example 2.9. A roving creel survey was used to estimate harvest of chinook salmon in the Ninilchik River in 1991. Fishing occurred by regulation only on Saturday, Sunday, and Monday beginning the weekend of 25-27 May and continuing through three subsequent weekends. At the end of the season, fishing was allowed during 19-24 June with some restrictions on the length of the fishing day. Sampling followed a stratified, two-stage design with sampling periods as first-stage units. Strata were defined as the following combination of dates (length of sampling periods are listed alongside):

25 May	4-hr	9-10 June	8-hr
26-27 May	8-hr	15 June	4-hr
1 June	4-hr	16-17 June	8-hr
2-3 June	8-hr	19-24 June	6-hr
8 June	4-hr		

With one exception, each stratum had six sampling periods (Q_h = 6) of which three were systematically selected for sampling ($q_{hi} = q_h$ = 3); the last stratum had 24 sampling periods (Q_h = 24) of which 5 ($q_{hi} = q_h$ = 5) were randomly selected for sampling. The first sampling period in a stratum always began at 0001 hours. This uncommon sampling design was chosen because fishing occurred around the clock in this fishery with considerably more fishing effort expended on Saturdays. The last stratum was an addendum to the original design to cover unexpected opportunities for fishing. Anglers were interviewed as they left the fishery ("completed-trip" interviews). The daily bag limit was one fish.

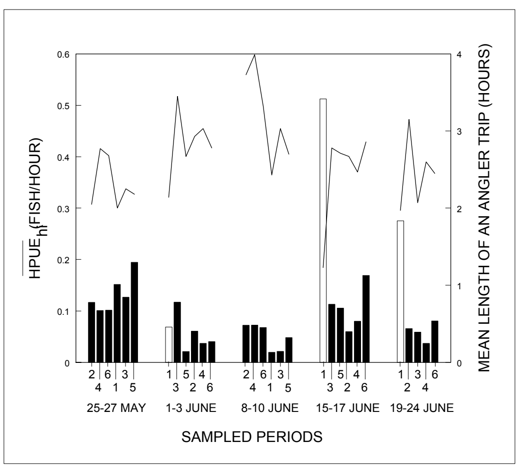


Figure 2.1 - Estimated mean harvest rate (bars) and mean length of angler-trips (lines) by sampled periods in the creel survey of the fishery for chinook salmon on the Ninilchik River in 1991. Clear bars correspond to periods immediately after resumption of fishing.

In the three instances in which anglers were interviewed during the first sampling period in a weekend (sampling period no. 1 on 25 May, 1, 8, 15, or 19 June), two produced unlikely estimates of mean harvest rate (15 and 19 June) (Figure 2.1). In both these instances, average length of angler trips was low, especially on 15 June. High estimates of mean harvest rate for first sampling periods on both dates resulted from early exit (within half an hour of the opening) of successful anglers, making harvest rate among anglers inversely related to length of their fishing trip. Estimated mean harvest rate for the first sampling period on Saturday, 1 June was not unusually high, and no inverse relationship was apparent.

* * * *

Lengthening the sampling period and catch cards are partial remedies for LOS bias caused by bag limits in roving-access surveys. If sampling periods are long relative to the length of most angler's fishing trips, LOS bias will occur, but will be negligible. This tactic has worked for some surveys in some Alaska fisheries (Bernard et al. 1998), but not others. The problem is that anglers' patience with lack of success can vary considerably. If most anglers are willing to fish up to half a fishing day with limited or no success, this partial remedy will not work. Because all anglers still fishing at the end of a sampled period have no chance of being interviewed during the period, handing out catch cards to fishing anglers at the end of the period provides a chance to get "completed-trip" interviews from these fishers, so long as anglers return the cards. Anglers are asked to record their harvest during their fishing trip, the length of time spent fishing, and where they exited the fishery. The last question is needed to adjust the catch-card sample to fit with the sample of interviews taken during the sampled period. Without the adjustment, estimated mean harvest rate for the period will be biased low. Catch cards are usually not a viable remedy for LOS bias because they are cumbersome to use and are subject to non-response and prestige bias (Pollock et al. 1994:71-72,129-130).

The only complete remedy for LOS bias when harvest is regulated with a daily bag limit is to equate the length of a sampled period with that of the fishing day. In this way, when an angler exits a fishery should not effect their chance of being interviewed. The comparison between estimated mean harvest rates in Example 2.8 is valid only because the sampled period in the whitefish fishery near the Elliot Highway Bridge is the same length as the fishing day (night). ***Equating lengths of sampling periods and fishing days also restricts stages in sampling designs to two, excludes the use of unequal probability sampling*** as described in Pollock et al. (1994:173-4), and ***excludes TOD (time of day) stratification*** (see Section 1.2).

2.2.2.2 Avoiding Opportunity Bias. In roving-access surveys, anglers not exiting the fishery through access points manned by technicians have no chance of being interviewed. If anglers exiting at different locations have different mean harvest rates on average, estimated mean harvest rate for the sampled period will be biased. In geographically large fisheries, distribution of fish abundance could produce this opportunity bias. Opportunity bias will also occur if different types of anglers, for instance guided and unguided anglers, have different success rates and exit the fishery at different locations. Opportunity bias can be detected by simultaneously interviewing anglers at more than one access point on a semi-regular schedule and comparing resulting estimates of mean harvest rate.

Opportunity bias can be reduced or eliminated by changing sampling sites randomly throughout the sampling period and by stratification. Random selection of sampling sites will reduce, but not eliminate opportunity bias. If potential for opportunity bias is small, randomization might be a sufficient remedy. In fisheries with geographic differences in harvest rates, groups of contiguous access points could be used as a dimension of stratification if where an angler entered the fishery determined where he or she exited. A single roving-access survey would become several independent surveys, one for each geographic stratum, with independent sampling schedules. For instance, participants in the recreational troll fishery for chinook salmon near Juneau who launch from Douglas head south to fish; those that launch from Auke Bay head north. Different stocks of chinook salmon pass through these two different fishing grounds. Opportunity bias from geographic differences in mean harvest rates across this fishery could be avoided through standard stratification. If some of the anglers fishing on the northern grounds exit the fishery from Douglas and vice versa for anglers exiting at Auke Bay, opportunity bias can still be avoided by post-stratifying anglers as to where they fished. This is the same situation and solution demonstrated in the access-point survey of the fishery for sockeye salmon in the Russian River (Examples 2.2 and 2.4). Post-stratification can also be used to avoid opportunity bias when different types of anglers with different harvest rates have different probabilities of being interviewed. Bernard et al. (1998) use the roving-access survey of the fishery for coho salmon on the Kenai River in 1992 to demonstrate post-stratification of anglers by expertise (guided vs. unguided) for just this reason. A different standard stratum of the same survey is provided as a demonstration in Example 2.11.

2.2.3 Harvest, Catch, Fishing Effort, and Angler-Trips. Harvest, fishing effort, catch, and the number of angler-trips from each sampled period are expanded to the stratum according to two-stage sampling designs for access-point surveys (Table 2.1) in which the final stage units are angler-trips or angler counts, and the first-stage units are fishing days. Because Alaska's sport fisheries are regulated with daily bag limits, only two-stage sampling designs are considered here to avoid the potential for LOS bias. Access point is no longer a stage in sampling designs even though anglers may be interviewed at a single location during a sampled period. Access point may be considered a dimension of stratification or post-stratification in some instances to avoid opportunity bias (Section 2.2.2.2). Equations in Table 2.1 can be used with the following substitutions:

$$\hat{Y}_{hi} = \left[\begin{array}{c} \hat{E}_{hi}\,\overline{hpue}_{hi} = \hat{H}_{hi} \\ \hat{E}_{hi}\,\overline{cpue}_{hi} = \hat{C}_{hi} \\ \hat{E}_{hi} \\ \hat{E}_{hi} / \bar{e}_{hi} = \hat{M}_{hi} \end{array} \right.$$

for estimated harvest, catch, fishing effort, or number of angler-tips. Contrary to statements in Pollock et al. (1994:175), estimated numbers of angler-trips will be unbiased so long as sampling periods are the same length as the fishing day (Bernard et al. 1998).

Example 2.10. The two-stage creel survey of the spear fishery for whitefish near the Elliot Highway Bridge on the Chatanika River continues as our example (see Examples 2.1, 2.6, 2.7, and 2.8). Although there was no stratification in this survey, the subscript *h* designating strata was added to this and subsequent equations to make demonstrations more general. Estimates fishing effort (25 h) and of mean harvest rate (5.1625) were calculated in Examples 2.6 and 2.7 for 10 October. Estimated harvest for that period is 129 whitefish ($=\hat{H}_{hi} = \hat{E}_{hi}\overline{hpue}_{hi}$). From the 16 completed-trip interviews that night (data are listed in Table 2.6), the average duration of an angler-trip is:

$$\bar{e}_{hi} = \frac{3+3+2+\ldots+1+3+3}{16} = 2.00$$

and the estimated number of angler trips 13 [$= \hat{M}_{hi} = \hat{E}_{hi}/\bar{e}_{hi} = 25/2$]. Note that the estimated number of angler-trips for 10 October is less than the number of anglers interviewed. Such inconsistency is caused by either imprecision in the estimate (as in this instance) or anglers having fishing trips shorter than the hiatus between counts.

Statistics for all sampled periods are listed in Table 2.8. Note that there are missing data for the nights of 18 and 26 September in the systematic sampling schedule. These were ignored when calculating estimates of harvest, fishing effort, or number of angler-trips (but not when estimating their variances). From Equation 2.1 in Table 2.1, the estimated mean harvest for first-stage sampling units is:

$$\hat{\bar{H}}_h = \frac{0+2+2+\ldots+129+182+41}{15} = 163$$

where $d_h = 15$. Estimated means for fishing effort and number of angler-trips are 71 hr and 31 trips per night, respectively. Expansion of these averages for the 35 ($= D_h$) first-stage sampling units (periods) in this stratum produced estimates of 5,705 whitefish harvested [=35(163)] in 2,485 hr of fishing [=35(71)] through 1,085 angler-trips [=35(31)].

2.2.4 Estimating Variance. Estimating variance for estimates of harvest, catch, fishing effort, or number of angler-trips from roving-access surveys centers around the sampled period. Since harvest or catch is the product of two independent variates (fishing effort in hours and harvest per hour of fishing), estimating variance of harvest or catch in a sampled period follows the method of Goodman (1960).

Table 2.8 - Estimates of fishing effort $\hat{E}_{hi}$, harvest $\hat{H}_{hi}$, and number of angler-trips $\hat{M}_{hi}$ by sampled period in the roving-access survey of the spear fishery near the Elliot Highway Bridge on the Chatanika River in 1989.

	$\hat{E}_{hi}$	$\hat{H}_{hi}$	$\hat{M}_{hi}$
September			
12	3.6	0	4
14	10.0	2	8
16	52.3	2	47
18			
20	24.0	1	14
22	55.7	10	29
24	101.1	293	39
26			
28	60.0	275	26
30	233.0	666	81
October			
2	48.0	175	22
4	70.0	176	34
6	135.0	285	44
8	105.4	206	39
10	25.0	129	13
12	41.0	182	18
14	100.3	41	45

Because the estimated number of angler trips $\hat{M}$ is a ratio, its estimated variance is approximated with the delta method (see Seber 1982:8). Table 2.9 contains equations for calculating estimated variances for these statistics for a stratified two-stage design with fishing days as sampling periods. Equations 2.9 and 2.10 describe estimating variances $\hat{E}_{hi}$ under these circumstances. Sample variance of mean length of an angler-trip is calculated with information from completed-trip interviews:

Table 2.9 - Equations to estimate or approximate sample-period variances and covariances for estimates of harvest, catch, fishing effort, and number of angler trips in stratified, two-stage roving-access surveys.

Variances:

$$var(\hat{H}_{hi}) = var(\overline{hpue}_{hi})\hat{E}_{hi}^2 + var(\hat{E}_{hi})\overline{hpue}_{hi}^2 - var(\hat{E}_{hi})var(\overline{hpue}_{hi}) \qquad \textbf{(2.12)}$$

$$var(\hat{C}_{hi}) = var(\overline{cpue}_{hi})\hat{E}_{hi}^2 + var(\hat{E}_{hi})\overline{cpue}_{hi}^2 - var(\hat{E}_{hi})var(\overline{cpue}_{hi}) \qquad \textbf{(2.13)}$$

$$var(\hat{M}_{hi}) \cong \hat{M}_{hi}^2\left[\frac{var(\bar{e}_{hi})}{\bar{e}_{hi}^2} + \frac{var(\hat{E}_{hi})}{\hat{E}_{hi}^2} - \frac{var(\bar{e}_{hi})var(\hat{E}_{hi})}{\bar{e}_{hi}^2\hat{E}_{hi}^2}\right] \qquad \textbf{(2.14)}$$

Covariances:

$$cov(\hat{H}_{h'i}, \hat{H}_{h''i}) = \overline{hpue}_{h'i}\,\overline{hpue}_{h''i}T^2 cov(\bar{x}_{h'i}, \bar{x}_{h''i}) \qquad \textbf{(2.15)}$$

$$cov(\hat{C}_{h'i}, \hat{C}_{h''i}) = \overline{cpue}_{h'i}\,\overline{cpue}_{h''i}T^2 cov(\bar{x}_{h'i}, \bar{x}_{h''i}) \qquad \textbf{(2.16)}$$

$$cov(\hat{E}_{h'i}, \hat{E}_{h''i}) = T^2 cov(\bar{x}_{h'i}, \bar{x}_{h''i}) \qquad \textbf{(2.17)}$$

$$cov(\hat{M}_{h'i}, \hat{M}_{h''i}) = \frac{T^2}{\bar{e}_{h'i}\bar{e}_{h''i}} cov(\bar{x}_{h'i}, \bar{x}_{h''i}) \qquad \textbf{(2.18)}$$

$$var(\bar{e}_{hi}) = \frac{\sum_{j=1}^{m_{hi}}(e_{hij} - \bar{e}_{hi})^2}{m_{hi}(m_{hi} - 1)} \qquad (2.19)$$

Estimating variance for mean harvest rate has been covered earlier in Section 2.2.2 (Equations 2.11).

Once variances have been estimated for each sampled period, calculations are the same as those for access-point surveys (Table 2.4). Estimated variances for estimated harvest, fishing effort, catch, and number of angler-trips for each period are substituted for the following expression in Equations 2.5 for two-stage sampling designs:

$$\left.\begin{array}{r} var(\hat{E}_{hi}) \\ var(\hat{H}_{hi}) \\ var(\hat{M}_{hi}) \end{array}\right] \text{ for } \left[M_{hi}^2(1 - f_{2hi})\frac{s_{hi}^2}{m_{hi}}\right]$$

There are no second-stage *fpc*s [(1- f_{2hi})] in two-stage designs for roving surveys because there is an infinite number of possible counts of anglers in each sampled period. Beyond these substitutions, calculations are the same for both roving and access-point surveys.

Example 2.11. This is Example 2.10 revisited to calculate variance of estimated harvest of whitefish in the fishery near the Elliot Highway Bridge. Statistics for this demonstration can be found in Examples 2.6, 2.7, and 2.10 for the night of 10 October. From Equation 2.12, estimated variance of estimated harvest that night is:

$$var(\hat{H}_{hi}) = 532 =$$
$$(0.1383)25^2 + (16.8)(5.1625)^2 - (0.1383)(16.8)$$

For the same night:

$$var(\bar{e}_{hi}) = 0.03125 =$$
$$\frac{\left[\begin{array}{l}(3.0-2.0)^2+(3.0-2.0)^2+(2.0-2.0)^2+\ldots\\ \ldots+(1.0-2.0)^2+(3.0-2.0)^2+(3.0-2.0)^2\end{array}\right]}{16(16-1)}$$

from Equation 2.19. Variance for the estimated number of angler trips is approximated with Equation 2.14:

$$var(\hat{M}_{hi}) \cong 5.4 =$$
$$(12.5)^2\left[\frac{0.03125}{2.^2}+\frac{16.8}{25^2}-\frac{0.03125(16.8)}{2.^2(25^2)}\right]$$

Estimated variances for estimates of fishing effort, harvest, and number of angler-trips for all sampled periods are listed in Table 2.10.

At this point, Equation 2.5 in Table 2.4 can be used to estimate variance for estimated harvest, fishing effort, or number of angler-trips for the stratum:

$$var(\hat{Y}_h) = (1-f_{1h})D_h^2\frac{S_{1h}^2}{d_h} + f_{1h}^{-1}\sum_{i=1}^{d_h} var(\hat{Y}_{hi}) \qquad (2.20)$$

Since calculations of estimated variances for all three statistics follow the same procedures, only calculations for estimated variance of estimated harvest are given here as an example. Remember scheduled sampling for 16 and 28 September did not occur. Since sampling periods (days) were scheduled systematically, sample variance for the first-stage sampling units is adjusted as with the same situation in access-point surveys (see Section 2.1.3 and Example 2.5):

$$S_{1h}^2 \cong \frac{\sum_{i=2}^{\tilde{d}_h}\phi_{hi}\phi_{h(i-1)}(\hat{H}_{hi}-\hat{H}_{h(i-1)})^2}{2\sum_{i=2}^{\tilde{d}_h}\phi_{hi}\phi_{h(i-1)}}$$

Table 2.10 - Estimated variances for estimates of fishing effort, harvest, and number of angler-trips by sampled period in the roving-access survey of the fishery for whitefish near the Elliot Highway Bridge in 1989.

	$var(\hat{E}_{hi})$	$var(\hat{H}_{hi})$	$var(\hat{M}_{hi})$
September			
12	16.2	0	16
14	15.6	2	12
16	222.0	6	187
18			
20	88.5	2	33
22	216.9	35	62
24	327.0	3,437	56
26			
28	109.8	2,978	24
30	777.0	17,189	108
October			
2	78.4	1966	18
4	229.8	2,867	59
6	287.3	2,000	32
8	74.6	1,284	14
10	16.8	532	5
12	70.8	2,431	17
14	136.3	315	34
Sums	2,667.0	35,044	677

where $\tilde{d}_h$ is the number of periods scheduled to be sampled, ϕ_i = 1 if scheduled period i was sampled and 0 if not. With data from Table 2.8:

$$S_{1h}^2 \cong \frac{\begin{bmatrix} (1)(1)(2-0)^2 + (1)(1)(2-2)^2 + \\ (0)(1)(?-2)^2 + (1)(0)(1-?)^2 + \\ (1)(1)(10-1)^2 + \ldots + (1)(1)(41-182)^2 \end{bmatrix}}{2[(1)1+(1)1+(0)1+(1)0+(1)1+\ldots+(1)1]}$$

$$\cong 21{,}703$$

The sum of estimated variances across second-stage sampling units (angler-trips) is 35,045 (Table 2.10). Estimated variance for the stratum is:

$$var(\hat{H}_h) =$$

$$(1 - 0.43)(35)^2 \frac{21{,}703}{15} + (0.43)^{-1}(35{,}044)$$

$$= 1{,}091{,}775$$

The SE for estimated harvest in the fishery near the Elliot Highway Bridge is 1,045 [=$(1{,}091{,}775)^{1/2}$]. From Example 2.10, the corresponding estimate of harvest is 5,705.

2.2.5 Post-stratification and Covariance. If there is post-stratification, calculations are adjusted through apportioning counts of fishing anglers according to these rules:

- If the correct stratum can be determined for fishing anglers as they are counted, no adjustment is needed.

- If the correct stratum can not be so determined and ***if average lengths of angler-trips are <u>the same</u> for all post-stratified categories***, count by category is a prorated approximation:

$$\bar{x}_{h'i} \approx \bar{x}_{hi} \left[\frac{m_{h'i}}{m_{hi}} \right]$$

  where $m_{h'i}$ is the number of anglers interviewed during sampled period i that belonged to stratum h', a new stratum based on a post-stratified category of anglers.

- If the correct stratum can not be determined while counting and ***if average lengths of angler-trips are <u>different for at least one post-stratified category</u>, anglers interviewed while exiting a fishery will not be representative of the distribution of the anglers being counted***. The hypothetical situation below is a demonstration of this problem:

Turnover Rate	Trip Length	Number Fishing	Distribution Actual	Distribution Perceived
Guided Anglers:				
20/hr	2 hr	40	67%	50%
Unguided Anglers:				
20/hr	1 hr	20	33%	50%

Interviews reflect the relationship between two turnover rates for two types of anglers (rates at which they enter and leave the fishery) instead of the distribution of fishing anglers across categories. There is as yet no post facto solution for this problem. ***Lengths of angler-trips across post-stratified categories should be compared to detect this situation.***

In most instances, combining estimates across strata in roving-access surveys follows the same procedures as in access-point surveys (see Section 2.1.2). For standard stratification, statistics are summed across strata as per Equations 1.1. How statistics are combined under a post-stratified scheme depends on the reason for the post-stratification. If anglers were originally post-stratified only to provide a useful presentation of statistics, data should be pooled across categories of stratification and statistics recalculated for reasons given in Section 2.1.2 and in Example 2.4. However, if anglers were post-stratified to avoid opportunity bias (Section 2.2.2.2) data can not be pooled across strata without biasing estimates.

Post-stratified estimates can be summed within sampled periods to avoid opportunity bias, but only at the cost of some covariance. Estimates of mean harvest rate by post-stratified groups of anglers and estimates of mean trip-length are independent, however, counted numbers of anglers by group are not (Bernard et al. 1998). If $x_{h'it}$ and $x_{h''it}$ are the numbers of anglers of type h' and h'' counted during count t in sampled period i in standard stratum h, $cov(x_{h'it}, x_{h''it}) = -x_{h'it}x_{h''it}/x_{hit}$ where x_{hit} is the number of anglers of all types counted. Covariances for mean numbers of counted anglers for the period are:

$$cov(\bar{x}_{h'i}, \bar{x}_{h''i}) = r^{-2}\sum_{t=1}^{r} cov(x_{h'it}, x_{h''it})$$

Equations for estimates of covariances between estimates of harvests, between estimates of catches, between estimates of fishing effort, and estimates of angler-trips are listed in Table 2.9. Derivation of covariance equations in this table are described in Appendix E. The general equation to calculate sample variances for estimates of harvest, catch, fishing effort, or numbers of angler-trips with post-stratification is:

(2.21)

$$var(\hat{Y}_h) = (1-f_{1h})D_h^2\frac{S_{1h}^2}{d_h} + f_{1h}^{-1}\sum_{i=1}^{d_h}\sum_{h'=1}^{L'} var(\hat{Y}_{h'i}) + f_{1h}^{-1}\sum_{i=1}^{d_h}\sum_{h'=1}^{L'}\sum_{h'<h''}^{L'} 2\,cov(\hat{Y}_{h'i}, \hat{Y}_{h''i})$$

where L' is the number of elements in the post-stratified dimension. Estimates summed across post-stratification would follow Equation 2.1 for two-stage, roving-access surveys with the adaptation $\hat{Y}_{hi} = \sum_{h'=1}^{L'}\hat{Y}_{h'i}$.

Example 2.12. In 1992 the roving-access creel survey of the fishery for coho salmon in the Kenai River was post-stratified by type of angler (guided vs. unguided) to avoid opportunity bias (a third category, anglers fishing from shore, is excluded to simplify this demonstration). The survey had a stratified two-stage design with days as first-stage sampling units and angler-trips and counts as second-stage units. Sampling periods were 8-hr long. The survey also had TOD stratification (morning vs. afternoon), seasonal (August vs. September), and DOW (weekend day vs. weekdays) stratification. Sampling periods were randomly scheduled for sampling. One technician passed through the fishery thrice during each sampled period counting fishing anglers while another technician interviewed exiting anglers. Guided and unguided anglers could be distinguished during counts.

Data and statistics for post-stratified groups of anglers fishing on weekdays in the afternoons of September are listed in Table 2.11. On 1 September, guided and unguided anglers harvested an estimated 102 and 132 salmon, respectively, and together 234 salmon (Table 2.12). Estimated fishing effort for that day is 221.6 and 653.6 hr by group totaling 875 hr. By Equation 2.1, estimated mean harvests and mean expenditure of fishing effort across all sampled days are 109.7 (=768/7) salmon and 664.9 hours (=4,654/7). Considering that September in 1992 had 21 week days ($=D_h$), estimates of harvest and fishing effort weekday afternoons in September are 2,304 salmon [=21(109.7)] and 13,963 hours [=21(664.9)].

For estimated harvest, second-stage component of estimated variance by Equation 2.21 with data in Table 2.12 is:

$$f_{1h}^{-1}\sum_{i=1}^{d_h}\sum_{h'=1}^{L'} var(\hat{H}_{h'i}) = \frac{21}{7}(11{,}005 + 6{,}145) = 51{,}450$$

(Note the order of summation is not important.) First-stage sampling variance with randomly selected units is :

$$S_{1h}^2 = 12{,}876 = \frac{(234-109.7)^2 + (258-109.7)^2 + \ldots + (0-109.7)^2}{7-1}$$

The first-stage component of estimated variance is:

$$(1-f_{1h})D_h^2\frac{S_{1h}^2}{d_h} = (1-\frac{7}{21})21^2\frac{12{,}876}{7} = 540{,}792$$

Estimated covariances between numbers of counted anglers from data collected on 1 September are:

Table 2.11 - Mean numbers of counted anglers, number of interviews, estimates of mean harvest rate, fishing effort, harvest, and their estimated variances for guided and unguided anglers fishing on weekdays in the afternoons of September, 1992 for coho salmon in the Kenai River.

	$\bar{x}_{h'i}$	$var(\bar{x}_{h'i})$	$m_{h'i}$	$\overline{hpue}_{h'i}$	$var(\overline{hpue}_{h'i})$	$\hat{E}_{h'i}$	$var(\hat{E}_{h'i})$	$\hat{H}_{h'i}$	$var(\hat{H}_{h'i})$
GUIDED ANGLERS:									
1-Sep	27.7	406.3	5	0.461	0.003	221.6	26,003	102	5,596
2-Sep	41.3	276.3	12	0.275	0.002	330.4	17,683	91	1,520
10-Sep	44.7	116.5	18	0.143	0.001	357.6	7,456	51	273
11-Sep	56.0	2,897.0	20	0.139	0.001	448.0	185,408	62	3,598
17-Sep	11.0	60.5	17	0.027	0.002	88.0	3,872	2	11
24-Sep	4.7	10.3	5	0.085	0.004	37.6	659	3	8
28-Sep	0.0	0.0	0	0.000	0.000	0.0	0	0	0
Sums	185		77			1,483	241,081	312	11,005
UNGUIDED ANGLERS:									
1-Sep	81.7	301.3	17	0.202	0.002	653.6	19,283	132	1,603
2-Sep	87.3	181.0	13	0.240	0.002	698.4	11,584	168	1,620
10-Sep	70.7	1.3	25	0.021	0.001	565.6	83	12	320
11-Sep	95.0	484.3	19	0.167	0.003	760.0	30,995	127	2,504
17-Sep	21.0	36.0	23	0.031	0.001	168.0	2,304	5	28
24-Sep	33.0	5.0	20	0.045	0.001	264.0	320	12	70
28-Sep	7.7	1.3	3	0.000	0.000	61.6	83	0	0
Sums	396		120			3,171	64,652	456	6,145

Table 2.12 - Estimated covariances for mean counts and estimates of fishing effort and harvest for guided (h') and unguided (h'') anglers separately and combined (h) fishing on weekdays in the afternoons of September, 1992 for coho salmon in the Kenai River.

	$\hat{E}_{hi}$	$\hat{H}_{hi}$	$cov(\bar{x}_{h'i}, \bar{x}_{h''i})$	$cov(\hat{E}_{h'i}, \hat{E}_{h''i})$	$\overline{hpue}_{h'i}$	$\overline{hpue}_{h''i}$	$cov(\hat{H}_{h'i}, \hat{H}_{h''i})$
1-Sep	875	234	-6.52	-417	0.461	0.202	-39
2-Sep	1,029	259	-9.19	-588	0.275	0.240	-39
10-Sep	923	63	-8.99	-576	0.143	0.021	-2
11-Sep	1,208	189	-11.08	-709	0.139	0.167	-16
17-Sep	256	8	-1.81	-116	0.027	0.031	0
24-Sep	301	15	-1.27	-81	0.085	0.045	0
28-Sep	62	0	0.00	0	0.000	0.000	0
Sums	4,654	768		-2,487			-96

	t=1	t=2	t=3
$x_{h'it}$	56	68	75
$x_{h''it}$	102	16	11
x_{hit}	158	84	86
$cov(x_{h'it}, x_{h''it})$	-36.15	-12.95	-9.59

Estimated covariance between mean numbers of anglers counted for 1 September is:

$$cov(\bar{x}_{h'it}, \bar{x}_{h''it}) = \frac{(-36.15)+(-12.95)+(-9.59)}{3^2} = -6.52$$

From Equation 2.15, estimated covariance among post-stratified estimates of harvest for the period sampled on 1 September is:

$$cov(\hat{H}_{h'i}, \hat{H}_{h''i}) = 0.461(0.202)(8)^2(-6.52) = -39$$

With estimated covariances for other sampled periods listed in Table 2.12, the component of estimated variance due to covariance arising from post-stratification is:

$$f_{1h}^{-1} \sum_{i=1}^{d_h} \sum_{h'=1}^{L'} \sum_{h'<h''}^{L'} 2\, cov(\hat{H}_{h'i}, \hat{H}_{h''i}) = \frac{21}{7} 2(-96) = -576$$

Estimated variance $var(\hat{H}_h)$ is the sum of these three components: 540,792 + 51,450 + (-576) = 591,666 (SE=769). With information in Tables 2.11 and 2.12 and with Equation 2.20, $var(\hat{E}_h)$ = 8,418,211 + 917,199 + (-14,922) = 9,320,488 (SE=3,053).

* * * *

Note that estimated covariances due to post-stratification are negligible in sample variances for estimated harvests and estimated fishing efforts. If covariance from post-stratification is ignored in calculations, ***SEs would increase by a fraction of a salmon and two hours of fishing.*** Most of the estimated variance in the example comes the first-stage components. If the seven periods sampled had been all the first-stage units in the standard stratum ($d_h = D_h = 7$), $var(\hat{H}_h)$ = 0 + 17,150 + (-192) = 16,958 (SE=130), and $var(\hat{E}_h)$ = 0 + 305,733 + (-4,976) = 300,757 (SE=548). If estimated covariances are ignored in this hypothetical situation, ***SEs would increase by only a salmon and by five hours of fishing***. Estimated covariances from post-stratification in a different standard stratum in the same roving-access survey are also negligible (Bernard et al. 1998). ***This situation of negligible covariance is germane to almost all roving-access surveys***. Only in fisheries where mean harvest rate is high and nearly uniform within each group of post-stratified anglers, when numbers of anglers are relatively constant through sampled periods, and when few anglers are counted will covariances from post-stratification be a significant reduction in estimated variance. ***Outside of these unique and rare circumstances, covariances from post-stratification can be ignored with little penalty in roving-access surveys***.

2.2.6. Missing Data. There are three kinds of "missing data" in onsite roving-access surveys:

- a scheduled period is not sampled;
- no anglers are counted during a sampled period while some are interviewed; and
- no anglers are interviewed while some are counted.

If failures to sample scheduled periods are random occurrences, no adjustment in calculations other than reducing sample size is necessary (see Section 2.1.3 on response to missing data in access-point surveys and Example 2.11 on roving-access surveys). Chances of missing data for the other reasons depend on the length of fishing trips and fishing effort during that sampled period, and on the extent anglers have access to the fishery. If angler-trips are short relative to the time between counts and access limited, most (and occasionally all) anglers might begin and end their fishing trip between counts. If angler-trips are few and access extensive, anglers might be counted, but none may exit the fishery in front of a technician. Of the latter two types of missing data, the first (no anglers counted, but some interviewed) is extremely rare while the second (counted anglers, but no interviews) is common. For this reason, we concentrated on procedures for the latter situation.

The fix for missing information on mean harvest rate and trip-length in a sampled period is to impute, that is "plug-in" a substitute formed with information from other sampled periods within the same or from another stratum. Setting "missing" estimates to zero would bias overall estimates downward because absence of interviews is more likely when fishing effort, catch, and harvest are low. Ignoring periods with "missing" information would bias overall estimates upward.

