Part 5. Properties of Procedures

5.1. Introduction

We here assess the performance of the various statistical methods. Theory indicates that the desirable properties of each method become increasingly better as sample size increases. In these methods, sample size is not a single value; rather, it is related to the size of the releases (R_{ui}), the capture probabilities (p_{ui}), the number of sampling occasions (k), and (to a lesser degree) the survival probabilities (ϕ_{ui}). Furthermore, the methods perform best when the assumptions are fully met (see Manly 1970, 1971b).

Here we examine performance of estimators in terms of bias, precision, robustness, relative efficiency, and the effects of heterogeneity (terms to be defined later). The power of tests used in selecting the proper model and their nominal significance levels are also assessed. We use the Monte Carlo method and a method based on expectations in Part 5; however, some analytical results on power of certain tests are given in Part 3.

It is not possible to perform Monte Carlo studies on all, or even a significant fraction, of the cases of potential interest. A sequence of models has been developed under each of the four protocols. Each model has several estimators and associated tests. In each instance, k can reasonably range from two to at least six or eight; and we have seen studies where the numbers released vary from as few as 60 to as many as 100,000 individuals. Previous studies have had capture probabilities ranging from 0.01 to 0.8, whereas treatment survival rates have ranged from 0.3 to nearly 1.0. The parameter space suggested above is far too large for exhaustive Monte Carlo study (or even analytical-numerical investigation). Furthermore, we could not afford to tabulate these results, even if adequate computer time allowed the study of a large number of cases. Therefore, our approach here has been to tabulate the results of a few representative cases of potential interest.

Our approach in evaluating the statistical properties of various methods is twofold. First, a powerful Monte Carlo simulation procedure (PROC SIMULATE) is included in program RELEASE (see Part 9). Researchers are encouraged to use this procedure in both the design and analysis phases of their experiments. Second, we present the results of 48 Monte Carlo studies to give the reader an impression of the performance of various methods. This material focuses on experiments with two groups where the treatment effect is acute. The parameters for various simulated models are presented in Table 5.1. In all studies, k = 5, 1,000 replicate data sets were generated, and there were no unplanned losses on capture. These simulated data sets constitute the basis for much of the material presented in Part 5.

Table 5.1. – Summary of the parameters used in the Monte Carlo studies for four cases. Data were simulated under both the first and complete capture history protocols.

Case and parameters	True model[a]	Subpopulation[b]	ϕ_{t1} and ϕ_{c1}	p_{t1} and p_{c1}
Case A: $R_{t1} = R_{c1} = 5{,}000$				
Homogeneous	H_0		All 0.9	All 0.1
Homogeneous	$H_{1\phi}$		All 0.9, except $\phi_{t1} = 0.8$	All 0.1
Homogeneous	H_{2p}		All 0.9, except $\phi_{t1} = 0.8$	All 0.1, except $p_{t2} = 0.15$
Heterogeneous	H_0	1	All 0.95	All 0.15
		2	All 0.85	All 0.05
Heterogeneous	$H_{1\phi}$	1	All 0.95, except $\phi_{t1} = 0.85$	All 0.15
		2	All 0.85, except $\phi_{t1} = 0.75$	All 0.05
Heterogeneous	H_{2p}	1	All 0.95, except $\phi_{t1} = 0.85$	All 0.15, except $p_{t2} = 0.225$
		2	All 0.85, except $\phi_{t1} = 0.75$	All 0.05, except $p_{t2} = 0.075$

Case B: Same as case A, except $R_{t1} = R_{c1} = 1{,}000$
Case C: Same as case A, except $R_{t1} = R_{c1} = 200$
Case D: Same as case C, except all p_{t1} and p_{c1} are increased by 0.7

[a]In the heterogeneous cases, the parameter structure follows models H_0, $H_{1\phi}$, and H_{2p}; however, the assumption of independence is violated.

[b]Each subpopulation is of equal size, $R_{v1}/2$.

5.2. Estimator Bias

Bias in estimators of model parameters is undesirable. In this chapter we examine the statistical bias of MLEs under models for the first and complete capture history protocols for cases where the parameters are homogeneous. Bias is defined as

$$\text{bias} = E(\hat{\theta}) - \theta,$$

where θ is a particular parameter (e.g., S or ϕ). The exact value of θ is known in each of the Monte Carlo studies and the expected value can be estimated with good precision from the results of the Monte Carlo studies. The expected value is estimated as

$$\hat{E}(\hat{\theta}) = \frac{1}{1{,}000} \sum_{i=1}^{1{,}000} \hat{\theta}_i,$$

where $\hat{\theta}_i$ is the estimate computed from the ith Monte Carlo trial. We used 1,000 replicate data sets for each Monte Carlo study to ensure that the expected values are estimated precisely.

