Part 4. Importance of Replication

4.1. Introduction

The importance of replication and methods for the analysis of replicated experiments are covered in Part 4. The statistical estimation and testing theory for dealing with a single lot ("replicate") appears in Parts 2 and 3. Here, a variety of methods is given to extend the analysis methods to cover replicated experiments. This material is presented in a separate section of the monograph because of the importance of replication in experimental research.

4.1.1. Need for Replication and Multiple Lots

We now deal with issues of replication and multiple lots, statistical methods for combining results over lots, estimation of empirical variances, and some formulae for estimation of variance components. Repeatability, within measured statistical limits of precision, is fundamental for scientific credibility. When sampling variation is substantial, as in these fisheries experiments, there should be many repeated releases of treatment and control fish.

We define a lot to be the pairing of a batch of treatment fish and a batch of control fish released (almost) simultaneously. Prior to release, all the fish in a lot should be as similar as possible (in, for example, species, strain, age, and condition); they should have arisen from a common source, been handled, marked, and transported as a unit and assigned to treatment group randomly. A lot, so defined, is analogous to two paired experimental units in standard statistical terminology. Thus, we are here defining the experimental unit as a batch of fish to which there is a single application of one treatment condition (see Steel and Torrie 1980 for a general discussion of an experimental unit).

To demonstrate repeatability (and estimate precision), two sources of variation must be dealt with. First, there is the within-lot sampling variation, which manifests itself in $\text{var}(\hat{S} \mid S)$. This variation is reduced by increasing either the lot size or the recapture effort or both. Second, there is the possible statistical variation, from lot to lot, in S itself. The effect of this variation on the precision of $\hat{\overline{S}}$ is reduced (and better estimated) only by having multiple lots.

Multiple lots are desirable because S may vary with environmental conditions. If so, a sample of those conditions should be observed. This sample of conditions can either be treated as a random sample, or a design can be imposed in terms of external variables that might affect S.

Multiple lots are also desirable to detect equipment failures or other methodological problems. For example, if nine of 10 lots all released under similar conditions produced estimates of S ranging from 0.83 to 0.94, but one of the 10 lots produced $\hat{S} = 0.5$, you can check to see if a methodological problem occurred (and they can occur). If that lot can be shown to be aberrant, it can be deleted. It is not wise to put all one's fish in a single lot.

Finally, some form of repeated experimentation, such as replicated experimental units, is needed to get empirical estimates of variance. It is not scientifically desirable to use unquestioningly a theoretical variance in environmental studies. Theoretical variance estimators often underestimate the amount of variance occurring in a study. This failure of theory can be attributed to the failure of one or more assumptions. Lack of independence of fish fates might inflate variances (failure of assumption 7), and thus cause the estimated theoretical variances to be too low. Parameter heterogeneity (failure of assumption 12) may result in increased variation. Finally, if one's choice of model is not correct, theoretical variances may not be reliable (assumption 9). With respect to assumptions 7 and 12, the bias of a parameter estimator is less affected by the assumption failures than is its theoretical variance.

4.1.2. Replication and Multiple Lots

4.1.2.1. Replication. – Replicated experimental units are needed in order to estimate residual (or pure) error. However, it is very difficult (if not impossible) to define replication unambiguously because it is really a concept about random selection of experimental units, and the relevant population of experimental units is elusive. Perhaps because of this difficulty, many texts on statistics do not define replication. Some texts, such as Steel and Torrie (1980:Chapter 6), define it so generally that replication is a generic term for almost any sort of repeated experimental unit. Such a general definition is useful; it allows discussion of different levels of replication that are associated with different sources of variation. We decided to use a narrow definition of replication despite some inherent difficulties with it.

We define replication to mean repeated, independent experimental units treated identically so that there are no known differences (which would affect response) between the units before their responses to the experiment are observed. Differences in the responses of n replicate units must then reflect only residual (unexplained) sampling variation. If one looks closely enough, however, there will *always* be identifiable differences among experimental units before a study. Thus, what constitutes replicate units is really a function of our level of ignorance about external variables that might affect the outcome for those experimental units.

It is difficult to achieve replication (as we have defined it) in fisheries survival experiments, especially for the purpose of estimating within-lot sampling variation. Replicate lots should be released at the same time, which often is not possible. If they have been held in separate holding facilities, such facilities may have had their own effects.