All substitutes can be expressed as linear combinations of other statistics (Bernard et al. 1998). A substitute could come from another stratum, either another post-stratified group in the same sampled period or from different sampled period in a nearby or contemporary stratum. For instance, if no guided anglers were interviewed during a sampled period but unguided anglers were, estimated mean harvest rate for the latter group could be ***scaled*** as a substitute for missing information on the former group. A substitute could be an average of statistics from other sampled periods in the same stratum, such as, the average of mean harvest rates for the day before and the day after the sampled period with missing information. Equations for these common substitutions are listed in Table 2.13. When the scaling constant a_{hi} = 0, sampled period i in stratum h is not involved in the substitution. When a_{hi} > 0, the scaling constant represents some underlying relationship between statistics, calculation of an average, or both.

Once substitutions have been made, calculations proceed according to Equations 2.1 and 2.5 to produce estimates $\hat{H}$ or $\hat{M}$ and ***minimum*** estimated variances $var(\hat{H})_{min}$ or $var(\hat{M})_{min}$. These estimated variances are minimums because there is a penalty for substitutions. Bernard et al. (1998) developed corrections in estimated variance for imputed substitutes of estimated mean harvest rate. Those corrections along with corrections for substitutes for missing information on mean trip lengths are listed in Table 2.14. Derivations of these corrections using the method of moments can be found Bernard et al. (1998) and in Appendix F. Unbiased estimates of sample variance are:

$$var(\hat{H}) = var(\hat{H})_{min} + \textit{Correction for Harvest}$$
$$var(\hat{M}) = var(\hat{M})_{min} + \textit{Correction for Trips}$$

Example 2.13. No guided anglers were interviewed during three sampled periods in a stratum composed of August weekend mornings in the roving-access creel survey to estimate the harvest of coho salmon in the Kenai River in 1991. The survey that year had the same design as the survey a year later (see Example 2.12). Data and statistics for this stratum are listed in Table 2.15.

Bernard et al. (1998) used information from the same stratum in the survey a year later in 1992 to demonstrate corrections in estimated variance from substitutions for missing data. Because coho salmon are a migratory species, Bernard et al. scaled information from unguided anglers interviewed in the same periods where information on guided anglers was missing and used Equations 2.22 and 2.28 to estimate harvest and correct its estimate of variance. In our example here, we used average harvest rates from the same stratum as substitutes for missing data (Equations 2.26 and 2.32). This approach is less realistic for a migratory population than that taken by Bernard et al. (1998). However, our purpose here is to demonstrate another method that can be more realistic in other situations, such as when populations are not migratory.

An average of estimated mean harvest rates for the weekend mornings of 10 and 11 August, 1991 stratum (Equations 2.26) were used as a substitute for missing data on 3, 4, and 25 August. Information from 17 August was not part of the average (linear combination) because the bulk of that year's migration was present on the fishing grounds that weekend. From Equations 2.26:

$$\overline{hpue}_{hi'} = (1/2)(0.215) + (1/2)(0.101) = 0.158$$
$$var(\overline{hpue}_{hi'}) = 0.003 = (1/2)^2(0.008) + (1/2)^2(0.004)$$

Vectors $\{a_{hi}\}$ = {0, 0, 0.5, 0.5, 0, 0} and $\{b_{hi'}\}$ = {1, 1, 0, 0, 0, 1}. From Equation 2.1, $\hat{\hat{H}}_{hi'}$ = 375 (=2,251/6), and from Equation 2.5:

$$S_{1h}^2 = 352{,}296 = \frac{(76-375)^2 + (39-375)^2 + \ldots + (100-375)^2}{(6-1)}$$

$$var(\hat{H}_h)_{min} = (1-\frac{6}{8})8^2\frac{352{,}296}{6} + \frac{8}{6}58{,}897 = 1{,}017{,}107$$

The correction for using the imputed substitutions is 36,780 (from Table 2.16) making the unbiased sample variance for estimated harvest 1,053,887 (SE=1,027) for this stratum. Estimated harvest is 3,000 salmon [=8(375)]. The correction for using an imputed substitution is a relatively minor component of estimated variance in this example. In the example in Bernard et al. (1998), the correction is a significant portion of estimated variance.

Table 2.13 - Common substitutions for missing information from interviews in a sampling period i' stratum h' in an onsite roving-access creel survey with information from sampled periods i stratum h. Scaling constant $a = 0$ if sampled period is not involved in substitution, and $a > 0$, and 0 if it is. Note that when post-stratification is involved, $i' = i$.

Across Post-Stratification within a Period:

Harvest: $\overline{hpue}_{h'i'} = a_{hi'}\overline{hpue}_{hi'}$ $\quad v(\overline{hpue}_{h'i'}) = a_{hi'}^2 v(\overline{hpue}_{hi'})$ **(2.22)**

Mean Length of an Angler-trip: $\bar{e}_{h'i'} = a_{hi'}\bar{e}_{hi'}$ $\quad v(\bar{e}_{h'i'}) = a_{hi'}^2 v(\bar{e}_{hi'})$ **(2.23)**

Across Time or Location Defined (TLD) Strata and Periods:

Harvest: $\overline{hpue}_{h'i'} = a_{hi}\overline{hpue}_{hi}$ $\quad v(\overline{hpue}_{h'i'}) = a_{hi}^2 v(\overline{hpue}_{hi})$ **(2.24)**

Mean Length of an Angler-trip: $\bar{e}_{h'i'} = a_{hi}\bar{e}_{hi}$ $\quad v(\bar{e}_{h'i'}) = a_{hi}^2 v(\bar{e}_{hi})$ **(2.25)**

Averaging across Periods within a TLD Stratum:

Harvest: $\overline{hpue}_{hi'} = \sum_{i=1}^{d_h} a_{hi}\overline{hpue}_{hi}$ $\quad var(\overline{hpue}_{hi'}) = \sum_{i=1}^{d_h} a_{hi}^2 var(\overline{hpue}_{hi})$ **(2.26)**

Mean Length of an Angler-trip: $\bar{e}_{hi'} = \sum_{i=1}^{d_h} a_{hi}\bar{e}_{hi}$ $\quad var(\bar{e}_{hi'}) = \sum_{i=1}^{d_h} a_{hi}^2 var(\bar{e}_{hi})$ **(2.27)**

Table 2.14 - Corrections in sample variances for estimated harvest and number of angler trips when there are imputed substitutions of mean harvest rates for a sampling period i' stratum h' in an onsite roving-access creel survey. Substitutes are linear combinations of estimates of mean harvest rates for sampled periods i stratum h (see Table 2.13). Constant $b = 1$ if sampled period has substitute, and 0 if not. Note that when post-stratification is involved, $i' = i$.

Across Post-Stratification within a Period:

Harvest:

$$2\frac{D_h}{d_h}\sum_{i'=1}^{d_h} a_{hi} b_{h'i'} v(\overline{hpue}_{hi})\hat{E}_{hi}\hat{E}_{h'i'} \quad \textbf{(2.28)}$$

Mean Length of an Angler-trip:

$$2\frac{D_h}{d_h}\sum_{i'=1}^{d_h} a_{hi} b_{h'i'} \frac{var(\bar{e}_{hi})}{\bar{e}_{hi}^{\,4}}\hat{E}_{hi}\hat{E}_{h'i'} \quad \textbf{(2.29)}$$

Across TLD Strata and Periods:

Harvest:

$$2\sum_{h'=1}^{L}\left(\frac{D_h}{d_h}\frac{D_{h'}}{d_{h'}}\right)^{1/2}\sum_{i'=1}^{d_{h'}} a_{hi} b_{h'i'} v(\overline{hpue}_{hi})\hat{E}_{hi}\hat{E}_{h'i'} \quad \textbf{(2.30)}$$

Mean Length of an Angler-trip:

$$2\sum_{h'=1}^{L}\left(\frac{D_h}{d_h}\frac{D_{h'}}{d_{h'}}\right)^{1/2}\sum_{i'=1}^{d_{h'}} a_{hi} b_{h'i'} \frac{var(\bar{e}_{hi})}{\bar{e}_{hi}^{\,4}}\hat{E}_{hi}\hat{E}_{h'i'} \quad \textbf{(2.31)}$$

Averaging across Periods within a TLD Stratum:

Harvest: **(2.32)**

$$2\frac{D_{h'}}{d_{h'}}\sum_{i=1}^{d_{h'}} a_{h'i} var(\overline{hpue}_{h'i})\left[\hat{E}_{h'i}\sum_{i'=1}^{d_h} b_{h'i'}\hat{E}_{h'i'} + a_{h'i}\sum_{i'=1}^{d_{h'}} b_{h'i'}\hat{E}_{h'i'}\left(\sum_{i''>i'}^{d_{h'}} b_{h'i''}\hat{E}_{h'i''}\right)\right]$$

Mean Length of an Angler-trip: **(2.33)**

$$2\frac{D_{h'}}{d_{h'}}\sum_{i=1}^{d_{h'}} a_{h'i} \frac{var(\bar{e}_{h'i})}{\bar{e}_{h'i}^{\,4}}\left[\hat{E}_{h'i}\sum_{i'=1}^{d_h} b_{h'i'}\hat{E}_{h'i'} + a_{h'i}\sum_{i'=1}^{d_{h'}} b_{h'i'}\hat{E}_{h'i'}\left(\sum_{i''>i'}^{d_{h'}} b_{h'i''}\hat{E}_{h'i''}\right)\right]$$

Table 2.15 - Original statistics with substitutions (in bold italics) for missing data for the stratum of guided anglers fishing on weekend days in the mornings of August for coho salmon in the Kenai River in 1991. No exiting anglers were interviewed on 3, 4, or 25 August.

	$\bar{x}_{h'i}$	$var(\bar{x}_{h'i})$	$m_{h'i}$	$\overline{hpue}_{h'i}$	$var(\overline{hpue}_{h'i})$	$\hat{E}_{h'i}$	$var(\hat{E}_{h'i})$	$\hat{H}_{hi'}$	$var(\hat{H}_{h'i})$
3-Aug	60	94	0	***0.158***	***0.003***	480	6,016	***76***	***823***
4-Aug	31	0	0	***0.158***	***0.003***	248	0	***39***	***185***
10-Aug	196	171	11	0.215	0.008	1,568	10,944	337	20,087
11-Aug	162	131	8	0.101	0.004	1,296	8,384	131	6,770
17-Aug	184	375	8	1.065	0.001	1,472	24,000	1,568	29,364
25-Aug	79	334	0	***0.158***	***0.003***	632	21,376	***100***	***1,668***
Sums						5,696	70,720	***2,251***	***58,897***

Table 2.16 - Expression and calculation of the correction for missing information on mean harvest rates on 3, 4, and 25 August in estimated variance (Equation 2.32) for estimates of coho salmon harvested in the Kenai River, 1991 by guided anglers fishing during mornings in August. Estimates from 10 and 11 August were averaged to produce imputed substitutions.

Correction =

$$= 2\frac{D_{h'}}{d_{h'}}\left\{ a_{h'(10)} var\left(\overline{hpue}_{h'(10)}\right) \begin{bmatrix} \hat{E}_{h'(10)}\left(\hat{E}_{h'(3)} + \hat{E}_{h'(4)} + \hat{E}_{h'(25)}\right) + \\ a_{h'(10)}\left(\hat{E}_{h'(3)}\hat{E}_{h'(4)} + \hat{E}_{h'(3)}\hat{E}_{h'(25)} + \hat{E}_{h'(4)}\hat{E}_{h'(25)}\right) \end{bmatrix} \right.$$

$$\left. + a_{h'(11)} var\left(\overline{hpue}_{h'(11)}\right) \begin{bmatrix} \hat{E}_{h'(11)}\left(\hat{E}_{h'(3)} + \hat{E}_{h'(4)} + \hat{E}_{h'(25)}\right) + \\ a_{h'(11)}\left(\hat{E}_{h'(3)}\hat{E}_{h'(4)} + \hat{E}_{h'(3)}\hat{E}_{h'(25)} + \hat{E}_{h'(4)}\hat{E}_{h'(25)}\right) \end{bmatrix} \right\}$$

Correction = 36,780 =

$$= 2\frac{8}{6}\left\{ 0.5(0.008) \begin{bmatrix} 1{,}568(480 + 248 + 632) + \\ 0.5[480(248) + 480(632) + 248(632)] \end{bmatrix} \right.$$

$$\left. + 0.5(0.004) \begin{bmatrix} 1{,}296(480 + 248 + 632) + \\ 0.5[480(248) + 480(632) + 248(632)] \end{bmatrix} \right\}$$

* * * *

Problems with missing data can be reduced in future surveys through better planning. When in the past no exiting anglers have been interviewed during a particular period, yet anglers have been counted while they fished, more sampling resources should be spent to simultaneously cover more access points. If many sampling periods are still expected to have "missing data", the survey needs serious redesign or the fishery is too small to warrant an onsite creel survey at all.

Effects of plug-in substitutions for mean harvest rates and covariances from post-stratification are independent and additive (Bernard et al. 1998; see Appendix E). In short, $var(\hat{Y}_h)$ = Equation 2.18 + *Covariance* + *Correction*. This independence exists because covariance from post-stratification is a function of counts, while corrections for plug-in substitutions arise from linear combinations of estimated harvest rates. Although these two sets of statistics are gathered during the same sampled periods, mean harvest rate and numbers of anglers counted are considered independent within each period.

As a final aside, please note that roving-access surveys in Examples 2.11 and 2.12 were subject to LOS bias. Both surveys had TOD stratification with a 16-hr fishing day split into two 8-hr sampling periods, and in both years harvest was regulated with a 3-fish daily bag limit. As reported in Bernard et al. (1998), LOS bias in estimated harvest from the survey in 1992 was negligible (no more than 1.3%) in comparison to estimated harvest as reported in the statewide harvest survey (Mills 1993). From Mills (1992), estimated harvest of coho salmon in the same area is 54,391 while that from the onsite survey is 20% higher at 65,000 coho salmon. ***A long sample period can blunt LOS bias, but will not eliminate it when fishing is regulated with daily bag limits.***

2.3 REFERENCES FOR EXAMPLES

Boyle, L., and M. Alexandersdottir. 1992. Contribution of stocked chinook salmon to the Ninilchik River sport fishery, 1991. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 92-30. Anchorage.

Hammarstrom, S. L. 1992. Angler effort and harvest of coho salmon during the recreational fisheries in the lower Kenai River, 1991. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 92-36. Anchorage.

Marsh, L. E. 1992. Catch and effort statistics for the sockeye salmon sport fishery during the late run to the Russian River with estimates of escapement. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 92-39. Anchorage.

Merritt, M. F., A. E. Bingham, and N. Morton. 1990. Creel surveys conducted in interior Alaska during 1989. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 90-54. Anchorage.

Mills, M. J. 1992. Harvest, catch, and participation in Alaska sport fisheries during 1991. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 92-40. Anchorage.

Mills, M. J. 1993. Harvest, catch, and participation in Alaska sport fisheries during 1992. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 93-42. Anchorage.

Schwager-King, M. A. 1993. Angler effort and harvest of coho salmon during the recreational fisheries in the lower Kenai River, 1992. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 93-31. Anchorage.

CHAPTER THREE: Harvest Composition

Harvest is often divided into categories based on sex, age, length, species, and origin to facilitate management of sport fisheries in Alaska. Estimated harvest by length group has been used to set length limits for Arctic grayling taken from the Chena River. Estimates of harvest by age have been used to build brood tables for stocks of coho salmon, chinook salmon, and sockeye salmon taken from the Kenai River. Information gained by sampling harvested rockfish has been used to apportion information from the Statewide Harvest Survey (see Howe et al. 1996) into harvest by species for fisheries based in Kodiak, Homer, Valdez, Whittier, and Seward. Obligations set forth in a treaty with Canada have been met through estimates of how many hatchery-produced chinook salmon have been harvested in recreational fisheries in southeast Alaska. When harvest is so divided, the divisions collectively are called ***harvest composition***.

Often harvest composition is estimated from an onsite creel survey designed originally to estimate harvest. If every (or nearly every) fish inspected by technicians is sampled, harvest composition can be estimated without any additions or changes to the original stratified, multistage sampling design. However, sampling every fish inspected is sometimes not practical. If a large portion of creels are not inspected, a ***stratified random sampling design*** to estimate harvest composition can be superimposed over the original design to estimate harvest, so long as ***relative harvest composition*** does not appreciable change within strata. Fortunately, such homogeneity is the norm. Unlike harvest or fishing effort, relative composition of a fished population does not change from morning to night or from day to day. If relative composition of a fished population does change from place to place or from week to week, it usually does so gradually over several weeks or across a wide area.

Sometimes harvest composition is estimated through two, independent sampling programs, an offsite survey to estimate harvest and an onsite survey ***designed solely to estimate relative harvest composition***. In this instance, estimated harvest composition is the product of statistics from both surveys. The offsite survey used to estimate harvest in Alaska is the Statewide Harvest Survey (see Howe et al. 1996). Estimates of harvest from this mail survey are cheap and accurate, but delayed. When delay is not a concern, an onsite survey to estimate relative harvest composition is a cheaper alternative than an onsite survey to simultaneously estimate both harvest and its composition. When an onsite survey to estimate harvest is prohibitively expensive or would produce biased estimates of harvest, estimating relative harvest composition is the only reasonable alternative.

This chapter covers methods to estimate harvest composition or relative harvest composition with onsite creel surveys. The degree that creels are subsampled determines whether harvest and its composition can be estimated with the same sampling design, or if modifications should be superimposed over the original design. Sampling designs for onsite surveys to estimate only relative harvest composition are given along with conditions under which estimates will be unbiased. Examples include surveys of the spear fishery for whitefish on the Chatanika River, of marine fisheries for halibut out of Homer and Seward, of a fishery for sockeye salmon in the Kenai River, and of fisheries for chinook salmon near Ketchikan and along Prince of Wales Island. The chapter ends with methods for estimating harvest composition when not all members of a group can be distinguished during sampling, the usual situation when estimating contribution to harvest with coded wire tags. The chapter begins with a discussion of how to estimate harvest composition while simultaneously estimating harvest in a single, unaltered sampling design.

3.1 SIMULTANEOUS ESTIMATION

3.1.1 All Inspected Harvest Sampled. When all creels encountered by technicians are sampled, estimating harvest and its composition with the same sampling design is possible by ***post-stratifying harvest***. Estimating harvest by category begins by selecting one category (g) and disregarding all other harvest by setting all harvest not from category g observed during interviews to zero. Calculations proceed as described in Chapter 2 with the result that estimated harvest is now solely for category g. The procedure is repeated for each category in turn, producing estimates of harvest and estimates of their variances. Obviously these estimates are not independent. However, if there is a need to lump two or more categories together, the categories are redefined, data pooled, and calculations repeated for the new, redefined category. Onsite surveys of marine fisheries in southeast Alaska are examples of post-stratifying harvest to estimate harvest by species, species like salmon, halibut, and crab.

Post-stratifying harvest is possible only when all (or nearly all) creels of contacted anglers have been inspected and their contents sampled. Inspecting the creel of every interviewed angler is usually not a problem when categorizing is quick, such as when determining species, sex (sometimes), or when measuring length of harvested fish. However, determining age, cohort, and sometimes sex of harvested fish often requires a biological sample, such as a scale, an otolith, or a fish head. If biological sampling appreciably slows an interview, sample sizes to estimate overall harvest in the sampled period might decline. Taking biological samples only from a systematically selected subset of interviewed anglers would lessen the decline in the number of anglers interviewed. Also, some anglers will refuse to permit inspection of their creels. Whatever the reason for not inspecting some creels, sample size m (number of interviews) should be reduced to the number of interviews with biological samples if estimating harvest composition with post-stratification.

Because post-stratifying harvest is repetitive use of the methods described in the Chapter 2, no specific example of the technique will be given here. However, harvest composition estimated from post-stratification will be used later to demonstrate the robustness of other methods.

3.1.2 Inspected Harvest Subsampled. When only a fraction of inspected creels can be sampled, a ***stratified random sampling design*** to estimate harvest composition can be superimposed over the original design. Estimated harvest composition (a total) is the product of estimates of harvest and relative harvest composition across all strata:

$$\hat{H}_g = \sum_{h=1}^{L} \hat{H}_h \hat{p}_{gh} \tag{3.1a}$$

where $\hat{p}_{gh}$ is the estimated fraction of the harvest in stratum h composed of fish in category g. If estimates of relative harvest composition and harvest are treated as independent within each stratum, estimated variance of harvest by category can be calculated as (from Goodman 1960):

$$var(\hat{H}_g) = \sum_{h=1}^{L}\Big[\hat{H}_h^2 var(\hat{p}_{gh}) + \hat{p}_{gh}^2 var(\hat{H}_h) - var(\hat{p}_{gh}) var(\hat{H}_h)\Big] \tag{3.1b}$$

Estimates $\hat{p}_{gh}$ and estimates of their variances for each stratum are calculated as if data were randomly drawn from a binomial distribution (from Cochran 1977:51-52):

$$\hat{p}_{gh} = \frac{\sum_{i=1}^{n_h}\sum_{j=1}^{d_{hi}}\sum_{k=1}^{m_{hij}} H_{ghijk}}{\sum_{i=1}^{n_h}\sum_{j=1}^{d_{hi}}\sum_{k=1}^{m_{hij}} H_{hijk}} = \frac{y_{gh}}{\eta_h} \tag{3.2a}$$

$$var(\hat{p}_{gh}) = (1-\hat{f}_h)\frac{\hat{p}_{gh}(1-\hat{p}_{gh})}{\eta_h - 1} \tag{3.2b}$$

where m_{hij} here is the number of creels in the subsample on day j at location i in stratum h, η_h is the number of fish sampled in stratum h, y_{gh} is the number fish in category g within η_h, and $\hat{f}_h$ fraction of the harvest in stratum h sampled ($=\eta_h/\hat{H}_h$). The term $(1-\hat{f}_h)$ is an estimate of the correction for sampling without replacement from a finite population. A coin flip serves as an illustration of why biological samples can be treated as random samples within each stratum. Every flip of a coin happens at a specific time and place, but if the outcome of each flip is not related to when or where the flip occurs, the outcome is the result of a simple random process. ***While relative harvest composition might be related to when and where samples were taken across the season or across the fishing grounds, there should be no such relationship within each stratum.***

There are three options when superimposing strata to estimate harvest composition over the original strata used to estimate harvest:

- Keep the original strata used to estimate harvest intact;
- Combine adjacent strata in the original design and pool their harvest samples; or
- Split original strata and their harvest samples.

Keeping the original stratification is the most desirable option because stratified, unbiased harvest estimates (the $\hat{H}_h$) can be taken directly from the onsite survey without further calculation. The possibility of keeping the original stratification can be enhanced if estimating harvest composition is considered when the onsite survey is planned, instead of considering harvest composition as an afterthought.

The second option, combining original strata and pooling samples, is attractive only when sample sizes (η_h) within original strata are inadequate (too few to produce a sufficiently precise estimate). Estimates of harvest and its variance for several original strata are summed as per Equations 1.1 to represent a new stratum. For this option to produce unbiased estimates, relative harvest composition should change little across the original strata that will be combined.

The third option, splitting original strata, is the least preferred. Considerably more calculation is involved, and the subsequent "substrata" are no longer independent. Only when relative harvest composition changes dramatically within some of the original strata is this option considered, and only when sample sizes (the η_h) in the original strata are large.

Once data have been collected in an onsite survey, plotting estimated relative harvest composition across fishing days or locations can be helpful in determining if original strata should be "***redefined***". The cardinal rule is that ***relative harvest composition should vary little within redefined strata***.

Example 3.1. Humpback whitefish and least cisco are harvested in a spear fishery on the Chatanika River. The daily bag limit to restrict harvest is based collectively on all whitefish because participants can not determine species of their quarry until the fish is speared and held in hand. Like any mixed-stock fishery, estimates of its productivity are predicated on understanding the dynamics of each population exploited in it. This understanding is partially gained through knowledge of the harvest of each species.

In 1989, onsite surveys to estimate harvest by species were conducted near Olnes Pond, the Elliot Highway Bridge, and along the Steese Highway. Fishing occurred from 1800 to 2400 hours each night from early September through mid October that year. Sampling followed stratified, two-stage designs with nights (sampling periods) as first-stage units systematically drawn. Although almost every creel was sampled, a stratified random design superimposed over the original design can be used to estimate harvest composition by species. More information on this survey in 1989 can be found in Examples 2.1, 2.3, 2.5-8, and 2.10-11.

Estimated relative harvest composition switched slowly from least cisco to humpback whitefish as the season progressed in 1989 (Figure 3.1). Humpback whitefish migrate slowly upstream into and through the fishing grounds near Olnes Pond and near the Elliot Highway Bridge from August through September. In contrast, the more abundant least cisco migrated quickly into these

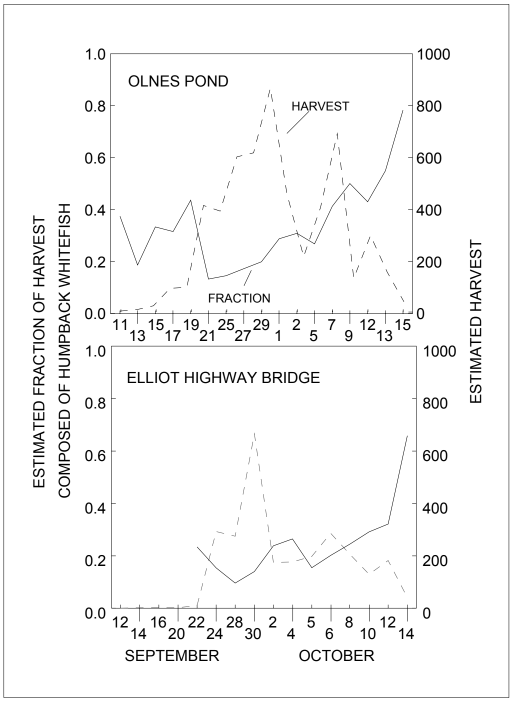


Figure 3.1 - Estimated harvest of whitefish (dashed line) from the Chatanika River in 1989 and the estimated proportion of harvest composed of humpback whitefish (solid line) by sampled day for two strata.

areas in late September. By mid October, least cisco had migrated back downstream while humpback whitefish reentered the lower two fishing areas after spawning upstream.

Data collected near Olnes Pond and the Elliot Highway Bridge were subdivided into five and three strata, respectively, to stabilize changes in relative harvest composition by species. Because estimated harvest along the Steese Highway was about 1% of all estimated harvest in 1989, data on relative harvest composition from this area was ignored. Sampling in the new "substratum" 8-12 October at the Elliot Highway Bridge occurred during 3 (=d_h) of the 5 (=D_h) sampling periods to produce $\hat{H}_{hi}$ of 206, 129, and 182 whitefish (Table 2.8) with $var(\hat{H}_{hi})$ of 1284, 533, and 2431 (Table 2.10). From Equations 2.1, 2.5, and 2.20:

$$\hat{\bar{H}}_h = \frac{\sum_{i=1}^{d_h} \hat{H}_{hi}}{d_h} = \frac{206+129+182}{3} = 172.33$$

$$\hat{H}_h = D_h \hat{\bar{H}}_h = 5(172.33) = 862$$

$$f_{1h} = \frac{3}{5} = 0.6$$

$$S_{1h}^2 = \frac{\sum_{i=2}^{d_h} (\hat{H}_{hi} - \hat{H}_{h(i-1)})^2}{2(d_h - 1)}$$
$$= \frac{(129-206)^2 + (182-129)^2}{2(3-1)} = 2{,}184.5$$

$$var(\hat{H}_h) = (1-f_{1h}) D_h^2 \frac{S_{1h}^2}{d_h} + f_{1h}^{-1} \sum_{i=1}^{d_h} var(\hat{H}_{hi})$$
$$= (1-0.6)(5)^2 \frac{2{,}184.5}{3} +$$
$$(0.6)^{-1}(1{,}284+533+2{,}431) = 14{,}362$$

Statistics for estimated harvest for the other newly drawn strata were calculated in the same manner. Statistics for all strata are listed in Table 3.1. Four hundred ninety-three harvested whitefish were inspected during 8, 10, and 12 October of which 153 were humpback whitefish. From Equation 3.1:

$$\hat{p}_{gh} = \frac{153}{493} = 0.31$$

$$var(\hat{p}_{gh}) = 0.000186 =$$
$$\left(1 - \frac{493}{862}\right) \frac{0.31(1-0.31)}{493-1}$$

Statistics for estimated relative harvest composition for the other newly drawn strata were calculated in the same manner. Statistics for all strata are listed in Table 3.1.

Of the estimated 16,667 whitefish harvested in 1989, 4,258 (SE=432) or 25.5% were estimated to have been humpback whitefish (Table 3.1). Because only two species of whitefish are caught in this fishery, least cisco comprise the balance of the harvest. Variance for estimated harvest of

Table 3.1 - Estimated harvest and relative harvest of humpback whitefish for eight strata superimposed on the original stratified, two-stage sampling design implemented in 1989 to estimate harvest of whitefish from the Chatanika River near Olnes Pond and the Elliot Highway Bridge.

	η_h	$\hat{H}_h$	$var(\hat{H}_h)$	Humpback Whitefish $\hat{p}_{gh}$	$var(\hat{p}_{gh})$	$\hat{H}_h\hat{p}_{gh}$	$\hat{H}_h^2 var(\hat{p}_{gh}) +$ $\hat{p}_{gh}^2 var(\hat{H}_h)$ - $var(\hat{p}_{gh})var(\hat{H}_h)$
Olnes Pond: (data on 2 October was substituted for missing data on 3 October)							
11-20 Sep	156	512	9,162	0.36	0.001030	184	1,448
21-30 Sep	1,023	5,796	156,792	0.17	0.000114	985	8,343
1-6 Oct	640	2,180	146,115	0.29	0.000227	632	13,334
7-12 Oct	883	2,258	503,535	0.43	0.000169	971	93,880
13-15 Oct	187	309	29,829	0.61	0.000504	188	11,132
Elliot Highway: (missing data on 18 and 26 September handled as per methods in Section 2.1.3)							
11 Sep - 7 Oct	1,204	4,627	1,267,773	0.20	0.000098	925	52,685
8-12 Oct	493	862	14,362	0.31	0.000186	267	1,516
13-15 Oct	22	123	586[1]	0.86	0.004605	106	4,377
Sums		16,667				4,258	186,716

[1] Because no estimate of variance was available for this stratum, the average of the CV for other strata was substituted for the missing statistic.

least cisco can be estimated by substituting 1- $\hat{p}_{gh}$ for $\hat{p}_{gh}$ in Table 3.1 and recalculating. For the curious, the SE for estimated harvest of least cisco is 1,086.

Note that estimated harvest of 16,667 whitefish in Table 3.1 is smaller than the 16,765 estimated in Examples 2.5 and 2.11. Such slight differences are common when original strata are subdivided. If original strata are used without subdivision, no differences in competing estimates of harvest will occur.

* * * *

There is a practical limit as to how far redefinition of strata should go. Although more redefined strata in space or time means more precision in estimated proportions and less bias, each additional stratum across a trend in harvest composition represents marginally less improvement. Usually six strata across a trend are the most that are practical (Cochran 1977:132-134). However, some fisheries will have more than one trend. Several recreational fisheries in Alaska exploit two separate stocks of salmon that sequentially migrate through fishing grounds (an early run and a later run). If each run has a single trend with older fish passing by earlier (as do runs of sockeye salmon), no more than six redefined strata during the passage of each stock (12 in all) would be judicious. Generally, only designs for situations with extreme changes in composition across time or space would warrant six redefined strata per trend.

3.2. SEPARATE ESTIMATION

Separate estimation implies that harvest and relative harvest composition are estimated with two, completely independent sampling programs. Estimated harvest and its estimated variance come from an offsite program, while relative harvest composition is estimated with an onsite survey designed solely for that purpose. Accurately estimating relative harvest composition from onsite sampling without simultaneously estimating harvest requires:

- sampling be proportional to harvest; or
- relative harvest composition have no spatial or temporal trends.

Under these circumstances, the onsite creel survey becomes a ***"self-weighted" survey***, a one-stage stratified survey in which strata (most or all of the strata) can be ignored and data pooled (see Section 1.5). From Cochran (1977:91) and in our notation, an unbiased estimate of each proportion representing relative harvest composition in a stratified sampling design would be:

$$\hat{p}_g = \sum_{h=1}^{L} \frac{H_h}{H} \hat{p}_{gh}$$

where $\hat{p}_{gh} = y_{gh} / \eta_h$. If sampling is proportional, $H_h / H = \eta_h / \eta$, and:

$$\hat{p}_g = \sum_{h=1}^{L} \frac{\eta_h}{\eta} \frac{y_{gh}}{\eta_h} = \frac{\sum_{h=1}^{L} y_{gh}}{\eta}$$

If samples are self-weighted (proportionally drawn), statistics from pooled data are mathematically equivalent to statistics calculated according to a stratified design.