Examination of the estimated expected values of the treatment survival probabilities in Tables 5.2 and 5.3 indicate that bias is negligible if the model assumptions are met. The bias is generally ≤ 0.004 if the numbers released were $\geq 1{,}000$. An exception is $\hat{S} = \hat{\phi}_{ti}/\hat{\phi}_{ci}$ when homogeneous first capture history data are generated under model H_{2p}. Here, the biases are -0.049 and -0.040 under $\hat{S}$ for model $H'_{2\phi}$ (Table 5.2). No estimator for model H_{2p} exists under the first capture history protocol, thus, model H'_{2p} is used as an approximation. Generally the estimator $\hat{S}$ performs similarly for both first and complete capture history protocols. When the true model is $H_{1\phi}$, the estimators of S are algebraically identical under the first and complete capture history protocols. However, estimated values of S under the first and complete capture history protocols in Table 5.2 (case A) under $H_{1\phi}$ are not identical because they are not based on exactly the same Monte Carlo data.

With only 200 releases in each group, bias in $\hat{S}$ is larger if the capture probability is low (Table 5.3, case C). However, the bias remains small (about 1%) if the assumptions of the model are realized. Comparison of bias in cases C and D in Table 5.3 shows the importance

Table 5.2. – Summary of estimated expected values and standard errors (computed empirically and theoretically) from Monte Carlo studies of complete and first capture history protocols for case A, R_{v1} = 5,000 and case B, R_{v1} = 1,000. The survival (ϕ_i) and capture (p_i) probabilities were allowed to be homogeneous or heterogeneous. Parameters used to simulate the data are given in Table 5.1.

True model	Capture history protocol	Parameter variation	Estimator performance under model $H_{1\phi}$ $\hat{E}(\hat{S})$	$\hat{se}(\hat{S})_e$	$\hat{se}(\hat{S})_t$	Estimator performance under model H_{2p} or $H'_{2\phi}$ $\hat{E}(\hat{S})$[a]	$\hat{se}(\hat{S})_e$	$\hat{se}(\hat{S})_t$
			Case A					
H_0	Complete	Homogeneous	1.000	0.033	0.033	1.002	0.039	0.038
(S = 1.0)		Heterogeneity	1.001	0.030	0.032	1.001	0.034	0.036
$H_{1\phi}$	Complete	Homogeneous	0.889	0.031	0.030	0.888	0.037	0.035
(S = 0.889)		Heterogeneity	0.891	0.028	0.030	0.891	0.032	0.034
H_{2p}	Complete	Homogeneous	1.004	0.034	0.033	0.889	0.036	0.036
(S = 0.889)		Heterogeneity	0.996	0.030	0.032	0.890	0.032	0.033
H_0	First	Homogeneous	1.000	0.033	0.033	1.001	0.045	0.043
(S = 1.0)		Heterogeneity	1.000	0.030	0.032	1.000	0.039	0.042
$H_{1\phi}$	First	Homogeneous	0.890	0.030	0.030	0.891	0.040	0.039
(S = 0.889)		Heterogeneity	0.891	0.028	0.030	0.891	0.037	0.039
H_{2p}	First	Homogeneous	1.006	0.033	0.033	0.840	0.039	0.038
(S = 0.889)		Heterogeneity	0.997	0.030	0.032	0.825	0.034	0.037
			Case B					
H_0	Complete	Homogeneous	1.003	0.074	0.074	1.004	0.086	0.086
(S = 1.0)		Heterogeneity	1.004	0.066	0.073	1.003	0.074	0.082
$H_{1\phi}$	Complete	Homogeneous	0.893	0.068	0.068	0.889	0.078	0.079
(S = 0.889)		Heterogeneity	0.894	0.061	0.067	0.895	0.067	0.076
H_{2p}	Complete	Homogeneous	1.009	0.073	0.074	0.889	0.077	0.078
(S = 0.889)		Heterogeneity	0.999	0.066	0.072	0.895	0.066	0.075
H_0	First	Homogeneous	1.001	0.077	0.074	1.002	0.099	0.096
(S = 1.0)		Heterogeneity	1.009	0.069	0.073	1.006	0.069	0.073
$H_{1\phi}$	First	Homogeneous	0.892	0.068	0.068	0.898	0.087	0.089
(S = 0.889)		Heterogeneity	0.895	0.065	0.068	0.897	0.085	0.088
H_{2p}	First	Homogeneous	1.011	0.074	0.074	0.849	0.086	0.086
(S = 0.889)		Heterogeneity	0.998	0.069	0.072	0.826	0.080	0.083

[a]Estimates were computed under model H_{2p} for the complete capture history protocol and under model $H'_{2\phi}$ for the first capture history protocol.

Table 5.3. – Summary of estimated expected values and standard errors (computed empirically and theoretically) from Monte Carlo studies of complete and first capture history protocols for case C, $R_{\tau 1}$ = 200 and case D, $R_{\tau 1}$ = 200. The survival (ϕ_i) and capture (p_i) probabilities were allowed to be homogeneous of heterogeneous. Parameters used to simulate the data are given in Table 5.1.

