Estimating the Population Size of Razor Clams Using a Model Assisted Sampling Design and Analysis

by

Denise Babineau

B.Math., University of Waterloo, 1998

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE
in the Department
of
Mathematics and Statistics

SIMON FRASER UNIVERSITY
August 2000

APPROVAL

Name: Denise Babineau

Degree: Master of Science

Title of project: Estimating the Population Size of Razor Clams Using a Model Assisted Sampling Design and Analysis

Examining Committee: Dr. G.A.C. Graham
Chair

____

Dr. Carl J. Schwarz
Senior Supervisor
Simon Fraser University

____

Dr. Jinko Graham
Simon Fraser University

____

Dr. Tim B. Swartz
Simon Fraser University

____

Dr. K.L. Weldon
External Examiner
Simon Fraser University

Date Approved: __________________________________

Abstract

For several decades, North Beach near Masset, British Columbia has been used for commercial and non-commercial harvesting of razor clams. In the early 1990's, there was some concern among environmentalists that the health of the stock seemed to be failing. To address this concern, the size of the population needed to be accurately estimated. Usually, standard sampling methods, like those outlined in Gillespie and Kronlund (1999), are used to estimate the population size of clams in a certain area. However, because methods like simple random sampling, stratified random sampling, two-stage sampling and stratified two stage sampling are designed to be applied to any population, such general methods fail to incorporate any biological knowledge that may be discovered concerning the distribution of razor clams. To use such information to an advantage, a model assisted sampling design and analysis must be carried out.

This type of analysis was used to estimate the population size of razor clams on North Beach and surrounding beaches using a three stage sampling design. Once the beaches of interest were chosen, transects were then randomly sampled along each beach. Once distances were systematically sampled along each sampled transect, sampling plots called quadrats were then replicated at each distance. Two methods of estimation were then used to estimate the number of razor clams on each transect considered. The first method used a straight line interpolation while the second method used a cubic smoothing spline. A preliminary analysis of the data also showed that the population density for each distance considered had an approximate Poisson distribution. This information was also integrated into the two separate analyses as well. To estimate the overall population size for each beach considered, both a ratio

estimator and a simple inflation estimator were then used. Results from this analysis provided knowledge of the clam density relative to chart datum (lowest possible tide). These results were then used to obtain a sampling design that provided the best estimate of population size using optimum numbers of transects, beach elevations, and samples of razor clams.

Acknowledgements

I would first like to thank my supervisor, Dr. Carl Schwarz. I have spent many an hour in his office and it still amazes me that he didn't kick me out! I can not convey how thankful I am to him.

I would also like to thank my supervisory committee, Dr. Jinko Graham, Dr. Tim Swartz and Dr. Larry Weldon. Their input was extremely valuable and improved my project a thousand fold!

I would also like to thank Dr. Nancy Heckman for her expertise in the area of smoothing splines. As this area was new to me, her suggestions were very helpful.

I would also like to thank both faculty and staff in the Department of Mathematics and Statistics, especially those in Statistics. Everyone has always been quite happy to answer any questions, about statistics or otherwise, that I have had. I don't think I will ever find a more supportive department!

I would also like to thank all of the graduate students that I have come to know throughout my time at SFU. In particular, they include Laurie Ainsworth, Dr. Derek Bingham, Ellen Chan, Grace Chiu, Nick Costanzino, Phil Dauk, Travis Gray, Ruth Joy, Farah Khandwala, May Lee, Shabnam Kavousian, Jason Loeppky, Dr. Michael Lunney, Colin Macleod, Jason Nielsen, Chuck Paltiel, Greg Pond, Jason Sutherland, Wanlin Wei, Dr. Changbao Wu and Ruihua Yin. Every one of these people have helped me in my studies or for much needed euchre breaks. I will miss all of you!

I would also especially like to thank my "step-sisters" Nicole McCutchen and Viviane Saleh-Hanna. Both have supported me on so many issues that I can't even begin to count. They have always been there when I needed a smile or just a hug. I only hope that other people can find such good friends as these.

Last, but certainly not least, I wish to thank my parents, my sister Cindy and my brother Bryan. Although they tend to complain about how little I see them, they have fully supported any decisions I have made and have been there whenever I needed someone to talk to.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Pacific razor clams (*Siliqua patula* Dixon), seen in Figure 1.1, are found on the Pacific beaches from Pismo Beach, California to the Aleutian Islands in Alaska. Throughout this area, there are eight major concentrations of razor clams that support commercial and non commercial fisheries. In 1993, not only was it discovered that commercial fishery landings were low but also that there were high proportions of undersized razor clams in the commercial catches. To environmentalists, this indicated that the health of the stock seemed to be failing. Thus, a quantitative assessment of razor clam stocks needed to be carried out.

Figure 1.1: Pacific razor clam (*Siliqua patula* Dixon)

To carry out an assessment such as this, intertidal clam surveys are conducted by digging small plots of beach called quadrats and observing the number, weight and age of the razor clams that are removed. Data of this type are then used to estimate the total number of razor clams, the total biomass and the growth rates of the razor clams in the survey area. An important feature of intertidal clam surveys is the type of sampling technique that is used to sample the clams. Gillespie and Kronlund (1999) review the following four sampling strategies that are commonly used.

The simplest way to conduct an intertidal clam survey is through the use of simple random sampling. This strategy requires quadrats to be randomly sampled from the entire population of quadrats on the beach being surveyed. However, this type of sampling may lead to a large variance if there are both high and low densities of clams on the beach. Hence, although simple random sampling is not appropriate for most of the intertidal clam surveys conducted, it is often used at some stage in more complicated designs.

To improve the efficiency of a simple random sampling design, a strategy called stratified random sampling can be used. In this case, quadrats are first divided into non-overlapping groups called strata by ensuring that all quadrats within a stratum are as similar as possible. A simple random sample is then taken within each stratum to obtain the sampled data. This type of sampling not only increases the precision of an estimate and its associated variance but also decreases the cost of the survey because quadrats within each stratum are in close proximity.

Another sampling design that is commonly used is called a two stage sampling design. This type of design partitions the beach into large plots called first stage units and then randomly samples from these units. Each sampled first stage unit is then partitioned into quadrats where a random sample of quadrats is then taken. In this case, there are two sources of variability. One source of variability is due to the different quadrats sampled within each of the first stage units. The second source of variability is due to the different first stage units that are sampled. Although this is similar to a stratified random sample, the difference lies in the two stages of sampling.

To incorporate both stratified sampling and two stage sampling in a design, stratified two stage sampling designs are also used in intertidal clam surveys. In this case, the beach is first partitioned into non-overlapping strata. Each stratum is divided into first stage units which are then divided into quadrats. A sample is then obtained by selecting a simple random sample of first stage units within each stratum followed by a simple random sample of quadrats within each sampled first stage unit. Advantages of this sampling strategy include lower surveys costs due to ease of quadrat location, better coverage of survey area due to stratification and two randomization stages and increased precision of estimates due to stratification.

Using the sampling strategies outlined above, intertidal populations have also been estimated using a variety of more complicated techniques like marked recapture, tagging and hydraulic core sampling. In particular, Bourne (1969) estimated the density of razor clams by tagging clams on the beaches, reburying them, conducting repeated digs and analyzing the counts of those clams that were sampled repeatedly. The Washington Department of Fisheries and Wildlife (Ayres and Simons, 1988) also used marked recapture but with stratified random digs. However, because marked recapture methods of this type require a large amount of work when tagging clams, other methods should be considered. This was done in Szarzi (1991) using hydraulic core sampling and a three stage sampling design to collect data from Cook Inlet, Alaska. However, a two stage design, which ignored some information gained from the three stage sampling design, was adopted for the final analysis. Hence, another analysis of data obtained from this type of sampling procedure should be employed to ensure that all of the information gained from the sampling procedure is used. This was done on beaches near Masset, British Columbia (Schwarz et al., 19xx). In this study, estimation of the number of razor clams on a beach used specific ranges of the beach elevation to coincide with how digging was done commercially. To obtain estimates and associated estimated standard errors of the number of razor clams on a beach for elevations above 0 m, above 1 m and above 2 m, it then used transects, distances and quadrats as random effects and the different levels of elevation as a fixed effect to fit an unbalanced mixed effect model using SAS.

The analysis about to be presented will use the same set of data as used in the unbalanced mixed effect model but will model the relationship between distance from chart datum (or elevation above chart datum) and the number of razor clams in a quadrat more explicitly. A review of the sampling method used to collect this data is now given.

The beaches that were surveyed to collect the data for this analysis are located near Masset, British Columbia and are called North Beach, South Beach and Agate Beach. However, North and South Beach were the only beaches sampled due to high clam density, beach accessibility and high usage by clam diggers. As such, Agate Beach is excluded from the population of interest in all analyses of this data. Furthermore, South Beach was divided into two beaches to simplify the sampling procedure. Hence, the three sections of beaches that were sampled between 1994 and 1996, as seen in Figure 1.2, are referred to as North Beach, South 1 Beach and South 2 Beach with lengths 7.2 km, 4.6 km and 6.75 km and areas $1,440,000\ m^2$, $598,000\ m^2$ and $540,000\ m^2$ respectively. A three stage sampling design, as seen in Figure 1.3, that sampled transects, distances from chart datum and quadrats was then used to sample razor clams from the beaches surveyed.

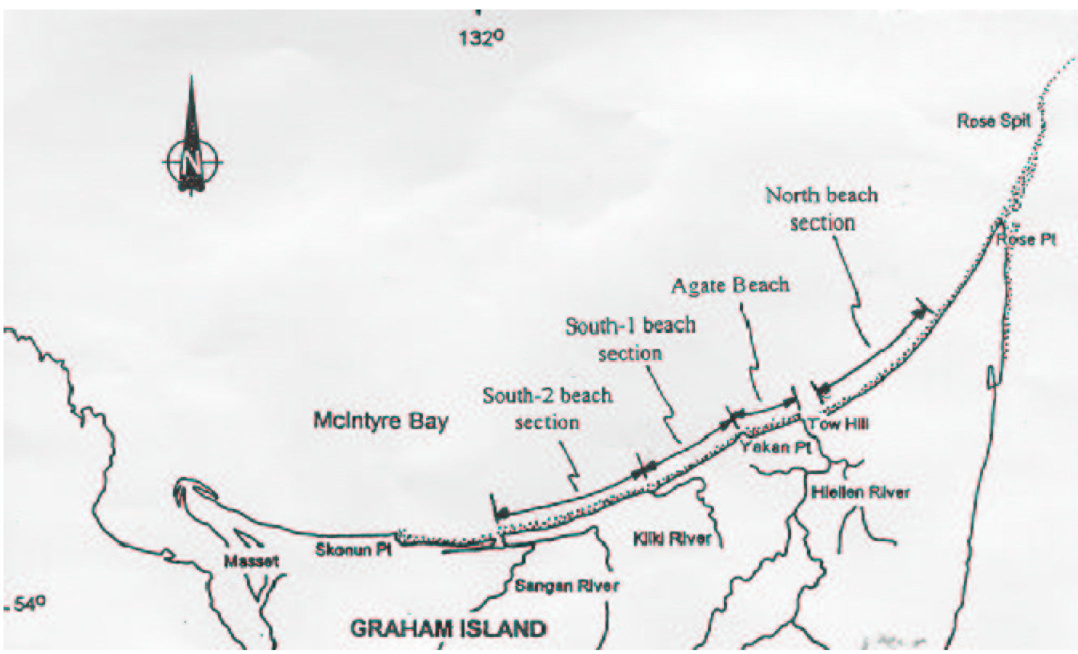

Figure 1.2: Northeastern shore of Graham Island, showing the locations of surveyed beaches.