Samples taken from harvest exiting a fishery at a particular place and time become "self-weighted" if the same fraction is sampled from ***all harvest regardless of when or where it exited the fishery*** (see Cochran 1977:91). If every *kth* fish harvested was systematically sampled throughout the fishing season, the resulting sample would be self-weighted. While this procedure theoretically will work every time, its application is usually impractical. A more realistic solution is to sample harvest by systematically selecting sampling periods in a pattern that is repeated throughout a fishing season and across fishing grounds. Sampled periods could be part of a day, a whole day, or several contiguous days. There is considerable latitude in setting this pattern so long as the pattern of sampling continues across the fishing grounds and throughout the fishing season. Each sampled period would represent a "stratum" in the self-weighted survey. Of course, if relative harvest composition is invariant, how and when fish are sampled are moot.

There is one rule for sample size in a sampled period: every fish exiting a fishery in front of a technician should be sampled. ***Whenever landing rate of harvest exceeds the maximum sampling rate of a technician(s), sampling becomes depensatory and is no longer self-weighted***. Depensatory sampling occurs when a smaller fraction of the harvest is sampled in those sampled periods with the greater harvest. This phenomenon often arises as a consequence of logistics. A technician has a maximum rate at which he/she can interview anglers and sample their creels. Once the rate of anglers leaving a fishery exceeds this sampling rate, the technician becomes swamped, and his/her sampling becomes inversely related to the harvest exiting during that period (i.e., depensatory). If harvest composition is different during sampled periods with greater numbers of fish harvested, depensatory sampling will produce biased estimates of harvest composition.

Strict adherence to proportional sampling is only as critical as is the intensity of trends in relative harvest composition. As mentioned before, when and where samples are drawn have no effect on estimates if there are no spatial or temporal trends in relative harvest composition. Conversely, the more pronounced the spatial and temporal trends, the more effect deviations from proportional sampling will have on bias in estimates. Long fisheries will tend to have shallow trends in relative harvest composition and are therefore more forgiving; short, intense fisheries with steep trends have great potential for bias if sampling is not strictly proportional. The potential for bias from deviations in strictly proportional sampling can be gauged for individual fisheries by calculating and plotting proportions in the compositions for each sampled period over time or space (as was done in Example 3.1).These plots are post facto information for last year's survey and a priori information for next year's. *If expected temporal and spatial trends in relative harvest composition are severe, as with some migratory species, statistics from self-weighted surveys will be too biased to be useful.* If so, only simultaneous estimation of harvest and its composition is the only realistic solution. Our experience has been that so long as sampling is somewhat proportional to harvest and there are moderate or shallow trends in relative harvest composition, bias in estimates from self-weighted surveys will be negligible.

Calculating statistics from self-weighted surveys begins with pooling data. From Cochran (1977:51-52) and in our notation, an estimated proportion of the harvest represented by group g and its estimated variance from pooled data are:

(3.3a)

$$\hat{p}_g = \frac{\sum_{i=1}^{n}\sum_{j=1}^{d_i}\sum_{k=1}^{m_{ij}} H_{gijk}}{\sum_{i=1}^{n}\sum_{j=1}^{d_i}\sum_{k=1}^{m_{ij}} H_{ijk}} = \frac{y_g}{\eta}$$

(3.3b)

$$var(\hat{p}_g) = (1-\hat{f})\frac{\hat{p}_g(1-\hat{p}_g)}{\eta-1}$$

where η is the sample size, y_g is the number of fish in category g in η, and $\hat{f}$ is the estimated fraction of the harvest sampled ($=\eta/\hat{H}$). Again, $1-\hat{f}$ is an estimate of the *fpc*.

In separate estimation, each estimate of harvest by category is a product of estimated relative harvest by category from the onsite survey and an estimate of harvest from an offsite survey:

(3.4a)

$$\hat{H}_g = \hat{H}\,\hat{p}_g$$

Because estimates of harvest and its relative composition come from independent sampling programs, estimated variance of harvest by category is calculated as (from Goodman 1960):

(3.4b)

$$var(\hat{H}_g) = \hat{H}^2 var(\hat{p}_g) + \hat{p}_g^2 var(\hat{H}) - var(\hat{p}_g)var(\hat{H})$$

Example 3.2. Age composition of halibut landed in the sport fishery out of Resurrection Bay (Seward) was estimated in 1991 with a self-weighted creel survey. Almost all anglers participating in this fishery use a single boat harbor. Signs exhorting fishermen to discard their halibut carcasses in large barrels were placed (along with the barrels) at all cleaning stations in the harbor every Thursday, Friday, Saturday, Sunday, and Monday. Signs and barrels were removed late each Monday night and were returned early each Thursday morning. Sampling began 18 June and ended 17 September. Age was determined from otoliths removed from discarded carcasses.

Table 3.2 - Estimated relative composition (the $\hat{p}_g$) by age of halibut landed in Seward in 1991.

Age (g)	η	$\hat{p}_g$	$var(\hat{p}_g)$ (x1000)
4	4	0.008	0.01620
5	16	0.033	0.06319
6	54	0.111	0.19600
7	82	0.169	0.27835
8	108	0.222	0.34301
9	73	0.150	0.25331
10	49	0.101	0.17991
11	35	0.072	0.13262
12	35	0.072	0.13262
13	15	0.031	0.05936
14	2	0.004	0.00813
15	4	0.008	0.01620
16	4	0.008	0.01620
17	3	0.006	0.01218
18	0	0.000	0.00000
19	1	0.002	0.00407
20	1	0.002	0.00407
Sum	486		

Estimated ages of the 486 halibut sampled in 1991 range from 4 to 20 years of age with most age 8 (Table 3.2). Data were pooled as per Equation 3.3 to produce estimates of relative harvest composition and their estimated variance. With an estimated 12,961 halibut landed in 1991, $\hat{f}$ = 486/12,961 = 0.0375 and the estimated *fpc* = 0.9625. For 8 year-olds:

$$\hat{p}_g = \frac{108}{486} = 0.222$$

$$var(\hat{p}_g) = = (1-0.0375)\frac{0.222(1-0.222)}{486-1} = 0.00034301$$

from Equations 3.3. From Equations 3.4, the estimated harvest of 8 year-old halibut in 1991 and its estimated variance are:

$$\hat{H}_g = 12{,}961(0.222) = 2{,}880$$

$$var(\hat{H}_g) = (12{,}961)^2(0.00034301) + (0.222)^2(889{,}921) - (0.00034301)(889{,}921) = 101{,}763$$

where $var(\hat{H}) = 889{,}921$. Calculating estimates for other age groups follows the same pattern.

Sampling at Resurrection Bay would have been proportional if 1) the same fraction of the week's harvest of halibut ended up in the barrels, 2) the barrels were never full, and 3) technicians processed all discarded halibut each day. There is no means to determine if the first presumption is true. The second condition was met while the third was not. Technicians could not work their way through the day's discard about 20% of the time. In these instances, sampling was depensatory, and some bias was imparted to the statistics.

As is often the case elsewhere, bias in statistics due to depensatory sampling of halibut landed at Seward in 1991 is negligible because there was little change in age composition as the season progressed. Artificially stratifying data by month shows that there was little or no variation in estimated relative age composition of harvest across the season (Figure 3.2). What differences there were are not significant (χ^2 = 10.08, df = 12, P = 0.63 for age groups 4-5, 6-7, 8-9, 10-11, and 12<). Under these circumstances, depensatory sampling is irrelevant.

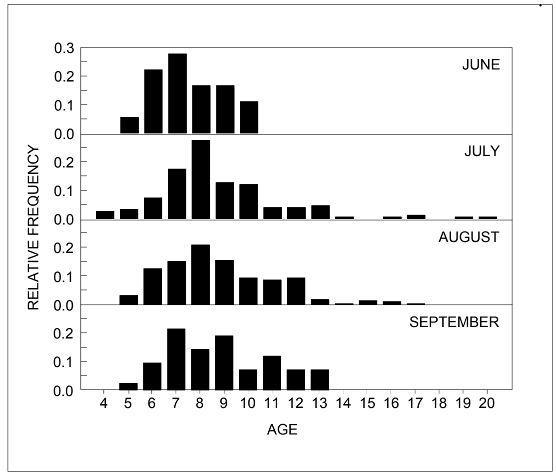


Figure 3.2 - Estimated relative age composition of harvested halibut landed at the boat harbor at Seward Alaska in 1991.

Example 3.3. Although sampling the spear fishery on the Chatanika River was not designed to be proportional in 1989, data collected that year can be analyzed as if they had been. Participants were interviewed approximately every other night during the short, five-week fishery at the two major access locations (near Olnes Pond and near the Elliot Highway Bridge). When data are pooled across interviews, across days, and locations and processed as per Equations 3.3, the estimated relative species composition for the season's harvest is 27% humpback whitefish (SE=0.6%) and 73% least cisco (SE=0.7%). When the same data were analyzed as per a stratified random design, an estimated 4,258 or 25.5% of 16,667 whitefish harvested were humpback whitefish in 1989 (Example 3.1). When the same data were post-stratified by species (see Section 3.1), an estimated 4,006 or 24% of 16,765 whitefish harvested were humpback whitefish (from Examples 2.5 and 2.10).

Estimates of relative harvest composition by species from pooling data are similar to estimates from both stratifying and post-stratifying data in spite of trends and depensatory sampling. There was considerable daily and seasonal fluctuation of both estimated harvest and the estimated fraction of that harvest sampled in 1989 at both Olnes Pond and near the Elliot Highway Bridge (Figure 3.3). Correlations between these two statistics are negative (Pearson's $r = -0.50$ at Olnes Pond and -0.41 near the Elliot Highway Bridge) indicating that sampling was mildly depensatory. Estimated fraction of the harvest composed of humpback whitefish rose steadily through time (Figure 3.1). The same phenomenon occurred in the survey near the Elliot Highway Bridge, only the trend was weaker.

Example 3.4. In 1991 most sockeye salmon harvested at the confluence of the Russian and Kenai Rivers exited the fishery through parking lots near the Ferry or at Grayling Campground. Both harvest and its relative age composition were estimated with an onsite survey during the late run of sockeye salmon (late July through late August). Although days were scheduled systematically for sampling to estimate harvest, technicians sampled harvest for scales only on irregularly scheduled days at both sampling locations. Age was later determined from examining these scales. In this example, age is defined as the number of years sockeye salmon spent at sea with salmon being two-ocean or three-ocean fish. Examples 2.2 and 2.4 contain more details on this survey in 1991.

Four strata defined by location and season are superimposed over the original sampling design for this example (Table 3.3). Season was used to redefine strata in the original survey because of trends in estimated harvest, its relative age composition, and sample sizes. Because anglers exiting near the Ferry or through the Grayling Campground fished the same area, there was no difference in the relative composition of harvest exiting at these two access points. During the first 10 days of the season (through 7 August), two-ocean fish represented an estimated 54% of the harvest (Figure 3.4, Table 3.4); two-ocean fish comprised a significantly larger portion (80%) of the remaining harvest (χ^2=22.00, df=1, P<0.001). From this superimposed stratified random design, an estimated 16,646 [=(0.54)17,075+(0.80) 9,282]

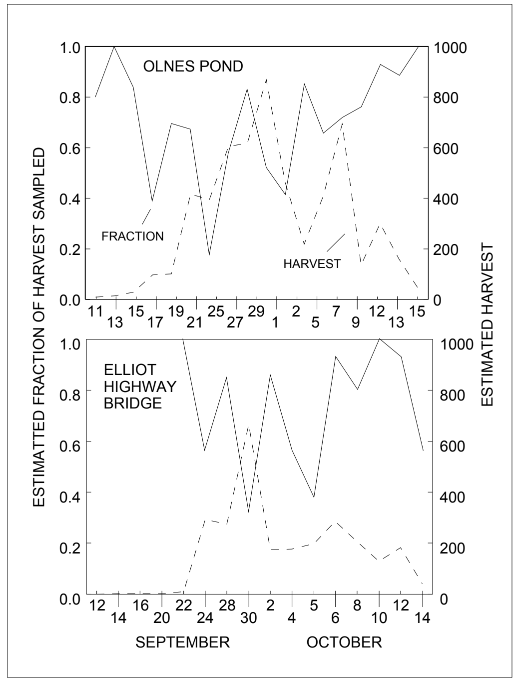


Figure 3.3 - Estimated harvest of whitefish (dashed line) and the proportion of estimated harvest sampled (solid line) by day in 1989 at the Chatanika River. Because of its small size, the stratum associated with the Steese Highway is not included.

Table 3.3 - Estimated harvest of sockeye salmon by redefined strata at the confluence of the Russian and Kenai rivers from late July through late August, 1991.

	29 Jul - 7 Aug	8 Aug - 28 Aug
Ferry	13,484	6,323
Grayling Campground	3,591	2,959
Sums	17,075	9,282
Grand Sum	26,357	

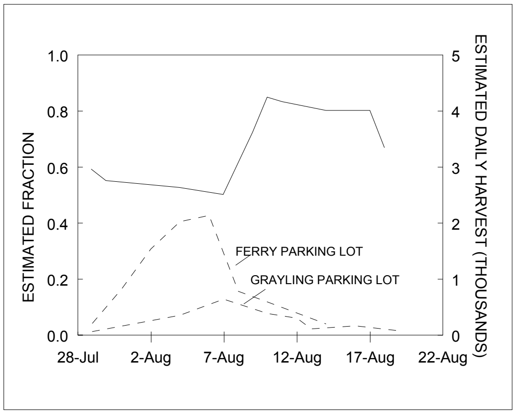


Figure 3.4 - Estimated harvest of sockeye salmon and its estimated relative age composition at the confluence of the Russian and Kenai rivers in 1991. The solid line represents the estimated fraction two-ocean fish, while dashed lines represent estimated harvest by anglers exiting the fishery from access points (parking lots).

two-ocean fish were harvested, about 63% (=16,646/(26,375)100] of all sockeye salmon harvested at the confluence that year (Table 3.3).

Sampling to estimate relative age composition was demonstrably depensatory. Fewer fish were sampled during the early strata when most of the estimated harvest occurred (Table 3.4). If sampling the harvest had been proportional, the relative distribution of samples across seasonal

Table 3.4 - Fractions of estimated harvest, samples, and estimated harvest composition from the confluence of the Russian and Kenai rivers from late July through late August, 1991.

	29 Jul - 7 Aug	8 Aug - 28 Aug
Estimated Harvest	0.65	0.35
Fraction Sampled	0.28	0.72
Fraction two-ocean	0.54	0.80

strata would be the same as the distribution of harvest. If data had been pooled instead of stratified, a biased estimate of 73% [=(0.28)0.54 + (0.72)0.80] of the harvest would have been composed of two-ocean fish, an estimated 19,254 [=0.73(26,375)] sockeye salmon. In contrast, the unbiased estimate from the superimposed, stratified design is 16,646.

* * * *

Examples in this section demonstrate some of the breaks and some of the pitfalls in onsite surveys designed to estimate relative harvest composition. The fates are more forgiving if the target species is non-migratory, such as halibut off Seward. While members of this species do migrate along the coast of the eastern Pacific Ocean, age compositions of local populations do not appreciably change throughout summer months, as indicated in Figure 3.2. Under this circumstance, self-weighted surveys provide accurate estimates of harvest composition.

More problems arise when the target population is migratory, as with whitefish on the Chatanika River in 1991. Change in estimated species composition through time was quite sharp in this instance. Although sampling was not designed in this survey to be proportional, sampling was spread systematically across the duration of the fishery. Sampling was mildly depensatory, but bias in estimated relative harvest composition by species was relatively mild. "Unbiased" estimates of relative harvest composition are 24% humpback whitefish from post-stratifying data and 26% from a superimposed stratified random design; the corresponding "biased" estimate from pooling all data is 27%.

The fates would not have been so forgiving if data collected in 1991 had to be pooled to estimate relative age composition of sockeye salmon harvested at the confluence of the Russian and Kenai rivers that year. Change in estimates of relative age composition of harvested sockeye was just as sharp as the change observed in species composition on the Chatanika River. However, sampling at parking lots to estimate age was not spread out across the fishing season, but was tardy, occurring when little harvest was left to sample. As a result, sampling was strongly depensatory, and estimated age composition of the harvest from pooling would have been strongly biased (73% versus 63% two-ocean salmon).

3.3 INCOMPLETELY MARKED COHORTS

Methods described in Section 3.1-2 are appropriate when every sampled fish in a harvest can be assigned to a group. Sex, age, length, or species describe every fish, so harvest composition based on these natural attributes can be estimated with the aforementioned methods. These methods can also be used to estimate harvest of arbitrarily defined groups, such as fish reared in a hatchery***, if every fish in such a group is marked***. However, different methods are available when only ***some of the fish in a group carry marks***.

Estimating harvest from cohorts with coded wire tags is an instance in which not all harvested members of a group (cohort) can be detected. A coded wire tag is a ≤1-mm long, small diameter wire with a code etched on its surface. Tags are inserted into the snouts of juvenile salmonids to identify their cohort; tagged fish are also marked by excising their adipose fin. Because of expense, not all juveniles in a cohort are tagged and marked. Harvest of returning adults is inspected, and heads from inspected fish missing their adipose fins are sent to a laboratory and dissected. If a tag is found, it's decoded. The tally of recovered tags is expanded upward for the fraction of harvest inspected, for the fraction of a group with tags, and for any marked fish sampled in the harvest whose dissected head did not yield a tag.

3.3.1 Stratified "Random" Sampling. Methods similar to those described in Section 3.1.2 are followed when estimating harvest composition by cohort is superimposed on an onsite creel survey to estimate harvest. From Bernard and Clark (1996), estimated relative harvest composition of ***tagged members*** of cohort g in stratum h is:

(3.5a)

$$\hat{p}_{gh} = \frac{\sum_{i=1}^{n_h}\sum_{j=1}^{d_{hi}}\sum_{k=1}^{m_{hij}} H_{ghijk}}{\lambda_h \sum_{i=1}^{n_h}\sum_{j=1}^{d_{hi}}\sum_{k=1}^{m_{hij}} H_{hijk}} = \frac{y_{gh}}{\lambda_h \eta_h}$$

where η_h is the size of the sample drawn from stratum h, y_{gh} is the number of tags from cohort g recovered from η_h, and λ_h is the compound fraction for stratum h defining the probability that the head from an inspected fish with a missing adipose fin will yield a readable tag from ***any*** cohort. For instance, if in a sample 100 fish were missing adipose fins, 90 heads reached the labs, tags were detected in 80 heads, and 70 of these tags successfully decoded, $\lambda_h = 0.7875 = (90/100)(70/80)$. From Bernard and Clark (1996), a simple, large-sample approximation of variance for $\hat{p}_{gh}$ is:

(3.5b)

$$var(\hat{p}_{gh}) = \frac{\hat{p}_{gh}(1-\lambda_h \hat{f}_h \theta_g)}{\lambda_h \eta_h}$$

where θ is the fraction of cohort g with tags. Bernard and Clark (1996) found this approximation to be within a few fish of unbiased estimates of variance over a representative set of examples pertaining to recreational fisheries. Relationships among Equations 3.2, 3.3, and 3.5 are described in Appendix G.

Harvest of ***tagged and untagged*** fish from cohort g in stratum h is estimated as a function of $\hat{p}_{gh}$, the estimated harvest in the stratum, and the fraction θ of cohort g with tags:

(3.6a)

$$\hat{H}_{gh} = \hat{H}_h \hat{p}_{gh} \theta_g^{-1}$$

From Goodman's (1960) method, estimated variance is:

(3.6b)

$$var(\hat{H}_{gh}) = \hat{H}_{gh}^2 \left[G(\hat{p}_{gh}) + G(\hat{H}_h) - G(\hat{H}_h) G(\hat{p}_{gh}) \right]$$

where G() is the square of the coefficient of variation. If the fraction θ is not known, but is estimated with significant imprecision:

(3.7a)

$$\hat{H}_{gh} = \hat{H}_h \hat{p}_{gh} \hat{\theta}_g^{-1}$$

(3.7b)

$$var(\hat{H}_{gh}) = \hat{H}_{gh}^2 \left[G(\hat{H}_h) + G(\hat{p}_{gh}) - G(\overline{\theta_g^{-1}}) - G(\hat{H}_h)G(\hat{p}_{gh}) - G(\hat{H}_h)G(\overline{\theta_g^{-1}}) - G(\hat{p}_{gh})G(\overline{\theta_g^{-1}}) + G(\hat{H}_h)G(\hat{p}_{gh})G(\overline{\theta_g^{-1}}) \right]$$

Estimates of harvest from a cohort for the season and its estimated variance are calculated as per Equations 1.1. Often an estimate of harvest from several cohorts is desired, such as an estimate of all hatchery-produced fish harvested in an area. The estimate of harvest for several cohorts is sum of harvests by cohorts. From Bernard and Clark (1996), the estimated variance of this statistic contains a covariance which under some circumstances can be sizable:

(3.8)

$$var(\Sigma\hat{H}_g) = \sum_g var(\hat{H}_g) + 2\sum_h G(\hat{H}_h)\sum_g \sum_{g'>g} \hat{H}_{gh}\hat{H}_{g'h}$$

Example 3.5. Harvest of chinook salmon reared in Alaska's hatcheries and caught in recreational fisheries near Ketchikan is estimated annually. In 1995 an onsite stratified, three-stage design was used to estimate harvest every two weeks (the stratification). Some harvest was inspected as an adjunct to counting and interviewing anglers.

Four strata were superimposed over the 12 in the original design by summing estimates of harvest and their estimates of variances across old strata, and by pooling data and recalculating other statistics (Table 3.5).

Coded wire tags were recovered from 18 cohorts released by six hatcheries (Table 3.6). Using the two tags recovered from cohort 04-35-31 (Deer Mountain Hatchery) from 19 June through 30 July as the start of the example (from Equations 3.5):

$$\hat{p}_{gh} = \frac{y_{gh}}{\lambda_h \eta_h} = \frac{2}{(0.9752)159} = 0.0129$$

$$\hat{H}_{gh} = \hat{H}_h \hat{p}_{gh} \theta_g^{-1} = $$

$$= 1{,}611(0.0129)(0.3226)^{-1} = 64.4$$

$$var(\hat{p}_{gh}) = \frac{\hat{p}_{gh}(1-\lambda_h \hat{f}_h \theta_g)}{\lambda_h \eta_h} =$$

$$\frac{0.0129[1-0.9752(0.0987)0.3226]}{(0.9752)159} = 0.0000806$$

From Table 3.5, $CV(\hat{H}_h) = 0.182$ for the stratum 19 June - 30 July, so $G(\hat{H}_h) = 0.0331$. Continuing the calculations:

$$G(\hat{p}_{gh}) =$$

$$\frac{var(\hat{p}_{gh})}{\hat{p}_{gh}^2} = \frac{\hat{p}_{gh}(1-\lambda_h \hat{f}_h \theta_g)}{\lambda_n \eta_h \hat{p}_{gh}^2} = \frac{1-\lambda_h \hat{f}_h \theta_g}{y_{gh}} =$$

Table 3.5 - Stratified statistics from the onsite creel survey to estimate harvest of chinook salmon in the marine recreational fishery near Ketchikan in 1995.

	$\hat{H}_h$	$var(\hat{H}_h)$	$CV(\hat{H}_h)$	η_h	$\hat{f}_h$	λ_h
28 Apr - 18 Jun	1,218	32,672	0.148	129	0.1059	0.8310
Derby	415	2,283	0.115	246	0.5928	1.0000
19 Jun - 30 Jul	1,611	85,873	0.182	159	0.0987	0.9752
31 Jul - 24 Sep	232	6,381	0.344	22	0.0948	1.0000

Table 3.6 - Number of tags recovered (y), estimates of relative and absolute harvest composition and estimates of their variances for 18 hatchery-produced cohorts of chinook salmon harvested in three strata near Ketchikan in 1995. No tags were recovered from 31 July through 24 September.

Hatchery or Site of Release:		24 April - 18 June					Derby					19 June - 30 July				
Tag Code	θ	y	$\hat{p}_{gh}$ x100	$var(\hat{p}_{gh})$ x10,000	$\hat{H}_{gh}$	$var(\hat{H}_{gh})$	y	$\hat{p}_{gh}$ x100	$var(\hat{p}_{gh})$ x10,000	$\hat{H}_{gh}$	$var(\hat{H}_{gh})$	y	$\hat{p}_{gh}$ x100	$var(\hat{p}_{gh})$ x10,000	$\hat{H}_{gh}$	$var(\hat{H}_{gh})$
Bell Island Net Pens:																
04-35-60	0.9671	1	0.9328	0.8138	12	129										
04-37-51	0.9747	1	0.9328	0.8132	12	127										
Deer Mountain:																
04-35-31	0.3226											2	1.2898	0.8060	64	2,081
04-37-46	0.8071	1	0.9328	0.8263	14	188										
04-37-50	0.7905						1	0.4065	0.0878	2	2	1	0.6449	0.3843	13	160
04-38-57	0.9452											1	0.6449	0.3781	11	110
04-38-58	0.9407						1	0.4065	0.0731	2	1					
04-39-04	0.5280						1	0.4065	0.1135	3	7					
Tamgas Creek:																
47-16-10	0.1376	1	0.9328	0.8787	82	6,884										
47-16-13	0.1584	1	0.9328	0.8771	72	5,185						1	0.6449	0.4096	66	4,239
47-16-54	0.2911						1	0.4065	0.1367	6	28					
Carroll Inlet:																
04-37-08	0.0911	1	0.9328	0.8823	125	15,768										
04-41-02	0.1003											1	0.6449	0.4119	104	10,630
Neets Bay:																
04-37-02	0.2857						1	0.4065	0.1373	6	29	1	0.6449	0.4045	36	1,287
04-39-38	0.1321											1	0.6449	0.4106	79	6,110
Whitman Lake:																
04-35-04	0.7407	1	0.9328	0.8315	15	225	1	0.4065	0.0927	2	3	1	0.6449	0.3863	14	183
04-37-04	0.5048						1	0.4065	0.1158	3	8					
04-41-01	0.4943						1	0.4065	0.1168	3	8					
Sums					332	28,506				27	86				387	24,800

$$= \frac{1 - 0.9752(0.0987)0.3226}{2} = 0.4845$$

Because θ is known with negligible error (cohorts were from hatcheries), Equations 3.6 were used to estimate variance:

$$var(\hat{H}_{gh}) = 2{,}081 =$$
$$(64.4)^2[\ 0.4845 + 0.0331\ -\ 0.4845(\ 0.0331)\ ]$$

From Equation 1.1a, the estimated harvest composition for hatchery-produced chinook salmon in the recreational fishery near Ketchikan in 1995 is 746 (=332+27+387). Decomposing Equation 3.8:

$$\sum_g var(\hat{H}_g) = 28{,}506 + 86 + 24{,}800 = 53{,}392$$

For the stratum running from 19 June through 30 July:

$$\sum_g \sum_{g'>g} \hat{H}_{gh}\hat{H}_{g'h} = 61{,}239 =$$

$$\begin{aligned} 64(13) + 64(11) + 64(66) + \ldots + 64(14) &+ \\ 13(11) + 13(66) + \ldots + 13(14) &+ \\ 11(66) + \ldots + 11(14) &+ \\ &\vdots \\ 79(14) & \end{aligned}$$

Similar statistics for the other three strata are 40,991; 309; and 0. With CVs from Table 3.5, estimated covariance is:

$$\sum_h G(\hat{H}_h) \sum_g \sum_{g'>g} \hat{H}_{gh}\hat{H}_{g'h} = 2{,}933.4 =$$

$$(0.148)^2(40{,}991) + (0.115)^2(309) +$$
$$(0.182)^2(61{,}329) + (0.344)^2(0)$$

Recombining the parts of Equation 3.8 provided an estimate of variance for estimated harvest of hatchery-produced chinook salmon:

$$var(\Sigma\hat{H}_g) = 53{,}392 + 2(2{,}933.4) = 59{,}259$$

The estimated harvest of 746 chinook salmon has an estimated CV of 33% (SE=243).

3.3.2 Self-weighted Surveys. When harvest is estimated in an offsite sampling program, such as the Statewide Harvest Survey, and ***relative*** harvest composition in an onsite program, onsite sampling must be proportional if estimates are to be unbiased. Methods similar to those described in Section 3.2 are followed when estimating relative harvest composition for incompletely marked cohorts in an onsite creel survey. Again, we will use cohorts marked with coded wire tags to demonstrate these methods. From Bernard et al. (submitted), estimated relative harvest composition of ***tagged members*** of cohort g and its estimated variance for the season are:

(3.9a)

$$\hat{\hat{p}}_g = \frac{\sum_{i=1}^{n}\sum_{j=1}^{d_i}\sum_{k=1}^{m_{ij}} \frac{H_{gijk}}{\lambda_{ij}}}{\sum_{i=1}^{n}\sum_{j=1}^{d_i}\sum_{k=1}^{m_{ij}} H_{ijk}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{d_i} \frac{y_{gij}}{\lambda_{ij}}}{\eta}$$

(3.9b)

$$var(\hat{\hat{p}}_g) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{d_i} \frac{y_{gij}}{\lambda_{ij}^2}(1 - \lambda_{ij}\hat{f}\,\theta_g)}{\eta^2}$$

where i and j define a sampled period, η is the pooled sample size, y_{gij} is the number of recovered tags from cohort g in a sampled period, and $\hat{f}$ is the estimated sampled fraction of the season's harvest. Estimated harvest of ***tagged and untagged*** fish from cohort g and its estimated variance are:

(3.10a)

$$\hat{H}_g = \hat{H}\,\hat{\hat{p}}_g\theta_g^{-1}$$

(3.10b)

$$var(\hat{H}_g) = \hat{H}_g^2\left[G(\hat{\hat{p}}_g) + G(\hat{H}) - G(\hat{\hat{p}}_g)G(\hat{H})\right]$$

if θ is known with negligible error. If θ is estimated, $\hat{\theta} \rightarrow \theta$ in the equations above and:

(3.10c)

$$var(\hat{H}_g) = \hat{H}_g^2 \left[G(\hat{H}) + G(\hat{\bar{p}}_g) + G(\overline{\theta_g^{-1}}) - G(\hat{H})G(\overline{\theta_g^{-1}}) - G(\hat{H})G(\overline{\theta_g^{-1}}) - G(\hat{\bar{p}}_g)G(\overline{\theta_g^{-1}}) + G(\hat{H})G(\hat{\bar{p}}_g)G(\overline{\theta_g^{-1}}) \right]$$

If an estimate of harvest from several cohorts is desired, ***the estimate of harvest for several cohorts is the sum of harvests by cohorts***. From Bernard et al. (1996), the estimated variance of summed harvests contains a covariance:

(3.11)

$$var(\Sigma\hat{H}_g) = \sum_g var(\hat{H}_g) + 2G(\hat{H})\sum_g \sum_{g'>g} \hat{H}_g \hat{H}_{g'}$$

with the last term representing the covariance.

Example 3.6. An onsite creel survey at Craig in 1995 was designed to estimate the relative cohort composition of chinook salmon produced in Alaska's hatcheries and harvested along the outer shore of Prince of Wales Island. One technician inspected landings of chinook salmon simultaneously at two harbors from 1100 hrs to 2000 hrs Thursday through Sunday from 1 May through 8 September. When on duty, the technician inspected virtually every landed chinook salmon by the offshore marine recreational fishery. While not insuring that sampling had been proportional, the regular scheduling and lack of depensatory sampling probably produced self-weighted samples. Sampling in other years at other sites had shown that relative harvest composition was similar along the west-side of Prince of Wales Island.

From the statewide harvest survey, an estimated 8,368 chinook salmon were harvested (estimated variance 525,625) from the west side of Prince of Wales Island in 1995 of which 1,693 were sampled at Craig ($\hat{f}$ = 0.202 = 1,693/8,368). No heads were lost and all detected tags were decoded, meaning λ = 1. Recovered coded wire tags represented seven cohorts from three Alaskan hatcheries (Table 3.7). Using the tag recovered from cohort 03-22-28 (Little Port Walter) to start this example:

Table 3.7 - Number of tags recovered (y), estimates of relative and absolute harvest and estimates of their variances for seven hatchery-produced cohorts of chinook salmon for recreational fisheries along the west side of Prince of Wales Island in 1995.