True model	Capture History protocol	Parameter variation	Estimator performance under model $H_{1\phi}$ $\hat{E}(\hat{S})$	$H_{1\phi}$ $se(\hat{S})_e$	$H_{1\phi}$ $se(\hat{S})_t$	H_{2p} or $H'_{2\phi}$ $\hat{E}(\hat{S})$[a]	H_{2p} or $H'_{2\phi}$ $se(\hat{S})_e$	H_{2p} or $H'_{2\phi}$ $se(\hat{S})_t$
			Case C					
H_0	Complete	Homogeneous	1.011	0.174	0.168	1.014	0.199	0.196
(S = 1.0)		Heterogeneity	1.015	0.157	0.166	1.015	0.182	0.187
$H_{1\phi}$	Complete	Homogeneous	0.911	0.161	0.158	0.912	0.186	0.183
(S = 0.889)		Heterogeneity	0.905	0.154	0.154	0.908	0.170	0.174
H_{2p}	Complete	Homogeneous	1.032	0.173	0.171	0.912	0.184	0.181
(S = 0.889)		Heterogeneity	1.012	0.167	0.166	0.908	0.169	0.172
H_0	First	Homogeneous	1.016	0.171	0.169	1.034	0.231	0.225
(S = 1.0)		Heterogeneity	1.006	0.154	0.164	1.022	0.216	0.219
$H_{1\phi}$	First	Homogeneous	0.896	0.158	0.155	0.905	0.209	0.204
(S = 0.889)		Heterogeneity	0.908	0.148	0.155	0.918	0.201	0.204
H_{2p}	First	Homogeneous	1.024	0.170	0.170	0.870	0.200	0.199
(S = 0.889)		Heterogeneity	1.005	0.157	0.164	0.836	0.183	0.190
			Case D					
H_0	Complete	Homogeneous	1.000	0.038	0.038	1.000	0.039	0.038
(S = 1.0)		Heterogeneity	1.001	0.037	0.038	1.000	0.037	0.038
$H_{1\phi}$	Complete	Homogeneous	0.889	0.041	0.041	0.889	0.042	0.041
(S = 0.889)		Heterogeneity	0.890	0.040	0.041	0.890	0.040	0.042
H_{2p}	Complete	Homogeneous	0.895	0.041	0.041	0.889	0.041	0.041
(S = 0.889)		Heterogeneity	0.894	0.040	0.041	0.888	0.040	0.041
H_0	First	Homogeneous	0.998	0.037	0.037	1.032	0.261	0.243
(S = 1.0)		Heterogeneity	0.998	0.038	0.038	1.030	0.262	0.247
$H_{1\phi}$	First	Homogeneous	0.889	0.042	0.041	0.918	0.232	0.223
(S = 0.889)		Heterogeneity	0.892	0.041	0.041	0.915	0.237	0.228
H_{2p}	First	Homogeneous	0.895	0.040	0.041	0.682	0.193	0.182
(S = 0.889)		Heterogeneity	0.900	0.041	0.041	0.654	0.183	0.180

[a]Estimates were computed under model H_{2p} for the complete capture history protocol and under model $H'_{2\phi}$ for the first capture history protocol.

of a high capture probability if the numbers released are small. A general conclusion from the results in Tables 5.2 and 5.3 is that bias in $\hat{S}$ is probably acceptable in most instances if the model assumptions are satisfied. In addition, if no treatment effect occurs (model H_0) and one tries to estimate the treatment survival rate by using models $H_{1\phi}$ or H_{2p}, the estimate of S will be close to 1.00 on the average.

Although any bias is undesirable, it must be related to the standard error to assess more correctly its importance. The following table, adapted from Cochran (1963:14), shows the effect of the ratio bias/standard error on expected 95% CI coverage:

Bias/standard error	Expected coverage
0.02	0.95
0.04	0.95
0.06	0.95
0.08	0.95
0.10	0.95
0.20	0.95
0.40	0.93
0.60	0.91
0.80	0.87
1.00	0.83
1.50	0.68

As an example, consider the bias in $\hat{S}$ under case C, model $H'_{2\phi}$ for first capture history data for the constant-parameter case in Table 5.3, where the true model is H_{2p}. The bias is 0.870 - 0.889 = -0.019 and its average standard error is 0.200. Thus, the ratio of bias to standard error is 0.095; and we determine that a 95% CI would still have an expected coverage of about 95%.

The general conclusion is that bias in $\hat{S}$ is relatively small if the assumptions of a particular model are satisfied. An option (UNBIAS) in PROCs CHMATRIX, LMREAD, and SIMULATE of program RELEASE allows the computation of bias-adjusted estimators. However, all results in this chapter are for the unadjusted MLEs. Use of the UNBIAS option substantially reduces the already small statistical bias (see Chapter 3.4). Bias can be severe if a poor model, whose assumptions are not met, is used (e.g., case C, model $H_{1\phi}$ for first capture history data for the constant-parameter case in Table 5.3 where the true model is H_{2p}). Such model bias can be seen by examining Tables 5.2 and 5.3.