To begin the sampling of the selected beach sections, transects (i.e. lines laid perpendicular to chart datum (lowest possible tide)) were first allocated by spacing

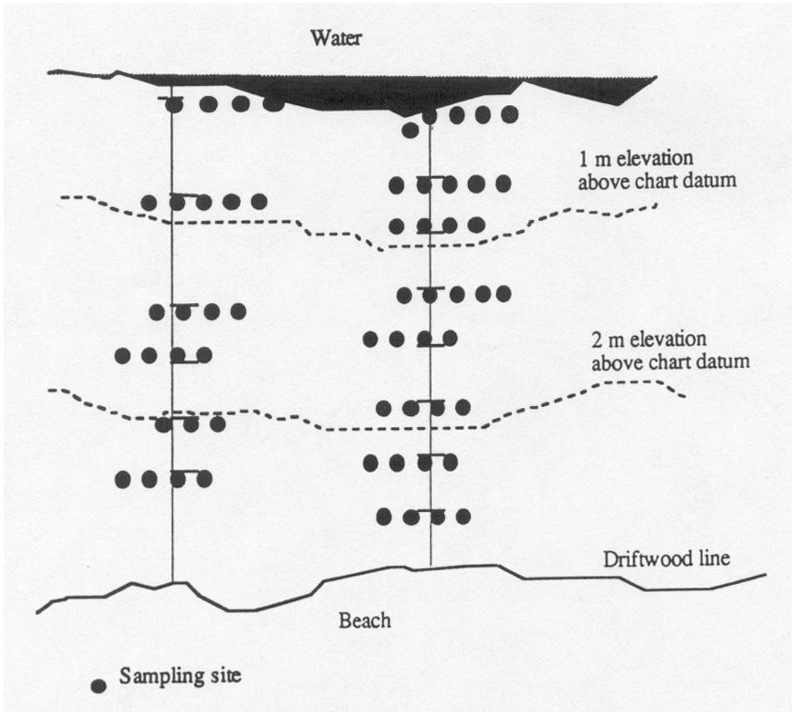

Figure 1.3: Illustration of sampled beach.

them equally along the beach. This ensured that no transects crossed one another. Transects were then randomly sampled and located by driving a fixed distance from the access point as measured by a truck odometer. In particular, a transect that was sampled 6.6 km along the beach was named Transect 6.6. The transect was then laid out using a rebar marked with flagging tape. Each transect took approximately 70 minutes to sample and mark. Because sampling can only occur during daylight low tides of no more than 1 m above chart datum and sampling one transect took most of this time, only one transect could be sampled per tide.

Once a transect was randomly selected, distances from chart datum along the transect were located systematically, using a metal rebar stake, with sampling beginning where the surfline was located at the time of arrival. Depending on the rate of tidal approach, the space between each sampled distance ranged between 15 m and

25 m. Because this type of systematic sampling ensured that the sampled distances were quite far apart, observations made at different distances were assumed to be independent. For each distance sampled, the elevation above chart datum was also recorded with a surveyor's level and rod, using tidewater level as a reference. Each distance took approximately 5 minutes to sample and mark.

For each distance sampled, several quadrats pertaining to that distance were then randomly sampled. Quadrats were first located by drawing a line parallel to chart datum that was within 7 m of the transect line. Quadrats were then sampled with approximately the same number of quadrats on either side of the transect line. Although the quadrats sampled were quite close to one another, all quadrats were assumed to be independent. At each selected quadrat location, a galvanized steel circular sampling ring with a diameter of 0.79 m and an area of 0.5 m^2 was drilled into the sand to a depth of 0.5 m. Pressurized seawater was used to emulsify the substrate to dislodge the razor clams. Dipnets were then used to capture the razor clams. In general, this procedure required 4 minutes per quadrat when a five person crew was being used. For a more detailed description of this procedure and the equipment used, refer to Szarzi (1991).

Once the sample was taken, all razor clams found within that sample were placed into separate labelled bags. After field sampling was completed, characteristics such as the number of razor clams in the sample, length of each razor clam in the sample and the weight of the entire sample were measured. Analysis based on this data focused on those razor clams with lengths greater than 4 mm, greater than 20 mm and greater than 90 mm. Information such as this is used to determine the number of harvestable razor clams (i.e. razor clams with a length greater than 90 mm). Table 1.1 indicates the number of transects, the number of distances, the number of quadrats sampled and the total number of clams sampled for 1994, 1995 and 1996 on all beaches surveyed [1].

For the current analysis, the data found in Schwarz et al. (19xx) is re-analyzed using other techniques. Chapter 2 discusses the estimation of the number of razor clams on a beach using a straight line interpolation of the average number of razor

[1] South 2 Beach was not surveyed in 1995.

Year	Beach	Transects	Distances	Quadrats	Clams
1994	North	7	52	269	3537
	South 1	4	30	176	856
	South 2	3	17	88	175
1995	North	7	80	235	1564
	South 1	4	36	106	608
1996	North	6	56	168	1233
	South 1	4	31	93	367
	South 2	2	12	36	82

Table 1.1: Number of transects, distances, quadrats and razor clams sampled for all surveys conducted.

clams sampled at a quadrat versus the distance from chart datum to estimate the total number of razor clams on a transect. Ratio and inflation estimators are then developed to estimate the total number of razor clams on a beach. Chapter 3 discusses how cubic smoothing splines can be used in lieu of straight line interpolation techniques when estimating the total number of razor clams on a transect. Chapter 4 discusses how to estimate the total number of razor clams over specific ranges of elevation using estimators found in Chapter 2 and Chapter 3. Chapter 5 develops strategies to determine the optimal allocation of effort among transects, distances and quadrats sampled. To illustrate the techniques developed in Chapter 2 to Chapter 5, application of these strategies to the data obtained from the 1994 survey of North Beach for razor clams with lengths greater than 4 mm is discussed throughout each of the chapters. Chapter 6 then determines estimates of the number of differently sized razor clams on all beaches for all years surveyed. These estimates are then compared to what was found using the unbalanced mixed effect model used in Schwarz et al. (19xx).

Chapter 2

Estimation Using A Straight Line Interpolation

This analysis estimates the number of razor clams on a beach using a straight line interpolation. One important feature is that the density of razor clams is approximately a monotonic decreasing function of the distance from chart datum. Using Taylor's power law (Taylor, 1961), it is also known that the total number of razor clams in a quadrat is approximately Poisson distributed. With this information, an estimate of the total number of razor clams per transect and its associated standard error is found. Using the transect estimates and two different types of estimators, estimates and associated standard errors for the total number of razor clams on a specific beach are then easily determined. These concepts are developed and applied to data collected in the 1994 survey of North Beach for razor clams with lengths greater than 4 mm.

2.1 Preliminary Plots

2.1.1 Average Density Versus Distance From Chart Datum

Due to environmental factors and the physiological requirements of the organism, razor clams are more likely to be found closer to chart datum. To show that this is consistent with the data used in this study, a transect plot of the average number of razor clams sampled per quadrat by distance from chart datum is shown in Figure 2.1[1].

Most transects show a general downward trend in the relationship between density and distance from chart datum. It is also clear from these plots that for some transects, the density of razor clams reaches zero at shorter distances than others and the range of the density of razor clams over all transects is somewhat variable. Important biological knowledge such as this is used to determine the total number of razor clams on North Beach.

2.1.2 Distribution of Counts

It is also useful to determine if the data can be modeled parametrically. Taylor (1961) showed that various biological organisms follow a simple power law given by

$$\sigma^2 = a\mu^b \quad \text{or} \quad \log \sigma^2 = \log a + b \log \mu$$

where σ^2 is the variance and μ is the population mean. Applying this concept to the razor clam data, σ^2 is the variance of the quadrat counts that were sampled at a particular distance along a transect and μ is the mean number of razor clams at that distance. Because the number of razor clams varies considerably over the length of the beach and also varies with the distance from chart datum, it is also suspected that the number of razor clams in quadrats sampled at each distance follows a Poisson distribution but with a parameter that varies according to the distance of the sampled

[1] If the last sampling point along a transect has a non-zero average, a pseudo sampling point is created at a distance equal to the last sampling interval past the last sampling point and a value of 0 is assigned for the average count.

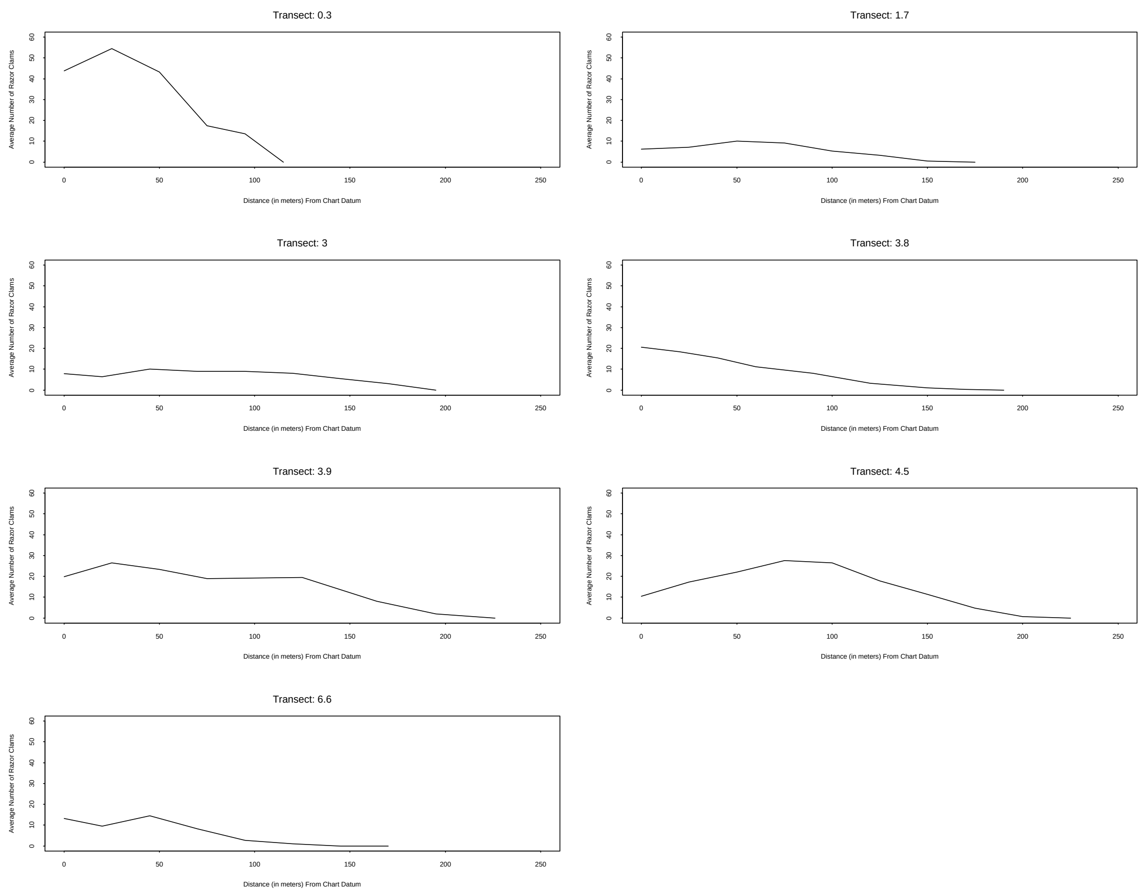

Figure 2.1: Average number of razor clams sampled per quadrat versus distance (in meters) from chart datum for each of the 7 transects sampled on North Beach in 1994.

quadrats from chart datum. Consequently, a plot of $\log s^2$ versus $\log \overline{x}$ should produce a plot with a slope of approximately 1 and intercept 0 because $\sigma^2 = \mu$ for Poisson distributed data. From Figure 2.2, the relationship between the log of the sample