For two, or more, lots to be replicates, they must have the same underlying survival (ϕ) and capture (p) parameters. Thus, the same treatment effect S applies to all n replicates and the variation among $\hat{S}_1, ..., \hat{S}_n$ is strictly sampling variation. It follows that $\hat{S}_1, ..., \hat{S}_n$ are independent and identically distributed random variables if replication has been successful.

We would like to have replicate treatment-control releases as a basis for reliably estimating empirical sampling variances and variance components. Because true replication in these experiments is so difficult, theoretical sampling variances have often been used. This is not necessary; it is possible to obtain empirical variances without true replication. Moreover,

it is more important to have multiple lots than to have true replicates. It suffices to have, for example, 10 treatment-control lots released one a day every 3 days. Such a design includes no true replication, but could include either sublots, some natural partitioning of the recapture data, or both. We discuss these approaches below.

4.1.2.2. Sublots. – A lot consists of many fish, often thousands, so potentially there is information on empirical variation from just a single lot. To exploit this potential, one must use multiple, if not unique, marks within a lot.

The optimal case, statistically, occurs when uniquely numbered tags are used. Then the releases in that lot can be partitioned randomly by tag number into, say, five or 10 sublots and parameter estimates from each sublot can be computed. A good approach with large lots would be to partition on the last digit of the number, thus, creating 10 sublots. From these, we get 10 estimates, $\hat{S}_1, ..., \hat{S}_{10}$. Because the releases were all from one lot, each $\hat{S}_i$ estimates the single parameter S. Therefore, the variation among these 10 $\hat{S}_i$ reflects only within-lot sampling variation.

It is necessary to assume the estimators $\hat{S}_1, ..., \hat{S}_n$, based on sublots, are independent. By assumption 7 (fish fates are independent), this is true. However, this assumption could fail. It is safer to assume lots are independent (assumption 8), but it is difficult to believe one can achieve true replicates of lots. There is a tradeoff here. We believe sublots will be the better approximation to (true) replication because then $S_i = S$ is very believable and independence of sublots seems reasonable to us if sublots are large (say 500 to 1,000 or more fish).

4.1.2.3. Multiple lots. – Typically, lots are released at different times, so they cannot be assumed to be replicates for the purpose of estimating the within-lot sampling variation. This is because, for the ith treatment-control release, the treatment effect S_i may differ from other lots. We must now consider how these potentially different S_i are related to one another in order to analyze the data. The simplest case is if we can treat the (unknown) $S_1, ..., S_n$ as a random sample from a distribution with average value $E(S)$ and $\text{var}(S) = \sigma^2$. The case $\sigma^2 = 0$ corresponds to all S_i being the same, which is what we would expect if the n lots were replicates.

The unknown S_i may be viewed as random variables if the multiple lots are released at random with respect to conditions (environmental and engineering) that affect S. In this case, we want to estimate the average, $\overline{S}$. Indeed, if we know the true S_i, $i = 1, ..., n$, we would use

$$\overline{S} = \frac{1}{n} \sum_{i=1}^{n} S_i$$

as our point estimator, and

$$\hat{\text{se}}(\overline{S}) = \left[\frac{\sum_{i=1}^{n} (S_i - \overline{S})^2}{n(n-1)} \right]^{1/2}.$$

The complication of having only an estimator $\hat{S}_i$, often subject to substantial within-lot sampling variation, is not difficult to deal with if the lots are random with respect to variations in the treatment effect S.

The more complex case is if the multiple lots are released as part of a "designed" study. They might be released at known river flows, or at predetermined turbine operating conditions. Other design factors can be fish size or releases at different turbines. These design factors potentially influence the S_i and we want to test for that influence and possibly incorporate it into a model. If the S_i were known and were statistically independent over lots, the analysis could be done as a standard analysis of variance, regression, or covariance. However, we only have estimates, $\hat{S}_i$, subject to possibly large sampling variances, $\text{var}(\hat{S}_i \mid S_i)$, relative to σ^2.

4.1.3. Empirical Variance Estimation without Replication

4.1.3.1. Quasi-likelihood theory. – Likelihood theory relies on a completely specified probability model. Given that model, one derives a theoretical variance of the ML estimators. Often, especially with count data, the variation in the data clearly exceeds the postulated theoretical variation. This is so even when the structure (expected values) of the model seems quite appropriate for the data. Consequently, the point estimators are still acceptable, but not their theoretical variances.