Approximations to the bias of an estimator can be computed using the EXPECT option in program RELEASE. Results can be obtained in a few seconds using this option, whereas an adequate Monte Carlo simulation can frequently take several hours. This subject is discussed briefly in Chapter 5.4 and more fully in Chapter 5.9

5.3. Measures of Precision

The sampling variance and its square root, the standard error, are measures of the precision of an estimator. True sampling variance can be estimated in different ways (e.g., empirically or by using theoretical formulae). The ML method assures us that the estimated variance or standard error based on theoretical formulae will have good properties as the sample size increases, if model assumptions hold. Here we assess the performance of the theoretical formulae for estimating sampling variance over a range of smaller sample sizes by first computing an empirical standard error of the treatment survival rate $\hat{se}(\hat{S})_e$ based on 1,000 replications from the Monte Carlo studies. Thus,

$$\hat{se}(\hat{S})_e = \left[\frac{1}{999}\sum_{i=1}^{1,000}(\hat{S}_i - \overline{S})^2\right]^{1/2},$$

where $\overline{S}$ is the average of the 1,000 estimates of S. This quantity is a measure of precision of the estimator of treatment survival rate under a particular model. In contrast, a theoretical estimate of the sampling variance and standard error is available, based on the model and ML theory (Section 1.2.1.2). The average of these,

$$\hat{se}(\hat{S})_t = \frac{1}{1,000}\sum_{i=1}^{1,000}\hat{se}(\hat{S}_i),$$

computed from the Monte Carlo studies, provides a basis for assessing average performance of the theoretical variance as a measure of precision.

Examination of the estimated standard errors in Tables 5.2 and 5.3 indicates that, on the average, the theoretical standard errors are satisfactory as a measure of estimator precision, even in case C and the heterogeneous populations, if the assumptions of the model are met.

5.4. Estimator Robustness

Estimator robustness relates to a large collection of issues. Many questions regarding the robustness of an estimator are specific to a particular study. In this chapter we present some general results that may be useful.

Experience with the general Jolly-Seber model and its extensions indicates that the survival rates (ϕ_i) are relatively well estimated, in terms of bias and precision, compared to the number of births and the population size. Estimates of population size or the estimated number of marked animals at a particular time may be seriously biased by failure of model assumptions but the corresponding estimate of survival is not often affected. Generally, the survival estimators (ϕ_{wi}) have good properties. Estimators of treatment survival rate (S) are

ratios of the $\hat{\phi}_i$, and therefore might also be expected to have good statistical properties.

In general, if a model such as H_0 or $H_{1\phi}$ is the correct model, then estimators of S or the ϕ_i from other models in the sequence (e.g., H_{2p}, $H_{2\phi}$, H_{3p}) will be little biased; however, some precision will be lost. Problems with substantial bias occur when the true model is H_{2p}, $H_{2\phi}$, or a more general model in the sequence (e.g., $H_{3\phi}$), but the estimation is conducted under a more restrictive model such as $H_{1\phi}$. Cases A, B, and C in Tables 5.2 and 5.3 illustrate the situation where data sets were generated under model H_{2p} (S = 0.889) and estimates of treatment survival were made under model $H_{1\phi}$. The $\hat{E}(\hat{S})$ under model $H_{1\phi}$ for the homogeneous populations was about 1.00 to 1.030, a bias of 12 to 16%.

The above example indicates the problem of using an inappropriate model. The lack of estimator robustness forces additional attention on the tests for model selection. Finally, although estimators of treatment survival were biased by 12-16% in the example above, it must be noted that this bias was caused by the substantial inequality of capture probabilities (p_{t2} being 50% higher than p_{c2}). Thus, in this sense, $\hat{S}$ may be considered to be somewhat robust.

We have not attempted to present measures of robustness when mortality due to a treatment is chronic or indirect. These issues are left to the reader for specific studies. Clearly, potential bias exists and, in some cases, a better measure of treatment effect is $\hat{\phi}_{c1} - \hat{\phi}_{t1}$ (Section 2.5.2.5).

5.5. Heterogeneity

The models and estimation procedures presented in this monograph assume that all members of the population have homogeneous parameters at a given time and place. For example, all animals in the control group have the same probability of survival from dams 3 to 4, ϕ_{c3}. Alternatively, one could postulate that each animal has a unique parameter, i.e., $\theta_{(i)}$ for i = 1, ..., N at a particular time and place. This situation is referred to as the heterogeneous case where the parameter θ itself has a distribution. This distribution is, of course, unknown to the investigator but the distribution has a conceptual mean and a variance. An example is an annual survival probability where individuals in the population vary in their innate ability to survive. The theory for model building to incorporate heterogeneity directly is quite complex and requires information that is rarely known (e.g., the form and variance of the distribution of θ_i).

We here present information on the performance of the models developed for the homogeneous case when they are used in the analysis of simulated data sets generated to incorporate heterogeneity in the parameters among animals. We generated data sets to allow heterogeneity in both the survival (ϕ_{vi}) and recapture (p_{vi}) parameters (Table 5.1), using program RELEASE. Use of this approach enables us to ask if the MLEs are robust (insensitive) to heterogeneity.

Recent studies (e.g., Nichols et al. 1982; Pollock and Raveling 1982; Vaupel and Yashin 1985) have shown that it is the variance among members of the population that is important, rather than the shape of the distribution. We included this information in the Monte Carlo

studies by defining two subpopulations of equal size, each with different parameter values (see Table 5.1).