Hatchery: Tag Code	θ_g	$\sum_i \sum_j y_{gij}$	$\hat{\bar{p}}_g$ x1000	$var(\hat{\bar{p}}_g)$ x10^6	$G(\hat{\bar{p}}_g)$	$\hat{H}_g$	$var(\hat{H}_g)$
Deer Mountain:							
04-37-48	0.6698	1	0.5907	0.3016	0.8730	7	47
Little Port Walter:							
03-22-26	0.9690	1	0.5907	0.2805	0.8039	5	21
03-22-28	0.6270	1	0.5907	0.3046	0.8730	8	54
03-22-31	0.9911	1	0.5907	0.2789	0.7993	5	20
Medvejie:							
04-36-44	0.0719	1	0.5907	0.3438	0.9853	69	4,652
04-36-48	0.0838	1	0.5907	0.3430	0.9830	59	3,422
04-36-53	0.0349	1	0.5907	0.3464	0.9928	142	19,882
Sums						295	28,098

$$\hat{\hat{p}}_g = \frac{\sum_{i=1}^{n}\sum_{j=1}^{d_i}\frac{y_{gij}}{\lambda_{ij}}}{\eta} = \frac{1/1.0}{1{,}693} = 0.0005907$$

$$var(\hat{\hat{p}}_g) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{d_i}\frac{y_{gij}}{\lambda_{ij}^2}(1-\lambda_{ij}\hat{f}\,\theta_g)}{\eta^2} =$$

$$\frac{\frac{1}{(1.0)^2}\left[1-1.0(0.202)0.6270\right]}{(1{,}693)^2} = 0.0000003046$$

Note that when $\lambda_{ij} = \lambda = 1$, all recoveries can be pooled. Estimated harvest and an estimate of its variance for this cohort from Equation 3.10 is:

$$\hat{H}_g = \hat{H}\,\hat{\hat{p}}_g\theta_g^{-1} = \frac{8{,}368(0.0005907)}{0.6270} = 7.9$$

$$var(\hat{H}_g) = \hat{H}_g^2\Big[G(\hat{\hat{p}}_g) + G(\hat{H}) - G(\hat{\hat{p}}_g)G(\hat{H})\Big]$$

$$= (7.9)^2\left[0.8730 + 0.0075 - 0.8730(0.0075)\right] = 54$$

Calculated statistics for this and other cohorts are listed in Table 3.7. An estimated 295 chinook salmon produced in Alaskan hatcheries were harvested along the west side of Prince of Wales Island. Calculating an estimate of variance for this statistic begins by calculating the covariance in Equation 3.11 with the sum of unique products:

$$\sum_{g}\sum_{g'>g}\hat{H}_g\hat{H}_{g'} = 29{,}228 =$$

$$\begin{aligned}7(5) + 7(8) + 7(5) + \ldots\ldots + 7(142) &+\\ 5(8) + 5(5) + \ldots\ldots + 5(142) &+\\ 8(5) + \ldots\ldots + 8(142) &+\\ &\vdots\\ 59(142)&\end{aligned}$$

Putting the parts of Equation 3.11 back together again with the sum of the estimated variances from Table 3.7 and $G(\hat{H}) = 0.0075$:

$$var(\Sigma\hat{H}_g) = 28{,}536 = 28{,}098 + 2(0.0075)\,29{,}228$$

In summary, an estimated 295 chinook salmon from Alaskan hatcheries were harvested along the west side of Prince of Wales Island. Estimated variance of this statistic is 28,536 making the SE = 169 and CV = 57%.

* * * *

While methods in Section 3.3 are usually applied to surveys of salmon fisheries and cohorts that are incompletely marked with coded wire tags, these methods are appropriate whenever a subset of a cohort is marked. A short list of works in which statistical methods for estimating harvest composition by partially marked cohorts includes Clark and Bernard (1987), Geiger (1990), Newman (1990), Schnute (1992), Bernard and Clark (1996), and Bernard et al. (submitted).

3.4 OPPORTUNITY BIAS

Estimates of harvest composition will be inaccurate because of opportunity bias if:

- creels from different types of anglers have unequal probabilities of being sampled, **AND**

- harvests by these types of anglers have different compositions.

Unequal probabilities of sampling creels can arise because of different probabilities of encountering anglers (see Section 2.2.2.2), or because different types of anglers have different ways of handling their harvest. Harvest composition can also differ among types of anglers because of spatial differences in the distribution of exiting anglers across access points or because different types of anglers tend to keep fish of different sizes.

Opportunity bias can be avoided through post-stratification by type of angler or by stratification by location. Knowledge of where different types of anglers are likely to fish, how they select captured fish for their creel, how they likely will exit a fishery, and how they are likely to process their harvest can be used to design sampling that will increase the likelihood that ***estimates of harvest composition will be relatively unbiased.***

Example 3.7. The best way to find out how anglers behave is to ask them. The halibut fishery near Homer has been surveyed to estimate relative harvest composition by age for the International Pacific Halibut Commission. Participants have been contacted as they exited the fishing grounds at the Homer Spit, while they attended meetings of the local association of charter boat operators, on visits to our office in Homer, and informally while patronizing local, commercial establishments. Over the years, these contacts revealed a distinct difference where guided and unguided anglers fish, how they processed their harvest, and where they exited the local fishery (Table 3.8). During the last few years, crews on chartered boats began to clean smaller halibut and discard the carcasses at sea. In contrast, private anglers tended to land all halibut. Since the only opportunity for technicians to sample the harvest was ashore, cooperation was solicited from several charter boat owners to retain carcasses of all halibut cleaned at sea. Otoliths from these carcasses and from halibut cleaned ashore by unguided anglers indicated that in some years age composition of halibut harvested by guided and unguided anglers were different. Opportunity bias was avoided through post-stratification. Relative age composition of harvest was estimated independently for each of these two types of anglers in the onsite survey. Relative age composition of the harvest was estimated for guided anglers from carcasses returned by cooperating charter-boat operators. Sampling at Homer Spit was used to calculate relative harvest composition for unguided anglers. Harvest was estimated separately for guided and unguided anglers in the Statewide Harvest Survey.

* * * *

Know your fish and know your fishery. Good advice for designing any creel survey, but especially relevant when designing a survey to estimate relative harvest composition. Without concomitant estimates of harvest, there is no way after data have been collected to avoid inaccurate estimates from depensatory sampling or from opportunity bias. Fortunately, failure to strictly sample harvest proportionally usually produces negligibly biased estimates of harvest composition, especially when relative composition of the population is reasonably stable or

Table 3.8 - Questions and their common answers from two types of anglers fishing for halibut in waters near Homer.

	Chartered Anglers	Private Anglers
Where did they fish?	Farther from Homer Spit	Nearer to Homer Spit
When did they exit the fishery?	About the same time each day	At various times throughout the day
Where did they exit the fishery?	At Homer Spit	At Homer Spit and elsewhere
How did they process their harvest?	Small halibut were cleaned at sea with carcasses pitched over board; large halibut were cleaned onshore after being photographed	All halibut were cleaned onshore

homogenous. Still, there's no substitute for doing the homework. Knowledge of the fished population can be obtained from literature on the species, past stock assessment programs, and past creel surveys. Information on a fishery can also be obtained from informal contacts, as occurred in Example 3.7, or from more formally designed demographic surveys. ***The annual statewide harvest survey is another source of valuable information on a fishery; respondents can be contacted to provide insightful information, such as where they had exited a fishery and how they had processed their fish.***

3.5 REFERENCES FOR EXAMPLES

Howe, A. L., G. Fidler, A. E. Bingham, and M. J. Mills. 1996. Harvest, catch, and participation in Alaska sport fisheries during 1995. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 96-32. Anchorage .

Hubartt, D. J., A. E. Bingham, and P. M. Suchanek. 1996. Harvest estimates of selected marine sport fisheries in southeast Alaska during 1995. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 96-28. Anchorage.

Marsh, L. E. 1992. Catch and effort statistics for the sockeye salmon sport fishery during the late run to the Russian River with estimates of escapement. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 92-39. Anchorage.

Merritt, M. F., A. E. Bingham, and N. Morton. 1990. Creel surveys conducted in interior Alaska during 1989. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 90-54. Anchorage.

Meyer, S. C. 1991. Biological characteristics of the sport harvest of marine groundfishes in southcentral Alaska, 1991. Alaska Department of Fish and Game, Sport Fish Division 92-41. Anchorage.

Meyer, S. C. 1996. Recreational halibut fishery statistics for southcentral Alaska (Area 3A), 1994: A report to the International Pacific Halibut Commission. Alaska Department of Fish and Game, Sport Fish Division Special Publication 96-1. Anchorage.

CHAPTER FOUR: Angler Success

Angler success is an expression of harvest rates measured in fish harvested by day of fishing. A day here is a 24-hr calendar day as specified in codified regulations. If from 0010 hrs to 2400 hrs an angler fishes only 15 min, fishes constantly for 24 hr, or fishes sporadically throughout the day, their fishing corresponds to a single ***angler-day*** of fishing effort. That angler-day may have produced no harvest or a harvest of one, two, three, or more fish. ***Angler success is then expressed as a set of proportions with each proportion being the fraction of angler-days that produced a harvest of one fish, a harvest of two fish, three fish, no fish, etc. such that as a set these proportions sum to one.***

Estimates of angler success can be used to evaluate the efficacy of current and proposed daily bag limits to restrict harvest. A bag limit of five fish per day is not restricting harvest if the average ***daily*** harvest is four or less for 99% of angler-days. Conversely, if four or fewer fish are harvested on 50% of angler-days and five fish on the remaining "days" of fishing effort, this daily bag limit is restricting harvest. If a daily bag limit of three fish controls harvest as in the upper panel of Figure 4.1, a reduction of 25% in harvest could be anticipated with a proposed limit of two fish a day. However, the same reduction in the daily bag limit would reduce harvest only 5% if the original bag limit had been ineffective (Figure 4.1). In all such analyses, anticipated reductions in harvest from lowering bag limits are conditioned on affected anglers behaving in a predictable manner.

Angler success is not so much estimated to gain knowledge of what ***has*** happened, but to anticipate what ***will*** happen. Unanticipated changes in fish abundance and angler behavior are two common factors that often shift the future into new, unperceived directions. Under these circumstances, variances for past estimates of angler success do not reflect the uncertainty in predicting the effects of new bag limits. Although not a guarantee of the future, accurate, precise estimates of past angler success do represent the necessary conditions for good prediction.

This chapter begins with the basic procedures behind estimating proportions that describe angler success and how these estimated proportions can be used to evaluate bag limits. Stratified "random" and self-weighted surveys to estimate angler success are described along with procedures to promote proportional sampling. Basic designs are compared through example, and post-stratification of statistics is discussed. How to avoid or detect opportunity and length-of-stay (LOS) bias are also described.

4.1 THE BASICS

Only information from those anglers finished fishing for the day should be used to estimate angler success. An interviewed angler is asked how many fish he/she harvested during the day's fishing and how many fishing trips were made during the day. If an angler had only one fishing trip during the day (exited the fishery once), their

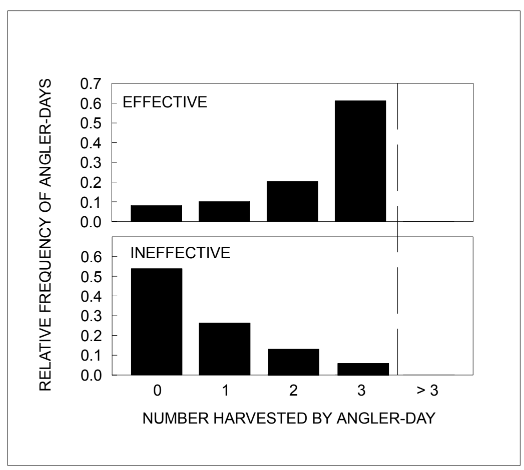


Figure 4.1 - Measure of angler success under an effective and under an ineffective daily bag limit of three fish.

angler made more than one trip (exited more than once during the day), the sum of harvests from all trips is the important statistic. When interviewed, anglers should be asked how many fish they had harvested that day and whether or not they have finished fishing for the day. For many fisheries in Alaska, illegal "party" fishing occurs and the enforceable bag limit is the product of the daily bag limit and the number of anglers in the party. In this case, the number fishing in the party must be asked during the interview, and harvest "divided" evenly among anglers in the analysis.

The fraction of angler-days resulting in a harvest of g fish can be estimated as a binomial proportion p_g by treating the sample of interviewed anglers as having been randomly drawn. From Cochran (1977:52), the estimated fraction $\hat{p}_g$ and an estimate of its variance are:

$$\hat{p}_g = \frac{\sum_{i=1}^{n}\sum_{j=1}^{d_i} m_{gij}}{\sum_{i=1}^{n}\sum_{j=1}^{d_i} m_{ij}} \tag{4.1a}$$

$$var(\hat{p}_g) = \frac{\hat{p}_g(1-\hat{p}_g)}{\left(\sum_{i=1}^{n}\sum_{j=1}^{d_i} m_{ij}\right)-1} \tag{4.1b}$$

where m_{ij} is the number of anglers interviewed at location i on day j after having completed fishing for the day, and m_{gij} is the number of m_{ij} who had a bag of g fish. When some angler-days produce a harvest greater than one fish (more than two fractions to estimate), an estimate of the resulting covariance between any two fractions would be:

$$cov(\hat{p}_g, \hat{p}_{g'}) = -\frac{\hat{p}_g\,\hat{p}_{g'}}{\left(\sum_{i=1}^{n}\sum_{j=1}^{d_i} m_{ij}\right)-1} \tag{4.1c}$$

where g and g' specify two different bags. Covariance is a consideration because all proportions will be involved in predicting the effects of new bag limits. Note that there are no corrections for sampling from finite populations (***fpc***s) in Equations 4.1. Usually the sample of angler-days is negligible in comparison to fishing effort, making the *fpc* an inconsequential reduction in estimated variance or covariance.

Predicting the relative reduction in harvest from lowering a bag limit involves averaging estimated fractions. The expected harvest per angler-day (the average harvest per angler-day) is $\Sigma g\hat{p}_g$ across all possible harvests (g = 0, 1, 2, 3, ...). If the p_g are estimated with data collected when the daily bag limit was B_o fish, the anticipated relative reduction in harvest with a bag limit of B fish would be:

$$\hat{\Delta}(B,\omega) = \frac{\sum_{g=B+1}^{B_o}(g-B\omega)\hat{p}_g}{\sum_{g=1}^{B_o} g\,\hat{p}_g} \tag{4.2}$$

where $\hat{\Delta}(B,\omega)$ is the anticipated relative reduction in harvest given a new bag limit B and an expected change ω in fishing effort by those anglers affected by the change in the limit. For instance, if affected anglers are expected to double the number of days they fish, $\omega = 2$; if affected anglers increase their days fishing by 10% or lower them by 20%, $\omega = 1.1$ or 0.8. The simplest way to estimate the potential effect of reducing the bag limit on number of days fished is to ask anglers their response to a reduction. Note that when $\hat{\Delta}(B,\omega)$ is negative, an ***increase*** in the harvest is anticipated.

The simplest way to estimate a variance for $\hat{\Delta}(B,\omega)$ is with simulation. Because some of the $\hat{p}_g$ are common to both the numerator and denominator in Equation 4.2, there is no closed-form method to estimate variance. An approximate estimate can be calculated with the delta method (from Seber 1982:7-8):

$$var(\hat{\Delta}(B,\omega)) \cong \sum_{g=1}^{B_O} var(\hat{p}_g)\left(\frac{\partial\Delta}{\partial\hat{p}_g}\right)^2 + 2\sum_{g=1}^{B_O}\sum_{g'>g}^{B_O} cov(\hat{p}_g,\hat{p}_{g'})\left(\frac{\partial\Delta}{\partial\hat{p}_g}\right)\left(\frac{\partial\Delta}{\partial\hat{p}_{g'}}\right)$$

However, this approximation is only marginally easier to compute than a simulation and under some circumstances would be a much poorer estimate of variance. A parametric bootstrap simulation can be based on harvest by angler-day being distributed according to a multinomial distribution with sample size and the $\{\hat{p}_g\}$ as parameters. A simulated sample of angler-days is drawn randomly with replacement from such a multinomial distribution, and the fractions of the simulated sample with simulated harvests of 0, 1, 2, 3, ... (=*g*) are calculated. These simulated fractions are plugged into Equation 4.2 to produce a simulated estimate of $\Delta(B,\omega)$. This whole process is repeated a large number of times, say *X* times, producing a set of simulated estimates for $\Delta(B,\omega)$. From Efron and Tibshirani (1993:45-49), an estimate of variance for $\hat{\Delta}(B,\omega)$ would be:

$$var(\hat{\Delta}(B,\omega)) \cong \frac{\sum_{b=1}^{X}[\hat{\Delta}_b(B,\omega) - \hat{\bar{\Delta}}(B,\omega)]^2}{X-1}$$

where $\hat{\Delta}_b(B,\omega)$ is a simulated estimate, $\hat{\bar{\Delta}}(B,\omega)$ is the average of simulated estimates. The difference between the original estimate $\hat{\Delta}(B,\omega)$ and the average of simulated estimates $\hat{\bar{\Delta}}(B,\omega)$ is a measure of the sampling-induced bias in the original (Efron and Tibshirani 1993:124-126). Appendix H contains a description of a QuickBasic™ program to perform a simulation for all prospective bag limits such that $1 \leq B < B_O$. A compiled version of this program can be obtained from the senior author upon request.

4.2 SELF-WEIGHTED SURVEYS

A self-weighted survey to produce unbiased estimates of angler success can be designed if:

- angler success has no spatial or temporal trend; or
- sampling is proportional to the number of angler-days of fishing effort.

Any sample should produce unbiased estimates if angler success does not change throughout or across the fishing grounds or across types of anglers. If there are trends in angler success, sampling must be regimented. If every ***k***th angler-day could be systematically sampled throughout a fishing season, resulting statistics would be unbiased regardless of trends. While this procedure theoretically will work every time, its application is impractical. A more realistic solution is to ***proportionally*** sample anglers by systematically selecting sampling periods in a pattern that is repeated throughout a fishing season and fishing grounds. During those days that sampling occurs, every exiting angler should be interviewed. If all fishing effort is sampled, sampling by definition is proportional at least across all days sampled (100%). There is considerable latitude in setting the sampling pattern so long as it continues across the fishing grounds and throughout the fishing season.

Depensatory sampling is a common problem with attaining a proportional sample. A technician has a maximum rate at which he/she can interview anglers. Once the rate of anglers leaving a fishery at a site exceeds this sampling rate, the technician(s) becomes swamped, and sampling becomes inversely related to the number of anglers exiting during that period. This phenomena is the same depensatory sampling noted in Section 3.2. If angler success is different during sampling periods with greater fishing effort, depensatory sampling will produce biased estimates of angler success for the season.

Deviations from planned proportional sampling are only as critical as are the intensity of trends in angler success in the fishery. When and where samples are drawn have no effect on estimates if there are no spatial, temporal, or demographic trends in angler success. Conversely, the more pronounced these trends, the more effect deviations from proportional sampling will have on bias in estimates. Long fisheries tend to have shallow trends in angler success and are therefore more forgiving of deviations from proportional sampling. Short, intense fisheries with steep trends have great potential for bias if sampling is not strictly proportional. The potential for bias from deviations in self-weighted sampling can be gauged for individual fisheries by calculating and plotting angler success for each sampling period over time or space (see Example 4.1). These plots are post facto information for last year's survey and a priori information for next year's. Since most temporal or spatial changes in angler success are gradual, statistics on angler success should follow near linear trends. Erratic statistics over space or time can result from interviewing too few anglers each sampling period. Sample sizes (number of angler days) must be sufficiently large to detect trends that have the potential to seriously bias statistics.

Example 4.1. The spear fishery for whitefish on the Chatanika River starts in early September and lasts to mid October. There are two major points of access to the fishery: Olnes Pond and the Elliot Highway Bridge. Fishing occurs at night at about 1800 hrs with almost all fishing completed by 2400 hrs. The angler-trip is essentially the angler-day; anglers who exited the fishery once during an evening hardly ever return that day. In 1989, the daily bag limit was 15 whitefish. More detail on this fishery is available in Examples 2.1, 2.3, 2.5-8, 2.10-11, 3.1, and 3.3.

While the creel survey on the Chatanika River in 1989 was not designed to be self-weighted, data from that survey can be pooled to demonstrate a self-weighted survey. Sampling was mildly depensatory in 1989 (Figure 4.2) with correlations between numbers of estimated angler-days and fraction of those angler-days with interviews equal to - 0.50 for sampling near Olnes Pond and -0.61 near the Elliot Highway Bridge. Estimated success rates increased with the season (Figure 4.3). Even though these circumstances indicate estimates of angler success based on pooled data will be biased, we will concentrate on pooled data for our demonstration of self-weighted sampling. Creels of 788 completed-trip anglers were inspected that year near Olnes Pond and near the Elliot Highway Bridge (Table 4.1); anglers fishing along the Steese Highway were excluded from this analysis. Note that 26 (3.3%) of anglers interviewed had exceeded the bag limit and that

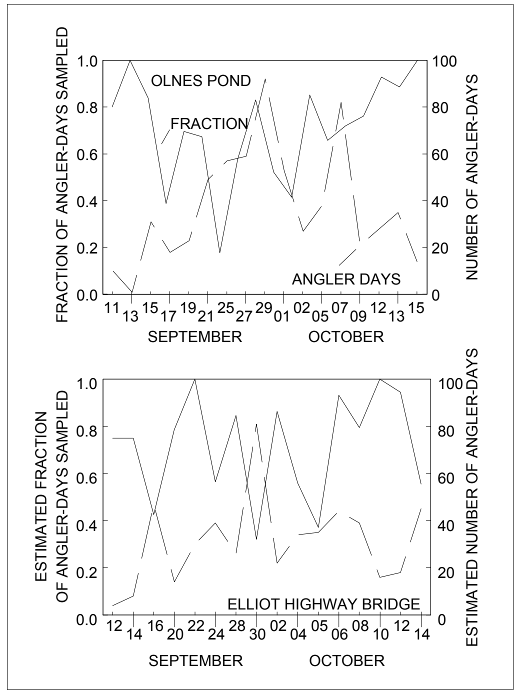


Figure 4.2 - Number of angler-days of fishing effort (dashed line) and the fraction of angler-days sampled (solid line) by day in 1989 at the Chatanika River. Because of its small size, the stratum associated with the Steese Highway is not included.

Table 4.1 - Number and fraction of exiting anglers interviewed near Olnes Pond and near the Elliot Highway Bridge with g whitefish in their bag in 1989.

g	$\Sigma\Sigma m_{ghi}$	$\hat{p}_g$	g	$\Sigma\Sigma m_{ghi}$	$\hat{p}_g$
0	139	0.176	9	25	0.032
1	54	0.069	10	50	0.063
2	46	0.058	11	21	0.027
3	41	0.052	12	38	0.048
4	44	0.056	13	23	0.029
5	59	0.075	14	29	0.037
6	39	0.049	15	85	0.108
7	35	0.044	>15	26	0.033
8	34	0.043			

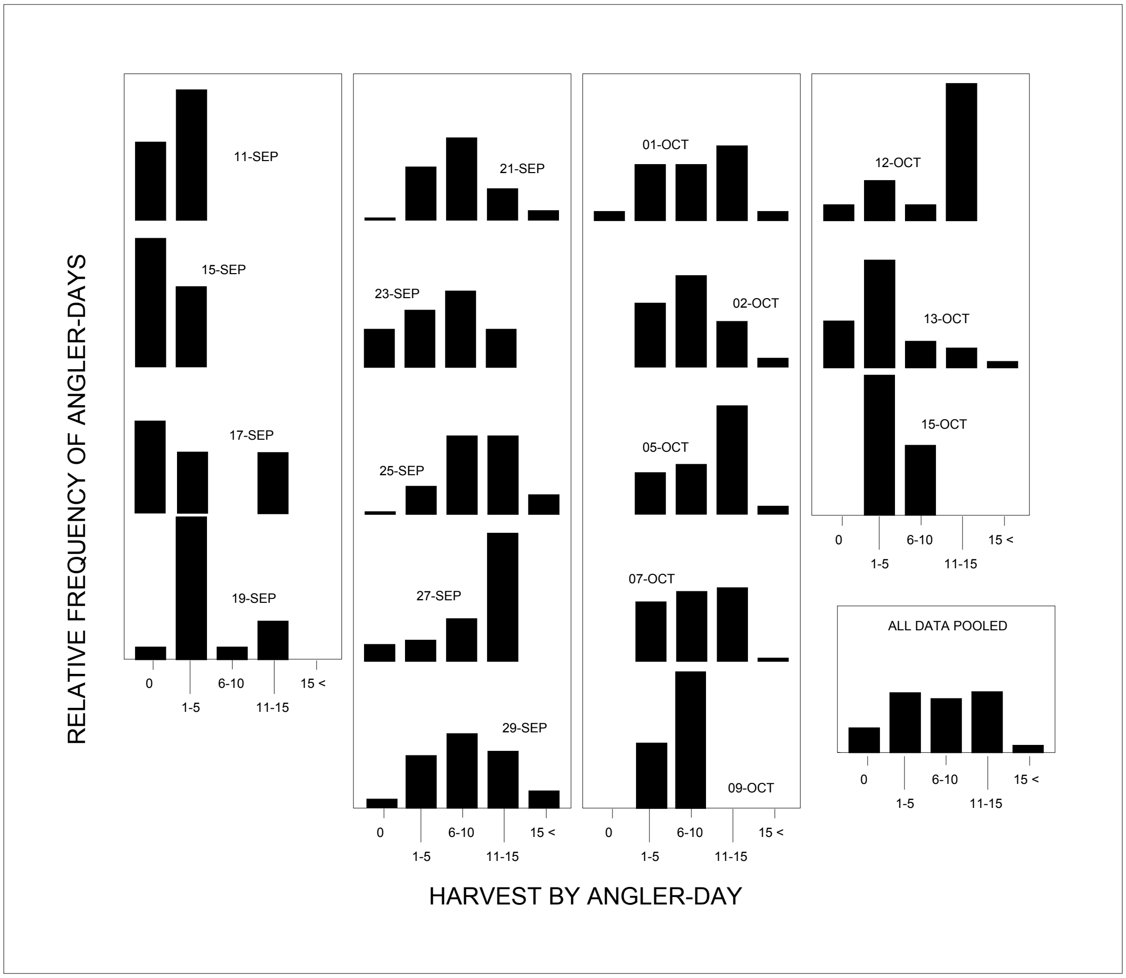


Figure 4.3 - Angler success as the relative frequencies of angler-days by their harvest of whitefish for all days sampled in 1989 at the Chatanika River near Olnes Pond. Scale on all histograms is 0.0 to 0.7.

14.1% of all anglers had harvested 15 or more whitefish in a night.

Anticipated reductions in harvest from Equation 4.2 with all possible bag limits below 15 whitefish along with simulated statistics were obtained with the compiled QUICKBASIC™ program DELTA.EXE (Table 4.2). Simulated statistics were based on 100 simulated samples each of 788 interviews randomly drawn from the original data with replacement. The weighting factor ω was set to one (no anticipated change in fishing effort by affected anglers). A 16% (0.159x100%) reduction in harvest is anticipated by lowering the daily bag limit from 15 to 10 whitefish. A lowering of the limit to 5 whitefish would reduce the harvest an anticipated 48%. Note that little bias in anticipated reductions is indicated ($\hat{\Delta}(B,\omega)$ - $\hat{\hat{\Delta}}(B,\omega)$ ≅ 0) and that estimated precision is quite good. ***Any bias associated with depensatory sampling is not included in either of these estimates of bias or precision***. Lack of bias and good precision apparent in Table 4.2 indicating that ***bias and imprecision from small sample sizes have been avoided in this instance***.

Table 4.2 - Anticipated and simulated reductions and in harvest of whitefish along with simulated SEs for the spear fishery on the Chatanika River based on data collected in 1989 ($\omega = 1$).

B	$\hat{\Delta}(B,\omega)$	$\hat{\bar{\Delta}}(B,\omega)$	$SE(\hat{\Delta}(B,\omega))$
1	0.877	0.877	0.00280
2	0.764	0.764	0.00513
3	0.660	0.660	0.00671
4	0.564	0.563	0.00780
5	0.476	0.475	0.00842
6	0.400	0.399	0.00846
7	0.330	0.329	0.00822
8	0.267	0.267	0.00770
9	0.211	0.210	0.00691
10	0.159	0.159	0.00602
11	0.117	0.117	0.00508
12	0.079	0.079	0.00406
13	0.048	0.048	0.00284
14	0.021	0.021	0.00148

* * * *

A special comment is in order about angler success by different types of anglers. If all anglers have an equal chance of being interviewed, different harvest rates by different types of anglers will not bias estimates of angler success for the season. However, ***if different types of anglers with different rates of success are interviewed at different rates, estimates of angler success from pooling data will be biased***. Guided anglers often will have different success rates than will unguided anglers; local anglers will be more successful than non-local anglers, etc. If guided anglers access the fishery at different locations than unguided anglers (as in Example 3.7), or if local and non-local anglers tend to fish on different days of the week, pooling data will bias estimates.

4.3 STRATIFIED "RANDOM" SURVEYS

When sampling can not be proportional and angler success changes across the fishing season, across the fishing grounds, or across different types of anglers, unbiased estimates of angler success can still be obtained through stratification. Angler success and anticipated reduction in harvest are estimated just as before with Equations 4.1 and 4.2, but now they are estimated for each stratum. Sampling within each stratum is considered random and angler success invariant because strata are drawn to be so. The more strata involved in the design, the less variable angler success will be within each stratum, although the marginal reduction in bias lessens as the number of strata increases.

When estimates of angler success have been stratified, predicting relative change in harvest from lowering bag limits ($\hat{\Delta}$) requires estimates of harvest. As per methods in Cochran (1977:91):

$$\hat{\Delta} = \frac{\sum_{h=1}^{L} \hat{\Delta}_h \hat{H}_h}{\sum_{h=1}^{L} \hat{H}_h} \tag{4.3a}$$

where $\hat{\Delta}_h$ is the predicted, relative change in harvest for a stratum, L is the number of strata in the survey and $\hat{H}_h$ are stratified estimates of harvest. An offsite survey, such as the Statewide Harvest Survey, can provide estimates of harvest when onsite sampling to estimate angler success has been post-stratified by type of angler, such as guided vs. unguided or anglers fishing from the shore vs. those fishing from boats. Estimates of harvest from onsite surveys can also be used when anglers have been post-stratified, but unlike estimates from the statewide harvest survey, harvest estimates from onsite surveys can also be temporally or spatially stratified. ***This opportunity to stratify by time and area is a prerequisite to accurately estimating reductions in harvests from lowering bag limits on migratory species,*** such as salmon and whitefish. When an onsite survey is used to estimate harvest, a stratified "random" sampling design to estimate angler success can be superimposed over the original sampling design, and both harvests and angler success estimated simultaneously. Regardless of how estimates of harvest are obtained, through the Statewide Harvest Survey or with an onsite creel survey, an approximate variance for the anticipated reduction from lowered bag limits can be found with the delta method (Seber 1982:7-8):

$$var(\hat{\Delta}) \cong \sum_{h=1}^{L} var(\hat{\Delta}_h)\left(\frac{\hat{H}_h}{\hat{H}}\right)^2 + \sum_{h=1}^{L} var(\hat{H}_h)\left(\frac{\hat{\Delta}_h - \hat{\Delta}}{\hat{H}}\right)^2 \tag{4.3b}$$

where $\hat{H} = \Sigma\hat{H}_h$.

Example 4.2. Our demonstration of stratified "random" sampling to estimate angler success uses the stratified estimates of whitefish harvest on the Chatanika River in 1989 (Example 3.1, Table 3.1). The access-point survey near Olnes Pond and the roving-access survey near the Elliot Highway Bridge were divided into eight strata, and harvest was estimated for each stratum (Table 4.3). Information from interviewed anglers was segregated according to the same strata, and anticipated reductions from lowering the bag limit from 15 whitefish per day were estimated with the compiled QUICKBASIC™. Program DELTA.EXE with no anticipated change in fishing effort by affected anglers (ω = 1). Table 4.3 contains statistics for a daily bag limit of 10 fish (= *B*) from eight runs of the program, one for each stratum in the survey. From Equation 4.3a, the predicted reduction by lowering the limit to 10 fish is:

$$\hat{\Delta}(10,1) = \frac{2{,}642.8}{16{,}667} = 0.159$$

$$var(\hat{\Delta}(10,1)) \cong 0.00005237147 + 0.000006787645 = 0.00005915912$$

Similarity of predicted reduction in harvests from this stratified "random" sampling design (16%) and from pooling the data in Example 4.1 (again 16%) indicates that deviations from proportional sampling and trends in angler success were not

Table 4.3 - Statistics used to anticipate a lowering of the daily bag limit of 15 to 10 whitefish from the Chatanika River near Olnes Pond and the Elliot Highway Bridge in 1989 (B = 10, ω = 1).