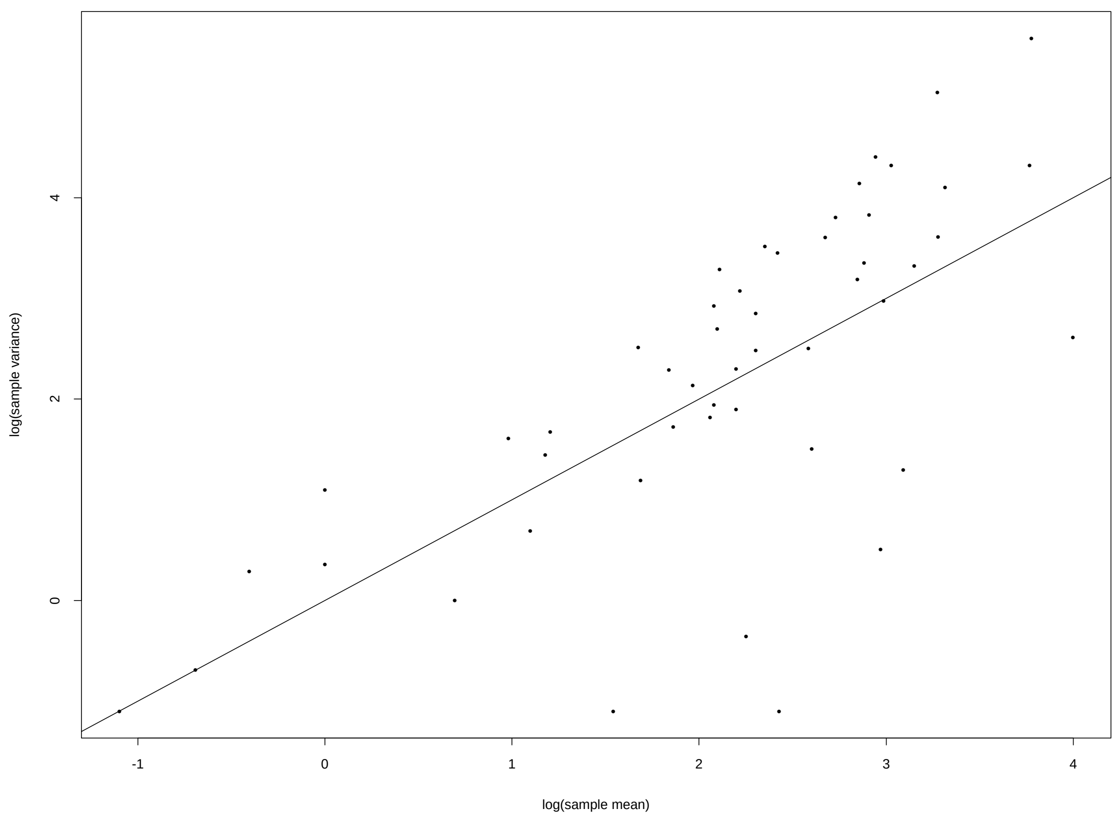

Figure 2.2: Plot of the log of the sample variance versus the log of the sample mean at distances along all sampled transects in the 1994 survey of North Beach. The line drawn is the theoretical line of a Poisson distribution where appropriate.

mean and the log of the sample variance seems to be slightly non linear which does cast some doubt on the assumption that the quadrat counts are Poisson distributed. In fact, because most points lie above the theoretical line, this suggests that the data is overdispersed relative to a Poisson distribution. It should also be noted that those points that lie below the theoretical line seem to be scattered over all transects.

However, to illustrate this method for any beach surveyed, it is helpful to note that when a straight line is fit to the data, the slope is approximately 1 with intercept 0. Thus, the number of razor clams at each distance can be modelled using a Poisson distribution with a mean that varies according to the distance from chart datum.

2.2 Estimation of the Transect Total

An estimate of the total number of razor clams for a particular transect is found by estimating the total area under the curve of density versus distance (i.e. like those curves given in Figure 2.1). In this method, the area under the curve is determined by cutting the curve into several trapezoids and then summing the area of all trapezoids together. This produces the estimate of the total number of razor clams for that transect. Letting $\overline{Y}_{ij}$ represent the mean number of razor clams located at distance j along transect i, the estimated number of razor clams along transect i is given by

$$\begin{aligned} \hat{T}_i &= \frac{\sum_{j=1}^{n_i-1} (\overline{Y}_{ij} + \overline{Y}_{i,j+1}) b_j}{2} \\ &= \frac{\overline{Y}_{i1} b_1}{2} + \frac{\sum_{j=2}^{n_i-1} \overline{Y}_{ij}(b_{j-1} + b_j)}{2} + \frac{\overline{Y}_{in_i} b_{n_i-1}}{2} \end{aligned} \tag{2.1}$$

where n_i is the number of distances sampled along transect i and b_j is the distance between the j^{th} distance and the $(j+1)^{st}$ distance. Note that the n_i's include pseudo-distances with zero counts if needed.

The variance of the transect estimate shown in (2.1) is given by

$$
\begin{aligned}
\mathrm{Var}(\hat{T}_i) &= \mathrm{Var}\left(\frac{\overline{Y}_{i1} b_1}{2} + \frac{\sum_{j=2}^{n_i-1} \overline{Y}_{ij}(b_{j-1}+b_j)}{2} + \frac{\overline{Y}_{in_i} b_{n_i-1}}{2} \right) \\
&= \frac{{b_1}^2}{4}\mathrm{Var}(\overline{Y}_{i1}) + \frac{\sum_{j=2}^{n_i-1} \mathrm{Var}(\overline{Y}_{ij})(b_{j-1}+b_j)^2}{4} + \frac{b_{n_i-1}^2}{4}\mathrm{Var}(\overline{Y}_{in_i}) \quad (2.2)
\end{aligned}
$$

by independence of the $\overline{Y}_{ij}$'s. Because $\overline{Y}_{ij} = \sum_{k=1}^{n_{ij}} Y_{ijk}/n_{ij}$ where Y_{ijk} is the number of razor clams sampled in quadrat k located at distance j along transect i and n_{ij} is the number of quadrats sampled at distance j in transect i, then

$$
\begin{aligned}
\mathrm{Var}(\overline{Y}_{ij}) &= \frac{\sum_{k=1}^{n_{ij}} \mathrm{Var}(Y_{ijk})}{n_{ij}^2} \\
&= \frac{\sigma_{ij}^2}{n_{ij}} \quad (2.3)
\end{aligned}
$$

where $\sigma_{ij}^2 = \mathrm{Var}(Y_{ijk})$. An estimate of (2.3) is found using the variance of the counts of the sampled quadrats and is given by

$$
\widehat{\mathrm{Var}}(\overline{Y}_{ij}) = \frac{s_{ij}^2}{n_{ij}} \quad \text{where} \quad s_{ij}^2 = \frac{\sum_{k=1}^{n_{ij}} (Y_{ijk} - \overline{Y}_{ij})}{n_{ij}-1} \quad (2.4)
$$

Using (2.2) and (2.4), an estimate of the variance of the transect estimate is then

$$
\widehat{\mathrm{Var}}(\hat{T}_i) = \frac{{b_1}^2\, s_{i1}^2}{4\, n_{i1}} + \frac{\sum_{j=2}^{n_i-1} \frac{s_{ij}^2}{n_{ij}}(b_{j-1}+b_j)^2}{4} + \frac{b_{n_i-1}^2\, s_{in_i}^2}{4\, n_{in_i}} \quad (2.5)
$$

At this point, no distributional assumptions have been made about the data. However, Figure 2.2 showed that the number of razor clams sampled per quadrat could be modelled using a Poisson distribution (i.e. $Y_{ijk} \sim \text{Poisson}(\mu_{ij})$). Hence, another estimate of (2.3) is given by

$$\widehat{\text{Var}}(\overline{Y}_{ij}) = \frac{\overline{Y}_{ij}}{n_{ij}} \tag{2.6}$$

Using (2.2) and (2.6), a model assisted estimate of the variance of the transect estimate is then

$$\widehat{\text{Var}}(\hat{T}_i) = \frac{{b_1}^2\, \overline{Y}_{i1}}{4\, n_{i1}} + \frac{\sum_{j=2}^{n_i-1} \frac{\overline{Y}_{ij}}{n_{ij}} (b_{j-1} + b_j)^2}{4} + \frac{b_{n_i-1}^2\, \overline{Y}_{in_i}}{4\, n_{in_i}} \tag{2.7}$$

Applying these concepts to the data collected from North Beach in 1994 for razor clams with lengths greater than 4 mm, the transect estimates and estimated standard errors using a parametric and non parametric analysis are given in Table 2.1. A comparison of the estimated standard errors shows that standard errors are somewhat lower when the data is assumed to be Poisson distributed. This indicates that the data is overdispersed relative to a Poisson distribution. However, this is not surprising since overdispersion was present in Figure 2.2. One method that could be used to counteract this problem is to determine the amount of overdispersion and incorporate it into the estimated transect standard errors. However, it may also be helpful to determine a more suitable model. Elliott (1977) explains why negative binomial distributions may be more suitable for data of this type.

2.3 Estimation of the Beach Total

Once all estimates of the total number of razor clams per sampled transect are calculated, a method must be found to estimate the total number of razor clams on a beach. Two methods that are used in this analysis involve ratio estimators and inflation estimators.

Transect	Estimate	Estimated Standard Error	Estimated Standard Error (Poisson Assumption)
0.3	3,652.3	159.2	142.2
1.7	964.2	92.9	67.8
3.0	1,336.5	70.8	69.4
3.8	1,532.9	129.0	89.4
3.9	3,413.1	230.2	156.3
4.5	3,325.0	137.9	126.6
6.6	1,006.7	76.1	56.3

Table 2.1: 1994 North Beach transect estimates using a straight line interpolation.

2.3.1 Ratio Estimation

One method that will be used stems from the fact that longer transects contain higher numbers of razor clams than shorter transects; as shown in Figure 2.3. By Cochran (1977), two correlated variables can be used to obtain a more precise estimate of the total number of razor clams on a beach through the use of ratio or regression estimators. Figure 2.3 shows that a regression estimator may be more suitable than a ratio estimator because a straight line fitted to this plot would not go through the origin. In fact, the relationship between the length of the transect and the transect estimate is not even linear and also includes an outlier that may drastically affect the beach estimates if a ratio or regression estimator were to be used for estimation purposes. However, for the purposes of this analysis, a ratio estimator will be used keeping in mind that estimates may be biased and have a higher variance. Consequently, the ratio estimator that is used to estimate the total number of razor clams on a beach is given by

$$\hat{B}_{ratio} = \frac{2A\sum_{i=1}^{n}\hat{T}_i}{\sum_{i=1}^{n}L_i} \tag{2.8}$$

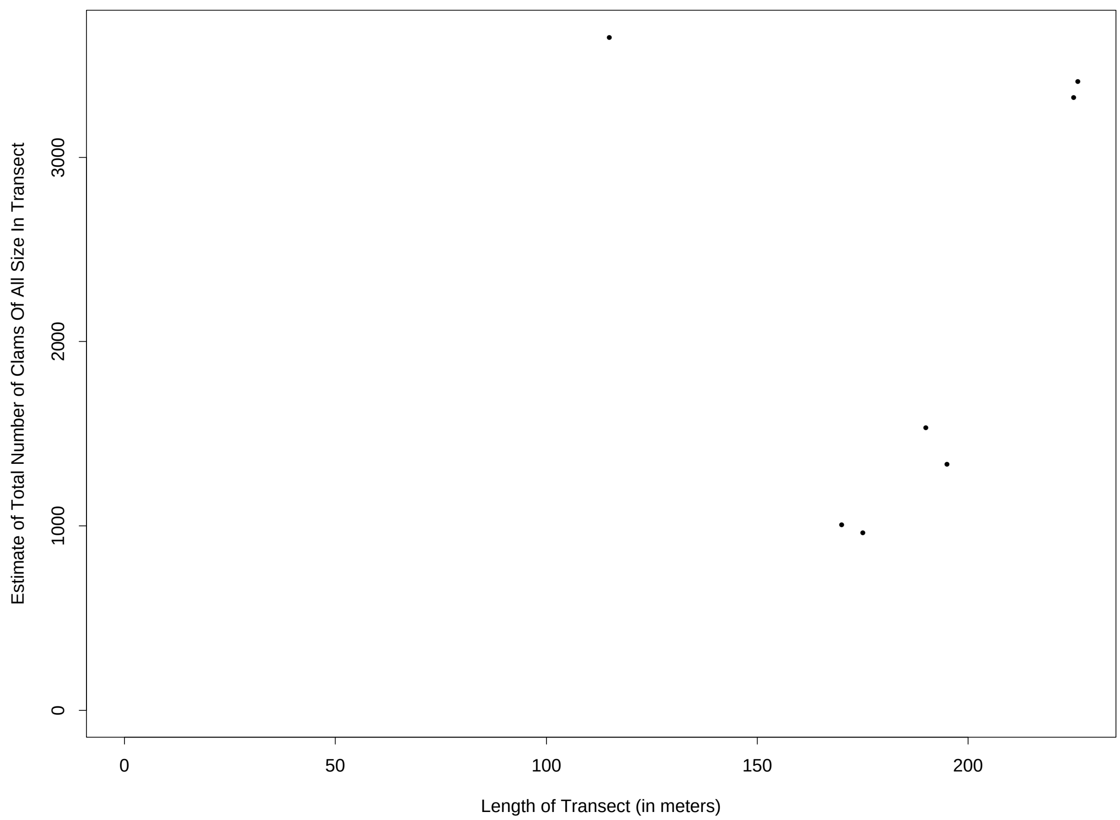

Figure 2.3: Plot of transect length (in meters) and estimated total number of razor clams for all sampled transects in the 1994 survey of North Beach.

where $\hat{T}_i$ is the estimated total number of razor clams along transect i, L_i is the length of transect i in m, n is the number of transects sampled on the beach and A is the area of the beach in m^2. Interpretation of this estimator begins by noting that quadrats, having an area of 0.5 m^2, are used to sample razor clams along the transects. Thus, $\sum_{i=1}^{n} \hat{T}_i / \sum_{i=1}^{n} L_i$ approximately represents the number of razor clams per 0.5 m^2. To determine the total number of razor clams per m^2, this ratio must then be multiplied by two. This quantity is then multiplied by the total area of the beach to estimate the total number of razor clams on the beach.