The estimators $\hat{S}$ are often less variable with heterogeneous data then with homogeneous data (see every comparison between the empirical variance estimates in cases A, B, and C, Tables 5.2 and 5.3). This result stems from the fact that the effective capture probability, with heterogeneity, exceeds the mean of the p_{vi}, $i = 2, ..., k$, for the two subpopulations (this is a manifestation of Jensen's inequality). As an example, under the complete capture history protocol, case A, the mean value of the $p_{vi} = 0.1$ for model $H_{1\phi}$ (Table 5.1); however, the effective (mean) capture probability is 0.13 under $H_{1\phi}$ (i.e., $\hat{p}_{v2} = \hat{p}_{v3} = 0.13$). Thus, the estimates under heterogeneity are less variable then the theoretical ML variance estimates would indicate. This result is important. In general, heterogeneity has little effect on the performance of the estimators of S or its standard error (Tables 5.2 and 5.3). We did not investigate the possibility of heterogeneity in the treatment effect itself. This might be caused if extreme heterogeneity existed in the marked releases in each group and each subpopulation. We suspect that bias in $\hat{S}$ could be expected in these extreme situations.

In the presence of heterogeneity, the parameters are weakly dependent upon previous capture histories. In addition, the assumption of independence is violated, but the effect on the statistical properties is often quite small. We tentatively conclude that a reasonable amount of heterogeneity in the survival and capture process will not seriously affect the performance of estimators of treatment survival, if the correct model is selected. This pattern is also generally true for the estimators of the ϕ_{vi} parameters.

5.6. Estimator Efficiency

An advantage in considering the sequence of models $H_{1\phi}$, H_{2p}, ..., $H_{k-1,\phi}$ under the complete capture history protocol is that assumptions about the equality of parameters across groups can be relaxed. Under model H_{2p}, one need not assume that treatment and control fish have the same capture probability on the second sampling occasion. Similarly, model H_{3p} also allows $p_{t3} \neq p_{c3}$ on the third sampling occasion. If $p_{t2} = p_{c2}$ and $p_{t3} = p_{c3}$ are incorrectly assumed (e.g., model $H_{1\phi}$), bias will result in $\hat{\phi}_{v1}$ and S. Thus, the choice among alternative models is important.

A disadvantage in selecting a model that allows relaxed assumptions is that the sampling variance of the estimators is larger than that for models entailing stronger assumptions. This loss in precision leads to a consideration of relative efficiency of estimators. The relative efficiency between two models is measured here by the ratio of their standard errors, with the convention that the model with the most assumptions (and, therefore, the smaller standard error) appears in the numerator. An example of the meaning of relative efficiency is the entry 0.79 in the first row in Table 5.4 under case A. The value 0.79 indicates that if model $H_{1\phi}$ is true, but one uses $\hat{S}$ under model H_{2p}, then the estimated standard error is 1/0.79 = 1.27 times larger than it should be (under $H_{1\phi}$).

Some general conclusions can be drawn from the efficiency information in Tables 5.4-5.5. First, relative efficiency is similar for large or small numbers released (R_{v1} = 5,000 or 200, cases A and C, respectively). Second, if p_{vi} is nearly constant over sampling occasions, there is a substantial loss of efficiency when models with increasingly relaxed assumptions are considered. Third, the loss of efficiency is severe in models $H_{2\phi}$, H_{3p}, ..., $H_{k\text{-}1,\phi}$ compared to models $H_{1\phi}$ or H_{2p} for the complete capture history protocol. Consideration of the trade-off between bias and efficiency in the sequence of models, $H_{1\phi}$, H_{2p}, $H_{2\phi}$, ..., $H_{k\text{-}1,\phi}$, has implications in the design of experiments (Part 6).

Table 5.4. – Summary of relative efficiencies (the ratio of standard errors) for $\hat{S}$ under cases A and C (see Table 5.1). The Monte Carlo data were generated under model $H_{1\phi}$ for the first capture history protocol, assuming homogeneous parameters.

	Case A				Case C			
Model	$H_{1\phi}$	$H'_{2\phi}$	$H'_{3\phi}$	$H'_{4\phi}$	$H_{1\phi}$	$H'_{2\phi}$	$H'_{3\phi}$	$H'_{4\phi}$
$H_{1\phi}$	1	0.79	0.58	0.38	1	0.76	0.55	0.31
$H'_{2\phi}$		1	0.74	0.48		1	0.71	0.41
$H'_{3\phi}$			1	0.65			1	0.57
$H'_{4\phi}$				1				1

Table 5.5. – Summary of estimator efficiency (the ratio of standard errors) of $\hat{S}$ for cases A and C (see Table 5.1). The Monte Carlo data were generated under model $H_{1\phi}$ for the complete capture history protocol, assuming homogeneous parameters.

	Case A				Case C			
Model	$H_{1\phi}$	H_{2p}	$H_{2\phi}$	H_{3p}	$H_{1\phi}$	H_{2p}	$H_{2\phi}$	H_{3p}
$H_{1\phi}$	1	0.86	0.27	0.27	1	0.86	0.21	0.21
H_{2p}		1	0.32	0.32		1	0.24	0.24
$H_{2\phi}$			1	0.30			1	0.28
H_{3p}				1				1

5.7. Power of Tests

Statistical tests aid in the rejection of models that are poor for a particular data set. The Monte Carlo studies generally indicate that bias in the estimators is reasonably small if the correct model is used. In contrast, it is easy to show substantial bias if a model does not fit the data (e.g., an estimator $\hat{S}$ under $H_{1\phi}$ has poor properties if the data are from model H_{3p}). Therefore, it is important to assess the power of specific between-model tests. Power is defined as the probability of rejecting the null hypothesis.