	$\hat{H}_h$	$var(\hat{H}_h)$	$\hat{\Delta}_h$	$var(\hat{\Delta}_h)$	$\hat{\Delta}_h\hat{H}_h$	$var(\hat{\Delta}_h)\left(\frac{\hat{H}_h}{\hat{H}}\right)^2$ x 10^6	$var(\hat{H}_h)\left(\frac{\hat{\Delta}_h - \hat{\Delta}}{\hat{H}}\right)^2$ x 10^6
Olnes Pond: (data on 2 October was substituted for missing data on 3 October)							
11-20 Sep	512	9,162	0.159	0.001480	81.4	1.39665	0.000001
21-30 Sep	5,796	156,792	0.176	0.000189	1,020.1	22.85618	0.171575
1-6 Oct	2,180	146,115	0.175	0.000322	381.5	5.50876	0.142076
7-12 Oct	2,258	503,535	0.175	0.000272	395.2	4.99231	0.489614
13-15 Oct	309	29,829	0.068	0.001022	21.0	0.35128	0.880734
Elliot Highway: (missing data on 18 and 26 September handled as per methods in Section 2.1.3)							
11 Sep - 7 Oct	4,627	1,267,773	0.127	0.000208	587.6	16.03051	4.547139
8-12 Oct	862	14,362	0.181	0.000462	156.0	1.235781	0.026023
13-15 Oct	123	586	0.000	0.000000	0.0	0.0	0.530483
Sums	16,667				2,642.8	52.37147	6.787645

severe enough to bias estimates from pooled data. This result is somewhat surprising for a fishery on migratory species, such as whitefish, and indicates that relatively unbiased estimates of angler success would be possible from self-weighted surveys of many fisheries.

* * * *

When regulations are to be tailored to restrict one group of anglers more than another, or when on group is interviewed at a different rate, some post-stratification of angler success is required. For instance, bait may be banned to reduce angler success as has been done in the fishery for chinook salmon on the Kenai River. Fishing time may be restricted for guided anglers to limit their success as has been done in this same fishery. In either case, fishing gear (bait vs. artificial lures) and expertise of anglers (guided vs. unguided) would be dimensions of post-stratification. Calculations would follow the same equations as before (Equations 4.1-3), only the raw data would be post-stratified, and calculations would be repeated for each element in the post-stratification. Post-stratification on expertise of anglers would produce two sets of proportions, the first describing the angler success of guided anglers and the second the angler success of their unguided contemporaries.

4.4 AVOIDING BIAS

4.4.1 LOS Bias. Estimates of angler success can be subject to length-of-stay (LOS) bias. Successful anglers in a fishery regulated with a daily bag limit will tend leave the fishery early in the fishing day while less successful anglers will tend to leave late in the day. Under these circumstances, anglers interviewed upon exiting a fishery in the morning will tend to have an atypically high rate of angler success while those exiting in the evening will have an atypically low rate of success. In either instance, samples of angler-days will produce biased estimates of angler-success when the fishing day is subsampled.

The only way to avoid LOS bias is to interview anglers throughout the fishing day (Bernard et al. 1998)**.** So long as sampling is not depensatory within a fishing day, sampling all day long will provide a proportional sample of interviews. An angler's chance of being interviewed that day will be the same regardless of when they stopped fishing. In the three common circumstances involved in estimating angler success, all or some sampling periods must be the same length as the fishing day:

- In ***self-weighted surveys*** designed to estimate angler success through proportional sampling, all sampling periods should be the same length as a fishing day.

- If angler success is to be estimated simultaneously with harvest through an onsite ***roving-access creel surveys***, a sampling period should cover a fishing day to avoid LOS bias in estimates of harvest (Bernard et al. 1998) and in estimates of angler success. Accurate estimates of harvest and angler success can both be based on data from the same sampled periods.

- If angler success is to be estimated simultaneously with harvest through an onsite ***access-point survey***, special sampling should be scheduled so estimates of angler success are based on extended sampling periods that are each equal to a fishing day. Estimates of harvest from access-point surveys can still be based on the shorter sampling periods, however, estimates of angler success must be calculated only with data from the longer periods.

Note that sampling periods in access-point and roving-access surveys in Examples 4.1-2 were the same length as the fishing day, thereby avoiding LOS bias in estimates of angler success.

4.4.2 Opportunity Bias. As with many other types of sample surveys, estimates of angler success are subject to ***opportunity bias***. Success of anglers exiting a fishery at times and locations not manned by technicians have no chance of being sampled. If unsampled anglers have consistently better or worse success than sampled anglers, estimates will be biased. Potential for opportunity bias occurs when different types of anglers enjoy different rates of success and fish or exit the fishery in different locations and at different times. Resident and visiting anglers or guided and unguided anglers often have different levels of expertise, fish at different times, and exit fisheries at different locations. The remedy is to stratify sampling.

Opportunity bias can be detected through simultaneous sampling provided sample sizes are sufficient to expose differences that correspond to meaningful bias. Occasionally a second or third technician should be employed to intercept anglers at locations and times not usually sampled. Opportunity bias is indicated if statistics ***consistently*** differ between the standard and test samples. If lucky, no opportunity bias will be indicated, or data can be post-stratified by type of angler to avoid bias in the current survey. If unlucky, the sampling design can be altered for the next survey to avoid opportunity bias.

4.5 REFERENCE FOR EXAMPLES

Merritt, M. F., A. E. Bingham, and N. Morton. 1990. Creel surveys conducted in interior Alaska during 1989. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 90-54. Anchorage.

CHAPTER FIVE: CPUE as an Index of Abundance

Mean CPUE ($\overline{CPUE}$) from a fishery may be considered an index of abundance under the traditional linear model:

$$[c/e]_i = qA + \varepsilon_i \tag{5.1}$$

where *c/e* is the catch per unit of effort during the ith angler-day (= *cpue*), *A* is abundance, *q* is the catchability coefficient, and ε is random error with mean 0 and variance σ^2. In this case, each angler-day is considered a separate, replicated sample in a "test fishery". Because harvest is influenced by anglers' preferences as well by abundance, catch rate is used in calculating an index of abundance, not harvest rate. The main advantage of using anglers as a test fishery in this way is that it's cheap.

Estimates of $\overline{CPUE}$ are generally used to decide when or whether a fishery should be opened or closed during its normal progress. For fisheries on migrating populations, abundance on the fishing grounds will change as migration waxes, then wanes, and periodic estimates $\overline{cpue}$ can provide periodic indices of abundance as the migration progresses. In the case of fisheries on resident populations, abundance will change as functions of recruitment, mortality rates, and emigration.

Inherent in using $\overline{cpue}$ as an index of abundance is that the catchability coefficient *q* is constant. This can be a tenuous assumption at best. Ability obviously varies from angler to angler and with it *q*. Competition among anglers, if severe enough, will also effect catchability. Although competition and anglers' expertise can be factored into calculating indices, changes in environment and fish behavior usually can not, often compromising indices. For this reason, using $\overline{cpue}$ from recreational fisheries as an index of abundance carries a higher risk of management error (making the wrong decision) than using many other statistics, even if $\overline{cpue}$ is an accurate estimate of $\overline{CPUE}$.

Although accurate and precise estimates of $\overline{CPUE}$ in a recreational fishery may be a misleading index of fish abundance, inaccurate and imprecise estimates certainly will be. In this chapter we explore ways to produce accurate and precise estimates of mean catch per unit of effort from an onsite creel survey.

5.1 THE CPUE SURVEY

5.1.1 Planning and Calculation. Sampling in a creel survey to estimate $\overline{CPUE}$ as an index of abundance will generally be stratified by week, by fortnight, by month, by area, and/or by type of angler. Samples from each stratum represent a "snapshot" of the abundance at that place and time. Because abundance changes over time and place, the different "snapshots" are needed to expose trends in abundance. Anglers leaving Douglas (4 km south of Juneau) head south to fish for salmon while those leaving Auke Bay (26 km north of Juneau) head north. These two groups of anglers undoubtedly fish two separate groups of salmon separated by considerable distance and time. The presence of Douglas Island between these two harbors dictates both patterns of migration and fishing. Here, location is an obvious dimension of stratification for a CPUE survey.

Size, duration, and type of strata are determined

- steeper temporal and/or spatial trends in fish abundance require more strata;

Size, duration, and type of strata are determined according to the following rules:

- steeper temporal and/or spatial trends in fish abundance require more strata;
- strata should have the same or finer resolution as information needed to manage a sport fishery; and
- different abilities to capture fish (different q's) among groups of anglers and different probabilities of interviewing members of each group.

When a variable, such as fish abundance, has a trend related to an auxiliary variable, such as time, sampling should be organized across that trend to detect it. The steeper the trend, the more strata are needed. Migratory species will have steeper trends than will resident species, so there should be more and shorter strata in a CPUE survey for migratory populations. The same rule applies to spatial trends.

Often there are several fishing publics, each with different abilities in catching fish. So long as all groups can be sampled in ***proportion*** to their participation in the fishery, $\overline{cpue}$ will remain an unbiased estimate of $\overline{CPUE}$. When sampling is not self-weighted to size of the groups, or if there is a need to segregate statistics for these groups to manage the fishery, the groups should be post-stratified by keeping separate sets of statistics for each group. For example, anglers with guides often have different catch rates than anglers without, which translates into different catchability coefficients. Another common situation arises when anglers fishing from shore are less successful than those fishing from boats. In both cases, $\overline{CPUE}$ ***for each group would be a consistent index of abundance within that group across time and area, but would differ between groups.***

Sampling should be conducted as succinctly as possible within the "temporal middle" of each stratum. In a stratum several days long, anglers should be interviewed during one or more consecutive fishing days. These fishing days are in the middle of a stratum to maximize the information across any trend in abundance. For reasons given in Section 5.2:

- sampling periods should be the same length or whole multiples of a fishing day; and
- only anglers exiting a fishery should be interviewed.

Mean *cpue* from interviews will be an unbiased estimate of catch per unit of effort for ***all*** anglers exiting the fishery at a location if:

- All anglers exiting at that location are interviewed; or
- Time taken to interview an exiting angler is not related to his or her success.

So long as interviewed anglers are asked only how many fish they caught, how long they fished, and who they are (for post-stratification), interviews will take the same time regardless of how successful the angler.

Occasionally fishing is so sporadic that no angler-trips are sampled during a stratum, and no information is available on abundance for that stratum. This stratum should be ignored as if it was never chosen to be sampled. Strata with no angler-trips should be an infrequent occurrence in any survey. If they are frequent, the survey is in need of redesign, or the fishery is too small to be a good indicator of fish abundance.

Calculation of $\overline{cpue}$ for each stratum begins with dividing the catch made during an individual angler-trip by the hours needed for the angler to attain that catch:

$$cpue_{hi} = \frac{c_{hi}}{e_{hi}} \tag{5.2}$$

The next step is to average *cpue* within each stratum:

$$\overline{cpue}_h = \frac{\sum_{i=1}^{m_h} cpue_{hi}}{m_h} \tag{5.3a}$$

$$var(\overline{cpue}_h) = \frac{\sum_{i=1}^{m_h}(cpue_{hi} - \overline{cpue}_h)^2}{m_h(m_h - 1)} \tag{5.3b}$$

Example 5.1 Although the access-point creel survey at Olnes Pond on the Chatanika River in 1989 was designed to estimate harvest of whitefish (see Example 2.1), its design can also be viewed as one to estimate $\overline{CPUE}$ as an index of abundance. Since two species of whitefish migrate through the fishing grounds, periodic estimates $\overline{cpue}$ should follow changes in abundance as the fishery progresses. A sampling design using anglers as units of sampling effort would consist of sampling anglers every few days (evenings). Unlike under the original access-point survey, each night of sampling would represent a single, two-day stratum in a CPUE survey. In this instance, each sampled period would represent a datum in a trend. Sampling did occur every few nights in 1989 with sampling spaced relatively uniformly throughout the fishing season. More information about the whitefish spear fishery on the Chatanika River is given in the preamble to Example 2.1.

Mean *cpue* was calculated for humpback whitefish in this example. As before, the example begins with calculations for 17 September at Olnes Pond. Seven anglers were interviewed (seven angler-trips or $m_h = 7$) while they exited the fishery (Table 5.1). From Equations 5.3, $\overline{cpue}_h$ and its estimated variance for 17 September are:

$$\overline{cpue}_h = \frac{0+0+1.27+0.63+0+1.25+1.00}{7} = 0.59$$

$$var(\overline{cpue}_h) = 0.0484 = \frac{\left[\begin{array}{c}(0-0.59)^2 + (0-0.59)^2 + (1.27-0.59)^2 + (0.63-0.59)^2 \\ +(0-0.59)^2 + (1.25-0.59)^2 + (1.00-0.59)^2\end{array}\right]}{7(7-1)}$$

Table 5.1 - Number of humpback whitefish speared (c_{hi}), hours taken to spear them (e_{hi}), and $cpue_{hi}$ for anglers interviewed on 17 September, 1989 as they exited the fishery for whitefish at Olnes Pond near the Chatanika River.

c_{hi}	e_{hi}	$cpue_{hi}$	c_{hi}	e_{hi}	$cpue_{hi}$
0	1.25	0.00	0	1.58	0.00
0	1.25	0.00	5	4.00	1.25
2	1.58	1.27	4	4.00	1.00
1	1.58	0.63			

Statistics for other "strata" are listed in Table 5.2. Some of the data used to estimate harvest in Example 2.1were omitted in this example because species of whitefish were not distinguished in some interviews.

Abundance of humpback whitefish increased in 1989 at Olnes Pond from late September through October (Figure 5.1). A cohort of fish moved through the fishery at the end of the third week in September temporarily increasing abundance in this part of the river. From late September through mid October, abundance of humpback whitefish steadily increased with the downstream migration of spent fish leaving their spawning grounds upstream.

5.1.2 Graphics and Confidence Intervals. As is obvious from Figure 5.1, graphical displays and confidence intervals are an ideal means of comparing statistics in search of a trend in abundance. Confidence intervals for the $\overline{cpue}_h$ can be expressed as:

- Ranges
- Standard Approximations
- Bootstrap Approximations
- Profile Likelihoods
- Credibility Intervals

The simplest interval to calculate is the range as defined by the minimum and maximum values of $cpue_{hi}$ in stratum h. A slightly more difficult to calculate standard approximation to confidence intervals is:

$$\overline{cpue}_h \pm t_{\alpha/2,df} SE(\overline{cpue}_h)$$

These intervals are predicated on the central limit theorem and large sample sizes. When sample sizes are small and $\overline{cpue}$ is low, standard intervals may unrealistically encompass zero (see standard interval for 11 September in Figure 5.1). More realistic intervals can be obtained through bootstrapping the $cpue_{hi}$ for each stratum [see Efron and Tibshirani (1993:45-9) for a description of this resampling technique]. A large number of bootstrap samples (usually 1,000) are drawn from the original data in each stratum. The bootstrap process to approximate confidence intervals is to:

1) Create the first bootstrap sample by drawing m_h interviews with replacement from original sample.

2) Calculate a bootstrap estimate $\overline{cpue}^*$ using Equation 5.3a with data from the bootstrap sample.

3) Repeat the process to produce 999 other bootstrap estimates.

4) Sort all 1,000 bootstrap statistics (the $\overline{cpue}^*$) in ascending order.

5) The lower bound for a 95% confidence interval will be the 26th bootstrap estimate $\overline{cpue}^*$ in the order; the upper bound will be the 975th estimate.

6) Repeat the process for each stratum.

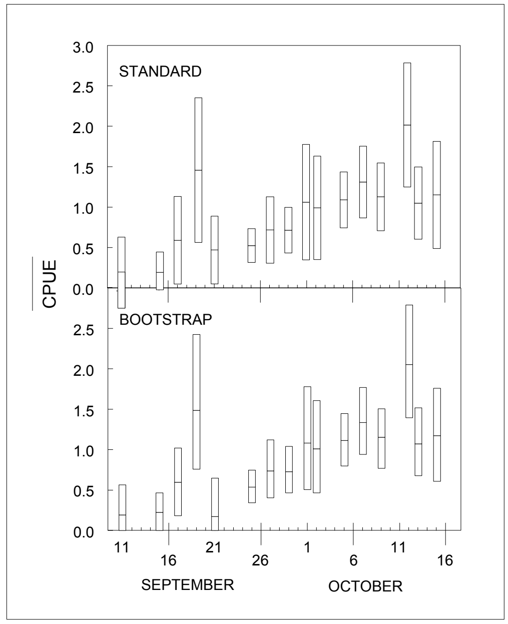


Figure 5.1 - Mean CPUE and 95% confidence intervals from standard and bootstrap methods for humpback whitefish caught near Olnes Pond on the Chatanika River in 1989.

Table 5.2 - Estimated catch per unit of effort of humpback whitefish, estimates of its standard error, and numbers of anglers interviewed in the fishery on the Chatanika River near Olnes Pond in 1989.

	m_h	$\overline{cpue}_h$	$SE(\overline{cpue}_h)$
September			
11	8	0.19	0.19
15	25	0.22	0.12
17	7	0.59	0.22
19	16	1.47	0.43
21	12	0.17	0.13
25	33	0.53	0.10
29	44	0.72	0.14
October			
1	22	1.07	0.35
2	21	1.00	0.31
5	25	1.10	0.17
7	57	1.32	0.22
9	16	1.14	0.20
12	26	2.03	0.38
13	30	1.06	0.22
15	14	1.16	0.31

This is the percentile method of approximating confidence intervals with bootstrapping (Efron and Tibshirani 1993:170-176). A bootstrap interval for 11 September in Figure 5.1 is 0 to 0.56 fish per hour. In contrast, the standard interval is -0.26 to 0.63 fish per hour. Generally, highly accurate confidence intervals can not be obtained through bootstrapping when sample sizes are small (DiCiccio and Romano 1988), however, confidence intervals from bootstrapping would still be a significant improvement over standard intervals that encompass zero. Confidence intervals based on profile likelihoods and Bayesian credibility intervals require considerably more computation than do standard or bootstrap intervals and require large sample sizes to be reasonably accurate. Hahn and Meeker (1991:434-444) gives an overview of these and other types of confidence intervals.

5.2 AVOIDING BIAS

5.2.1 LOS Bias. Estimates of $\overline{CPUE}$, like estimates of angler success and harvest from roving surveys, are subject to length-of-stay (LOS) bias. When anglers are more interested in bagging harvest than the "thrill" of the catch, harvest and a daily bag limit will influence when they quit fishing for the day. If so, anglers will quit fishing upon filling their daily limit. Under these circumstances, $\overline{cpue}$ ***will be <u>biased low</u> if anglers are interviewed while still fishing*** (the "incompleted-trip interview") (Pollock et al. 1994:251). The alternative is to interview anglers as they exit the fishery (the "completed-trip interview"), however, successful anglers will tend to leave the fishery early in the fishing day while less successful anglers will tend to leave late in the day. ***Unless exiting anglers are interviewed throughout the fishing day,*** $\overline{cpue}$ ***can be biased to some unknown degree*** (Bernard et al. 1998).

The only way to avoid LOS bias in estimates of $\overline{CPUE}$ is to ***interview all exiting anglers throughout the fishing day at whatever site chosen for sampling***. So long as sampling is not depensatory within a fishing day, sampling all day long will provide a proportional, representative sample of interviews. Every angler that exits at that location will have an equal chance of being interviewed regardless of when they stopped fishing. ***Depensatory sampling must be avoided.***

Depensatory sampling within the fishing day occurs when all exiting anglers are interviewed when few are exiting and not all are interviewed when many exit. Often a stampede of exiting anglers occurs late in a fishing day. If sampling is depensatory, these usually less successful anglers will be under represented in the sample, and that day's estimate of mean CPUE will be biased high. If depensatory sampling is expected during a part of a fishing day at a sampled site:

- add extra technicians at sampling locations to ensure that all exiting anglers are interviewed; or

- reschedule sampling to those fishing days or sites where fewer anglers exit the fishery.

When $\overline{CPUE}$ is to be estimated as a secondary consideration to harvest, catch, fishing effort, harvest composition, or angler success, some extra sampling may be required to avoid LOS bias in $\overline{cpue}$:

- ***Roving-access survey to estimate harvest, catch, and fishing effort***. LOS bias in $\overline{cpue}$ is avoided as long as sampling periods are the same length as a fishing day in these roving surveys and all exiting anglers are interviewed.

- ***Access-point survey to estimate harvest, catch, and fishing effort***. Extra sampling should be scheduled on some days to completely cover the fishing day to avoid LOS bias in $\overline{cpue}$. Also, <u>all exiting anglers should be interviewed</u> to avoid depensatory sampling.

- ***Survey to estimate harvest composition***. Extra sampling should be scheduled on some days to completely cover the fishing day to avoid LOS bias in $\overline{cpue}$. Also, <u>all exiting anglers should be interviewed</u> to avoid depensatory sampling.

- ***Survey to estimate angler success***. LOS bias in $\overline{cpue}$ is avoided as long as sampling periods are the same length as a fishing day in the roving survey and all exiting anglers are interviewed.

Obviously, ***calculating $\overline{cpue}$ as an afterthought in access-point surveys and surveys to estimate harvest composition is a bad idea.*** Some forethought is needed to schedule the extra sampling needed to avoid LOS bias in these surveys.

5.2.2 Opportunity Bias. Surveys to estimate $\overline{CPUE}$ as an index of abundance can be robust to opportunity bias. Sampling is opportunistic when technicians can not simultaneously work all access points to a fishery during a sampled period. Success of anglers exiting a fishery at times and locations not manned by technicians have no chance of being sampled. If unsampled anglers have consistently better or worse success than sampled anglers, estimates of $\overline{CPUE}$ for the for the entire fishery will be biased. However, $\overline{cpue}$ as an index of abundance is destined to be compared. So long as catchability of fish remains the same across time, temporal trends in $\overline{cpue}$ will be representative of temporal trends in abundance. The same would be true for spatial trends in abundance.

There are some simple rules that if followed, would make $\overline{cpue}$ almost immune to opportunity bias. They are:

- For temporal comparisons, always sample at the same locations.

- For spatial comparisons, always sample at the same times.

One final rule. If groups of anglers have different rates of success and probability of interviewing members of each group ***changes*** across time or area, $\overline{CPUE}$ ***should be estimated separately for both groups***. Each set of statistics (one for each group) can be used as an index of abundance.

5.3 REFERENCE FOR EXAMPLES

Merritt, M. F., A. E. Bingham, and N. Morton. 1990. Creel surveys conducted in interior Alaska during 1989. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 90-54. Anchorage.

CHAPTER SIX: One-Technician Surveys

"One-technician" surveys have all the attributes of any other survey, except a single technician is used to count and interview anglers and/or inspect harvest. Many creel surveys may have only one man-month of sampling during each calendar month, however, a "one-technician" survey literally involves only one technician. In communities with regional or large area offices with many research projects conducted annually, technicians can be shared among projects to promote flexibility of scheduling and efficiency of labor. However, in communities with small area offices, often only one person is available to collect data for a survey.

"One-technician" surveys can be planned just like other surveys, however, scheduling of sampling in "one-technician" surveys is more difficult and is sometimes impossible. Union rules in Alaska that complicate scheduling are:

- each technician must have two consecutive calendar days off each week;
- when scheduled work during a calendar day, a technician must be worked at least 4 consecutive hr;
- overtime pay must be given once a technician works more than 37.5 hr in a week.

Another scheduling problem occurs when sampling is required in two places at the same time. Any proposed survey with such conflicts can not be conducted successfully with only one technician.

Fatal problems with scheduling are more likely when harvest, catch, or fishing effort are to be estimated. Weekly and daily cycles in fishing effort produce weekly and daily cycles in harvest. In some fisheries, anglers prefer fishing in the afternoons or on weekends. When these cycles coexist alongside a heterogeneous distribution of fishing effort through different points of access to stream, lake, or sea, more than one technician is needed to sample at enough locations often enough to produce unbiased estimates of harvest and fishing effort. However, angler success, $\overline{cpue}$ as an index of abundance, or relative harvest composition are generally insensitive to weekly or daily variation in fishing effort and subsequent harvest. Shifts in relative harvest composition or changes in fish abundance, for instance, generally change slowly throughout a fishing season. Fewer periods need be sampled to achieve precise estimates of these statistics as opposed to estimates of harvest, catch or fishing effort.

The remainder of this chapter is a description of two types of schedules built around a single technician. The first is based on systematic sampling that can be used in many situations to produce any of the statistics commonly estimated in creel surveys. The second is a schedule based on weekly stratification and randomly selected sampling periods and is best used to estimate fishing effort and harvest in a fishery with weekly cycles in these parameters.

6.1 SYSTEMATIC SCHEDULING

6.1.1 Access-point Surveys. A systematic sampling schedule can be developed for access-point surveys with three or fewer strata based on location and/or time of day (TOD stratification), but not on days of the week (DOW stratification) (see Section 1.2). Scheduling follows a repeating four-day pattern rotating through strata followed by two consecutive days off (Table 6.1). Every third day is sampled in the first stratum while in the second and third every sixth day. The schedule "repeats" completely through the days of the week once every six weeks for the second and third strata and once every three weeks for the first. Every six weeks (42 calendar days), 14 days

Table 6.1 - Systematic schedule that will enable a single technician to completely cover a sampling design with three strata based on location or sampling period (TOD stratification).

		Strata			
Week	Day	1	2	3	
First	M	x			
	T		x		
	W			x	
	T	x			
	F				off
	S				off
	S	x			
Second	M		x		
	T			x	
	W	x			
	T				off
	F				off
	S	x			
	S		x		
Third	M			x	
	T	x			
	W				off
	T				off
	F	x			
	S		x		
	S			x	

Week	Day	1	2	3	
First	M	x			
	T				off
	W				off
	T	x			
	F		x		
	S			x	
	S	x			
Second	M				off
	T				off
	W	x			
	T		x		
	F			x	
	S	x			
	S				off
Third	M				off
	T	x			
	W		x		
	T			x	
	F	x			
	S				off
	S				off

(33% of the first-stage sampling units) are sampled in the first stratum and seven days (16% of the first-stage units) each in the second and third strata. Since the first stratum is scheduled to have the most sampling effort, location or TOD representing the greatest contribution to overall variance of estimates should be defined as that first stratum.

This systematic schedule is only partially systematic. The first day of sampling in the first stratum is randomly selected from the first three days of the fishery as per Cochran (1977:205). However, sampling in the other two strata does not begin with randomly selecting their first day of sampling, but follows sampling in the first stratum rigidly. This procedure breaks the independence of statistics from the three strata, so summing estimates across strata should include some negligible covariance within the variance of the sum. There is also potential for bias, particularly if the fishery was short (a few days long). However, these problems with covariance and bias will be greatest for the second and third strata. If the first stratum represents the largest contribution to variance, problems with covariance and bias in the overall estimates should be negligible and can be ignored.

Strata in this schedule represent access points (location), sampling periods (TOD stratification), or a combination of both. In a fishery with three access points, each access point represents a stratum (top design, Table 6.2); in a fishery with one access point, stratification is by TOD (bottom design, Table 6.2); and strata are hybrids between location and TOD when there are two access points to a fishery (middle designs, Table 6.2). ***One technician can not cover a fishery with four or more access-points in an access-point survey without overtime pay.*** Whenever a stratum is based on location and the fishing day (first-stage units) is long, the fishing day must be divided into two or more sampling periods of which two or more are sampled. If the fishing day is short enough to equal a sampling period, multistage designs in Table 6.2 are all reduced by one stage.

6.1.2 Other Surveys. A systematic sampling schedule for a single technician will cover roving surveys to estimate harvest., catch, and fishing effort ***so long as there is no stratification by TOD or by location.*** Time-of-day stratification

Table 6.2 - Possible stratified multistage sampling designs based on systematic scheduling of one technician (see Table 6.1) in an onsite access-point survey with three or fewer access locations over six weeks. Numbers in parentheses are the number of sampling units in that stage.

Three Locations:

Stratum:	First Location	Second Location	Third Location
First-stage Units:	Days (42)	Days (42)	Days (42)
Second-stage Units:	Periods (2<)	Periods (2<)	Periods (2<)
Third-stage Units:	Angler-trips	Angler-trips	Angler-trips

Two Locations:

Stratum:	First Location	Second Location/ Early Day	Second Location/ Late Day
First-stage Units:	Days (42)	Days (42)	Days (42)
Second-stage Units:	Periods (2<)	Angler-trips	Angler-trips
Third-stage Units:	Angler-trips		

Stratum:	First Location/ Early Day	First Location/ Late Day	Second Location
First-stage Units:	Days (42)	Days (42)	Days (42)
Second-stage Units:	Angler-trips	Angler-trips	Periods (2<)
Third-stage Units:			Angler-trips

One Location:

Stratum:	Mid Day	Early Day	Late Day
First-stage Units:	Days (42)	Days (42)	Days (42)
Second-stage Units:	Angler-trips	Angler-trips	Angler-trips

is unwise because of the potential for LOS bias. ***During each day sampled, the entire fishing day should be covered to avoid LOS bias*** (see Section 2.2.2.1). For the same reason, the ***entire fishing day should be sampled in surveys to estimate angler success*** (Section 4.4.1) ***and*** $\overline{CPUE}$ (Section 5.2.1). Roving surveys can be stratified by location with strata corresponding to segments of the fishing grounds. However, the need for location stratification in roving-access surveys of Alaska's fisheries is rare, and when needed, requires more than a single technician to implement.

Depending on the length of the fishing day, some overtime can be involved with a systematic schedule for a roving survey:

- When every ***third day*** is sampled (as in the first stratum in Table 6.1), sampling will involve some overtime pay whenever the fishing day is longer than 12 hrs, if overtime pay is given for any specific week in which work exceeds 37.5 hrs (Table 6.3). If overtime pay is given whenever overtime exceeds the number of regular work hours in a pay period (half a month), no overtime pay would be required until the fishing day exceeds 15 hrs.

- When every ***fourth day*** is sampled, no overtime pay is involved unless the fishing day is over 18 hrs long.

- No overtime pay at all would be involved if sampling was at a slower frequency than every fourth day.

Stamina of the technician and the potential for measurement error as he or she tires is always a concern with extended sampling.

Although the potential is less for overtime pay ***as frequency of systematic sampling slows, the potential for "frequency" bias in estimates of harvest, catch, and fishing effort rises in roving surveys***. Any weekly trends in harvest, catch, and fishing effort, will be missed as sampling frequency drops to twice, and then to once a week. Bias results when the "periodic" frequency of systematic sampling in roving surveys is similar to the "periodic" frequency of the parameter to be estimated (see Cochran 1977:217-219). For example, if fishing effort is higher on weekends, a sampling frequency of seven days (sampling only one day per week) will result in an estimate biased either high (a weekend day sampled) or biased low (a weekday sampled). Because they usually do not have weekly trends, ***angler success and*** $\overline{CPUE}$ ***can be accurately estimated with scheduling sampling for only one fishing day (sampling period) per week with little fear of "frequency" bias.***

Systematic scheduling in one-technician surveys to estimate ***harvest composition***, ***angler success, and/or*** $\overline{CPUE}$ as an index of abundance can be "clustered". Sampling in these surveys is spread temporally across the season in a regular fashion to promote proportional sampling (see Sections 3.2, 3.3.2, and 4.2) or to expose trends (Section 5.1.1). Systematic scheduling here is not so much part of a sampling design as a strategy to spread

Table 6.3 - Relative increase in costs due to overtime in a one-technician roving survey with systematic scheduling every third day as the length of a fishing day increases.

Length of Fishing Day (hrs)	Increase in Cost	Length of Fishing Day (hrs)	Increase in Cost
13	6%	19	82%
14	18%	20	110%
15	30%	21	138%
16	42%	22	166%
17	54%	23	194%
18	66%	24	222%

out sampling. Sampling for two consecutive fishing days and not for the next five has the same effect on calculations as sampling for five days and not for two. For surveys to estimate harvest or angler success, all data would be pooled; for surveys to calculate $\overline{cpue}$ as an index of abundance, data would be pooled within each cluster of consecutively sampled periods. Such "clustering" can maximize the sample size given the available sampling effort, or can minimize the occurrence of depensatory sampling. The potential benefits from clustering are greatest in onsite surveys to estimate harvest composition because unlike surveys to index abundance or estimate angler success, ***not all of a fishing day need be sampled in surveys to estimate harvest composition*** (see Section 3.2) The survey described in Example 3.3 has a "clustered", systematic schedule for a single technician.