In a large class of problems, the ML estimators and their theoretical variances actually depend only on the expected values and the structure of the theoretical variances implicit in the probability model. That is, one does not actually need the likelihood to get the ML estimates. In these cases, which include all multinomial models, the ML estimators can be computed by iteratively reweighted least squares; this has been known for a long time (see Jennrich and Moore 1975; Green 1984). The term quasi-likelihood is due to Wedderburn (1974) and formally extends standard likelihood theory to allow for excess variation, and empirical estimation of an variance-inflation factor, c. Quasi-likelihood theory justifies the usual ML estimators as optimal point estimators of the parameters, even when there is excess variation (over-dispersion) in the data. Recent papers and references on theoretical properties of quasi-likelihood inference include Healy (1981), Williams (1982), Cox (1983), McCullagh (1983), McCullagh and Nelder (1983), McCullagh and Pregibon (1985), and Royall (1986). Other investigations have shown likelihood methods are robust in the face of certain failures of the assumed model (Cox 1961; Huber 1967; Kent 1982; Sprott 1982). Finally, there are numerous published instances of encountering and coping with over-dispersion in count data (e.g., Bartlett 1936; Fisher 1949; Armitage 1957; Finney 1971).

Getting replicate estimators $\hat{S}_1, ..., \hat{S}_n$ is one route to empirical sampling variances. The other approach, justified by quasi-likelihood theory, is analogous to using residuals from a fitted model. We give here the general idea for multinomial models. Let $n_1, ..., n_k$ be a multinomial with a sample size $n_. = n_1 + ... + n_k$ and cell probabilities $\pi_1(\underline{\theta}), ..., \pi_k(\underline{\theta})$. Then the theoretical variance-covariance matrix is $\Sigma(\underline{\theta})$ with elements

$$\text{var}\,(n_i) = n_.\pi_i(1 - \pi_i)\,, \quad i = 1, ..., k\,;$$

$$\text{cov}\,(n_i, n_j) = -n_.\pi_i\pi_j\,, \quad i \neq j\,.$$

The quasi-likelihood approach uses the model as

$$\begin{bmatrix} n_1 \\ . \\ . \\ . \\ n_k \end{bmatrix} = n_. \begin{bmatrix} \pi_1(\underline{\theta}) \\ . \\ . \\ . \\ \pi_k(\underline{\theta}) \end{bmatrix} + \underline{\varepsilon}\,,$$

where $\underline{\varepsilon}$ has variance-covariance matrix

$$c\,\Sigma(\underline{\theta})\,.$$

Here, c is an unknown variance-inflation factor (a parameter to be estimated) and $\underline{\theta}$ is an "a"-dimensional vector of parameters. Using a generalized inverse of $\Sigma(\underline{\theta})$ and iteratively reweighted least squares, one finds the generalized least-squares estimator of $\underline{\theta}$, which is also the ML estimator when $c = 1$ (see McCullagh and Nelder 1983). Given this $\hat{\underline{\theta}}$, one can compute the $\hat{\pi}_i = \pi_i(\hat{\underline{\theta}})$ and $\hat{\Sigma} = \Sigma(\hat{\underline{\theta}})$. The usual chi-square goodness of fit can be computed (it is analogous to a residual sum of squares), call it χ^2_{GOF}; this chi-square test statistic has $k - 1 - a$ df. The estimator of c is

$$\hat{c} = \frac{\chi^2_{GOF}}{k - 1 - a}\,.$$

Using this $\hat{c}$, we get an estimator of the actual (empirical) sampling variances, e.g.,

$$\hat{\text{v}}\text{ar}\,(n_i) = \hat{c}n_.\hat{\pi}_i(1 - \hat{\pi}_i)\,.$$

This quasi-likelihood method also leads to estimators of the variances and covariances of the $\hat{\theta}_1, ..., \hat{\theta}_a$ as

$$\hat{c}[\hat{\text{var}}\,(\hat{\theta}_i)]$$

and

$$\hat{c}[\hat{\text{cov}}\,(\hat{\theta}_i, \hat{\theta}_j)]\,,$$

where $\hat{\text{var}}\,(\hat{\theta}_i)$ and $\hat{\text{cov}}\,(\hat{\theta}_i, \hat{\theta}_j)$ are the likelihood-theory estimators. Because $\hat{c}$ is estimated based on k - 1 - a df, the adjusted variances are also now based on k - 1 - a df (for example, for purposes of setting confidence intervals).

Quasi-likelihood theory justifies using all the ML theory we present in this monograph and then, if overdispersion is a problem, estimating $\hat{c}$ by a method such as the one above and multiplying $\hat{c}$ times the theoretical likelihood estimators of variances and covariances of $\hat{\theta}$. In general, the variance-inflation method involves estimating c from some residual variation in the recapture data after a structural model has been selected.