If the null hypothesis is true, the power of the test should be equal to the significance level chosen (e.g., if $\alpha = 0.05$, then the power of the test should be 0.05). Table 5.6 provides summaries of the power of between-model tests for cases A to D, respectively. The first general conclusion is that heterogeneity has little effect on the performance of the between-model tests. This result is important as we believe heterogeneity is common in field studies of many biological populations. The power of the between-model tests is excellent for large samples (e.g, $R_t = R_c = 5{,}000$), as shown by the results for case A in Table 5.6. In releases of 1,000 animals per group (case B), the power drops to 0.31 for the test of H_0 versus $H_{1\phi}$ and the power is 0.79 for the test of $H_{1\phi}$ versus H_{2p} for both homogeneous and heterogeneous populations (Table 5.6).

In the worst example (case C, Table 5.6), the power of tests H_0 versus $H_{1\phi}$ and $H_{1\phi}$ versus H_{2p} decreases to 0.10 and 0.24, respectively, for both homogeneous and heterogeneous populations. With releases as small as 200 in each group and low capture probabilities, the power of between-model tests is poor. However, power for experiments with few animals released may be reasonable if the capture probabilities are high (case D, Table 5.6).

Interpretation of tests must be done with caution. For example, in case A the test of H_0 versus $H_{1\phi}$ has power of 0.05 when H_{2p} is the correct model. The power is poor because neither the null nor alternative hypothesis is true; however, there is a treatment effect. Other model parameters could have been chosen for model H_{2p} such that a treatment effect would have been much more easily detected (e.g., $p_{t2} = 0.05$). Thus, the results regarding power of tests must be interpreted with an understanding of the specific parameter values chosen. Knowledge of the approximate power of tests can increase the ability to interpret experimental results (e.g., Chapter 7.4).

The test for an overall treatment effect H_0 versus $H_{k\text{-}1,\phi}$ is powerful for data sets with large numbers of releases (e.g., case A) or where the capture probabilities are high (e.g., case D) for both first and complete capture history protocols (Table 5.7). The power for tests of overall treatment effect is poor for Cases B and C. The results in Table 5.7 show little difference in test power between homogeneous and heterogeneous parameters. These estimates of power are highly specific to the parameter values chosen and the magnitude of the treatment effect; thus, Table 5.7 gives only a rough impression of the power for some specific examples.

Table 5.6. – Estimated power of statistical tests at the 0.05 level of significance for cases A, B, C, and D of the complete capture history protocol. Parameters used to simulate the data are given in Table 5.1.

		True model					
		Homogeneous parameters			Heterogeneous parameters		
Test	Models tested[a]	H_0	$H_{1\phi}$	H_{2p}	H_0	$H_{1\phi}$	H_{2p}
			Case A				
1.*R*1	H_0 versus $H_{1\phi}$	0.05	0.93	0.06	0.05	0.94	0.03
1.*T*2	$H_{1\phi}$ versus H_{2p}	0.05	0.05	1.00	0.04	0.05	1.00
1.*R*2	H_{2p} versus $H_{2\phi}$	0.04	0.05	0.05	0.05	0.05	0.04
1.*T*3	$H_{2\phi}$ versus H_{3p}	0.06	0.05	0.05	0.06	0.06	0.06
1.*R*3	H_{3p} versus $H_{3\phi}$	0.04	0.06	0.05	0.05	0.05	0.05
			Case B				
1.*R*1	H_0 versus $H_{1\phi}$	0.06	0.31	0.05	0.03	0.31	0.29
1.*T*2	$H_{1\phi}$ versus H_{2p}	0.04	0.05	0.79	0.05	0.04	0.79
1.*R*2	H_{2p} versus $H_{2\phi}$	0.05	0.05	0.06	0.06	0.05	0.05
1.*T*3	$H_{2\phi}$ versus H_{3p}	0.05	0.05	0.05	0.06	0.04	0.04
1.*R*3	H_{3p} versus $H_{3\phi}$	0.05	0.06	0.06	0.05	0.05	0.05
			Case C				
1.*R*1	H_0 versus $H_{1\phi}$	0.06	0.10	0.05	0.04	0.10	0.04
1.*T*2	$H_{1\phi}$ versus H_{2p}	0.05	0.05	0.24	0.06	0.06	0.24
1.*R*2	H_{2p} versus $H_{2\phi}$	0.05	0.04	0.05	0.04	0.05	0.04
1.*T*3	$H_{2\phi}$ versus H_{3p}	0.05	0.06	0.06	0.06	0.05	0.05
1.*R*3	H_{3p} versus $H_{3\phi}$	0.05	0.06	0.05	0.05	0.04	0.04
			Case D				
1.*R*1	H_0 versus $H_{1\phi}$	0.05	0.75	0.71	0.05	0.74	0.71
1.*T*2	$H_{1\phi}$ versus H_{2p}	0.05	0.05	0.19	0.05	0.03	0.21
1.*R*2	H_{2p} versus $H_{2\phi}$	0.05	0.06	0.05	0.05	0.06	0.05
1.*T*3	$H_{2\phi}$ versus H_{3p}	0.05	0.06	0.06	0.06	0.05	0.05
1.*R*3	H_{3p} versus $H_{3\phi}$	0.05	0.04	0.04	0.05	0.04	0.04

[a]Null hypothesis versus the alternative hypothesis.