6.2 RANDOM SCHEDULING

6.2.1 The Basic Schedule. Random scheduling of sampling periods in one-technician surveys permits day-of-the week (DOW) stratification. Saturday and Sunday (and sometimes Friday or Monday) are considered weekends (a stratum) and both days are always sampled. Monday through Friday are considered weekdays, and two of these days are selected randomly without replacement for sampling. Table 6.4 is an example of a random schedule for a single technician for six weeks. Note days off are not consecutive in the third week. There are 10 possible combinations of picking two sampling days from Monday through Friday, only one of which will fail to leave two consecutive days off for the technician. This difficulty should be solved with overtime or use of permanent staff, not with compromising the design.

Do not pick two consecutive days off during week days, then sample around them. Under this method, sampling would occur on Mondays or Fridays 25% of the time and on Tuesdays, Wednesdays, or Thursdays 16%. Since many "weekend" fishing holidays begin on a Friday or end on Monday, fishing effort, catch, and harvest on these days will be higher than on Tuesday through Thursday. Under these circumstances, directly picking two consecutive days will bias (inflate) estimates of fishing effort, harvest, and catch.

6.2.2 Access-point Surveys. For access-point surveys, ***random scheduling of sampling by a single technician is possible only when anglers have a single point of access to the fishery***. In fisheries with more than one access point, a single technician can not sample in two or more places at the same time. If a second or third technician is not available, sampling should be scheduled systematically or the access-point survey design abandoned for a roving survey.

No overtime pay would be involved with random scheduling of a single technician in an access-point survey of a fishery with one access point. Under the schedule in Table 6.4, a technician could spend just under 9.4 hrs sampling per day. If fishing days are ≤ 9.4 hrs long, the sampling design will be stratified (DOW stratification) with two stages: days and angler-trips (Section 2.1). If fishing days are >9.4 hrs, the fishing day can be divided into three or more sampling periods of which two can be randomly chosen for sampling. The sampling design will be stratified (DOW stratification) with three stages: days, sampling periods, and angler-trips (Section 2.1). In a random schedule, time-of-day (TOD) stratification is impossible without frequently breaking the union rule for two consecutive days off each week for a single technician.

Table 6.4 - Example of a random schedule that will enable a single technician to completely cover a sampling design with weekly DOW stratification.

Week	**Day**		**Week**	**Day**	
First	M		**Fourth**	M	off
	T	off		T	off
	W	off		W	off
	T	off		T	
	F			F	
	S			S	
	S			S	
Second	M	off	**Fifth**	M	
	T	off		T	off
	W			W	off
	T			T	
	F	off		F	off
	S			S	
	S			S	
Third	M		**Sixth**	M	off
	T	off		T	off
	W			W	off
	T	off		T	
	F			F	
	S			S	
	S			S	

6.2.3 Roving Surveys. Random scheduling of a technician in a one-technician, roving survey hinges around the length of the fishing day. The entire fishing day needs to be sampled in a roving survey to avoid LOS bias in estimates of harvest, catch, and fishing effort regulated with a daily bag limit (see Section 2.2.2.1). In the basic random schedule, week is a dimension of stratification with each week having two strata requiring a minimum schedule of two sampled periods (days) per strata and four per week. A single technician can cover four sampled periods per week if the fishing day is ≤ 9 (=37.5/4) hrs long (Table 6.5). If the fishing day is longer than nine hours, the basic random schedule will not cover the sampling design.

The fix is to drop weekly stratification in favor of biweekly stratification. DOW stratification would still remain, only now there would be 4 (instead of 2) weekend days and 10 (instead of 5) week days in the two strata. A single technician now represents 75.0 hrs of sampling effort that can be spread over the fishing days in the two strata. Table 6.5 contains generic allocations of sampling effort that will meet the design requirements of roving surveys given different lengths of fishing days up to almost 19 hrs. When a fishing day is longer than 12.5 hrs, three or two days are scheduled randomly without replacement from the four Saturdays and Sundays in the biweek (fortnight). Regardless of the length of a fishing day, two days are scheduled randomly without replacement from the 10 Mondays, Tuesdays, Wednesdays, Thursdays, and Fridays in the biweek. The union rule of two consecutive days off will always be met with biweekly stratification in one-technician surveys.

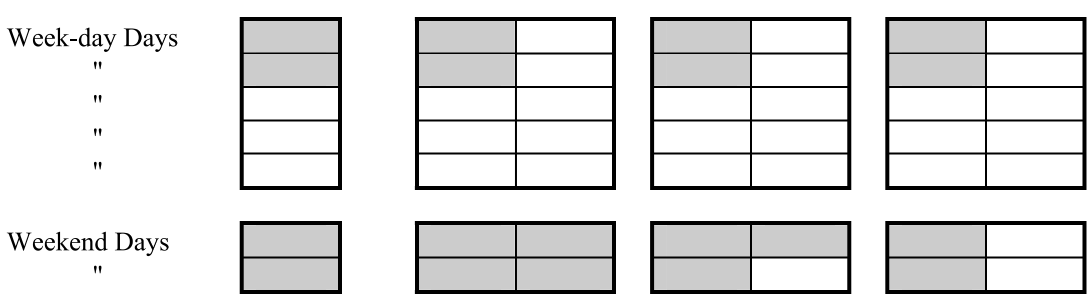
6.3 COMPARISONS, COMPROMISES, AND BIAS

Random scheduling of sampling periods is superior to systematic scheduling when there is a need for day-of-the-week (DOW) stratification (Section 1.2). When most fishing effort occurs on weekends, DOW stratification can greatly improve precision in estimates of ***harvest, catch, and fishing effort***. Systematic scheduling in one-technician surveys to estimate harvest composition, angler success, and mean CPUE is preferred over random scheduling because these parameters are not influenced by weekly cycles in fishing effort.

Other schedules may be used effectively in "one-technician" surveys with more complex designs when some informed "stream-lining" can be accomplished. If some strata represent negligible harvest, fishing effort, or catch, they can be ignored with little penalty, thereby providing considerably more flexibility in scheduling. For instance, strata could be ignored if they represent a negligible source of variation. Rainbow Campground on the Russian River (Example 2.2) represented only 2% of the estimated harvest of sockeye salmon that exited the fishery in 1991. Since information from previous years followed the same trend, loss of sampling at Rainbow Campground represents a relatively small bias in estimated harvest. ***Any stream-lining of sampling designs should be defensible with a rigorous analysis of information from past surveys.***

Table 6.5 - Allocation of sampling effort among fishing days (first-stage units) under a random schedule that permits a single technician to cover the sampling design in a roving survey with weekly and biweekly DOW stratification. Boxes represent possible sampling periods (fishing days), and shaded boxes represent sampled days (periods).

	Weekly Stratification	Biweekly Stratification	Biweekly Stratification	Biweekly Stratification
Length of a Fishing Day (hrs)	≤ 9.4	9.4 - 12.5	12.5 - 15.0	15.0 - 18.8



CHAPTER SEVEN: Counts, Interviews, and Creels

Counts and interviews of anglers and inspection of their creels are at the core of onsite creel surveys. Individual anglers, groups of anglers, or mixed groups of anglers and their non-fishing companions are interviewed in onsite surveys as they exit fisheries or while they are fishing. Data are collected through observation prior to an interview, through inquiry during an interview, and through inspection of harvest. If errors are to be avoided when collecting data, technicians must be instructed how to recognize an angler, how to count anglers, what kind of interview to conduct; and how to sample creels to get biological information on harvests. These instructions vary among onsite surveys according to how and what is to be estimated.

This chapter covers sampling protocols that guide technicians when collecting data in access-point, roving-access, and self-weighted surveys. Instructions on how to identify solitary or "party" anglers from recreationists, how to count anglers in roving surveys, what to do with unsolicited information from anglers, and how to sample a creel are given in detail. Different measures of fishing effort, when they are used, and what information is needed to calculate each are also discussed in this chapter, along with a more rigorous distinction between catch and harvest. The chapter ends with an editorial on the role of planning in avoiding errors in collecting and editing data.

7.1 COUNTING ANGLERS

7.1.1 Identifying Anglers. Counting anglers becomes ambiguous when anglers fish in groups. A solitary person exiting a fishery with rod or fish in hand or actively fishing is most likely an angler. However, friends, families, and acquaintances often fish together in groups ("parties"). When a party is encountered on the fishing grounds, or when leaving the fishing grounds, identifying anglers in the group is often problematical, especially when access to the fishing grounds is by boat as in most marine fisheries, or when people can only be encountered as they leave an area in automobiles.

For access-point surveys, the solution is for technicians to count parties as the basic unit of sampling. A party is one or more persons recreating together. There will be no sample variance associated with anglers within parties because all anglers within each contacted, cooperative party will be represented in the interview. Counting parties follows these two rules:

- If parties with no anglers (parties of sightseers, picnickers, water skiers, etc.) can be recognized as such without an interview, ***only those exiting parties with anglers need be counted.***
- If an interview is necessary to identify parties without anglers, ***<u>all</u> exiting parties must be counted.***

In this latter situation, harvest, fishing effort, and catch by all parties without anglers would be recorded as zero. In either case, equations to estimate harvest, catch, and fishing effort in hours listed in previous chapters can still be used, only M is redefined as the number of parties counted, m the number of parties interviewed, and c, H, and e redefined as the catch, harvest, and fishing effort for the party as a whole. If information from individual anglers is available, $\overline{cpue}$ can be calculated for use as an index of fish abundance.

Example 7.1. Harvest of whitefish from the Chatanika River was estimated with an access-point survey in 1993. During sampled periods, technicians counted all vehicles just they pulled onto the Elliot Highway from the two roads providing access to the fishery. Some vehicles were stopped, and their occupants were interviewed. If at least one occupant of a vehicle had been fishing, harvest of all occupants in the car was recorded. If no one in a vehicle had been fishing, a harvest of zero was recorded for the party in that vehicle. Harvest for a sampled period was the harvest by party averaged over all vehicles stopped multiplied by the number of vehicles exiting onto the Elliot Highway during the sampled period.

* * * *

For roving creel surveys, individual anglers and their angler-trips must remain the basic sampling units. Information from interviews is not expanded by counts of exiting anglers, but by roving counts of individual anglers. Anglers must therefore be identified as such while technicians rove through the fishery. Since most roving surveys in Alaska are roving-access surveys, there is no opportunity for technicians to ask people encountered if they had been fishing. A person judged to be fishing during the count will exhibit one of the following characteristics:

- actively fishing, either casting or soaking a line;
- working with fishing gear, unsnarling a line, baiting a hook, changing a lure, etc.; or
- landing a fish

A person(s) walking along the stream bank or boating across the water should not be counted as an actively fishing angler.

7.1.2 Duration of Counts. The count of anglers in access-point surveys continues through a sampled period, but ***counts in roving surveys must be "instantaneous" where instantaneous is relative to rates of change in the number of anglers fishing during a sampled period.*** If time each angler spends fishing is long relative to time taken for a technician to pass once through a fishery on a count, counted numbers of fishing anglers will reflect fishing effort. However, if time spent by anglers fishing is short relative to the duration of a count, estimates of fishing effort and subsequently
catch and harvest from the roving survey will be biased low. This problem can be detected by comparing average length of a fishing trip across anglers interviewed throughout the fishing day (see Appendix A) with duration of a roving count. The solution is to have more technicians participate in a count to shorten its duration. If this solution is unfeasible, the onsite survey should be canceled.

7.1.3 Geography of Counts. Because counts in most roving surveys are progressive counts that must start at some geographical point, that point should be selected randomly (see Pollock et al. 1994:176-177) for each sampled period. Once a starting point has been selected, the technician should pass through the fishery at an even pace towards the far geographical point. The same route should be followed for subsequent counts through the fishery in that sampled period. If anglers (or parties) can be counted from a single location while they fish, selecting a geographical starting point for counts is a moot issue.

7.2 INTERVIEWING ANGLERS

7.2.1 Avoiding Bias. There is a potential for bias in estimates from creel surveys because technicians can not randomly encounter anglers. Spatial and temporal distributions of anglers limit opportunity to interview anglers during a sampled period. In access-point and roving-access surveys based on "completed-trip" interviews, anglers choose when they can be sampled, not technicians. Even in roving-roving surveys in which anglers are counted and interviewed "simultaneously", the sample is at best systematically drawn.

Interviewing all (or nearly all) anglers exiting a fishery at an access point during a sampled period guarantees an unbiased sample in access-point surveys to estimate harvest, catch, or fishing effort, but not in other surveys. Catch rates and subsequently harvest rates can be consistently different for anglers exiting at different access points. These rates are of no consequence to estimating harvest, catch, or fishing effort in access-point surveys, but is of paramount importance in roving surveys, CPUE surveys, and surveys to estimate angler success. Consistent trends in $\overline{HPUE}$ can be detected with hypothesis tests on statistics gathered at several locations. Extra technicians may be employed in a short observational study to obtain simultaneous information at several access-points during a subset of sampled periods. If these hypothesis tests show no significant trends when sample sizes are adequate to do so, evidence is that samples taken at the usual location are unbiased. If trends are indicated, sampling designs should be changed in the current and in future surveys.

If a significant fraction of exiting anglers can not be interviewed at a location during a sampled period, sampling can produce biased information if technicians have a higher probability of interviewing one type of angler over another. This problem can be detected with a short observational study during a subset of sampled periods. A second technician is instructed to only interview anglers when the first technician is engaged in an interview with another angler. Responsibilities of the second technician are reduced so this person can contact as many exiting anglers as possible, but again, only while the first technician is engaged with an angler. Statistics gathered by both technicians are compared, and if found to be disparate, evidence is that at least two technicians are needed to produce a representative sample of exiting anglers.

Unless all anglers exiting at a location can be interviewed, technicians should set aside information from unsolicited anglers that request an interview. Many successful anglers will volunteer to be interviewed because they know technicians have a scale to weigh fish. Technicians should be courteous and record the volunteered information, but should keep it separate from data collected through solicited interviews. If all exiting anglers can be interviewed during a sampled period, information from unsolicited interviews need not be set aside.

7.2.2 Fishing Effort. Fishing effort can not be tallied in the appropriate units unless technicians ask anglers about the chronology of their fishing. If fishing effort is to be measured in ***angler-hours*** (an angler-hour is one hour of fishing with one unit of gear), technicians must ask each angler how long in hours they +fished that day. This measure of fishing effort is used in roving surveys to estimate catch or harvest, or in CPUE surveys to index abundance. An hour of fishing effort should be an hour spent actively engaged in fishing (e.g., line in the water). However, calculating fishing effort in such detail is seldom feasible because anglers do not generally keep close track how they spend their time. ***Asking an angler to recall the approximate time he or she started fishing, and by subtraction estimate how long they had fished, is a more realistic approach to estimating hours spent fishing.*** Time spent fishing can also be discounted by answers to questions on ***how much time was spent traveling to and from the fishing grounds or the length of any long breaks in fishing while on the grounds.*** If fishing effort is being estimated as part of a roving or CPUE survey, technicians must also ask each angler how many fish they caught (and kept) while fishing, regardless if that angler is alone or a member of a party.

As a measure of fishing effort, an ***angler-day*** occurs whenever an angler fishes within a 24-hour day. ***Exiting anglers must be asked if they have finished fishing for the day.*** In an access-point survey, anglers counted exiting the fishery (*M*) must be discounted by the fraction of interviewed anglers that responded "no". No discount is needed to estimate harvest or catch. In a roving survey, fishing effort in angler-days is estimated with the same procedures to estimate effort as angler-trips (Section 2.2.3), only the time spent fishing that day is calculated with information only from those interviews in which anglers responded "yes" to the question. Angler-days from onsite surveys can be compared with a angler-days of fishing effort from the statewide harvest survey (i.e., Howe et al. 1996). An angler-day is also the unit of fishing effort used to determine angler success (Section 4.1) which is calculated only with information from anglers that had completed fishing for the day.

If fishing effort is to be recorded in ***angler-trips*** (angler-trips are counts of anglers exiting a fishery), no additional question needs be asked exiting anglers. Interviews in both access-point and roving-access surveys are "completed-trip" interviews (see Sections 2.1 and 2.2.1) which only occur at the ends of fishing trips. Angler-trips are usually the basic sampling units in all surveys; however, their number becomes a measure of fishing effort on rare occasions. When angler-trips are being estimated as fishing effort, sampling periods become the basic sampling units.

If statistics, including fishing effort, are to be post-stratified, a technician must also obtain information about an angler and his or her fishing experience. If a technician can obtain pertinent information through casual observation of an angler, no questions need be asked. If not, technicians must ask direct questions so information from interviews can be assigned to the correct stratum.

7.2.3 Catch and Harvest. For onsite creel surveys in Alaska, ***an angler's <u>catch</u> are all fish captured during his or her fishing trip while his or her <u>harvest</u> is a subset of the catch not returned alive to the water***. A fish is considered captured if the decision to keep it is made by the angler and not the fish. "The one that got away" should not be considered part of the catch. Because anglers may not share our definitions of catch and harvest, technicians should define these terms in the questions they ask.

Example 7.2. This is another excerpt from the manual given to technicians employed in 1993 by the Juneau Marine Creel Survey:

> *.... ask them* (the angler) *what they caught and kept, and also caught and released. Do not record strikes or fish that got off the line by themselves as "released" fish.*

* * * *

Information on the catch will always be less reliable than information on harvest. Some anglers will not accurately recall how many fish they released nor have accurately identified them by species. In contrast, harvest is usually available for inspection by a technician. Exceptions are anglers who have cleaned, processed, and sometimes eaten their harvest prior to the interview.

7.3 SAMPLING THE CREEL

Before the start of a survey, technicians should be instructed on how to correctly identify species of fish that might be caught or harvested in the fishery to be surveyed. The *Alaska Sport Fish Identification Handbook*, *Alaska's Saltwater Fishes and Other Sea Life* (Kessler 1985), plus *Guide to Northeast Pacific Rockfishes* (Kramer and O'Connell 1988) can be used for instruction and for reference. Each manual has a color photo section on species identification (mainly close-ups of fish tails) plus photos of chinook salmon in progressive states of maturity. During an onsite survey, accuracy of identifications by technicians should be checked periodically with visits to sampling sites.

Technicians should be instructed to ***sample all available fish in each creel sampled***. Sample sizes are maximized under this rule while any biases from a non-random subsample are avoided. If most creels are large and sampling is time-consuming, a subsample can be systematically drawn from each creel to avoid potential bias, albeit at a reduced sample size. Biases may still occur if some of the harvest has been cleaned prior to the interview. For instance, smaller halibut harvested by chartered anglers are often cleaned at sea with carcasses thrown overboard while larger fish are returned to port. Under these circumstances, post-stratification of harvest can negate bias in estimates of age and size composition (see Example 3.2).

7.4 AVOIDING ERRORS

Data are captured by recording answers, counts, and measurements onto water repellent paper, either in script for later transcription by hand, or in code for machine reading, or by entry into electronic notepads. Electronic files of captured data are edited to remove obvious and not so obvious errors from recording and transcription. Files are finally reformatted to ease calculation of statistics.

At every step in this process, there is opportunity for error. Technicians can misidentify a species of fish or miscount exiting anglers. Data can be misrecorded during measurement or during transcription. Anglers may be offended and become uncooperative when facing a technician with an attitude. A list of what can go wrong with data collected during an onsite survey seems almost endless. Once hidden in the data, some errors can not be identified or corrected. The only real solution is not to make these errors at all.

If errors are to be avoided, technicians must understand and follow their instructions. Technicians should be conscientious, intelligent, personable, and have a good work ethic. Their training should begin before the survey and should encompass definitions, schedules, protocols, public relations, taxonomy, and enough of the sampling design to understand what information is desired from the survey and how it will be obtained. During the survey, each technician should occasionally be observed while at work and their performance evaluated. Feedback from technicians on potential problems with their instructions should be solicited.

Many sources of error listed above can be avoided through good planning. Alaska has a policy of planning for its research into recreational fisheries (Bernard et al. 1993), research that includes onsite creel surveys. Each creel survey must have a written plan completed before data are collected. Part of this operational plan, the DATA REDUCTION Section, is a list of steps taken to capture, edit, and transform data for analysis. A thorough transcription of these steps in the DATA REDUCTION Section forces those in charge of onsite creel surveys to think about what can go wrong with collecting data and subsequently how to avoid errors.

7.5 REFERENCES FOR EXAMPLES

Hallberg, J. E., and A. E. Bingham. 1994. Creel surveys conducted in interior Alaska during 1993. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 94-27. Anchorage.

Hubartt, D. J., A. E. Bingham, and P. M. Suchanek. 1994. Harvest estimates for selected marine sport fisheries in southeast Alaska, 1993. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 94-33. Anchorage.

LITERATURE CITED

Bernard, D. R., W. H. Arvey, and R. A. Holmes. 1993. Operational planning: the Dall River and rescue of its sport fishery. Fisheries 18(2):6-13.

Bernard, D. R., A. E. Bingham, and M. Alexandersdottir. (1998). Robust harvest estimates from onsite roving-access creel surveys. Transactions of the American Fisheries Society 127:(May).

Bernard, D. R., and J. E. Clark. 1996. Estimating salmon harvest with coded-wire tags. Canadian Journal of Fisheries and Aquatic Sciences 53:2323-2332.

Bernard, D. R., R. P. Marshall, and J. E. Clark. (submitted). Planning programs to estimate salmon harvest with coded-wire tags. Canadian Journal of Fisheries and Aquatic Sciences.

Clark, J. E., and D. R. Bernard. 1987. A compound multivariate binomial-hypergeometric distribution describing coded microwire tag recovery from commercial salmon catches in southeastern Alaska. Alaska Department of Fish and Game, Division of Commercial Fisheries Informational Leaflet 261. Juneau.

Cochran, W. G. 1977. Sampling techniques, 3rd ed. John Wiley and Sons, Inc. New York.

DiCiccio, T. J., and J. P. Romano. 1988. Review of bootstrap confidence intervals. Journal of the Royal Statistical Society B 50:338-354.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics Publication No. 38. Philadelphia.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman Hall. New York.

Geiger, H. J. 1990. Parametric bootstrap confidence intervals for estimating contributions to fisheries from marked salmon populations. American Fisheries Society Symposium 7:667-676.

Goodman, L. A. 1960. On the exact variance of products. Journal of the American Statistical Association 55:708-713.

Hahn, G. J., and W. G. Meeker. 1991. Statistical intervals: a guide for practitioners. John Wiley and Sons, Inc. New York.

Hoenig, J. M., C. M. Jones, K. H. Pollock, D. S. Robson, and D. L. Wade. 1997. Calculation of catch rate and total catch in roving surveys of anglers. Biometrics 53:306-317.

Howe, A. L., G. Fidler, A. E. Bingham, and M. J. Mills. 1996. Harvest, catch, and participation in Alaska sport fisheries during 1995. Alaska Department of Fish and Game, Division of Sport Fish Fisheries Data Series Report 96-32. Anchorage.

Jones, C. M., D. S. Robson, H. D. Lakkis, and J. Kressel. 1995. Properties of catch rates used in analysis of angler surveys. Transactions of the American Fisheries Society 124:911-928.

Kessler, D. W. 1985, Alaska's saltwater fishes and other sea life. Alaska Northwest Publishing Company, Anchorage.

Kramer, D. E., and V. M. O'Connell 1988. Guide to northeast Pacific rockfishes. Alaska Sea Grant Marine Advisory Bulletin No. 25. Fairbanks.

Malvestuto, S. P. 1983. Sampling the recreational fishery, p. 397-419 *in* L. A. Nielsen and D. L. Johnson, editors. Fisheries Techniques, American Fisheries Society, Bethesda, Maryland.

Malvestuto, S. P., W. D. Davies, and W. L. Shelton. 1978. An evaluation of the roving creel survey with nonuniform probability sampling. Transactions of the American Fisheries Society 107:255-262.

Mills, M. J. 1992. Harvest, catch, and participation in Alaska sport fisheries during 1991. Alaska Department of Fish and Game, Division of Sport Fish Fisheries Data Series Report 92-40. Anchorage.

LITERATURE CITED (Continued)

Mills, M. J. 1993. Harvest, catch, and participation in Alaska sport fisheries during 1992. Alaska Department of Fish and Game, Division of Sport Fish Fisheries Data Series Report 93-42. Anchorage.

Newman, K. B. 1990. Variance estimation for stock-contribution estimates based on sample recoveries of coded-wire-tagged fish. American Fisheries Society Symposium 7:677-685.

Pollock, K. H., C. M. Jones, T. L. Brown. 1994. Angler survey methods and their applications in fisheries management. American Fisheries Society Special Publication 25.

Pollock, K. H., J. M. Hoenig, C. M. Jones, D. S. Robson, and C. J. Greene. 1997. Catch rate estimation for roving and access point surveys. North American Journal of Fisheries Management 17:11-19.

Robson, D. S. 1961. On the statistical theory of a roving census of fishermen. Biometrics 17:415-437.

Schnute, J. T.1992. Statistical analysis of embedded replicates in mark-recovery experiments. Canadian Journal of Fisheries and Aquatic Sciences 49:432-442.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters, 2nd ed. Charles Griffin. London.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and C. Asok. 1984. Sampling theory of surveys with applications, third ed. Iowa State University Press, Ames, Iowa.

Thompson, S. K. 1992. Sampling. John Wiley and Sons, New York.

Wolter, K. M. 1985. Introduction to variance estimation. Springer-Verlag, New York.

APPENDIX

A: Estimating Mean Length of an Angler-Trip

B: Selecting Sampling Units in Proportion to their Size

C: Variance Equations for Unequal Sampling Fractions

D: Jackknife Estimator for Mean Harvest Rate in Roving Surveys

E: Covariance among Post-Stratified Estimates of Harvest in Roving Surveys

F: Roving Surveys with Missing Data on Harvest Rates

G: Relationships among Equations to Estimate Relative Harvest Composition

H: Source Code for the QuickBasic™ Program "DELTA.BAS"

APPENDIX A

Estimating Mean Length of an Angler-Trip

Accurate estimates of mean length of an angler-trip for a stratum or in an entire fishery are only possible when sampling periods in an onsite creel survey are the same length as the fishing day (Bernard et al. 1998). Statistics from sampled periods shorter than the fishing day are subject to LOS bias. Because the fishing day is subsampled in most access-point surveys, mean length of an angler-trip estimated from these surveys will be biased, especially if fishing is regulated with a daily bag limit. In contrast, sampling periods are the same length as the fishing day in all roving-access surveys (see Section 2.2.2.1), and estimates of mean length of an angler-trip should be free of LOS bias in this latter type of survey.

Because sampled periods in onsite surveys in Alaska are chosen with equal probability (see Section 1.4.2), estimated means for each sampled period must be weighted by the relative size of the sampled period and stratum (number of angler-trips) (see Cochran 1977:299, Method II in Table 11.5). During each sampled period, a sample of anglers are asked at the completion of their fishing trip how long they fished. Their answers are multiplied by ***sample weights*** and the products averaged over the sampled period. Means for each sampled period are averaged over all sampled periods to produce statistics for the next higher stage in the sampling design. This process continues through all sampling stages. Finally, means for each stratum are multiplied by ***stratum weights***, and the products summed to produce one estimated mean for the fishery.

Application of ***sample weights*** to individual data for two- and three-stage survey designs are, respectively:

$$y_{hij} = \frac{M_{hi}}{\hat{\bar{M}}_h} e_{hij}$$

$$y_{hijk} = \frac{M_{hij}}{\hat{\bar{M}}_{hi}} e_{hijk}$$

where e is the length of an angler-trip, M the number of anglers exiting during a sampled period, $\hat{\bar{M}}_{hi}$ the estimated mean numbers exiting across all sampled periods, y is the weighted statistic, h denotes a stratum, and i, j, and k first through third-stage sampling units. In this formulation, the sample weights (i.e., $M_{hij} \big/ \hat{\bar{M}}_{hi}$) are applied to answers supplied directly by anglers even though sample weights would traditionally be applied to estimated means for sampled periods (see Sukhatme et al. 1984:326-8). Coding basic data with sample weights instead of penultimate units (sampled periods) simplifies the expression of equations without changing results. In access-point surveys, the M_{hij} are obtained through counting anglers as they exit the fishery during the sampled period. In roving-access surveys, the M_{hi} are estimated by dividing estimated fishing effort $\hat{E}_{hi}$ by estimated mean length of angler-trip for a ***sampled period*** $\bar{e}_{hi}$ where:

$$\bar{e}_{hi} = \frac{\sum_{j=1}^{m_{hi}} e_{hij}}{m_{hi}}$$

Once data have been so coded with sample weights, coded statistics (y) are put through equations in Tables A.1 or A.2 to estimate mean length of an angler-trip for each stratum.

Statistics for each stratum are multiplied by stratum weights, and the products summed to estimate mean length for the fishery:

$$\hat{W}_h = \frac{\hat{M}_h}{\sum_{h'=1}^{L} \hat{M}_{h'}} = \frac{\hat{M}_h}{\hat{M}}$$

When stratum weights are estimated (as here), the estimate of the true mean $\bar{\varepsilon}$ is biased by the quantity $\sum(\hat{W}_h - W_h)\bar{\varepsilon}_h$ and its estimate of variance by $[\sum(\hat{W}_h - W_h)\bar{\varepsilon}_h]^2$ (Cochran 1977:117-8). With this bias and with a small bias from estimating sample weights, estimated variance becomes a mean square error (MSE):

$$MSE(\bar{e}) = \sum_{h=1}^{L}\left[\hat{W}_h^2 var(\bar{e}_h) + var(\hat{W}_h)\bar{e}_h^2 - var(\hat{W}_h)var(\bar{e}_h)\right]$$

based on methods in Goodman (1960). Using the delta method in Seber (1982:8-9), an approximate estimate of variance for $\hat{W}_h$ is:

$$var(\hat{W}_h) \cong \hat{W}_h^2\left[\frac{var(\hat{M}_h)}{\hat{M}_h^2} + \frac{var(\hat{M})}{\hat{M}^2} - \frac{2var(\hat{M}_h)}{\hat{M}_h\hat{M}}\right]$$

where $var(\hat{M}) = \sum var(\hat{M}_h)$. Equations to calculate $\hat{M}_h$ and estimate its variance for roving-access surveys are given in Table 2.9 (Equation 2.14). For access-point surveys, the following substitutions should be made in the listed equations to calculate statistics:

Statistic	Equation	Plug-in:
Two-stage design:		
$\hat{M}_h$	2.1	$\hat{Y}_{hi} \leftarrow M_{hi}$
$var(\hat{M}_h)$	2.5	$\hat{Y}_{hi} \leftarrow M_{hi}$
Three-stage design:		
$\hat{M}_h$	2.3	$\hat{Y}_{hij} \leftarrow M_{hij}$
$var(\hat{M}_h)$	2.7	$\hat{Y}_{hij} \leftarrow M_{hij}$

Three or four-stage sampling designs based on subsampling a fishing day are not covered because estimates of mean length of an angler trip from such surveys will be biased.

Table A.1 - Equations to estimate mean length of an angler-trip by stratum and to estimate its variance in ***two-stage sampling designs*** used in ***roving-access*** and ***access-point creel surveys*** of sport fisheries in Alaska. Individual data have been weighted by relative size of sampling periods in angler-trips prior to these calculations. Number of fishing days (sampling periods) in a stratum is D, and d is the number of these periods sampled. Choice of formulations for s^2 correspond to sampling units being selected randomly or systematically.

$$\bar{e}_h = \frac{\sum_{i=1}^{d_h} \bar{y}_{hi}}{d_h} \qquad \bar{y}_{hi} = \frac{\sum_{j=1}^{m_{hi}} y_{hij}}{m_{hi}}$$

$$f_{1h} = \frac{d_h}{D_h} \qquad s_{1h}^2 = \frac{\sum_{i=1}^{d_h} (\bar{y}_{hi} - \bar{y}_h)^2}{d_h - 1} \quad \text{or} \quad \frac{\sum_{i=2}^{d_h} (\bar{y}_{hi} - \bar{y}_{h(i-1)})^2}{2(d_h - 1)}$$

Roving-access surveys:

$$var(\bar{e}_h) = (1 - f_{1h})\frac{s_{1h}^2}{d_h} + f_{1h}\sum_{i=1}^{d_{hi}} var(\bar{e}_{hi})$$

Method to calculate $var(\bar{e}_{hi})$ is given in Equation 2.19.