4.1.3.2. Variance-inflation factor method. – In practice, there can be difficulties deciding on what structural model we should use. However, once that model is chosen, quasi-likelihood theory justifies the ML estimates and the practice of adjusting their variances and covariances with $\hat{c}$. For example, if we selected model $H_{1\phi}$ under the complete capture history protocol, then the goodness of fit statistic for this model is the sum of the chi-squares of TESTs 1.*T*2, 1.*R*2, ..., 1.*Rk*-1 and TESTs 2 and 3. Divide that total chi-square by its degrees of freedom to obtain $\hat{c}$. (An hypothesis test can be made to test whether c is significantly greater than one.) This sort of procedure has long been used in probit analysis (Finney 1971), and it is recommended by McCullagh and Pregibon (1985). It provides one approach for getting an empirical sampling variance-inflation adjustment with capture-recapture data when the design does not include lots or sublots. Exactly how to proceed depends on the design used.

The most restrictive design is to have only one lot and no sublots; hence, only a single batch mark is used for controls and a different single batch mark for treatment fish. No release-time replication of any sort is then available. However, the recapture process will usually lead to multiple counts of some type. These multiple counts can be used to assess the empirical variation in the experimental data. If there are multiple recapture sites, there then are at least k - 1 counts. In principle, some components of TEST 1 can be used as a basis for estimating the variance-inflation factor, c. For example, if we judged model $H_{1\phi}$ to be the correct model in a first capture history protocol, then we can pool the chi-squares of TEST 1 components 1.*T*2, 1.*T*3, ..., 1.*Tk* - 1 (k - 3 df) as a basis to estimate c. The problem with this is that there are too few degrees of freedom unless there are 10 or more sites or occasions.

A practical alternative arises when fish are recaptured in gatewells at large dams: put recapture effort at all gatewells and use the variation in the treatment-to-control ratios across gatewells, within a dam, to assess the degree to which variation conforms to theory. If there are 10 gatewells at a dam, we have a 2×10 contingency table of counts for that dam. This gives 9 df. When data are pooled over even two or three recapture dams, there will be enough degrees of freedom from large-scale studies to estimate c, and test $c = 1$. The estimate of c is of the form

$$\hat{c} = \frac{\text{pooled chi-square}}{\text{pooled df}} .$$

When theoretical variances such as $\hat{\text{var}}(\hat{S})$ are then modified to be $\hat{c}\hat{\text{var}}(\hat{S})$, these new, empirically adjusted, sampling variance estimates have the degrees of freedom involved in $\hat{c}$.

Because we judge it so likely that $c > 1$, we recommend seriously considering such an empirical adjustment to theoretical sampling variances if the pooled degrees of freedom exceed 10 and the pooled chi-square is significant at even the $P = 0.2$ level. With only a few (pooled) degrees of freedom available, say, less than five, one wants more stringent evidence of the need for such an adjustment, i.e., the "usual" 0.05 or even 0.01 significance level.

4.1.4. Example with Multiple Lots

4.1.4.1. Description of the study. – As part of a large study, Long et al. (1975) released 12 pairs of treatment-control lots of young coho salmon *Oncorhynchus kisutch* at Lower Monumental Dam on the Snake River, Idaho. The experiment was for the purpose of measuring survival of small salmonids passing through operating turbines with and without perforated bulkheads. Part of that study involved the release of four lots of young fish on each of 3 days (13, 17, and 21 April 1974). Test fish were released upstream, and control fish downstream, of the test structure. This was an unknown capture history protocol with recaptures made at Ice Harbor and McNary dams (in that order), hence, $k = 3$. Marking was by freeze branding; marks batch-identified each treatment and release date and time. Fish were marked at a common facility, transported to Lower Monumental Dam, and then randomly allocated (within each treatment group) to separate holding tanks for each of the four lots released that day. The four lots released each day are not true replicates because of the possibility of an equipment effect. However, we will refer to the lots released within a day as reps (not replicates but reps in the sense of repeating the basic experiment).

4.1.4.2. Example data. – In order to illustrate some analysis methods we have extracted, from Table 1 of Long et al. (1975), the data for turbine unit 2. On a given day, releases were made in the morning from about 0700 to 0900 hours, one lot after the other. It was later found

there were equipment problems with the control batch of rep 3. The treatment batch of rep 3 could still be used; however, we are not doing a definitive reanalysis of these data. Rather, we chose to drop both treatment and controls for rep 3 on each day, and thus have nine paired releases to illustrate methods. Note that there is also strong evidence that reps 1 and 4 differed from rep 2, but there did not appear to be a day effect (Table 4.1).