The ability of the statistical tests to select the correct model is largely dependent upon the number of fish released and the magnitude of the capture probabilities. For cases A and D, tests have high power when a treatment effect exists; therefore, the correct model is

Table 5.7. – Power of the test (α = 0.05) for treatment effect H_0 versus $H_{4\phi}$ (TEST 1) for the Monte Carlo data simulated for homogeneous and heterogeneous populations. Parameters used to simulate the data are given in Table 5.1.

	True model					
	First capture history protocol			Complete capture history protocol		
Case	H_0	$H_{1\phi}$	H_{2p}	H_0	$H_{1\phi}$	H_{2p}
	Homogeneous populations					
A	0.06	0.80	1.00	0.05	0.71	1.00
B	0.05	0.19	0.52	0.05	0.16	0.46
C	0.06	0.08	0.12	0.04	0.06	0.12
D	0.05	0.52	0.60	0.06	0.43	0.48
	Heterogeneous populations					
A	0.03	0.79	1.00	0.04	0.70	1.00
B	0.05	0.18	0.60	0.05	0.14	0.45
C	0.05	0.06	0.12	0.04	0.05	0.12
D	0.06	0.52	0.60	0.04	0.40	0.47

selected (Table 5.8). Cases B and C were poor in that insufficient data were available to allow reliable model selection. Model selection appears poorer for the complete compared to the first capture history protocol (Table 5.8). Much of this is only due to the fact that more models are available under the complete capture history protocol, thus, the chance for error is greater.

Table 5.8. – Summary of the probability of selecting the correct model for the four Monte Carlo cases (see Table 5.1). Simulation results are based on 1,000 replications, α = 0.05, for the homogeneous populations.

	True model					
	First capture history protocol			Complete capture history protocol		
Case	H_0	$H_{1\phi}$	H_{2p}	H_0	$H_{1\phi}$	H_{2p}
A	0.83	0.80	0.91	0.72	0.68	0.78
B	0.82	0.30	0.64	0.71	0.22	0.69
C	0.80	0.09	0.17	0.71	0.08	0.18
D	0.82	0.66	0.21	0.69	0.54	0.19

The analytic capabilities (EXPECT and SIMULATE) of program RELEASE should be used to assess the model selection and power of tests before field experiments are performed. The information in Table 5.8 provides only a glimpse of the performance of these statistical methods in selecting a good model, as these results are highly dependent on many factors. The comparable results for the heterogeneous cases were virtually identical with those for the homogeneous cases in Table 5.8.

5.8. Confidence Interval Coverage

Asymptotic theory assures that a 95% CI can be established as

$$\hat{S} \pm 1.96\, \mathrm{se}(\hat{S}).$$

However, the actual or achieved coverage may be less than the nominal 95% level due to bias in $\hat{S}$, nonnormal sampling distribution of $\hat{S}$, a poor estimate of se($\hat{S}$), or sample sizes that are too "small" (not asymptotic). The Monte Carlo simulations (see Table 5.1) provide a means of assessing the coverage achieved.

Achieved coverage is summarized in Table 5.9 for cases A-D for the first and complete capture history protocols for both homogeneous and heterogeneous data. These results indicate that nominal coverage is achieved, even for the small numbers released in case C, if the model is correct. Little difference between homogeneous and heterogeneous parameters was found. Coverage is frequently less than the nominal level if an incorrect model is used (e.g., use of model $H'_{2\phi}$ for the first capture history protocol, when the true model is H_{2p}.

5.9. Analytical-Numerical Approximations

A method based on expectations, presented in Chapter 3.6, provides a useful alternative to Monte Carlo simulations. This method assumes a model, say $H_{1\phi}$, specific values for the model parameters (p_{vi} and ϕ_{vi}), and numbers released R_{v1}. This method can best be described as analytical-numerical. The numerical step involves the computation of the expected m_{vij}-array (i.e., $E[m_{vij} \mid R_{v1}]$). The computed expected values in the m_{vij}-array are then treated as "data." The analytic step involves the computation of estimates and test statistics using either the same model ($H_{1\phi}$) or another model (e.g., $H'_{2\phi}$). Thus, the method can be said to be based on expectations. Papers by Nelson et al. (1980) and Anderson and Burnham (1980) used this convenient approach in closely related contexts. The entire method can be easily done using the EXPECT option in PROC SIMULATE of program RELEASE. A typical run often takes about 11 seconds of microcomputer time, compared with, perhaps, 30-1,000 minutes for a Monte Carlo simulation run.

Table 5.9. – Summary of achieved confidence interval coverage (in %) on $\hat{S}$ for several estimators under the first and complete capture history protocols. Parameters used to simulate the data are given in Table 5.1.