To find the variance of (2.8), conditional expectations are used by conditioning upon the set of transects sampled along the beach. This method results in the variance of (2.8) to be given by

$$\mathrm{Var}(\hat{B}_{ratio}) = \mathrm{Var}(\mathrm{E}(\hat{B}_{ratio}|\mathrm{transects})) + \mathrm{E}(\mathrm{Var}(\hat{B}_{ratio}|\mathrm{transects})) \tag{2.9}$$

To begin,

$$\mathrm{E}(\hat{B}_{ratio}|\mathrm{transects}) = 2A \, \frac{\sum_{i=1}^{n} T_i}{\sum_{i=1}^{n} L_i}$$

since $\hat{T}_i$ is unbiased for T_i. Then

$$\mathrm{Var}(\mathrm{E}(\hat{B}_{ratio}|\mathrm{transects})) = 4A^2 \; \mathrm{Var}\left(\frac{\sum_{i=1}^{n} T_i}{\sum_{i=1}^{n} L_i}\right) \tag{2.10}$$

Letting $X = \sum_{i=1}^{N} T_i / \sum_{i=1}^{N} L_i$, an estimate of the variance of this ratio estimator given in Cochran (1977) is

$$\mathrm{Var}\left(\frac{\sum_{i=1}^{n} T_i}{\sum_{i=1}^{n} L_i}\right) \approx \frac{(1-f)}{n\overline{L}^2} \frac{\sum_{i=1}^{N} (T_i - XL_i)^2}{N-1}$$

where N is the number of transects on the surveyed beach having a width of approximately 0.5 m, $f = n/N$ and $\overline{L}$ is the population mean of transect lengths. As a sample estimate of this variance, it is customary to take

$$\widehat{\text{Var}}\left(\frac{\sum_{i=1}^{n} T_i}{\sum_{i=1}^{n} L_i}\right) \approx \frac{(1-f)}{n\hat{\overline{L}}^2} \frac{\sum_{i=1}^{n}(\hat{T}_i - \hat{X}L_i)^2}{n-1} \tag{2.11}$$

where $\hat{X} = \sum_{i=1}^{n} T_i / \sum_{i=1}^{n} L_i$ and $\hat{\overline{L}}$ is the sample mean of transect lengths. Using (2.11), an estimate of (2.10) is

$$\widehat{\text{Var}}(\text{E}(\hat{B}_{ratio}|\text{transects})) = 4A^2 \frac{(1-f)}{n\hat{\overline{L}}^2} \frac{\sum_{i=1}^{n}(\hat{T}_i - \hat{X}L_i)^2}{n-1} \tag{2.12}$$

An estimate of $\text{E}(\text{Var}(\hat{B}_{ratio}|\text{transects}))$ must now be determined. Because

$$\text{Var}(\hat{B}_{ratio}|\text{transects}) = \left(\frac{2A}{\sum_{i=1}^{n} L_i}\right)^2 \sum_{i=1}^{n} \text{Var}(\hat{T}_i) \tag{2.13}$$

and a first order approximation of its expectation is

$$\text{E}(\text{Var}(\hat{B}_{ratio}|\text{transects})) \approx 4A^2 \frac{\frac{1}{n}\frac{\sum_{i=1}^{N}\text{Var}(\hat{T}_i)}{N}}{\left(\frac{\sum_{i=1}^{N} L_i}{N}\right)^2} \tag{2.14}$$

so the variance of $\hat{B}_{ratio}$ is found to be

$$\operatorname{Var}(\hat{B}_{ratio}) = 4A^2 \left(\frac{(1-f)}{n\overline{L}^2} \frac{\sum_{i=1}^{N}(T_i - XL_i)^2}{N-1} + \frac{\frac{1}{n}\frac{\sum_{i=1}^{N}\operatorname{Var}(\hat{T}_i)}{N}}{\left(\frac{\sum_{i=1}^{N} L_i}{N}\right)^2} \right)$$

and the estimated variance is

$$\widehat{\operatorname{Var}}(\hat{B}_{ratio}) = 4A^2 \left(\frac{(1-f)}{n\hat{\overline{L}}^2} \frac{\sum_{i=1}^{n}(\hat{T}_i - \hat{X}L_i)^2}{n-1} + \frac{\sum_{i=1}^{n}\widehat{\operatorname{Var}}(\hat{T}_i)}{\left(\sum_{i=1}^{n} L_i\right)^2} \right) \tag{2.15}$$

2.3.2 Inflation Estimator

Another method that may also be used to estimate the total number of razor clams on a beach uses a simple inflation estimator. For the razor clam data, the first stage of sampling involves a simple random sample of transects along the beach. Consequently, the inflation estimator is

$$\hat{B}_{inf} = N\frac{\sum_{i=1}^{n}\hat{T}_i}{n} \tag{2.16}$$

where $\hat{T}_i$ is the estimated total number of razor clams along transect i, n is the number of transects sampled on the beach and N, the total number of transects on the beach having a width of approximately 0.5 m, is used as the inflation factor. Examination of this estimate shows that the numerator consists of the sum of estimated transect totals while the denominator is the number of transects sampled along the beach. Once the sum of the transect estimates is divided by the number of sampled transects, an estimate of the average number of razor clams per transect is found. To determine the estimate of the total number of razor clams on the beach, this quantity is then multiplied by the total number of transects on the beach having a width of approximately 0.5 m.

To find the variance of (2.16), a similar method to that used to determine the variance of (2.8) is used. Using this method, we again condition on the transects sampled along the beach. Then the variance of (2.16) is given by

$$\text{Var}(\hat{B}_{inf}) = \text{Var}(\text{E}(\hat{B}_{inf}|\text{transects})) + \text{E}(\text{Var}(\hat{B}_{inf}|\text{transects})) \tag{2.17}$$

Note that

$$\text{E}(\hat{B}_{inf}|\text{transects}) = N\frac{\sum_{i=1}^{n} T_i}{n}$$

because $\hat{T}_i$ is unbiased for T_i. Then

$$\text{Var}(\text{E}(\hat{B}_{inf}|\text{transects})) = N^2\text{Var}\left(\frac{\sum_{i=1}^{n} T_i}{n}\right) \tag{2.18}$$

Because the T_i's are a simple random sample, then by Cochran (1977),

$$\text{Var}\left(\frac{\sum_{i=1}^{n} T_i}{n}\right) = \frac{S_T^2(1-f)}{n}$$

where $S_T^2 = \sum_{i=1}^{N}(T_i - \overline{T})/(N-1)$ and $f = n/N$. As a sample estimate of this variance, it is customary to take

$$\widehat{\text{Var}}\left(\frac{\sum_{i=1}^{n} T_i}{n}\right) \approx \frac{s_T^2(1-f)}{n} \tag{2.19}$$

where $s_T^2 = \sum_{i=1}^{n}(\hat{T}_i - \bar{\hat{T}})/(n-1)$. Using (2.19), an estimate of (2.18) is

$$\widehat{\text{Var}}(\text{E}(\hat{B}_{inf}|\text{transects})) = N^2 \frac{s_T^2(1-f)}{n} \tag{2.20}$$

An estimate of E(Var($\hat{B}_{inf}$|transects)) must now be determined. Because

$$\text{Var}(\hat{B}_{inf}|\text{transects}) = N^2 \frac{1}{n} \frac{\sum_{i=1}^{n} \text{Var}(\hat{T}_i)}{n} \tag{2.21}$$

it follows that

$$\text{E}(\text{Var}(\hat{B}_{inf}|\text{transects})) = N^2 \frac{1}{n} \frac{\sum_{i=1}^{N} \text{Var}(\hat{T}_i)}{N} \tag{2.22}$$

so an estimate of (2.22) is just given by (2.21). The variance of $\hat{B}_{inf}$ is then

$$\text{Var}(\hat{B}_{inf}) = N^2 \left(\frac{S_T^2(1-f)}{n} + \frac{1}{n} \frac{\sum_{i=1}^{N} \text{Var}(\hat{T}_i)}{N} \right) \tag{2.23}$$

and the estimated variance is

$$\widehat{\text{Var}}(\hat{B}_{inf}) = N^2 \left(\frac{s_T^2(1-f)}{n} + \frac{1}{n^2} \sum_{i=1}^{n} \widehat{\text{Var}}(\hat{T}_i) \right) \tag{2.24}$$

2.3.3 Application

Before applying the ratio and inflation estimators, it is useful to note that each estimator is more suitable for differently shaped beaches. Because the inflation estimator finds an estimate of the number of razor clams along any transect, it is more suitable for beaches that have transects with equal length. Thus, this estimator should be used when the beach is rectagular in shape. However, because the ratio estimator depends on the number of razor clams per m^2, it is more suitable for non-rectangular shaped beaches.

Method	Estimate (millions)	Estimated Standard Error (millions)	Estimated Standard Error (Poisson Assumption) (millions)
Ratio	33.8	7.60	7.58
Inflation	31.3	6.70	6.68

Table 2.2: 1994 North Beach estimates of the total number of razor clams with lengths greater than 4 mm using a straight line interpolation.

Upon applying the results for both estimation methods to the data collected from North Beach in 1994 for razor clams with lengths greater than 4 mm, the associated estimates and estimated standard errors are given in Table 2.2. Comparison of the ratio estimate to the inflation estimate shows that the inflation estimate has a slightly smaller estimate and estimated standard error. In most cases, this characteristic is not expected since estimation of the number of razor clams on the beaches surveyed is more suited to a ratio estimator due to their non-rectangular shape. However, from Figure 2.3, it appears that a ratio estimator would not perform well due to the plot's non-linearity and non-zero intercept.

Chapter 3

Estimation Using Smoothing Splines

Another model for the density of razor clams along a transect is fit using cubic smoothing splines. Variance estimates for each transect are then computed using bootstrap techniques. These concepts are developed and applied to data collected in a 1994 survey of North Beach for razor clams with lengths greater than 4 *mm*.

3.1 Smoothing Splines

From Figure 2.1, it is clear that the average number of razor clams sampled per quadrat along a transect depends non-linearly on distance from chart datum. To model this association, a smoother f_i can be used to provide a nonparametric estimate of the number of razor clams along transect i. A model for this type of analysis is given by

$$Y_{ijk} = f_i(d_{ij}) + \epsilon_{ijk}$$

where d_{ij} is the j^{th} distance sampled along transect i, the errors ϵ_{ijk} are independent of the d_{ij}'s, $\mathrm{E}(\epsilon_{ijk}) = 0$ and $\mathrm{Var}(\epsilon_{ijk}) = \sigma_{ij}^2$. Once a suitable smoother f_i is found for the transect, the area under this smooth curve is then calculated to find an estimate of the total number of razor clams along the transect.

3.1.1 Cubic Smoothing Splines

To find a smoother that provides a nonparametric estimate of the number of razor clams along a transect, the concept of a smoother is introduced. A smoother is used to nonparametrically model the trend of a response variate as a function of one or more covariates. Common types of smoothing techniques include bin smoothers, running mean smoothers, kernel smoothers and smoothing splines. Once a smoothing method is chosen, the amount of smoothing in a neighbourhood of each response value must be chosen. This is usually decided through the use of a smoothing parameter.