Table 4.2 shows the data in CH-matrix form for input into program RELEASE. From the analyses produced, we extracted the results shown in Tables 4.1 and 4.3; note, however, the estimate of $\bar{S}$, based on the totals in Table 4.1, required a separate computation. For purposes of analysis by program RELEASE, day 1, rep 1 is considered lot 1. For this lot 1, there were 28,739 treatment fish released. Table 4.2 shows m_{t2} = 1,014 (captures at Ice Harbor Dam) and m_{t3} = 766 (captures at McNary Dam), so $m_{t.}$ = 1,780, as shown in Table 4.1. The input number 26,959 (at h = {1..}) is the difference of 28,739 and 1,780.

The notation and methods used in the first part of the analysis here are from Chapter 2.3. In particular, for any treatment-control pair,

$$\hat{S} = \frac{m_{t.}/R_{t1}}{m_{c.}/R_{c1}},$$

and the theoretical $\hat{\text{var}}\,(\hat{S} \mid S)$ is that for the unknown capture history protocol.

Table 4.1. – Releases, $R_{\bullet 1}$, and total recaptures, $m_{\bullet .}$, for nine lots from Table 1 of Long et al. (1975), plus the computed $\hat{S}$ and theoretical $\hat{\text{se}}\,(\hat{S})$ based on the unknown capture history protocol and model $H_{1\phi}$.

			Treatment fish		Control fish			Theoretical
Lot	Day	Rep[a]	R_{t1}	$m_{t.}$	R_{c1}	$m_{c.}$	$\hat{S}$	$\hat{\text{se}}\,(\hat{S})$
1	1	1	28,739	1,780	13,724	970	0.876	0.0344
2	1	2	28,856	1,587	14,577	1,066	0.752	0.0293
3	1	4	28,558	1,722	14,590	1,006	0.875	0.0341
4	2	1	26,395	1,914	14,122	1,125	0.910	0.0335
5	2	2	23,710	1,336	14,121	1,229	0.647	0.0251
6	2	4	27,293	1,744	13,665	1,060	0.824	0.0315
7	3	1	31,929	2,270	15,100	1,161	0.925	0.0327
8	3	2	31,663	1,788	15,404	1,252	0.695	0.0251
9	3	4	30,951	2,128	14,856	1,173	0.871	0.0311
Total			258,094	16,269	130,159	10,042	0.817[b]	0.0102[b]

[a]Reps are repeated experiments but not true replicates. Rep 3 was not used due to equipment problems.
[b]These values were computed from the totals rather than being averages over the nine values.

Table 4.2. – Program RELEASE input for analysis of reps 1, 2, and 4, taken on 3 days for turbine unit 2, Lower Monumental Dam, as reported by Long et al. (1975).

```
proc title data from Table 1, Long et al. 1975, turbine unit 2;
proc chmatrix nodetail occasions=3, groups=2 lots=9;
/* group 1 is the treatment, group two is the control */
                lot 1 ;
                   11.        1014      548;
                   1.1         766      422;
                   1..       26959    12754;
                lot 2 ;
                   11.         969      606;
                   1.1         618      460;
                   1..       27269    13511;
                lot 3 ;
                   11.         966      578;
                   1.1         756      428;
                   1..       26836    13584;
                lot 4 ;
                   11.        1006      619;
                   1.1         908      506;
                   1..       24481    12997;
                lot 5 ;
                   11.         686      597;
                   1.1         650      632;
                   1..       22374    12892;
                lot 6 ;
                   11.         932      552;
                   1.1         812      508;
                   1..       25549    12605;
                lot 7 ;
                   11.        1198      643;
                   1.1        1072      518;
                   1..       29659    13939;
                lot 8 ;
                   11.         932      698;
                   1.1         856      554;
                   1..       29875    14152;
                lot 9 ;
                   11.        1162      651;
                   1.1         966      522;
                   1..       28823    13683;
                proc stop;
```

Table 4.3. – Results of model selection tests for the example data of Long et al. (1975); each individual test has 1 df; totals have 9 df.