	True model			
	First capture history protocol		Complete capture history protocol	
Case	$H_{1\phi}$	H_{2p}[a]	$H_{1\phi}$	H_{2p}
	Homogeneous populations			
A	97	71	94	94
B	95	89	94	95
C	95	92	95	95
D	95	67	95	95
	Heterogeneous populations			
A	97	58	96	96
B	97	86	97	97
C	96	90	96	95
D	95	63	95	95

[a]Coverage was assessed for $\hat{S}$ under model $H'_{2\phi}$ for the first capture history protocol.

The method can be used to obtain quick approximations of estimator bias and precision and to examine the power of certain tests. This method is quite useful and we recommend it. However, one would often want to refine insights with later Monte Carlo studies.

A comparison of results on estimator bias and precision is given in Table 5.10 for the Monte Carlo method and the method based on expectations. It is clear from the information in Table 5.10 that the quick method provides useful approximations in all cases. Computer runs required to prepare Table 5.10 took about 3 minutes of computer time, while the comparable runs for the Monte Carlo results took about 4,859 minutes (81 hours). Still, the Monte Carlo method has several distinct advantages. First, Monte Carlo studies can examine small-sample properties of estimators and test statistics, while the quick method is asymptotic. Second, the empirical sampling distribution can be studied only with Monte Carlo methods. Third, the numerical-analytical method cannot obtain the correct variances under conditions of heterogeneity. Finally, achieved confidence interval coverage can only be assessed by the Monte Carlo method.

Some results on the power of tests is given in Table 5.11, comparing the method based on expectations with the Monte Carlo method. Here, in particular, the quick method provides very useful results. The agreement between the two approaches is good. We recommend the use of EXPECT in program RELEASE for all but final, refined insights into issues regarding bias, precision, and test power.

Use of the option EXPECT provides chi-square statistics for a particular test. This chi-square value is a noncentrality parameter that can be used with its degrees of freedom and Table 3.4 to assess test power. Interpolation within Table 3.4 is usually satisfactory. Those wishing a copy of an expanded table can write the senior author.

Table 5.10. – Comparison of the Monte Carlo (MC) method with the method based on expectations (EXPECT) in estimating $E(\hat{S})$ and $\text{se}(\hat{S})$ for models H_0, $H_{1\phi}$, and H_{2p} for cases A and C for the complete and first capture history protocols.

			Model $H_{1\phi}$				Model H_{2p} or $H'_{2\phi}$			
			$\hat{E}(\hat{S})$		$\text{se}(\hat{S})$		$\hat{E}(\hat{S})$		$\text{se}(\hat{S})$	
Case	True model	Capture history	MC	EXPECT	MC[a]	EXPECT	MC	EXPECT	MC[a]	EXPECT
A	H_0	Complete	1.000	1.000	0.033	0.033	1.002	1.000	0.039	0.038
	$H_{1\phi}$	Complete	0.889	0.891	0.031	0.030	0.888	0.890	0.037	0.035
	H_{2p}	Complete	1.004	1.002	0.034	0.033	0.889	0.888	0.036	0.034
C	H_0	Complete	1.011	1.000	0.174	0.166	1.014	1.000	0.199	0.195
	$H_{1\phi}$	Complete	0.911	0.915	0.161	0.157	0.912	0.909	0.186	0.182
	H_{2p}	Complete	1.032	0.986	0.173	0.165	0.912	0.869	0.184	0.178
A	H_0	First	1.000	1.000	0.033	0.033	1.001	1.000	0.045	0.043
	$H_{1\phi}$	First	0.890	0.889	0.030	0.030	0.891	0.889	0.040	0.039
	H_{2p}	First	1.006	1.004	0.033	0.033	0.840	0.840	0.039	0.038
C	H_0	First	1.016	1.000	0.171	0.210	1.034	1.000	0.231	0.284
	$H_{1\phi}$	First	0.896	0.873	0.158	0.149	0.905	0.865	0.209	0.190
	H_{2p}	First	1.024	0.982	0.170	0.160	0.870	0.811	0.200	0.182

[a] $\text{se}(\hat{S})_e$ from case A, Table 5.2 and case C, Table 5.3.

Table 5.11. – Comparison of the Monte Carlo (MC) method with the method based on expectations (EXPECT) in estimating test power (α = 0.05) for cases A and C for the complete and first capture history protocols.

Case	True model	Capture history	Test 1.R1 H_0 versus $H_{1\phi}$		Test 1 H_0 versus $H_{4\phi}$	
			MC[a]	EXPECT	MC[b]	EXPECT
A	$H_{1\phi}$	Complete	0.93	0.92	0.71	0.69
	H_{2p}	Complete	0.06	0.05	1.00	1.00
C	$H_{1\phi}$	Complete	0.10	0.07	0.06	0.06
	H_{2p}	Complete	0.05	0.05	0.12	0.10
A	$H_{1\phi}$	First	0.93	0.93	0.80	0.79
	H_{2p}	First	0.05	0.05	1.00	1.00
C	$H_{1\phi}$	First	0.10	0.11	0.08	0.08
	H_{2p}	First	0.04	0.05	0.12	0.14

[a]Partially from cases A and C, Table 5.6.
[b]From Table 5.7.