For this analysis, a cubic smoothing spline is used to find an estimate of the total number of razor clams along a transect because it is a natural generalization to the straight line interpolation estimate used in Chapter 2. However, it should be noted that other smoothers, such as a kernel smoother, are suitable as well. To find a cubic smoothing spline, a function $f_i(d_{ij})$ must be found that minimizes the penalized least squares criterion that is given by

$$\sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} (Y_{ijk} - f_i(d_{ij}))^2 + \lambda \int_a^b (f_i''(t))^2 \, dt \tag{3.1}$$

where λ is the smoothing parameter and $a \leq d_{i1} \leq ... \leq d_{in_i} \leq b$. In order for such a function to exist, f_i must have a first derivative that is absolutely continuous and must also have $\int_a^b (f_i''(t))^2 \, dt < \infty$. In this case, $\hat{f}_i$ is a piecewise cubic polynomial with knots at each of the d_{ij}'s.

Before the function f_i can be found, the smoothing parameter must be specified. Examination of (3.1) indicates that the first term measures the closeness of the function f_i to the data while the second term penalizes any curvature in the function f_i. Intuitively, it can then be seen that as $\lambda \rightarrow \infty$, the second term dominates so it forces $f_i''(t) = 0$. Hence, the only way to minimize (3.1) is to set f_i to be the linear least squares line. Conversely, as $\lambda \rightarrow 0$, the first term dominates. In this case, to minimize (3.1), $f_i(d_{ij}) = \overline{Y}_{ij}$ so the solution becomes the interpolated line. Thus, it seems that larger values of λ produce smoother curves while smaller values of λ produce more wiggly curves.

A common method that is used to select the smoothing parameter λ uses cross validation techniques. Silverman (1985) indicated that the basic strategy behind cross validation is to leave each data point out one at a time and then choose the value of λ under which the missing data points are best predicted by the remainder of the data. A method related to cross validation that is also commonly used is generalized cross validation. This method approximates the cross validation technique using asymptotic properties. Silverman (1985) and Wahba (1990) both give detailed discussions of these and other related methods.

Unfortunately, techniques like these have been shown to fail by many researchers; see for example Hardle, Hall and Marron (1988) and Wahba (1990). In fact, Wahba suggests that because the generalized cross validation technique is asymptotically justified, good results can not be expected for very small sample sizes as the linear least squares line or a curve that interpolates the points is just as good as a smoothing spline. Wahba also found that through the use of Monte Carlo studies, there are reasonable estimates of λ when sample sizes are greater than 25. However, even for sample sizes of 50, it has been shown that some Monte Carlo replications produce extreme estimates of λ. Because the number of quadrats sampled at any one distance is quite small and there are very few distances sampled along each transect, the smoothing parameter was fixed throughout this analysis to avoid problems such as these.

To fix the smoothing parameter, it is easier to use the degrees of freedom associated with the smoother. The simplest definition of the degrees of freedom is $\mathrm{tr}(\mathbf{S})$ where $\mathbf{S}$ is the smoother matrix that produces the fit (ie. $\hat{Y}_{ijk} = \mathrm{S}(Y_{ijk})$) where $\mathbf{S} = (\mathrm{S}_{ij})$. To find an appropriate choice for the degrees of freedom, arbitrary values of the degrees of freedom are selected until one is found that gives a visually satisfying fit to the data. Some disadvantages to this approach is that there are no optimality properties that can be attributed to the selected value and that it is also very time consuming. However, problems with extreme values for the smoothing parameter are also avoided.

3.1.2 Weighted Smoothing Splines

It has been previously shown that the number of razor clams at each distance along a transect can be adequately modelled using a Poisson distribution with a mean that varies according to the distance from chart datum. In general, to fit a cubic smoothing spline to data with heterogeneous variances, unequal weight must be assigned to each of the observations. Eubank (1988) indicates that it is common to give each Y_{ijk} a weight of $1/Var(Y_{ijk}) = 1/{\sigma_{ij}}^2$ if the σ_{ij}'s are known. In this case, the penalized weighted least squares criterion is

$$\sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \frac{1}{{\sigma_{ij}}^2}(Y_{ijk} - f_i(d_{ij}))^2 + \lambda \int_a^b (f_i''(t))^2 \, dt$$

However, in most cases, the σ_{ij}'s are unknown and iterative weighted smoothing techniques, like those found in Silverman (1985), must be used to determine appropriate weights. Because the razor clam data is approximately Poisson distributed, techniques such as these will be used to calculate the appropriate weights given to each observation.

3.2 Estimation of Transect Total

Once an appropriate smoother $\hat{f}_i$ is found to fit the razor clam data along transect i, an estimate of the transect total is easily found using Riemann sums. To compute a Riemann sum, a partition of $[0, d_{in_i}]$ must be created. For this analysis, the partition P is given by $\{x_0, x_1, ..., x_n\}$ where

$$x_0 = 0 \text{ and } x_j = x_{j-1} + \frac{d_{in_i}}{10,000}$$

A Riemann sum of the smoother $\hat{f}_i$ over $[0, d_{in_i}]$ relative to P gives an estimate of the number of razor clams on transect i and is given by

$$\hat{T}_i = \sum_{j=1}^{n} \hat{f}_i(c_j)(x_j - x_{j-1})$$

where

$$c_j = \frac{x_{j-1} + x_j}{2}$$

3.3 Variance Estimation of Transect Total

It is common for the variance of a smoother to be given in terms of pointwise standard error bands. Because results from this analysis must be compared with the analysis used in Chapter 2, a method must be found to determine an estimated variance for the estimate of the total number of razor clams along a specific transect. Although information from the pointwise standard error bands can be used to develop variance estimates analytically, it is a rather complicated procedure. However, variance estimates can be easily found using bootstrap techniques.

To find the bootstrap estimate of the total number of razor clams along a transect, sampling with replacement must be done on each set of quadrat counts at each distance sampled along the transect. This produces a new set of counts at each distance sampled along the transect. Once this procedure is applied to all distances sampled along the transect, a cubic smoothing spline is then fitted to the new set of data. By repeating this procedure 250 times, 250 different estimates of the total number of clams along the transect are found. An estimate of the variance of the transect estimate is then found by determining the sample variance of the bootstrap transect estimates.

Transect	Estimate	Estimated Bootstrap Standard Error
0.3	3,624.1	150.3
1.7	983.5	88.8
3.0	1,358.6	69.3
3.8	1,512.8	120.0
3.9	3,406.0	200.1
4.5	3,340.3	128.5
6.6	1,018.5	70.2

Table 3.1: 1994 North Beach transect estimates using cubic smoothing splines.

3.4 Application

To illustrate the smoothing analysis just defined, cubic smoothing splines are fitted to each of the transects sampled in the 1994 survey of North Beach for razor clams with lengths greater than 4 *mm* using functions defined in S-plus. Once transect estimates and estimated standard errors are found, estimates of the total number of razor clams with lengths greater than 4 *mm* on North Beach in 1994 are computed using the estimators given in Section 2.3. A comparison of these estimates with those found in Table 2.1 and Table 2.2 is then given.

To determine an appropriate cubic smoothing spline that fits the razor clam data accurately, the smoothing parameter given by the degrees of freedom of the smooth was first determined. By selecting various values for the degrees of freedom, it was found that 3 degrees of freedom provided a good fit to the data for all transects sampled along North Beach. Through the use of Splus functions gam() and gam.predict(), a cubic smoothing spline was then fit for each transect sampled on North Beach in 1994. Figure 3.1 illustrates the number of razor clams counted at each sampled site and also gives the fitted cubic smoothing spline.

Using the transect estimates and standard errors given in Table 3.1, estimates of the total number of razor clams on North Beach using the ratio and inflation estimator were found and are given in Table 3.2. Clearly, the transect estimates and estimated

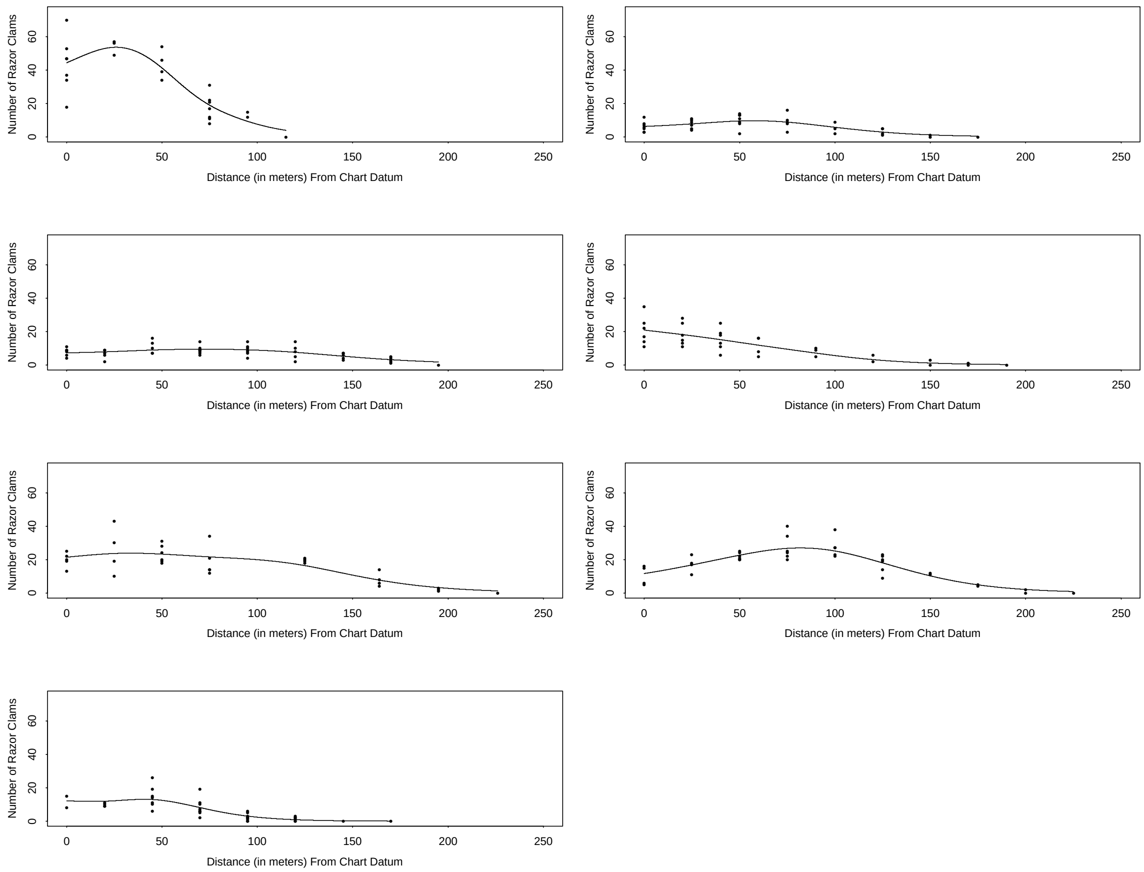

Figure 3.1: Number of razor clams sampled per quadrat versus distance (in meters) from chart datum for each of the 7 transects sampled on North Beach in 1994 with the fitted smoothing spline.

Method	Estimate (millions)	Estimated Bootstrap Standard Error (millions)
Ratio	33.9	7.52
Inflation	31.4	6.64

Table 3.2: 1994 North Beach estimates of the total number of razor clams with lengths greater than 4 mm using cubic smoothing splines.

bootstrap standard errors derived from the smoothing spline are comparable with those found in Table 2.1.

Chapter 4

Sub-domain Estimation

Because commercial diggers will often dig above 1 m from chart datum, it is of interest to determine the number of razor clams in this elevation range. To determine estimates of the number of razor clams for specific elevation ranges or sub-domains, the subset of data for each specific range must first be determined. Once the subset of data is found, the methodology given in Chapter 2 and Chapter 3 is applied. This chapter will specifically carry out such an analysis using the 1994 survey data for North Beach for elevations above 1 m for razor clams with lengths greater than 4 mm.

4.1 Data Modification

Throughout the analysis in Chapter 2 and Chapter 3, the covariate of interest is the distance from chart datum. To apply this analysis to a specific elevation range, the distances that cover this range must first be determined. To begin, the relationship between elevation above chart datum and distance from chart datum must be examined. A plot of distance versus elevation is plotted in Figure 4.1 for the 1994 survey data for North Beach.