Access-point surveys:

$$var(\bar{e}_h) = (1 - f_{1h})\frac{s_{1h}^2}{d_h} + f_{1h}\sum_{i=1}^{d_{hi}} \left[(1 - f_{2h)})\frac{s_{2hi}^2}{d_h^2 m_{hi}} \right]$$

$$f_{2hi} = \frac{m_{hi}}{M_{hi}} \qquad s_{2hi}^2 = \frac{\sum_{j=1}^{m_{hi}} (y_{hij} - \bar{y}_{hi})^2}{m_{hi} - 1}$$

Table A.2 - Equations to estimate mean length of an angler-trip by stratum and to estimate its variance in ***three-stage sampling designs*** used in ***access-point creel surveys*** of sport fisheries in Alaska. Individual data have been weighted by relative size of sampling periods in angler-trips prior to these calculations. Number of access points in the stratum is *N* and the number at which sampling occurred is *n*. Number of fishing days (sampling periods) in a stratum is *D,* and *d* is the number of these periods sampled. Choice of formulations for s^2 correspond to sampling units being selected randomly or systematically.

$$\bar{e}_h = \frac{\sum_{i=1}^{n_h} \bar{y}_{hi}}{n_h} \qquad \bar{y}_{hi} = \frac{\sum_{j=1}^{d_{hi}} \bar{y}_{hij}}{d_{hi}} \qquad \bar{y}_{hij} = \frac{\sum_{k=1}^{m_{hij}} y_{hijk}}{m_{hij}}$$

$$var(\bar{e}_h) = (1 - f_{1h})\frac{s_{1h}^2}{n_h} + f_{1h}\sum_{i=1}^{n_h}\left[(1 - f_{2hi})\frac{s_{2hi}^2}{n_h^2 d_{hi}}\right] + f_{1h}\sum_{i=1}^{n_h}\left[f_{2hi}\sum_{j=1}^{d_{hi}}\left[(1 - f_{3hij})\frac{s_{3hij}^2}{n_h^2 d_{hi}^2 m_{hij}}\right]\right]$$

$$f_{1h} = \frac{n_h}{N_h} \qquad f_{2hi} = \frac{d_{hi}}{D_{hi}} \qquad f_{3hij} = \frac{m_{hij}}{M_{hij}} \qquad s_{3hij}^2 = \frac{\sum_{k=1}^{m_{hij}}(y_{hijk} - \bar{y}_{hij})^2}{m_{hij} - 1}$$

$$s_{1h}^2 = \frac{\sum_{i=1}^{n_h}(\bar{y}_{hi} - \bar{y}_h)^2}{n_h - 1} \text{ or } \frac{\sum_{i=2}^{n_h}(\bar{y}_{hi} - \bar{y}_{h(i-1)})^2}{2(n_h - 1)}$$

$$s_{2hi}^2 = \frac{\sum_{j=1}^{d_{hi}}(\bar{y}_{hij} - \bar{y}_{hi})^2}{d_{hi} - 1} \text{ or } \frac{\sum_{j=2}^{d_{hi}}(\bar{y}_{hij} - \bar{y}_{hi(j-1)})^2}{2(d_{hi} - 1)}$$

APPENDIX B

Selecting Sampling Units in Proportion to their Size

In onsite creel surveys based on multistage sampling designs, probability of scheduling specific sampling units, such as locations, days, or sampling periods, can be the same across all units within a stage or proportional to the size of each unit. Procedures in this text are based on all sampling units except the last in the hierarchy (angler-trips) having been chosen either randomly or systematically with equal probability, a procedure called ***ep sampling*** (see Section 1.4.2). When sampling units within a stage are of unequal size, an alternative to *ep* sampling is often to select units with a probability proportional to their relative sizes. If sizes of units are known before the survey, this sampling is labeled ***pps sampling***, or is called ***ppz sampling*** if the sizes of units must be estimated (Cochran 1977:295-299). The size of a sampling unit is the number of subunits it contains, that is, the number of units in the next lower stage (Figure 1.2). Usually sampling periods are the only units in a sampling design for onsite creel surveys that have different sizes. The size of a sampling period is the number of angler-trips in the period. Because the number of angler-trips in a period is known only after sampling, *ppz* sampling is the appropriate alternative to *ep* sampling for onsite creel surveys. While *ep* sampling is based on selecting sampling periods without replacement, *ppz* sampling is based on selecting periods *with replacement.*

The question is which is better, *ep* or *ppz* sampling, for onsite creel surveys in Alaska. For ***access-point surveys to estimate catch, harvest, and fishing effort***, precision in estimates tends to be:

- Better with *ppz* sampling when harvest, catch, or fishing effort are correlated with size of sampling periods (from Cochran 1977:255-6).
- Better with *ep* sampling when most anglers exiting during a sampled period are interviewed (there is no *fpc* for the penultimate stage in *ppz* sampling because sampling is with replacement as noted in Cochran 1977:258).

In Alaska, harvest, catch, and fishing effort are usually strongly correlated to the number of anglers exiting a fishery during a sampled period, but in access-point surveys, almost all exiting anglers are interviewed. If all or nearly all exiting anglers are interviewed, *ep* sampling is the clear winner. ***With a census during each sampled period, the final component of estimated variance will disappear, and sampling periods in effect become the de facto basic sampling units.*** Under these circumstances, *ep* sampling and *pps* sampling are equivalent in that all higher sampling units are of the same size. Sampling with equal probability will produce more precise estimates because the finite number of sampling periods in a stratum is ignored by sampling with replacement when *ppz* sampling.

Some of the improvement in precision of estimated harvest, catch, and fishing effort gained through *ppz* sampling can be realized with *ep* sampling by stratifying sampling periods by size, then optimally allocating sampling effort among strata (see Cochran 1977:Sections 5.5-6). More *ep* sampling would occur during strata with more harvest, catch, or fishing effort. Unfortunately, this approach requires extra sampling effort (see Example B.1). However, ***if estimates of the sizes of sampling periods are biased, stratifying periods by size will still produce unbiased estimates of harvest, catch, or fishing effort with* ep *sampling***, though precision will suffer. In contrast, ***inaccurate estimates for the size of sampling periods will bias estimates under* ppz *sampling when harvest, catch, and fishing effort are correlated to size of periods***.

When estimating harvest composition through subsampling harvest simultaneously with estimating harvest (Section 3.1.2), *ppz* sampling will increase sample size if harvest is correlated to the number of exiting anglers. However, if sample sizes are large without this indirect benefit from *ppz* sampling, the marginal improvement in precision of estimated harvest composition from *ppz* sampling will be negligible.

Example B.1. In Alaska, *ppz* sampling has been used only in two access-point surveys, both marine recreational fisheries for chinook salmon in 1990. Sampling during surveys near Juneau and near Ketchikan in 1990 and a year earlier in 1989 followed a stratified three-stage design with days as first-stage units and harbors as second-stage units. There was seasonal (biweekly), TOD (early vs. late), and DOW (weekend vs. week day) stratification in both surveys in both years. Surveys in 1989 had additional stratification for location (high-use vs. low-use harbors). In 1990, *ppz* sampling of harbors was used to remove this last dimension of stratification. Four strata had to be sampled every two weeks in 1990 with a minimum of 16 periods sampled (two days each stratum with two harbors each sampled day) under *ppz* sampling. Under *ep* sampling in 1989, eight strata (add stratification for high-use vs. low-use harbors) and a minimum of 32 periods had to be sampled every two weeks . The same number of periods was sampled each year with the same level of staffing.

Sampled harbors were chosen with *ppz* sampling in 1990 with information from 1989. Sizes of sampling periods were estimated as the average number of "boat-trips" exiting each harbor within a stratum. Averages were normalized to sum to one within each stratum to provide estimates of probabilities. Different sets of probabilities were used for different strata throughout the season. About the same number of periods were sampled each year with the expectation of considerably better precision in estimates of harvest in 1990 from *ppz* sampling.

Instead estimated precision for estimates in 1990 was worse than for comparable statistics for 1989 (Table B.1). All or almost all exiting anglers were interviewed in both years, reducing the last component in estimated variance to zero or near zero in both years, essentially making sampling periods the basic sampling units. Although approximately the same number of harbors were sampled just as often in 1990 as in 1989, estimated variance in the later year did not benefit from an *fpc* for sampling second-stage units.

As expected, a larger proportion of harvest was inspected with *ppz* sampling in 1990 than with *ep* sampling in 1989. In Ketchikan during 1989, 9.3% of the harvest of chinook salmon was inspected for missing adipose fins, secondary marks on fish that have coded wire tags. During 1990, 10.1% of harvested chinook salmon were inspected in Ketchikan. Similarly, 7.0% of harvest was inspected in 1989 versus 11.0% in 1990 during surveys near Juneau.

Table B.1 - Estimates of fishing effort (angler-hours), harvest of chinook salmon, and estimates of their SEs and CVs from access-point surveys of marine fisheries near Juneau and Ketchikan in 1989 based on *ep* sampling and *ppz* sampling in 1990.

	1989: *ep* sampling	1990: *ppz* sampling
Ketchikan:		
Effort	228,798	203,020
SE	11,061	11,829
CV (%)	**4.8**	**5.8**
Harvest	4,675	7,814
SE	485	910
CV (%)	**10.4**	**11.6**
Juneau:		
Effort	307,361	333,157
SE	10,455	15,858
CV (%)	**3.4**	**4.8**
Harvest	6,293	6,546
SE	436	537
CV (%)	**6.9**	**8.2**

* * * *

The problem with *ppz* sampling in roving surveys is length-of-stay (LOS) bias more than precision. Proportional sampling has been used outside of Alaska to simplify scheduling by dividing the fishing day into two or more sampling periods (see Malvestuto et al. 1978; Malvestuto 1983; Pollock et al. 1994:43,247-250). Fewer technicians are needed to subsample the fishing day, thereby saving money. A single period is selected (sampled) per day with probability according the proportion of daily harvest, catch, or fishing effort expected in a period. After data have been collected, the estimate for the sampled period is expanded by that proportion to get the statistic for the day. Pollock et al. (1994:251) observes that statistics from roving-roving surveys are particularly susceptible to LOS bias and recommends use of roving-access surveys to estimate harvest, catch, or fishing effort. Unfortunately, Bernard et al. (1998) has shown that these statistics are still subject to LOS bias in roving-access surveys of fisheries regulated with a daily bag limit when the fishing day is subsampled. All recreational fisheries in Alaska are regulated with bag limits.

Which is better? Under typical sampling conditions in Alaska, *ep* sampling is the better choice. Because all or almost all exiting anglers are interviewed during sampled periods in our access-point surveys, *ep* sampling should provide estimates with better precision for these surveys. Estimates from *ppz* sampling are prone to bias because sizes of sampling periods must be estimated. When the same sampling period is selected more than once when *ppz* sampling with replacement in creel surveys, no new samples are actually collected. Subsequent samples are just replicates of the original sample. Because of the potential for LOS bias in estimates of mean harvest rates, *ppz* sampling should not be used in our roving-access surveys at all.

Suchanek, P. M., and A. E. Bingham. 1990. Harvest estimates for selected marine boat sport fisheries in southeast Alaska in 1989. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 90-51. Anchorage.

Suchanek, P. M., and A. E. Bingham. 1991. Harvest estimates for selected marine boat sport fisheries in southeast Alaska during 1990. Alaska Department of Fish and Game, Sport Fish Division Fishery Data Series 91-48. Anchorage.

APPENDIX C

Variance Equations for Unequal Sampling Fractions

Table C.1. - Equations to calculate estimated variance of estimated harvest, catch, or effort with ***access-point creel surveys*** based on stratified multistage sampling designs used in Alaska when sampling fractions in higher stages are unequal. Choice of formulations for sample variances S^2 corresponds to sampling units being selected randomly or systematically.

Three Stages (Day/Period/Trip): **(C.1)**

$$var(\hat{Y}_h) = (1-f_{1h})D_h^2\frac{S_{1h}^2}{d_h} + f_{1h}^{-1}\sum_{i=1}^{d_h}\left[Q_{hi}^2(1-f_{2hi})\frac{S_{2hi}^2}{d_h}\right] + f_{1h}^{-1}\sum_{i=1}^{d_h}\left[f_{2hi}^{-1}\sum_{j=1}^{q_{hi}}\left[M_{hij}^2(1-f_{3hij})\frac{s_{3hij}^2}{m_{hij}}\right]\right]$$

$$s_{3hij}^2 = \frac{\sum_{k=1}^{m_{hij}}(y_{hijk}-\bar{y}_{hij})^2}{m_{hij}-1} \qquad S_{2hi}^2 = \frac{\sum_{j=1}^{q_{hi}}(\hat{Y}_{hij}-\bar{\hat{Y}}_{hi})^2}{q_{hi}-1} \text{ or } \frac{\sum_{j=2}^{q_{hi}}(\hat{Y}_{hij}-\hat{Y}_{hi(j-1)})^2}{2(q_{hi}-1)}$$

$$S_{1h}^2 = \frac{\sum_{i=1}^{d_h}(\hat{Y}_{hi}-\bar{\hat{Y}}_h)^2}{d_h-1} \text{ or } \frac{\sum_{i=1}^{d_h}(\hat{Y}_{hi}-\hat{Y}_{h(i-1)})^2}{2(d_h-1)} \qquad f_{1h}=\frac{d_h}{D_h} \qquad f_{2hi}=\frac{q_{hi}}{Q_{hi}} \qquad f_{3hij}=\frac{m_{hij}}{M_{hij}}$$

Three Stages (Location/Day/Trip): **(C.2)**

$$var(\hat{Y}_h) = (1-f_{1h})N_h^2\frac{S_{1h}^2}{n_h} + f_{1h}^{-1}\sum_{i=1}^{n_h}\left[D_{hi}^2(1-f_{2hi})\frac{S_{2hi}^2}{n_h}\right] + f_{1h}^{-1}\sum_{i=1}^{n_h}\left[f_{2hi}^{-1}\sum_{j=1}^{d_{hi}}\left[M_{hij}^2(1-f_{3hij})\frac{s_{3hij}^2}{m_{hij}}\right]\right]$$

$$s_{3hij}^2 = \frac{\sum_{k=1}^{m_{hij}}(y_{hijk}-\bar{y}_{hij})^2}{m_{hij}-1} \qquad S_{2hi}^2 = \frac{\sum_{j=1}^{d_{hi}}(\hat{Y}_{hij}-\bar{\hat{Y}}_{hi})^2}{d_{hi}-1} \text{ or } \frac{\sum_{j=2}^{d_{hi}}(\hat{Y}_{hij}-\hat{Y}_{hi(j-1)})^2}{2(d_{hi}-1)}$$

$$S_{1h}^2 = \frac{\sum_{i=1}^{n_h}(\hat{Y}_{hi}-\bar{\hat{Y}}_h)^2}{n_h-1} \text{ or } \frac{\sum_{i=2}^{n_h}(\hat{Y}_{hi}-\hat{Y}_{h(i-1)})^2}{2(n_h-1)} \qquad f_{1h}=\frac{n_h}{N_h} \qquad f_{2hi}=\frac{d_{hi}}{D_{hi}} \qquad f_{3hij}=\frac{m_{hij}}{M_{hij}}$$

Table C.1. - Equations to calculate estimated variance of estimated harvest, catch, or effort with ***access-point creel surveys*** based on stratified multistage sampling designs used in Alaska when sampling fractions in higher stages are unequal (continued).

Four Stages (Location/Day/Period/Trip): (C.3)

$$var(\hat{Y}_h) = (1-f_{1h})N_h^2 \frac{S_{1h}^2}{n_h} + f_{1h}^{-1}\sum_{i=1}^{n_h}\left[D_{hi}^2(1-f_{2hi})\frac{S_{2hi}^2}{n_h}\right] + f_{1h}^{-1}\sum_{i=1}^{n_h}\left[f_{2hi}^{-1}\sum_{j=1}^{d_{hi}}\left[Q_{hij}^2(1-f_{3hij})\frac{s_{3hij}^2}{d_{hi}}\right]\right]$$

$$+ f_{1h}^{-1}\sum_{i=1}^{n_h}\left[f_{2hi}^{-1}\sum_{j=1}^{d_{hi}}\left[f_{3hij}^{-1}\sum_{k=1}^{q_{hij}}\left[M_{hijk}^2(1-f_{4hijk})\frac{s_{4hijk}^2}{m_{hijk}}\right]\right]\right]$$

$$s_{4hijk}^2 = \frac{\sum_{l=1}^{m_{hijk}}(y_{hijkl}-\bar{y}_{hijk})^2}{m_{hijk}-1}$$

$$S_{3hij}^2 = \frac{\sum_{k=1}^{q_{hij}}(\hat{Y}_{hijk}-\hat{\bar{Y}}_{hij})^2}{q_{hij}-1} \quad \text{or} \quad \frac{\sum_{k=2}^{q_{hij}}(\hat{Y}_{hijk}-\hat{Y}_{hij(k-1)})^2}{2(q_{hij}-1)}$$

$$S_{1h}^2 = \frac{\sum_{i=1}^{n_h}(\hat{Y}_{hi}-\hat{\bar{Y}}_h)^2}{n_h-1} \quad \text{or} \quad \frac{\sum_{i=2}^{n_h}(\hat{Y}_{hi}-\hat{Y}_{h(i-1)})^2}{2(n_h-1)}$$

$$S_{2hi}^2 = \frac{\sum_{j=1}^{d_{hi}}(\hat{Y}_{hij}-\hat{\bar{Y}}_{hi})^2}{d_{hi}-1} \quad \text{or} \quad \frac{\sum_{j=2}^{d_{hi}}(\hat{Y}_{hij}-\hat{Y}_{hi(j-1)})^2}{2(d_{hi}-1)}$$

$$f_{1h} = \frac{n_h}{N_h} \qquad f_{2hi} = \frac{d_{hi}}{D_{hi}} \qquad f_{3hij} = \frac{q_{hij}}{Q_{hij}} \qquad f_{4hijk} = \frac{m_{hijk}}{M_{hijk}}$$

APPENDIX D

Jackknife Estimator for Mean Harvest Rate in Roving Surveys

When information is gathered through "completed-trip" interviews, the best estimate of mean harvest rate is the ratio of means for harvest and trip length (Jones et al. 1995; Hoenig et al. 1997). For each sampled period:

$$\overline{hpue}^{u} = \frac{\hat{\bar{H}}}{\bar{e}} = \frac{\sum_{j=1}^{m} H_j}{\sum_{j=1}^{m} e_j} \tag{D.1}$$

where H_j is harvest by interviewed angler j, e_j is the length of his or her trip, and m is the number of anglers interviewed in the sampled period. The estimate $\overline{hpue}^{u}$ from Equation D.1 has an inherent sampling bias of order m^{-1} when anglers' harvests and trip-lengths are strongly correlated (Cochran 1977:162, 175). A jackknifed estimate has a smaller bias of order m^{-2} (Efron 1982:5-6) and is calculated in three steps:

$$hpue_j^{*} = \left[\frac{\sum_{\substack{k=1 \\ k \neq j}}^{m} H_k}{\sum_{\substack{k=1 \\ k \neq j}}^{m} e_k} \right] \tag{D.2a}$$

$$\overline{hpue}^{*} = \frac{\sum_{j=1}^{m} hpue_j^{*}}{m} \tag{D.2b}$$

$$\overline{hpue} \cong m(\overline{hpue}^{u} - \overline{hpue}^{*}) + \overline{hpue}^{*} \tag{D.2c}$$

Estimated variance of $\overline{hpue}$ is the estimated variance of $\overline{hpue}*$:

$$var(\overline{hpue}*) = \frac{m-1}{m} \sum_{j=1}^{m} (hpue_j^* - \overline{hpue}*)^2 \qquad \text{(D.3)}$$

Efron (1982:16-17) found this variance estimator to be slightly conservative with sample sizes of 10; Cochran (1977:179) found the same with sample sizes as low as 4.

Example D.1. In 1989, a roving-access, two-stage survey near the Elliot Highway Bridge across the Chatanika River was used to estimate harvest of whitefish in a spear fishery from early September through mid October (see Examples 2.6-8 for more details on the fishery and the creel survey in 1989). Fishing days (nights) were first-stage sampling units with each night comprised a single sampling period. Only

Table D.1 - Demonstration of the jackknife procedure to estimate mean harvest rate from 16 interviews of anglers exiting the fishery for whitefish near the Elliot Highway Bridge on the Chatanika River during the night of 10 October, 1989. Each row corresponds to an individual interview.

Data		j = 1		j = 2		j = 3		j = 4		j = 5		....	j = 16	
H_j	e_j	H_j	e_j	H_j	e_j	H_j	e_j	H_j	e_j	H_j	e_j	....	H_j	e_j
15	3.0	omit		15	3.0	15	3.0	15	3.0	15	3.0	....	15	3.0
22	3.0	22	3.0	omit		22	3.0	22	3.0	22	3.0	....	22	3.0
11	2.0	11	2.0	11	2.0	omit		11	2.0	11	2.0	....	11	2.0
8	2.0	8	2.0	8	2.0	8	2.0	omit		8	2.0	....	8	2.0
13	2.0	13	2.0	13	2.0	13	2.0	13	2.0	omit		....	13	2.0
11	2.0	11	2.0	11	2.0	11	2.0	11	2.0	11	2.0	....	11	2.0
11	2.0	11	2.0	11	2.0	11	2.0	11	2.0	11	2.0	....	11	2.0
11	2.0	11	2.0	11	2.0	11	2.0	11	2.0	11	2.0	....	11	2.0
14	2.0	14	2.0	14	2.0	14	2.0	14	2.0	14	2.0	....	14	2.0
3	1.5	3	1.5	3	1.5	3	1.5	3	1.5	3	1.5	....	3	1.5
3	1.5	3	1.5	3	1.5	3	1.5	3	1.5	3	1.5	....	3	1.5
7	1.0	7	1.0	7	1.0	7	1.0	7	1.0	7	1.0	....	7	1.0
4	1.0	4	1.0	4	1.0	4	1.0	4	1.0	4	1.0	....	4	1.0
5	1.0	5	1.0	5	1.0	5	1.0	5	1.0	5	1.0	....	5	1.0
12	3.0	12	3.0	12	3.0	12	3.0	12	3.0	12	3.0	....	12	3.0
15	3.0	15	3.0	15	3.0	15	3.0	15	3.0	15	3.0	....	omit	
Sums =														
165	32.0	150	29.0	143	29.0	154	30.0	157	30.0	152	30.0		150	29.0
$hpue_j^*$ =		5.1724		4.9310		5.1333		5.2333		5.0667		....	5.1724	

interviews of anglers leaving the fishery on 10 October were used for this example (Table D.1). Sixteen persons were interviewed as they left the river at the campground off the Elliot Highway that night (m = 16). Each pair of columns in Table D.1 has a row of omitted data corresponding to a different angler. For all 16 anglers, $\overline{hpue}^{u}$ = 5.1563 [=165/32] from Equation D.1, and $\overline{hpue}*$ is 5.1558 [=(5.1724 + 4.9310 + 5.1333 + 5.2333 + 5.0667 + ... + 5.1724)/16] from Equation D.2. The bias-corrected estimate of mean harvest rate $\overline{hpue}$ is 5.1638 [=16(5.1563 - 5.1558) + 5.1558], not appreciably different than the uncorrected statistic of 5.1563. As a demonstration of the sensitivity of first-order bias to sample size (bias related to m_i^{-1}), biases were calculated for all possible unique subsamples of 10, 5, 4, and 3 interviews from data collected on 10 October (Figure D.1). Even though harvest and trip length is strongly correlated ($|\rho|>0.8$) for this sampled period, first-order bias was negligible in simulations with four or more interviews. Estimated variance from Equations D.3 and 2.11b are 0.1405 and 0.1383, respectively.

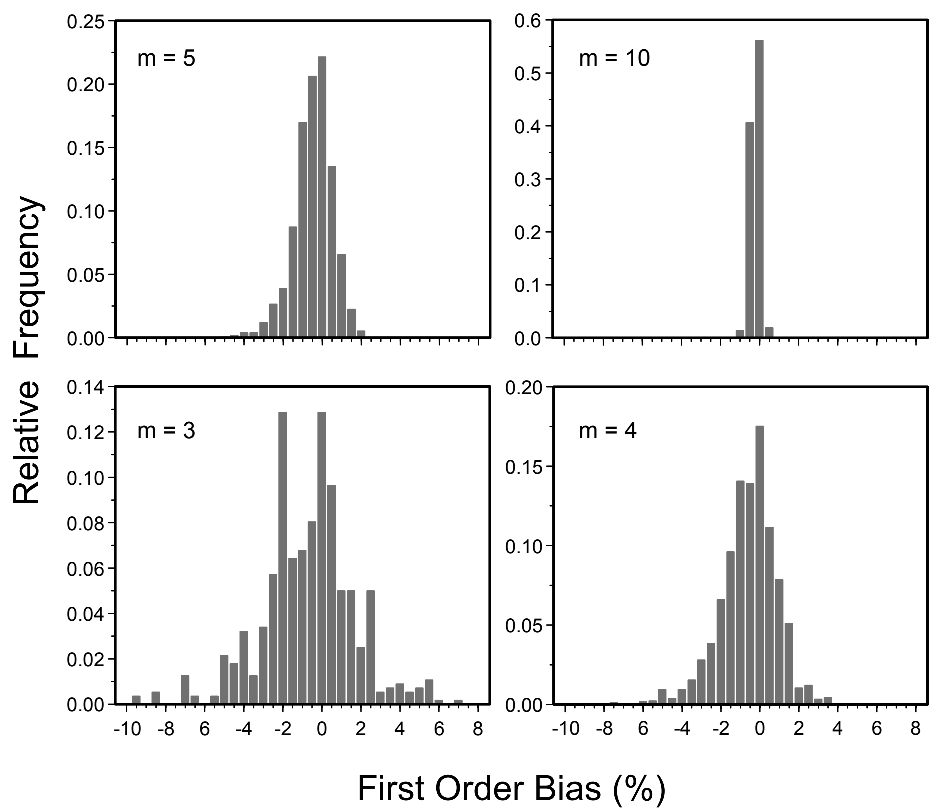


Figure D.1. Frequency of relative first-order bias in mean harvest rates when estimated as the ratio of means for all possible subsamples of size 3, 4, 5, and 10 from 16 fishers interviewed on 10 October, 1989 near the Elliot Highway Bridge on the Chatanika River, Alaska, as determined through jackknifing.

APPENDIX E

Covariance among Post-Stratified Estimates of Harvest in Roving Surveys

Because both plug-in substitutions for missing data (see Appendix F) and combined post-stratified estimates of harvests are linear combinations; sample variances, corrections for plug-in substitutions in those sample variances, and covariances between counts of post-stratified anglers are additive within the estimate of within-period variance. In the instance where anglers are post-stratified into two groups (second subscript), and estimates of harvest rates for both groups in one sampled period are plugged-in for missing estimates in other periods (first subscript represents period):

$$g = \overline{hpue}_{11}\hat{E}_{11} + \overline{hpue}_{12}\hat{E}_{12} + (a_{11}\overline{hpue}_{11})\hat{E}_{21} + (a_{12}\overline{hpue}_{12})\hat{E}_{22} + \ldots$$

with $a_{11} = a_{12} = 1$ and $\overline{hpue}_{21} =$ (?) and $\overline{hpue}_{22} =$ (?). For the sake of brevity, *v* is used instead of *var* to denote estimated variance. An unbiased estimate of variance *v*(*g*) of the function *g* can be found through the method of moments:

$$\mathrm{E}[g^2] - (\mathrm{E}[g])^2 =$$

$$= \mathrm{E}[\{\overline{hpue}_{11}\hat{E}_{11} + \overline{hpue}_{12}\hat{E}_{12} + (a_{11}\overline{hpue}_{11})\hat{E}_{21} + (a_{12}\overline{hpue}_{12})\hat{E}_{22} + \ldots\}^2] - \left(\mathrm{E}[\overline{hpue}_{11}\hat{E}_{11} + \overline{hpue}_{12}\hat{E}_{12} + (a_{11}\overline{hpue}_{11})\hat{E}_{21} + (a_{12}\overline{hpue}_{12})\hat{E}_{22} + \ldots]\right)^2$$

$$\begin{aligned}
= {} & \overline{hpue}_{11}^2\hat{E}_{11}^2 + 2\overline{hpue}_{11}\overline{hpue}_{12}\mathrm{E}[\hat{E}_{11}\hat{E}_{12}] + 2a_{11}\overline{hpue}_{11}^2\hat{E}_{11}\hat{E}_{21} + 2\overline{hpue}_{11}(a_{12}\overline{hpue}_{12})\hat{E}_{11}\hat{E}_{22} + \\
& \overline{hpue}_{12}^2\hat{E}_{12}^2 + 2\overline{hpue}_{12}(a_{11}\overline{hpue}_{11})\hat{E}_{12}\hat{E}_{21} + 2a_{12}\overline{hpue}_{12}^2\hat{E}_{12}\hat{E}_{22} + a_{11}^2\overline{hpue}_{11}^2\hat{E}_{21}^2 + \\
& + 2(a_{11}\overline{hpue}_{11})(a_{12}\overline{hpue}_{12})\mathrm{E}[\hat{E}_{21}\hat{E}_{22}] + a_{12}^2\overline{hpue}_{12}^2\hat{E}_{22}^2 + \ldots \\
& - \Big(\mathrm{E}[\overline{hpue}_{11}]^2\mathrm{E}[\hat{E}_{11}]^2 + 2\overline{hpue}_{11}\overline{hpue}_{12}\hat{E}_{11}\hat{E}_{12} + 2a_{11}\mathrm{E}[\overline{hpue}_{11}]^2\hat{E}_{11}\hat{E}_{21} + \\
& 2\overline{hpue}_{11}(a_{12}\overline{hpue}_{12})\hat{E}_{11}\hat{E}_{22} + \mathrm{E}[\overline{hpue}_{12}]^2\mathrm{E}[\hat{E}_{12}]^2 + 2\overline{hpue}_{12}(a_{11}\overline{hpue}_{11})\hat{E}_{12}\hat{E}_{21} + \\
& 2a_{12}\mathrm{E}[\overline{hpue}_{12}]^2\hat{E}_{12}\hat{E}_{22} + a_{11}^2\mathrm{E}[\overline{hpue}_{11}]^2\mathrm{E}[\hat{E}_{21}]^2 + \\
& 2(a_{11}\overline{hpue}_{11})(a_{12}\overline{hpue}_{12})\hat{E}_{21}\hat{E}_{22} + \\
& a_{12}^2\mathrm{E}[\overline{hpue}_{12}]^2\mathrm{E}[\hat{E}_{22}]^2 + \ldots\Big)
\end{aligned}$$

where $\mathrm{E}[\ ]$ is the expectation of the variable in the brackets. Substituting $y^2 - v[y]$ for $\mathrm{E}[y]^2$ in the equation above, canceling like terms, remembering that only estimates of post-stratified fishing effort in the same period are dependent, and rearranging produces:

$$
\begin{aligned}
v(g) = \quad & \overline{hpue}_{11}^{2} v(\hat{E}_{11}) + v(\overline{hpue}_{11})\hat{E}_{11}^{2} - v(\hat{E}_{11})v(\overline{hpue}_{11}) + \\
& \overline{hpue}_{12}^{2} v(\hat{E}_{12}) + v(\overline{hpue}_{12})\hat{E}_{12} - v(\hat{E}_{12})v(\overline{hpue}_{12}) + \\
& a_{11}^{2}\overline{hpue}_{11}^{2} v(\hat{E}_{21}) + a_{11}^{2} v(\overline{hpue}_{11})\hat{E}_{21}^{2} - v(\hat{E}_{21})a_{11}^{2} v(\overline{hpue}_{11}) + \\
& a_{12}^{2}\overline{hpue}_{12}^{2} v(\hat{E}_{22}) + a_{12}^{2} v(\overline{hpue}_{12})\hat{E}_{22}^{2} - v(\hat{E}_{22})a_{12}^{2} v(\overline{hpue}_{12}) + \\
& 2a_{11}v(\overline{hpue}_{11})\hat{E}_{11}\hat{E}_{21} + 2a_{12}v(\overline{hpue}_{12})\hat{E}_{12}\hat{E}_{22} + \\
& 2\overline{hpue}_{11}\,\overline{hpue}_{12}(\mathrm{E}[\hat{E}_{11}\hat{E}_{12}] - \hat{E}_{11}\hat{E}_{12}) + \\
& 2(a_{11}\overline{hpue}_{11})(a_{12}\overline{hpue}_{12})(\mathrm{E}[\hat{E}_{21}\hat{E}_{22}] - \hat{E}_{21}\hat{E}_{22}) + \ldots
\end{aligned}
$$

Note according to methods in Goodman (1960) the first four lines in the rhs of the equation above reduce to $\sum_{i=1}^{d}\sum_{j=1}^{L} v(\hat{H}_{ij})$, the fifth line reduces to the correction for using plug-in substitutions for harvest rate in both post-stratified groups (see Appendix F), and remembering that plug-in substitutions are $\overline{hpue}_{21} \leftarrow a_{11}\overline{hpue}_{11}$ and $\overline{hpue}_{22} \leftarrow a_{12}\overline{hpue}_{12}$, the last two lines equal $2\sum_{i=1}^{d}\sum_{j=1}^{L}\sum_{k>j}^{L} cov(\hat{H}_{ij}, \hat{H}_{ik})$ $= 2\sum\sum\sum \overline{hpue}_{ij}\,\overline{hpue}_{ik}\, cov(\hat{E}_{ij}, \hat{E}_{ik})$ $= 2\sum\sum\sum \overline{hpue}_{ij}\,\overline{hpue}_{ik} T^2 \operatorname{cov}(\bar{x}_{ij}, \bar{x}_{ik})$.