In each of the seven transect plots, it is obvious that there is a positively sloped linear relationship between elevation and distance. To determine the distances in a

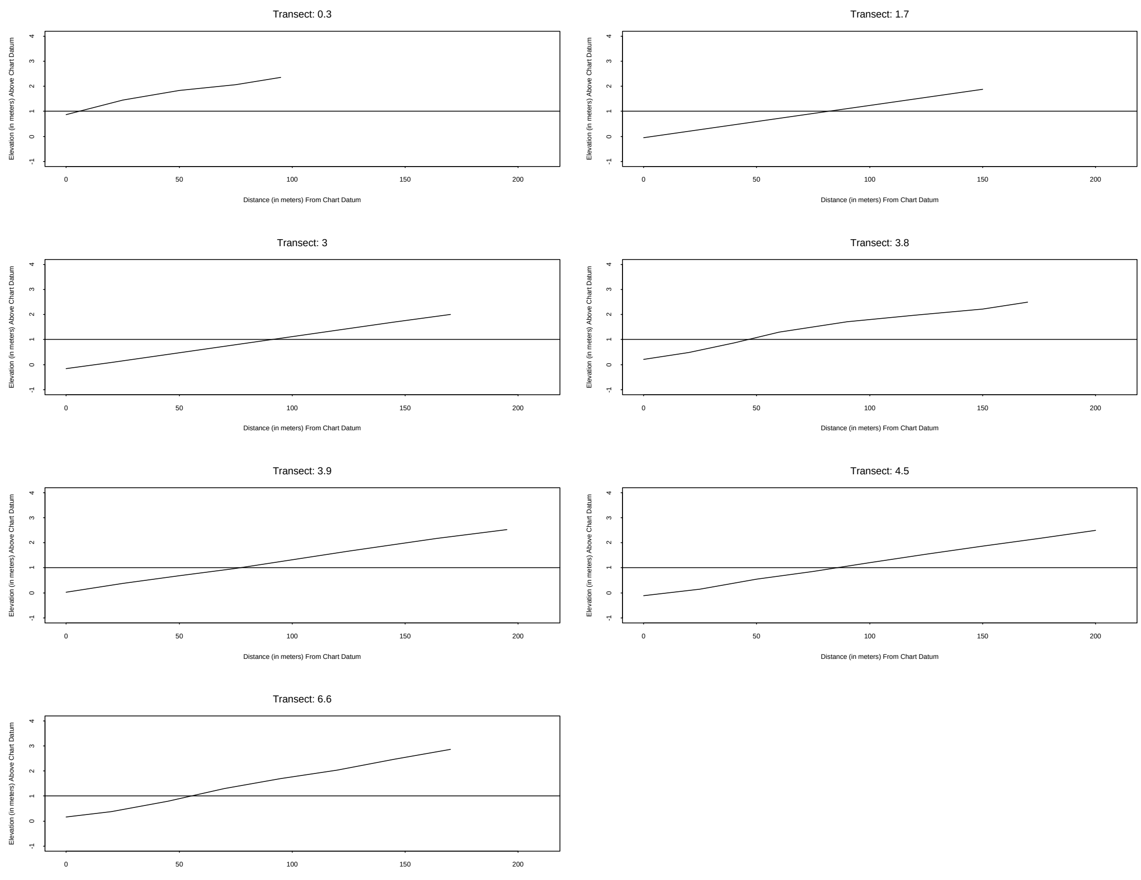

Figure 4.1: Elevation (in meters) above chart datum versus distance (in meters) from chart datum for each of the 7 transects sampled on North Beach in 1994. A line is drawn to indicate 1 m above chart datum.

transect that cover elevations above 1 m, the distance with an elevation of 1 m must first be found. This is not a problem if a distance that was sampled has an elevation of 1 m or if all distances sampled within the transect are above 1 m. In both cases, the data would not have to be modified and only the data with elevations at or above 1 m would be used. Another case where data modification is not needed is when the transect has no sampled quadrats with elevations above 1 m. In this case, transects of this type are not used for estimation purposes. However, for most transects, the distance from chart datum at an elevation of 1 m and the associated average number of razor clams and standard error is not exactly known and must be estimated using the data found for that particular transect.

Estimation of these quantities begins by defining $e_{i,l-1}$ as the elevation sampled in the i^{th} transect that is just below 1 m and $e_{i,l+1}$ as the elevation sampled in the i^{th} transect that is just above 1 m. Associated with these elevations are the distances from chart datum, given by $d_{i,l-1}$ and $d_{i,l+1}$ respectively and the average number of razor clams sampled at those distances, given by $\overline{Y}_{i,l-1}$ and $\overline{Y}_{i,l+1}$ respectively. To estimate the distance having an elevation of 1 m, given by $\hat{d}_{il}$, the equation of the line joining the two points $(d_{i,l-1}, e_{i,l-1})$ and $(d_{i,l+1}, e_{i,l+1})$ is first found. Let this line be given by

$$e = m_d d + b_d$$

where

$$m_d = \frac{e_{i,l+1} - e_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}$$

and

$$b_d = e_{i,l-1} - m_d d_{i,l-1}$$

Making obvious substitutions, the equation of the line between the points $(d_{i,l-1}, e_{i,l-1})$ and $(d_{i,l+1}, e_{i,l+1})$ is given by

$$e = \frac{e_{i,l+1} - e_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}(d - d_{i,l-1}) + e_{i,l-1} \tag{4.1}$$

To find an estimate of the distance having an elevation of 1 m, set $e = 1$ in (4.1) and solve for d. This gives

$$\hat{d}_{il} = \frac{(1 - e_{i,l-1})(d_{i,l+1} - d_{i,l-1})}{e_{i,l+1} - e_{i,l-1}} + d_{i,l-1} \tag{4.2}$$

Now that an estimate of the distance from chart datum having an elevation of 1 m is found, an estimate of the average number of razor clams at that distance must also be calculated. To do this, the equation of the line joining the two points $(d_{i,l-1}, \overline{Y}_{i,l-1})$ and $(d_{i,l+1}, \overline{Y}_{i,l+1})$ is used. Let this line be given by

$$\overline{Y} = m_Y d + b_Y$$

where

$$m_Y = \frac{\overline{Y}_{i,l+1} - \overline{Y}_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}$$

and

$$b_Y = \overline{Y}_{i,l-1} - m_Y d_{i,l-1}$$

Making obvious substitutions, the equation of the line between the points $(d_{i,l-1}, \overline{Y}_{i,l-1})$ and $(d_{i,l+1}, \overline{Y}_{i,l+1})$ is given by

$$\overline{Y} = \frac{\overline{Y}_{i,l+1} - \overline{Y}_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}(d - d_{i,l-1}) + \overline{Y}_{i,l-1} \tag{4.3}$$

To find an estimate of the average number of razor clams sampled at the estimated distance given by $\hat{d}_{il}$, set $d = \hat{d}_{il}$ in (4.3) and solve for $\overline{Y}$. This gives

$$\begin{aligned} \hat{\overline{Y}}_{il} &= \frac{\overline{Y}_{i,l+1} - \overline{Y}_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}(\hat{d}_{i,l} - d_{i,l-1}) + \overline{Y}_{i,l-1} \\ &= \left(\frac{\hat{d}_{i,l} - d_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}\right)\overline{Y}_{i,l+1} + \left(1 - \frac{\hat{d}_{i,l} - d_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}\right)\overline{Y}_{i,l-1} \end{aligned} \tag{4.4}$$

It follows that the variance of $\hat{\overline{Y}}_{il}$ is just

$$\text{Var}(\hat{\overline{Y}}_{il}) = \left(\frac{\hat{d}_{i,l} - d_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}\right)^2 \text{Var}(\overline{Y}_{i,l+1}) + \left(1 - \frac{\hat{d}_{i,l} - d_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}\right)^2 \text{Var}(\overline{Y}_{i,l-1})$$

by independence of the $\overline{Y}_{ij}$'s and by assuming that the estimated distance with an elevation of 1 m is the true value. An estimate of the variance for the estimated average number of razor clams at an elevation of 1 m is then

$$\widehat{\mathrm{Var}}(\hat{\overline{Y}}_{il}) = \left(\frac{\hat{d}_{i,l} - d_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}\right)^2 \widehat{\mathrm{Var}}(\overline{Y}_{i,l+1}) + \left(1 - \frac{\hat{d}_{i,l} - d_{i,l-1}}{d_{i,l+1} - d_{i,l-1}}\right)^2 \widehat{\mathrm{Var}}(\overline{Y}_{i,l-1}) \quad (4.5)$$

It should be noted that the estimated variance given in (4.5) is used to determine estimates of the non parametric transect totals and standard errors when $\mathrm{Var}(\overline{Y}_{ij})$ is estimated using the sample variance of the quadrats sampled. To carry out a parametric analysis (i.e. where the number of razor clams at each distance is assumed to be Poisson distributed), an alternate estimate of $\widehat{\mathrm{Var}}(\hat{\overline{Y}}_{il})$ uses estimates of $\widehat{\mathrm{Var}}(\overline{Y}_{i,l-1}) = \overline{Y}_{i,l-1}/n_{i,l-1}$ and $\widehat{\mathrm{Var}}(\overline{Y}_{i,l+1}) = \overline{Y}_{i,l+1}/n_{i,l+1}$ in (4.5).

By applying the above modification to all transects, the set of data for each transect to be used for sub-domain estimation becomes the estimated distance, estimated average number of razor clams and associated estimated variance having an elevation of 1 m and all data with an associated elevation above 1 m.

4.2 Estimation of Transect and Beach Totals

Application of the straight line interpolation method discussed in Chapter 2 to the data found for elevations above 1 m is quite simple. Using data that is modified according to the procedure outlined in Section 4.1, estimated transect totals of the number of razor clams above 1 m and estimated standard errors are then computed using the techniques outlined in Section 2.2. These results are given in Table 4.1.

Application of the cubic smoothing spline method discussed in Chapter 3 to the razor clam data is even simpler to calculate. This analysis involves fitting a cubic smoothing spline with no modification to the transect data. Once an appropriate smoother is found, the estimate of the transect total for the sub-domain is found using predict.gam(). Because the area underneath the smoother calculates the transect total, then the area underneath the smoother starting at the estimated distance given in (4.2) is needed for sub-domain estimation. Thus, prediction of the smoother is only

Transect	Estimate	Estimated Standard Error	Estimated Standard Error (Poisson Assumption)
0.3	3,400.6	137.65	147.6
1.7	283.0	52.3	46.5
3.0	561.2	47.3	44.9
3.8	717.3	90.5	75.5
3.9	1656.8	82.7	153.7
4.5	1587.8	76.4	92.4
6.6	343.1	42.1	36.3

Table 4.1: 1994 North Beach transect estimates for elevations greater than 1 m above chart datum using a straight line interpolation.

needed at distances at and above the distance estimated to have an elevation of 1 m. Variance estimation using bootstrap techniques follows in a similar way. The results of this analysis are given in Table 4.2.

Estimates of beach totals for the straight line interpolation method and smoothing spline method are then easily computed from the above transect estimates using the procedure outlined in Section 2.3. It should be noted that for the ratio estimator, the lengths of each transect had to be modified to reflect that only a section of the transect was used for sub-domain estimation. The results of this analysis are given in Table 4.3 and Table 4.4. Upon comparing the estimates using straight line interpolation and cubic smoothing splines, all estimates seem to be quite similar.

Transect	Estimate	Estimated Bootstrap Standard Error
0.3	3,402.8	131.5
1.7	275.2	50.6
3.0	564.8	47.9
3.8	706.8	84.2
3.9	1,710.3	65.4
4.5	1581.4	76.0
6.6	328.1	44.0

Table 4.2: 1994 North Beach transect estimates for elevations greater than 1 m above chart datum using cubic smoothing splines.

Method	Estimate (millions)	Estimated Standard Error (millions)	Estimated Standard Error (Poisson Assumption) (millions)
Ratio	18.0	6.12	6.13
Inflation	17.6	6.07	6.07

Table 4.3: 1994 North Beach estimates of the total number of razor clams with lengths greater than 4 mm for elevations greater than 1 m above chart datum using a straight line interpolation.

Method	Estimate (millions)	Estimated Bootstrap Standard Error (millions)
Ratio	18.1	6.15
Inflation	17.6	6.10

Table 4.4: 1994 North Beach estimates of the total number of razor clams with lengths greater than 4 mm for elevations greater than 1 m above chart datum using cubic smoothing splines.

Chapter 5

Determination of Optimal Allocation of Effort

The data used throughout this analysis was obtained by randomly sampling transects, systematically sampling distances from chart datum along the sampled transects and randomly sampling quadrats at each sampled distance. Clearly, this type of sampling allows the number of transects, the number of distances and the number of quadrats sampled at each distance to be chosen by the sampler. To determine the optimum allocation of effort among these three areas, the estimated variance of the beach total or the cost of conducting the survey can be minimized. This is done under various simplifying assumptions and using 1994 survey data collected on razor clams with lengths greater than 4 mm in North Beach.

5.1 Definition of The Optimality Problem

To determine the optimal allocation of effort among transects, distances and quadrats, either the variance of the beach total or the cost of conducting the survey is minimized. Such optimization problems must have a variance function and a cost function defined.

5.1.1 Variance Simplification

Two variances that could be used when determining the optimal allocation of effort are the variance of the ratio estimator and the variance of the inflation estimator. However, it can be shown that both variances are, in fact, equivalent under certain simplifying assumptions. Hence, the results of the analyses using either variance to determine the optimal allocation of effort among transects, distances and quadrats are equivalent. To show this characteristic, the variance given by the ratio estimator will first be simplified. It is given by

$$\mathrm{Var}(\hat{B}_{ratio}) = 4A^2 \left(\frac{(1-f)}{n\overline{L}^2} \frac{\sum_{i=1}^{N}(T_i - XL_i)^2}{N-1} + \frac{\frac{1}{n}\frac{\sum_{i=1}^{N}\mathrm{Var}(\hat{T}_i)}{N}}{\left(\frac{\sum_{i=1}^{N} L_i}{N}\right)^2} \right)$$

where

$$\mathrm{Var}(\hat{T}_i) = \frac{{b_1}^2}{4}\mathrm{Var}(\overline{Y}_{i1}) + \frac{\sum_{j=2}^{n_i-1}\mathrm{Var}(\overline{Y}_{ij})(b_{j-1}+b_j)^2}{4} + \frac{b_{n_i-1}^2}{4}\mathrm{Var}(\overline{Y}_{in_i})$$

To simplify the above variance, the following assumptions are made.

I The number of distances sampled along each transect is constant such that $n_i = n_d$.

II The length of each transect is constant such that $L_i = L_t$.

III The distance between each sampled distance along a transect is constant such that $b_j = b$. Consequently, $b = L_t/(n_t - 1)$.

IV The number of quadrats sampled at each distance along a transect is constant such that $n_{ij} = n_s$.

V The variance at each elevation is constant such that $\text{Var}(\overline{Y}_{ij}) = v/n_s$ where $Var(Y_{ijk}) = v$.

Making appropriate substitutions and simplifications, the variance for the estimate of the beach total using the ratio estimator becomes

$$\text{Var}(\hat{B}_{ratio}) = 4A^2 \left(\frac{(1-f)}{n{L_t}^2} S_T^2 + \frac{v(n_d - 1.5)}{nn_s(n_d-1)^2} \right) \tag{5.1}$$

where $f = n/N$ and $S_T^2 = \sum_{i=1}^{N} (T_i - \overline{T})^2/(N-1)$.

At this point, it should be noted that because Assumption II states that the length of the transects are equal, then the area of the beach, given by A, is equivalent to $(NL_t)/2$. Upon substituting this expression into (5.1), the variance for the estimate of the beach total using the ratio estimator becomes

$$\text{Var}(\hat{B}_{ratio}) = N^2 \left(\frac{(1-f)}{n} S_T^2 + \frac{L_t^2 v(n_d - 1.5)}{nn_s(n_d-1)^2} \right) \tag{5.2}$$

which is also the variance of the beach estimate using an inflation estimator once the above assumptions are applied to (2.23). As a result, either variance can be used to optimize the allocation of effort among transects, distances and quadrats.

To determine the fixed quantities in (5.2), data from the 1994 survey of North Beach is used. Known quantities include the total number of transects, N, along the beach $(14,400)$. However, all other fixed quantities are not exactly known and must be estimated. Estimation of transect length is easily done by averaging the lengths of all transects sampled. Thus,

$$\begin{aligned} \hat{L}_t &= \frac{\sum_{i=1}^{n} L_i}{n} \\ &\approx 185\ m \end{aligned}$$

Estimation of the variance among transects is done by taking the sample variance of the transect estimates. Thus,

$$\begin{aligned} s_T^2 &= \frac{\sum_{i=1}^{n}(\hat{T}_i - \hat{\overline{T}})^2}{n-1} \\ &\approx 1,500,000 \end{aligned}$$

Estimation of the variance at each quadrat is done by taking the average of the sample variances of the sampled quadrat counts at each distance. Thus,

$$\begin{aligned} \hat{v} &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} s_{ij}^2}{\sum_{i=1}^{n}\sum_{j=1}^{n_i} n_{ij}} \\ &\approx 23.3 \end{aligned}$$

where $s_{ij}^2 = \sum_{k=1}^{n_{ij}}(Y_{ijk} - \overline{Y}_{ij})/(n_{ij}-1)$. Using these estimates as the population values, all fixed quantities in (5.2) are now known. By substituting in all relevant quantities, the variance function given in (5.2) is

$$\text{Var}(\hat{B}_{ratio}) = 14,400^2 \left(\frac{\left(1 - \frac{n}{14,400}\right)}{n} 1,500,000 + \frac{(185^2)(23.3)(n_d - 1.5)}{n n_s (n_d - 1)^2} \right) \tag{5.3}$$

5.1.2 Cost Constraints

Based on data collected from the 1994 survey of North Beach, it took on average 70 minutes to sample a transect, 5 minutes to sample a distance along a transect and 4 minutes to sample a quadrat. Letting C represent the total cost (in minutes) to survey a beach, then the cost function is given by

$$C = 70n + 5nn_d + 4nn_dn_s \tag{5.4}$$

5.2 Variance Minimization

To optimally allocate effort among transects, distances and quadrats using variance minimization, an estimate of the total cost of the survey must be found. From the 1994 survey of North Beach, a total of 7 transects, 52 distances (approximately 7 distances sampled along each transect) and 269 quadrats (approximately 5 quadrats sampled at each distance) were sampled. Under this allocation, the standard error of the beach estimate given in (5.3) would be 7.26 million and is fairly close to the estimates of the standard error found in Table 2.2. Using the estimates of the amount of time taken to sample a transect, a distance and a quadrat, the average total cost of the 1994 survey was 1826 minutes. Thus, the optimization problem to be used is:

Minimize

$$\text{Var}(\hat{B}_{ratio}) = 14,400^2 \left(\frac{\left(1 - \frac{n}{14,400}\right)}{n} 1,500,000 + \frac{(185^2)(23.3)(n_d - 1.5)}{n n_s (n_d - 1)^2} \right)$$

subject to

$$\begin{aligned} 70n + 5nn_d + 4nn_dn_s &\leq 1826 \\ n &\geq 1 \\ n_d &\geq 2 \\ n_s &\geq 1 \end{aligned}$$

where $n, n_d, n_s \in Z$. Using a Newton-Raphson optimization procedure, the solution to this problem requires 19 transects, 2 distances along each transect and 2 quadrats per distance to be sampled. Such a solution gives a minimum standard error of 4.65 million with an estimated survey cost of 1824 minutes. Upon comparing this allocation with what was used in the 1994 survey of North Beach, it is clear that considerably higher numbers of transects and lower numbers of distances and quadrats are needed to be sampled to minimize the variance while keeping the cost of the survey at a similar level.

It should be noted that this allocation of effort among transects, distances and quadrats is not surprising because all of the assumptions that were made to simplify the variance of the beach estimate implied that the distribution of razor clams along a transect remained fairly constant. Intuitively, it is clear that sampling a larger number of distances and quadrats along a transect would not improve variance estimates while sampling more transects would. This characteristic is also seen when the cost of surveying a beach is minimized while a constraint is imposed on the estimated variance of the beach estimate.

5.3 Cost Minimization

To optimally allocate effort among transects, distances and quadrats using cost minimization, an estimate of the variance of the beach total must be found for the 1994 survey of North Beach. From Table 2.2, the estimated standard error was found to be 7.60 million. Thus, the optimization problem to be used is:

Minimize

$$\text{Cost} = 70n + 5nn_d + 4nn_dn_s$$

subject to

$$\begin{aligned}
\text{Var}(\hat{B}_{ratio}) = 14,400^2 \left(\frac{\left(1 - \frac{n}{14,400}\right)}{n} 1,500,000 + \frac{(185^2)(23.3)(n_d - 1.5)}{nn_s(n_d-1)^2} \right) &\leq 7,600,000^2 \\
n &\geq 1 \\
n_d &\geq 2 \\
n_s &\geq 1
\end{aligned}$$

where $n, n_d, n_s \in Z$. Using a Newton-Raphson optimization procedure, the solution to this problem requires 9 transects, 2 distances along each transect and 1 quadrat per distance to be sampled. Such a solution gives a minimum standard error of 7.15 million with an estimated survey cost of 792 minutes. Upon comparing this allocation with what was used in the 1994 survey of North Beach, it is clear that to achieve a similar variance but with a much smaller survey cost, a similiar number of transects must still be sampled but with much less sampling of distances and quadrats. However, upon comparison of this allocation to what was found in Section 5.2, it is found that although sampling of distances and quadrats is still low, the number of transects sampled is reduced. This is not surprising since this allocation minimizes the cost of the survey and the amount of time needed to sample a transect is quite high.

Chapter 6

Discussion and Summary

Because the health of the razor clam stock on beaches near Masset, British Columbia seemed to be failing in 1993, an assessment of the number of razor clams on these beaches needed to be employed. Although some of the more complicated sampling strategies outlined in Chapter 1 offered alternative sampling techniques, it was clear that a three stage sampling design offered many more advantages. More importantly, this design also allowed data to be collected on the distribution of razor clams as distance from chart datum varied. Because data of this type is heavily affected by the biological requirements of a razor clam, it followed that a biologically based statistical model of the distribution of razor clams could also be used to determine the total number of razor clams on the beaches surveyed.

The model that was first used on the data that was collected near Masset, British Columbia (Schwarz et al., 19xx) used an unbalanced mixed effect model. However, this analysis did not intuitively incorporate any biological knowledge that was gained from the three stage sampling design. To develop such a biologically based statistical model, the current methodology was developed.

Through the use of straight line interpolation and smoothing spline techniques, curves were first constructed to model the approximate monotone decreasing relationship that existed between razor clam density and distance from chart datum (or elevation above chart datum) along a transect. The total number of razor clams along a transect was then easily found using simple integration of the curve. Using sampling estimators, like those outlined in Cochran (1977), intuitive estimates for the total number of razor clams on a beach were derived. Because an analysis of this type is so intuitive and allows the relationship between the density of razor clams and the distance from chart datum to be modeled, this analysis may be preferred over the unbalanced mixed effect model.

To compare the current methodology to the unbalanced mixed effect model, the concepts developed in Chapter 2 and Chapter 3 were applied to the data obtained from surveys of North Beach, South 1 Beach and South 2 Beach between the years 1994 and 1996. The results from this analysis under the various estimation techniques are given in Table 6.1, 6.2 and 6.3. Because little modification is needed to develop estimates of the number of razor clams above a specifed elevation, concepts developed in Chapter 4 were not applied to the data collected from the surveys.

Examination of the estimates given in Tables 6.1, 6.2 and 6.3 reveal several features. Comparison of beach estimates using ratio and inflation estimators show that most estimates seem to agree. However, there seems to be a large difference in the estimates given for North Beach in 1996. This is not surprising because the plot of the transect estimates by transect lengths for North Beach in 1996 is scattered quite randomly. This indicates that the inflation estimator, and not the ratio estimator, should be used for estimation purposes. It is also interesting to note that although ratio estimates would seem to give more precise estimates than the inflation estimates due to the non rectangular shaped beaches that were surveyed, this is not the case for some of the surveys conducted. This is most likely due to the relationship between transect length and the estimated total number of razor clams along the transect being non linear (i.e. quadratic) or not passing through the origin when a linear relationship actually existed. However, it should be noted that where a linear relationship that passed through the origin did exist, estimated standard errors for the beach estimates using the ratio estimator were quite low compared to those found using the inflation estimator. Comparison of the parametric and non-parametric estimated standard errors using the straight line interpolation technique also shows no striking differences.