What happens when only anglers from one group are interviewed in a sampled period but anglers in both groups are counted? In the instance where anglers are post-stratified into two groups, and an estimate of harvest rate for only one group in one sampled period is plugged-in for a missing estimate in another period:

$$g = \overline{hpue}_{11}\hat{E}_{11} + \overline{hpue}_{12}\hat{E}_{12} + (a_{11}\overline{hpue}_{11})\hat{E}_{21} + \overline{hpue}_{22}\hat{E}_{22} + \ldots$$

with $a_{11} = 1$ and $\overline{hpue}_{21} =$ (?). An unbiased estimate $v(g)$ is again found through expectations:

$$\mathrm{E}[g^2] - (\mathrm{E}[g])^2 =$$

$$
\begin{aligned}
= \mathrm{E}[\{\overline{hpue}_{11}\hat{E}_{11} + \overline{hpue}_{12}\hat{E}_{12} + (a_{11}\overline{hpue}_{11})\hat{E}_{21} + \overline{hpue}_{22}\hat{E}_{22} + \ldots\}^2] - \\
\left(\mathrm{E}[\overline{hpue}_{11}\hat{E}_{11} + \overline{hpue}_{12}\hat{E}_{12} + (a_{11}\overline{hpue}_{11})\hat{E}_{21} + \overline{hpue}_{22}\hat{E}_{22} + \ldots]\right)^2
\end{aligned}
$$

$$
\begin{aligned}
= {} & \overline{hpue}_{11}^{2}\hat{E}_{11}^{2} + 2\overline{hpue}_{11}\overline{hpue}_{12}\mathrm{E}[\hat{E}_{11}\hat{E}_{12}] + 2a_{11}\overline{hpue}_{11}^{2}\hat{E}_{11}\hat{E}_{21} + 2\overline{hpue}_{11}\overline{hpue}_{22}\hat{E}_{11}\hat{E}_{22} + \\
& \overline{hpue}_{12}^{2}\hat{E}_{12}^{2} + 2\overline{hpue}_{12}(a_{11}\overline{hpue}_{11})\hat{E}_{12}\hat{E}_{21} + 2\overline{hpue}_{12}\overline{hpue}_{22}\hat{E}_{12}\hat{E}_{22} + a_{11}^{2}\overline{hpue}_{11}^{2}\hat{E}_{21}^{2} + \\
& + 2(a_{11}\overline{hpue}_{11})\overline{hpue}_{22}\mathrm{E}[\hat{E}_{21}\hat{E}_{22}] + \overline{hpue}_{22}^{2}\hat{E}_{22}^{2} + \ldots \\
& -\Big(\mathrm{E}[\overline{hpue}_{11}]^{2}\mathrm{E}[\hat{E}_{11}]^{2} + 2\overline{hpue}_{11}\overline{hpue}_{12}\hat{E}_{11}\hat{E}_{12} + 2a_{11}\mathrm{E}[\overline{hpue}_{11}]^{2}\hat{E}_{11}\hat{E}_{21} + \\
& 2\overline{hpue}_{11}\overline{hpue}_{22}\hat{E}_{11}\hat{E}_{22} + \mathrm{E}[\overline{hpue}_{12}]^{2}\mathrm{E}[\hat{E}_{12}]^{2} + 2\overline{hpue}_{22}(a_{11}\overline{hpue}_{11})\hat{E}_{12}\hat{E}_{21} + \\
& 2\overline{hpue}_{12}\overline{hpue}_{22}\hat{E}_{12}\hat{E}_{22} + a_{11}^{2}\mathrm{E}[\overline{hpue}_{11}]^{2}\mathrm{E}[\hat{E}_{21}]^{2} + \\
& 2(a_{11}\overline{hpue}_{11})\overline{hpue}_{22}\hat{E}_{21}\hat{E}_{22} + \\
& \mathrm{E}[\overline{hpue}_{22}]^{2}\mathrm{E}[\hat{E}_{22}]^{2} + \ldots\Big)
\end{aligned}
$$

Substituting $y^2 - v[y]$ for $\mathrm{E}[y]^2$ in the equation above, canceling like terms, and rearranging produces:

$$
\begin{aligned}
v(g) = {} & \overline{hpue}_{11}^{2}v(\hat{E}_{11}) + v(\overline{hpue}_{11})\hat{E}_{11}^{2} - v(\hat{E}_{11})v(\overline{hpue}_{11}) + \\
& \overline{hpue}_{12}^{2}v(\hat{E}_{12}) + v(\overline{hpue}_{12})\hat{E}_{12} - v(\hat{E}_{12})v(\overline{hpue}_{12}) + \\
& a_{11}^{2}\overline{hpue}_{11}^{2}v(\hat{E}_{21}) + a_{11}^{2}v(\overline{hpue}_{11})\hat{E}_{21}^{2} - v(\hat{E}_{21})a_{11}^{2}v(\overline{hpue}_{11}) + \\
& \overline{hpue}_{22}^{2}v(\hat{E}_{22}) + v(\overline{hpue}_{22})\hat{E}_{22}^{2} - v(\hat{E}_{22})v(\overline{hpue}_{22}) + \\
& 2a_{11}v(\overline{hpue}_{11})\hat{E}_{11}\hat{E}_{21} + \\
& 2\overline{hpue}_{11}\overline{hpue}_{12}(\mathrm{E}[\hat{E}_{11}\hat{E}_{12}] - \hat{E}_{11}\hat{E}_{12}) + \\
& 2a_{11}\overline{hpue}_{11}\overline{hpue}_{22}(\mathrm{E}[\hat{E}_{21}\hat{E}_{22}] - \hat{E}_{21}\hat{E}_{22}) + \ldots
\end{aligned}
$$

Note that the first four lines in the rhs of the equation above reduce to $\sum_{i=1}^{d}\sum_{j=1}^{L} v(\hat{H}_{ij})$, the fifth line reduces to the correction for using plug-in substitutions for harvest rate in both post-stratified groups, and remembering the plug-in substitution $\overline{hpue}_{21} \leftarrow a_{11}\overline{hpue}_{11}$, the last two lines reduce to $2\sum_{i=1}^{d}\sum_{j=1}^{L}\sum_{k>j}^{L} \overline{hpue}_{ij}\,\overline{hpue}_{ik}\,T^2\,cov(\bar{x}_{ij}, \bar{x}_{ik})$. These ideas are developed further in Bernard et al. (1998).

APPENDIX F

Roving Surveys with Missing Data on Harvest Rates

The method of moments can be used to produce an unbiased estimate of variance for the linear combination $\sum_{i=1}^{d} \hat{H}_i$ where some of the $\hat{H}_i$ have plug-in substitutions for missing estimates $\overline{hpue}$ of mean harvest rate (see Bernard et al. 1998). If this linear combination with its plug-in substitutions is renamed *g* (*g* is a function here, not a category), then an unbiased estimate of variance *v*(*g*) is $\mathrm{E}[g^2] - (\mathrm{E}[g])^2$ where E[*y*] is the statistical expectation of *y* (for the sake of brevity, *v* is used instead of *var* to denote estimated variance). The simplest substitution is a single statistic plugged-in for a single missing datum $g = \overline{hpue}_1 \hat{E}_1 + (a_1 \overline{hpue}_1)\hat{E}_2 + \ldots$ with $a_1 = 1$. Any sampled periods not involved in forming plug-in substitutes that do have unbiased estimates of harvest rate will have independent estimates of harvest with sample variances that are part of *v*(*g*), but that are not involved in estimating the correction for plug-in substitutions. Hereafter, presence of statistics in *g* from such independently sampled periods are implied with the understanding that their contribution to variance is in addition to variance arising from the derivations listed below. With that understanding in mind, sample variance *v*(*g*) for a single substitution for a missing datum is:

$$\mathrm{E}[g^2] - (\mathrm{E}[g])^2 = \mathrm{E}[\{\overline{hpue}_1 \hat{E}_1 + (a_1 \overline{hpue}_1)\hat{E}_2 + \ldots\}^2] - \left(\mathrm{E}[\overline{hpue}_1 \hat{E}_1 + (a_1 \overline{hpue}_1)\hat{E}_2 + \ldots]\right)^2$$

$$= \overline{hpue}_1^2 \hat{E}_1^2 + a_1^2 \overline{hpue}_1^2 \hat{E}_2^2 + 2a_1 \overline{hpue}_1^2 \hat{E}_1 \hat{E}_2 - \mathrm{E}[\overline{hpue}_1]^2 \mathrm{E}[\hat{E}_1]^2 - a_1^2 \mathrm{E}[\overline{hpue}_1]^2 \mathrm{E}[\hat{E}_2]^2 - 2a_1 \mathrm{E}[\overline{hpue}_1]^2 \mathrm{E}[\hat{E}_1]\mathrm{E}[\hat{E}_2] + \ldots$$

Substituting $y^2 - v[y]$ for $\mathrm{E}[y]^2$ in the equation above, canceling like terms, and rearranging produces:

$$\begin{aligned} v(g) = \; & \overline{hpue}_1^2 v(\hat{E}_1) + v(\overline{hpue}_1)\hat{E}_1^2 - v(\hat{E}_1)v(\overline{hpue}_1) + \\ & a_1^2 \overline{hpue}_1^2 v(\hat{E}_2) + a_1^2 v(\overline{hpue}_1)\hat{E}_2^2 - a_1^2 v(\hat{E}_2)v(\overline{hpue}_1) + \\ & 2a_1 v(\overline{hpue}_1)\hat{E}_1 \hat{E}_2 + \ldots \end{aligned}$$

Note that the first two lines in the right-hand side (rhs) of the equation above correspond to the calculation of variance with the plug-in substitutions alone following the methods in Goodman (1960) for estimating variance of the product of two independent variables. However, in this situation, estimates of harvest rate are somewhat dependent because some rates are "shared" across periods. The expression in the third line acts like a correction for this dependence when using specific plug-in substitutions. If in another situation, a plug-in statistic is substituted for *two* missing data such that $g = \overline{hpue}_1 \hat{E}_1 + (a_1 \overline{hpue}_1)\hat{E}_2 + (a_1 \overline{hpue}_1)\hat{E}_3 + \ldots$ again with $a_1 = 1$, the unbiased sample variance is:

$$\begin{aligned} v(g) = \quad & \overline{hpue}_1^2 v(\hat{E}_1) + v(\overline{hpue}_1)\hat{E}_1^2 - v(\hat{E}_1)v(\overline{hpue}_1) + \\ & a_1^2 \overline{hpue}_1^2 v(\hat{E}_2) + a_1^2 v(\overline{hpue}_1)\hat{E}_2^2 - a_1^2 v(\overline{hpue}_1)v(\hat{E}_2) + \\ & a_1^2 \overline{hpue}_1^2 v(\hat{E}_3) + a_1^2 v(\overline{hpue}_1)\hat{E}_3^2 - a_1^2 v(\overline{hpue}_1)v(\hat{E}_3) + \\ & 2a_1 v(\overline{hpue}_1)\left(\hat{E}_1\hat{E}_2 + \hat{E}_1\hat{E}_3 + a_1\hat{E}_2\hat{E}_3\right) + \ldots \end{aligned}$$

with the fourth line of the rhs in the above equation again acting as a correction for dependence from using plug-in substitutions. If a single statistic is substituted for *three* missing data, the correction can be shown with the method of moments to be:

$$Correction = 2a_1 v(\overline{hpue}_1)\left(\hat{E}_1\hat{E}_2 + \hat{E}_1\hat{E}_3 + \hat{E}_1\hat{E}_4 + a_1\hat{E}_2\hat{E}_3 + a_1\hat{E}_2\hat{E}_4 + a_1\hat{E}_3\hat{E}_4\right)$$

If a linear combination of *two* statistics, for instance an average, is substituted for *one* missing datum such that $g = \overline{hpue}_1\hat{E}_1 + \overline{hpue}_2\hat{E}_2 + (a_1\overline{hpue}_1 + a_2\overline{hpue}_2)\hat{E}_3 + \ldots$ this time with $a_1 = a_2 = 1/2$:

$$\begin{aligned} v(g) = \quad & \overline{hpue}_1^2 v(\hat{E}_1) + v(\overline{hpue}_1)\hat{E}_1^2 - v(\hat{E}_1)v(\overline{hpue}_1) + \\ & \overline{hpue}_2^2 v(\hat{E}_2) + v(\overline{hpue}_2)\hat{E}_2^2 - v(\hat{E}_2)v(\overline{hpue}_2) + \\ & \overline{\overline{hpue}}^2 v(\hat{E}_3) + v(\overline{\overline{hpue}})\hat{E}_3^2 - v(\hat{E}_3)v(\overline{\overline{hpue}}) + \\ & 2a_1 v(\overline{hpue}_1)\hat{E}_1\hat{E}_3 + 2a_2 v(\overline{hpue}_2)\hat{E}_2\hat{E}_3 + \ldots \end{aligned}$$

where $\overline{\overline{hpue}} = a_1\overline{hpue}_1 + a_2\overline{hpue}_2$ and $v(\overline{\overline{hpue}}) = a_1^2 v(\overline{hpue}_1) + a_2^2 v(\overline{hpue}_2)$. The fourth line in the rhs of the equation above acts as the correction in sample variance for using the plug-in substitute $\overline{\overline{hpue}}$. If a linear combination of *three* statistics (again named $\overline{\overline{hpue}}$) is substituted for *three* missing data such that $g = \overline{hpue}_1\hat{E}_1 + \overline{hpue}_2\hat{E}_2 + \overline{hpue}_3\hat{E}_3 + \overline{\overline{hpue}}\hat{E}_4 + \overline{\overline{hpue}}\hat{E}_5 + \overline{\overline{hpue}}\hat{E}_6 + \ldots$, the correction is:

$$\begin{aligned} Correction = \quad & 2a_1 v(\overline{hpue}_1)\left(\hat{E}_1\hat{E}_4 + \hat{E}_1\hat{E}_5 + \hat{E}_1\hat{E}_6 + a_1\hat{E}_4\hat{E}_5 + a_1\hat{E}_4\hat{E}_6 + a_1\hat{E}_5\hat{E}_6\right) + \\ & 2a_2 v(\overline{hpue}_2)\left(\hat{E}_2\hat{E}_4 + \hat{E}_2\hat{E}_5 + \hat{E}_2\hat{E}_6 + a_2\hat{E}_4\hat{E}_5 + a_2\hat{E}_4\hat{E}_6 + a_2\hat{E}_5\hat{E}_6\right) + \end{aligned}$$

$$2a_3 v(\overline{hpue_3})\left(\hat{E}_3\hat{E}_4 + \hat{E}_3\hat{E}_5 + \hat{E}_3\hat{E}_6 + a_3\hat{E}_4\hat{E}_5 + a_3\hat{E}_4\hat{E}_6 + a_3\hat{E}_5\hat{E}_6\right) + \ldots$$

where $\overline{\overline{hpue}} = a_1\overline{hpue_1} + a_2\overline{hpue_2} + a_3\overline{hpue_3}$ and $v(\overline{\overline{hpue}}) = a_1^2 v(\overline{hpue_1}) + a_2^2 v(\overline{hpue_2}) + a_3^2 v(\overline{hpue_3})$ with $a_1 = a_2 = a_3 = 1/3$. Derivation of corrections for situations used in the demonstrations above follow a pattern that can be used to form a general formula for calculating a correction for dependence caused by plug-in substitutions. If $a_i = 0$ for any sampled period whose statistics are not involved in forming the plug-in substitute $\overline{\overline{hpue}}$, and if there is a variable flag b_j such that $b_j = 1$ if a plug-in estimate for harvest rate was used for sampled period j or $b_j = 0$ if not, the general expression for the correction for dependence caused by using plug-in substitutions is:

$$Correction = \sum_{i=1}^{d} 2a_i v(\overline{hpue_i})\left[\hat{E}_i \sum_{j=1}^{d} b_j \hat{E}_j + a_i \sum_{j=1}^{d} b_j \hat{E}_j \left(\sum_{k>j}^{d} b_k \hat{E}_k \right)\right]$$

This general correction must be multiplied by D/d to represent a correction in the within-period sample variance and therefore to represent a correction in $v(\hat{H})$ for using plug-in substitutions.

APPENDIX G

Relationships among Equations to Estimate Relative Harvest Composition

When all fish in a cohort carry tags and all tagged fish in a sample are recognized, an estimate of the fraction p of a harvest comprised of fish from that cohort and its estimated variance are (from Cochran 1977:51-52):

$$\hat{p} = \frac{y}{\eta} \qquad\qquad var(\hat{p}) = (1-f)\frac{\hat{p}(1-\hat{p})}{\eta - 1}$$

where η is sample size, y the number in the sample from the cohort, and f the fraction of the harvest sampled. From Bernard and Clark (1996: Equations 8,10), an estimate of p and an exact estimate for the variance of $\hat{p}$ when not all members of a cohort carry tags and not all tags are recognized are:

$$\hat{p} = \frac{y}{\lambda\eta}$$

$$var(\hat{p}) = \left(\frac{1}{D}\frac{\eta}{\eta-1}\right)\left(\frac{\hat{p}}{\lambda\eta} - \left[1 - D\frac{\eta-1}{\eta}\right]\hat{p}^2\right)\left(\frac{1-y(1-C)}{E-y(E-C)} - \lambda f \theta \frac{C}{E-y(E-C)}\right)$$

where λ is a compound probability that a tag will be recognized, and D, C, and E are collected constants. When all fish carry tags (θ = 1) and all tags are recognized (λ = 1), these two sets of equations should be equivalent, which is obviously so for the estimates. When θ = 1 and λ = 1, D = 1 and C = 1, the variance equation immediately above reduces to:

$$var(\hat{p}) = \left(\frac{\eta}{\eta-1}\right)\left(\frac{\hat{p}}{\eta} - \left[1 - \frac{\eta-1}{\eta}\right]\hat{p}^2\right)\left(\frac{1}{E-y(E-1)} - f\frac{1}{E-y(E-1)}\right)$$

$$= \frac{\hat{p}(1-\hat{p})}{\eta-1}\left(\frac{1}{E-y(E-1)} - f\frac{1}{E-y(E-1)}\right)$$

Because $E = H/(H-1)$, E$\rightarrow$ 1 as H $\rightarrow \infty$. Therefore, when harvest is large:

$$var(\hat{p}) \cong \frac{\hat{p}(1-\hat{p})}{\eta-1}(1-f)$$

Note that the equation above can be rewritten as:

$$var(\hat{p}) \cong \frac{\hat{p}}{\eta - 1}(1 - f) - \frac{\hat{p}^2}{\eta - 1}(1 - f)$$

When p is very small, the second term represents a negligible part of $var(\hat{p})$. If ignored:

$$var(\hat{p}) \cong \frac{\hat{p}}{\eta - 1}(1 - f)$$

If as sample size becomes large, $\eta \cong \eta - 1$. With that substitution:

$$var(\hat{p}) \cong \frac{\hat{p}}{\eta}(1 - f)$$

which is equivalent to Equation 3.9b when $\theta = 1$ and $\lambda = 1$. Small p and large sample sizes are hallmarks of programs to estimate harvest of salmon from cohorts with coded wire tags (Bernard et al. 1996).

APPENDIX H

Source Code for the QuickBasic™ Program "DELTA.BAS"

```
10 CLS
20 RANDOMIZE
30 REM ----------------------------------Inputs
35 INPUT "Name the file to be opened for output"; NAME$
36 OPEN NAME$ FOR OUTPUT AS #1
37 INPUT "What is the weighting factor for fishing effort by affected anglers"; OMEGA
39 IF OMEGA = 0 THEN OMEGA = 1
40 INPUT "What is the current bag limit (Bo)"; B0
50 DIM P(B0), PC(B0), PS(B0), DELTASUM(B0), DELTASQ(B0)
115 INPUT "What fraction of angler-days ended with no harvest"; P(0)
117 PC(0) = P(0): PC(B0) = 1
120 FOR I = 1 TO B0 - 1
130 PRINT "What fraction of angler-days ended with a harvest of"; I;
140 INPUT "fish"; P(I)
150 PC(I) = PC(I - 1) + P(I)
160 NEXT I
165 P(B0) = 1 - PC(B0 - 1)
166 IF P(B0) >= 0 GOTO 170
167 PRINT "--------ERROR IN ENTERING FRACTIONS; TRY AGAIN--------"
168 GOTO 115
170 INPUT "How many angler-days were sampled to estimate fractions"; N
175 PUNIT = 1 / N
180 REM---------------------------------Simulate
185 INPUT "How many simulations do you want"; NSIMS
190 FOR I = 1 TO NSIMS
200 FOR G = 0 TO B0
210 PS(G) = 0
220 NEXT G
225 REM -------------------------------Simulate a sample
230 FOR J = 1 TO N
240 X = RND
250 FOR G = 0 TO B0
260 IF X > PC(G) GOTO 290
270 PS(G) = PS(G) + PUNIT
280 GOTO 300
290 NEXT G
300 NEXT J

310 REM--------------------------------Calculate denom for delta
```

```
315 DENOM = 0
320 FOR G = 1 TO B0
330 DENOM = DENOM + G * PS(G)
340 NEXT G
350 REM --------------------------------Calculate num for delta
360 FOR B = 1 TO B0 - 1
370 NUM = 0
380 FOR G = B + 1 TO B0
390 NUM = NUM + (G - B * OMEGA) * PS(G)
400 NEXT G
410 REM --------------------------------Tally Statistics
415 D = NUM / DENOM
420 DELTASUM(B) = DELTASUM(B) + D
430 DELTASQ(B) = DELTASQ(B) + D * D
440 NEXT B
450 NEXT I
460 REM ---------------------------------Calculate Anticipated Statistics
461 DENOM = 0
462 FOR G = 1 TO B0
464 DENOM = DENOM + G * P(G)
466 NEXT G
468 PRINT "------------------------------------------------------------"
470 PRINT "Daily bag limit/Anticipated delta/Simulated delta/Estimated SE"
475 FOR B = 1 TO B0 - 1
477 NUM = 0
478 FOR G = B + 1 TO B0
480 NUM = NUM + (G - B * OMEGA) * P(G)
482 NEXT G
483 DELTAA = NUM / DENOM
484 REM ---------------------------------Output Statistics
488 DELTAS = DELTASUM(B) / NSIMS
490 DELTAVS = (DELTASQ(B) - DELTASUM(B) * DELTASUM(B) / NSIMS) / (NSIMS - 1)
504 PRINT B;
505 PRINT USING "###.###"; DELTAA; DELTAS;
506 PRINT USING "###.#######"; SQR(DELTAVS)
508 PRINT #1, B;
509 PRINT #1, USING "###.###"; DELTAA; DELTAS;
510 PRINT #1, USING "###.#######"; SQR(DELTAVS)
520 NEXT B
530 END
```

GLOSSARY

Access point: A geographical location through which anglers physically enter and exit fishing grounds.

Access-point survey: A sampling program designed to estimate harvest, catch, or fishing effort for a fishery by expanding information gained by interviewing and counting anglers exiting the fishery at a subset of access points.

Angle-trip: A unit of fishing effort that begins with an angler entering a fishery through an access point and ends with his or her exit through the same or a different access point.

Angler-day: A unit of fishing effort corresponding to a single angler fishing for any amount of time in a fishery during a day regardless of the number of trips taken or the length of each trip that day.

Angler success: The relative distribution of anglers that harvested 0, 1, 2, 3, etc. fish during a day's fishing.

Avidity bias: Bias in estimates that arise when anglers who participate more frequently in a fishery (the more avid anglers) have a greater or lesser probability of being sampled than anglers who participate less.

Bag limit: The legal number of fish that can be harvested during an angler-day of fishing.

Basic sampling unit: The entity that is interviewed or measured to produce a datum (i.e., number fish harvested in a creel, number of anglers fishing, or number fish harvested by age).

Catch: The number of fish caught.

Catch cards: Cards given to actively fishing anglers with instructions to fill out the cards on completion of their fishing trip and to return the cards to creel survey personnel.

Catch rate: The number of fish caught per hour (a type of CPUE).

Cohort: A group of fish distinguished by having a subset of their members tagged or marked.

Completed-trip interview: Interviews of anglers at the end of their fishing trip, usually conducted as they exit a fishery.

Component of estimated variance: A sample variance for harvest, catch, or fishing effort in a multistage sampling design can be arithmetically divided into parts with each part representing all variation due to a separate sampling stage. A component of estimated variance is one of those parts.

Count: In a roving survey, a pass through the fishing grounds counting all persons actively fishing.

CPUE: Catch per unit of fishing effort, usually a number of fish caught in a fishing trip divided by the length of the fishing trip in hours.

Dimension of stratification: A temporal, spatial, or behavioral attribute of sampling units used to group them into strata. Each value of the attribute is considered an element of the dimension.

Day-of-week (DOW) stratification: A dimension of stratification based on grouping basic sampling units according to when they could be sampled during the week. Usually there are two elements in this dimension: weekend days (Saturday and Sunday) and weekdays (usually Monday through Friday).

Fishing Day: A period of continuous fishing by at least a small number of anglers. A fishing day is usually (but not always) contained within a calendar day with a hiatus in fishing between days occurring during the darkness of night.

Fishing effort: Time fishing measured in trips, days, or hours.

Full stratification: Every stratum represents only one element in each dimension of stratification. If there are K dimensions of stratification with k_1, k_2 .... k_K elements in each dimension, full stratification would imply there be $\prod_{i=1}^{K} k_i$ strata.

Harvest: The number of fish caught and kept.

Harvest composition: Harvest of fish divided into groups according to some attribute such as age, sex, or species.

Harvest rate: The number of fish caught and kept per hour (a type of HPUE).

HPUE: Harvest per unit of fishing effort, usually a number of fish caught in a fishing trip divided by the length of the fishing trip in hours.

Impute: Using a "plug-in" substitute for a missing datum.

Incompleted-trip interview: Interviews of anglers taken while they are still fishing.

Jackknife estimator: A method of estimating a ratio, such as $\overline{HPUE}$, by systematically eliminating each bivariate datum (harvest and fishing effort for an angler trip as measured in a single interview) in turn, estimating the ratio from the remaining data from each exclusion, then averaging these ratios to produce the estimate.

Location stratification: A dimension of stratification based on grouping basic sampling units according to where (what access-point) they could be sampled. Elements in this dimension can be a single or a group of access-points.

Length-of-stay (LOS) bias: Bias in an estimate that arises when the probability of interviewing an angler is related to how long he or she fished.

Opportunity bias: Bias in an estimate that arises when anglers or harvest exiting a fishery in front of technicians during a sampled period are not representative of the basic sampling units in that period.

Partial stratification: Given K dimensions of stratification with k_1, k_2 .... k_K elements in each dimension, not every element in every dimension is used to define some strata.

Plug-in substitute: A substitute for a missing datum from one sampled period that is some scaled function of data from other sampled periods or strata with similar characteristics.

Post-stratification: Stratification that is defined by dimensions based on behavior, demography, or morphology of basic sampling units. Contact with basic units is needed to determine their membership in a stratum defined with a post-stratified dimension.

Proportional sampling: The same proportion of basic sampling units is sampled in each stratum.

Random sampling: Samples are selected independently from one another without regard to any temporal or spatial order.

Ratio of means estimator: A method of estimating $\overline{HPUE}$ for a sampled period by dividing the sample mean for harvest by the sample mean for fishing effort (in hours).

Redefined strata: Information from a stratum is pooled across at least two consecutive strata to produce a new, larger stratum, or information from one stratum is split to produce two or more new, smaller strata.

Relative harvest composition: Fraction of harvest by groups according to some attribute such as age, sex, or species.

Roving-access survey: A sampling program designed to estimate harvest or catch as the product of estimated harvest or catch rate and estimated fishing effort for each sampled period. Harvest or catch rates are estimated from completed-trip interviews of anglers exiting from access-points while fishing effort is estimated from roving counts of fishing anglers. Estimates for sampled periods are expanded to cover periods not sampled.

Roving-roving survey: A sampling program to estimate harvest or catch as the product of estimated harvest or catch rate and estimated fishing effort for each sampled period. Harvest or catch rates are estimated from incompleted-trip interviews of anglers contacted by roving through the fishery. Fishing effort is estimated from counts by roving technicians as well. Estimates for sampled periods are expanded to cover periods not sampled.

Sample variance*:* Estimated variance associated with a sampling stage in a sampling design. Sample variances form the core of the components of estimated variance.

Sampled period: A sampling period from which basic units were sampled.

Sampled population: The set of basic sampling units that can be sampled (whether they are sampled or not). If the sampled population differs from the target population, estimates will be biased.

Sampling design: Hierarchical organization of basic sampling units into sampling stages and strata from which a sampling schedule will be derived. In the hierarchy for onsite creel surveys, basic sampling units always constitute the final sampling stage and sampling periods the penultimate stage.

Sampling frame: The number of sampling units in a stage or stratum.

Sampling period: A specific period of time at a specific location in which sampling can occur. Each basic sampling unit in the sampled population should be associated with one and only one sampling period.

Sampling schedule: Subset of sampling periods in a sampling design that will be sampled.

Sampling shadow: A reduction in the probability that an angler will be counted because the preceding angler was interviewed. A sampling shadow results from counting and interviewing anglers during a single rove through a fishery in a roving-roving survey.

Sampling stage: A group of like sampling units representing a level in the hierarchy of a sampling design. A subset of sampling units are sampled within each sampling stage.

Sampling units: Basic sampling units are persons or fish that can be sampled. All other sampling units higher in the hierarchy of a sampling design, such as first-stage sampling units, second-stage sampling units, etc., represent opportunities to sample (location, day, sampling period).

Seasonal stratification: A dimension of stratification based on grouping basic sampling units according to what month or run they could be sampled. Usually there is one element for every month or run represented in the fishery.

Self-weighted survey: Stratified “random” sampling program where basic sampling units are proportionally drawn (proportional sampling). Because each stratum will have the same weight in such surveys, data can be pooled across strata.

Size: The number of sampling units in a sampling stage or stratum.

Stratified “random” sampling design: Basic sampling units are exclusively into strata such that there is insignificant temporal or spatial variation within strata in parameters to be estimated.

Stratum: A group of sampling units representing the top level in the hierarchy of a sampling design. Strata are defined by one or more attributes (dimensions) of sampling units. Some sampling units are sampled from each stratum in the design.

Systematic sampling: Sampling units are selected according their temporal or spatial order. One of the first *k* units in the order is randomly selected, every *k*th unit in the order beyond the last chosen are sequentially sampled until the population is exhausted.

Target population: The set of basic sampling units from which knowledge is desired.

Time and location-defined (TLD) stratum: A stratum defined solely by dimensions based on when or where basic sampling units can be sampled such that no contact with the basic units is needed to determine their membership in the stratum.

Time-of-day (TOD) stratification: A dimension of stratification based on grouping basic sampling units according to when they could be sampled during a fishing day. Usually there are two or three elements in this dimension. For short fishing days, elements are mornings and afternoons; for long days mornings, afternoons, and evenings.

Weekly stratification: A dimension of stratification based on grouping basic sampling units according to what week they could be sampled. Usually there is one element for every week throughout the duration of a fishery.