Year	Beach	Method	Straight Line Interpolation			Smoothing Spline	
			Estimate	Estimated Standard Error	Estimated Standard Error (Poisson)	Estimate	Estimated Bootstrap Standard Error
1994	North	Ratio	33.85	7.60	7.58	33.88	7.51
		Inflation	31.33	6.70	6.68	31.36	6.64
	South 1	Ratio	5.64	2.37	2.37	5.67	2.42
		Inflation	5.96	2.78	2.78	6.00	2.83
	South 2	Ratio	2.39	1.05	1.02	2.38	1.06
		Inflation	2.59	1.69	1.67	2.58	1.69
1996	North	Ratio	20.33	2.00	1.95	20.04	1.91
		Inflation	13.73	1.57	1.54	13.53	1.48
	South 1	Ratio	4.30	0.69	0.68	4.26	0.70
		Inflation	3.72	0.44	0.42	3.69	0.44
	South 2	Ratio	2.21	0.90	0.91	2.17	0.93
		Inflation	2.28	0.73	0.74	2.24	0.76

Table 6.1: Estimates (in millions) of the total number of razor clams with lengths greater than 4 mm for all beaches surveyed between 1994 and 1996.

Year	Beach	Method	Straight Line Interpolation			Smoothing Spline	
			Estimate	Estimated Standard Error	Estimated Standard Error (Poisson)	Estimate	Estimated Bootstrap Standard Error
1994	North	Ratio	12.84	1.35	1.34	12.79	1.32
		Inflation	11.65	1.51	1.50	11.61	1.48
	South 1	Ratio	3.31	0.49	0.49	3.29	0.48
		Inflation	3.37	0.45	0.45	3.35	0.43
	South 2	Ratio	0.63	0.22	0.20	0.60	0.24
		Inflation	0.68	0.37	0.36	0.65	0.37
1995	North	Ratio	15.37	1.20	1.19	15.42	1.19
		Inflation	12.68	1.70	1.69	12.72	1.71
	South 1	Ratio	4.69	0.31	0.28	4.57	0.25
		Inflation	4.87	0.64	0.63	4.75	0.55
1996	North	Ratio	13.40	1.53	1.49	13.23	1.48
		Inflation	9.05	0.91	0.89	8.93	0.86
	South 1	Ratio	4.29	0.69	0.68	4.25	0.70
		Inflation	3.71	0.44	0.42	3.68	0.43
	South 2	Ratio	2.21	0.90	0.91	2.17	0.93
		Inflation	2.28	0.73	0.74	2.24	0.76

Table 6.2: Estimates (in millions) of the total number of razor clams with lengths greater than 20 *mm* for all beaches surveyed between 1994 and 1996.

Year	Beach	Method	Straight Line Interpolation			Smoothing Spline	
			Estimate	Estimated Standard Error	Estimated Standard Error (Poisson)	Estimate	Estimated Bootstrap Standard Error
1994	North	Ratio	4.73	0.76	0.76	4.68	0.75
		Inflation	4.02	0.78	0.78	3.97	0.76
	South 1	Ratio	1.50	0.39	0.39	1.46	0.37
		Inflation	1.53	0.44	0.44	1.49	0.42
	South 2	Ratio	0.26	0.063	0.055	0.23	0.065
		Inflation	0.28	0.10	0.095	0.25	0.084
1995	North	Ratio	3.72	0.58	0.56	3.66	0.57
		Inflation	3.07	0.52	0.51	3.02	5.14
	South 1	Ratio	1.86	0.38	0.38	1.86	0.37
		Inflation	1.82	0.33	0.33	1.82	0.32
1996	North	Ratio	4.49	0.42	0.44	4.40	0.38
		Inflation	2.98	0.36	0.37	2.92	0.33
	South 1	Ratio	2.14	0.55	0.55	2.10	0.55
		Inflation	1.79	0.35	0.35	1.76	0.35
	South 2	Ratio	1.72	0.69	0.70	1.69	0.71
		Inflation	1.77	0.55	0.57	1.74	0.58

Table 6.3: Estimates (in millions) of the total number of razor clams with lengths greater than 90 *mm* for all beaches surveyed between 1994 and 1996.

Similarily, a comparison of the estimates and estimated standard errors while using a straight line interpolation or smoothing spline to model the distribution of razor clams along a transect shows similar results. Comparison of the estimates for all beaches also reveals that North Beach has the most amount of razor clams, followed by South 1 Beach and then South 2 Beach. This is not surprising due to the area, lengths and locations of the beaches.

Based on the preceding conclusions, several recommendations for future analyses of the data obtained from the surveys of beaches surrounding Masset, British Columbia can be made. Because estimates using straight line interpolation or smoothing spline techniques are so similar, it is suggested that the smoothing spline technique developed in Chapter 3 be used because the relationship between the density of razor clams along a transect is more suitable to a smooth curve. However, it should be noted that the amount of time taken to apply the straight line interpolation technique is considerably less than the time needed to apply the smoothing spline technique. Hence, use of either technique is a trade off between the interpretation of the technique used and the amount of time given to the analysis of the data. It is also clear that ratio estimates gave more precise estimates of the total number of razor clams on a beach when a linear relationship passing through the origin was found between transect estimates and transect lengths. This suggests that techniques using a regression estimator be developed for those beaches not possessing such a relationship. However, if no other technique is developed, it is suggested that for those cases where a ratio estimator is unsuitable, the inflation estimator should be used. Lastly, a comparison of the estimates between the unbalanced mixed effect model, given in Appendix B, and the techniques developed in the preceding discussion show that the estimates seem to differ slightly. This is probably due to the fact that the unbalanced mixed effect model does not explicitly model the relationship between the transect estimates and the transect lengths. Thus, it is suggested that the current methodology be used to obtain a more realistic analysis of the data for future surveys of this type.

Appendix A

North Beach 1994 Raw Survey Data For Clams of All Sizes

Transect	Elevation	Distance	Quadrat Counts						
0.3	0.87	0	37	47	53	34	47	70	18
0.3	1.45	25	56	56	57	49			
0.3	1.84	50	34	39	54	46			
0.3	2.06	75	22	21	31	8	17	12	11
0.3	2.36	95	15	12					
1.7	-0.06	0	5	3	8	3	12	7	6
1.7	0.27	25	7	9	5	11	4	4	10
1.7	0.59	50	14	2	13	11	9	13	8
1.7	0.91	75	10	8	16	9	3		
1.7	1.24	100	2	5	9				
1.7	1.56	125	5	2	1	5			
1.7	1.88	150	1	0					
3.0	-0.16	0	11	8	9	4	6	9	
3.0	0.09	20	9	6	6	6	9	2	7
3.0	0.41	45	7	7	16	7	13	10	10

Transect	Elevation	Distance	Quadrat Counts						
3.0	0.73	70	9	14	6	7	8	10	9
3.0	1.05	95	10	4	7	14	9	8	11
3.0	1.37	120	2	10	10	8	8	5	14
3.0	1.69	145	6	7	4	3	7		
3.0	2.01	170	4	1	3	3	5	2	
3.8	0.21	0	14	11	35	22	25	17	
3.8	0.48	20	25	13	28	18	15	11	
3.8	0.87	40	11	19	18	25	6	13	
3.8	1.30	60	16	5	16	8			
3.8	1.72	90	10	5	9				
3.8	1.97	120	6	2	2				
3.8	2.22	150	3	0	0				
3.8	2.50	170	1	0	0				
3.9	0.03	0	19	22	13	25	20		
3.9	0.38	25	30	30	19	43	10		
3.9	0.68	50	24	28	19	18	31	20	
3.9	0.98	75	14	12	14	34	21		
3.9	1.66	125	21	19	20	18			
3.9	2.17	164	4	14	6	8			
3.9	2.53	195	3	2	1				
4.5	-0.12	0	5	6	16	15			
4.5	0.14	25	11	18	23	17			
4.5	0.54	50	20	22	22	20	25	21	24
4.5	0.86	75	22	20	40	24	25	34	
4.5	1.20	100	27	22	22	27	38	23	
4.5	1.54	125	9	14	20	19	22	23	
4.5	1.86	150	11	11	12				
4.5	2.18	175	4	5	5				
4.5	2.50	200	2	0	0				
6.6	0.16	0	15	15	15	8			

Transect	Elevation	Distance	Quadrat Counts								
6.6	0.38	20	9	9	9	10	9	11			
6.6	0.79	45	10	19	6	11	26	14	15	15	
6.6	1.30	70	5	2	10	11	6	7	6	19	
6.6	1.69	95	3	1	5	5	2	0	6	2	0
6.6	2.04	120	2	0	0	0	1	2	3	0	
6.6	2.46	145	0	0	0	0	0	0			
6.6	2.87	170	0								

Appendix B

Estimation of Beach Total Using An Unbalanced Mixed Effect Model

The estimates given in the following tables are taken from Schwarz et al. (19xx).

Year	Beach	Estimate	Estimated Standard Error
1994	North	34.01	10.22
	South 1	5.61	2.89
	South 2	1.91	1.16
1996	North	16.68	1.99
	South 1	3.63	0.80
	South 2	2.28	0.39

Table B.1: Estimates (in millions) of the total number of razor clams with lengths greater than 4 *mm* for all beaches surveyed between 1994 and 1996 using an unbalanced mixed effect model.

Year	Beach	Estimate	Estimated Standard Error
1994	North	12.38	1.57
	South 1	3.30	0.85
	South 2	0.61	0.21
1995	North	12.76	1.47
	South 1	3.98	0.42
1996	North	10.69	0.76
	South 1	3.62	0.80
	South 2	2.28	0.39

Table B.2: Estimates (in millions) of the total number of razor clams with lengths greater than 20 mm for all beaches surveyed between 1994 and 1996 using an unbalanced mixed effect model.

Year	Beach	Estimate	Estimated Standard Error
1994	North	4.07	0.63
	South 1	1.40	0.48
	South 2	0.34	0.09
1995	North	2.83	0.31
	South 1	1.41	0.25
1996	North	3.41	0.35
	South 1	1.66	0.56
	South 2	1.73	0.34

Table B.3: Estimates (in millions) of the total number of razor clams with lengths greater than 90 mm for all beaches surveyed between 1994 and 1996 using an unbalanced mixed effect model.

Bibliography

Ayres, D.L. and Simons, D.D. (1988). *The spring and fall 1988 recreational razor clam fisheries with a review of razor clam sampling methodology.* Washington Department of Fisheries. Progress Report 276.

Bourne, N. (1969). *Population Studies on the Razor Clam at Masset, British Columbia.* Fish. Res. Board Can. Tech. Rep. No. 232.

Cochran, W. G. (1977). *Sampling Techniques*, 3rd edition. New York: John Wiley.

Elliott, J. M. (1977). *Some methods for the Statistical Analysis of Samples of Benthic Invertebrates*, 2nd edition. Ambleside, England: Freshwater Biological Association.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* New York: Marcel Dekker, Inc..

Gillespie, G. E. and Kronlund, A. R. (1999). *The Manual for Intertidal Clam Surveys.* Canadian Technical Report of Fisheries and Aquatic Sciences, No. 2270.

Hardle, W., Hall, P. and Marron, S. (1988). How far are the optimally chosen smoothing parameters from their optimum? *Journal of the American Statistical Association* **83**, 86-95.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models.* New York: Chapman and Hall.

Schwarz, C.J., Jones, R.R. and Lee, L. (19xx). Intertidal population estimate of razor clams (*Siliqua patula*) at North Beach, Haida Gwaii/Queen Charlotte Islands and applications to fishery management). Unpublished manuscript.

Silverman, B.W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society* **47**, 1-21.

Szarzi, N.J. (1991). *Distribution and abundance of the Pacific razor clam (Siliqua patula Dixon), on the Eastside Cook Inlet beaches, Alaska.* Master's thesis. Fairbanks: University of Alaska.

Taylor, L.R. (1961). Aggregation, variance and the mean. *Nature* **189**, 732-735.

Wahba, G. (1990). *Spline Models for Observational Data.* Philadelphia: Society for Industrial and Applied Mathematics.