			TEST 1.*R*1		TEST 1.*T*2	
Lot	Day	Rep	χ^2	*P*	χ^2	*P*
1	1	1	11.72	<0.001	0.06	0.812
2	1	2	55.52	<0.001	4.69	0.030
3	1	4	12.21	<0.001	0.48	0.490
4	2	1	6.78	0.009	1.73	0.189
5	2	2	131.85	<0.001	1.97	0.161
6	2	4	26.69	<0.001	0.49	0.483
7	3	1	5.08	0.024	2.10	0.147
8	3	2	105.56	<0.001	3.89	0.049
9	3	4	15.63	<0.001	0.24	0.622
Total			371.03	<0.001	15.64	0.075

4.1.4.3. Model selection. – It is natural to look first at the data summary and estimators of S under model $H_{1\phi}$, which we presented in Table 4.1. However, prior to using model $H_{1\phi}$, we must examine the available tests of assumptions. For the unknown capture history protocol, the only within-lots tests available are TESTs 1.*R*1 and 1.*T*2. First, look at the nine chi-square results and the pooled results for TEST 1.*T*2, which here is testing that, for each lot,

$$\frac{p_{t2}}{p_{t2} + \phi_{t2}p_{t3}} = \frac{p_{c2}}{p_{c2} + \phi_{c2}p_{c3}}$$

(assuming no losses on capture). These test results are (overall) consistent with this null hypothesis. Moreover, TEST 1.*R*1, by comparison, shows that the overwhelming effect certainly is on ϕ_1 (this judgment is made in terms of the ratio 371.032/15.644 = 23.717, which roughly will behave like an *F*-statistic with 9 and 9 df under the null hypothesis that direct and indirect effects are equal).

More can be done with these data and test results to investigate a rep effect or a day effect. Analysis of variance can be used on the nine $\hat{S}_i$. The chi-squares can be pooled by day or rep. We proceed, however, to use model $H_{1\phi}$ and illustrate estimation of $\overline{S}$ and se($\overline{\hat{S}}$).

4.1.4.4. Point estimation. – From model $H_{1\phi}$ we have nine estimates, $\hat{S}_1$, ..., $\hat{S}_9$; let the corresponding true treatment effects be S_i, $i = 1, ..., 9$. We assume that either all S_i are the same ($S_i \equiv S$), or any variations in the S_i are random; hence, we treat the S_i as random

variables with mean $E(S)$ and unknown variance σ^2. In either of these cases, $\overline{S} = E(\hat{S})$ is the parameter of interest and the recommended estimator of $\overline{S}$ is obtained by pooling over lots and analyzing the pooled data. The resultant $\overline{S}_{\text{pooled}}$ is almost identical to an optimally weighted average of the separate $\hat{S}_i$. It is superior to such an optimally weighted average in that the optimal weights depend on unknown parameters that would have to be estimated. In this example, given the above assumption about $S_1, ..., S_9$ (which may not be true; this is only an example of methods),

$$\overline{S}_{\text{pooled}} = \frac{16{,}269/258{,}094}{10{,}042/130{,}159} = 0.817.$$

From the pooled data, totals by recapture dam are

v	Ice Harbor	McNary	$m_{v.}$
t	8,865	7,404	16,269
c	5,492	4,550	10,042

The estimate of $\hat{\text{var}}(\hat{\overline{S}})$ for the unknown capture history protocol uses these sums by dam (see Chapter 2.3); we find the theoretical standard error for the $\overline{S}_{\text{pooled}}$ to be 0.0102.

The use of standard error of $\overline{S}_{\text{pooled}}$ is based on restrictive assumptions: (1) all $S_i \equiv S$, so there is no variation in treatment effect attributed to unknown (or known) variations in test conditions over reps and days, and (2) for a given lot the assumption of binomial variation is true. Both assumptions are likely to fail. In general, theoretical variances tend to underestimate the real sampling variance, so even if all $S_i = S$ (this is equivalent to assuming $\sigma^2 = 0$), the $\hat{\text{se}}(\hat{S}_i)$ are too low.

Because sample sizes of treatment and controls are about the same over the nine lots, we have reason to believe the nine true var$(\hat{S})$ are all about the same. It then follows that a simple average of the $\hat{S}_i$ is going to give about the same value as the pooled estimator, 0.817. We find

$$\hat{\overline{S}} = \frac{1}{9}\left(\sum_{i=1}^{9} \hat{S}_i\right) = 0.819.$$

The corresponding standard error for this average is computed (assuming independence among lots) from

$$\hat{\text{var}}\left(\frac{1}{9}\sum_{i=1}^{9}\hat{S}_i\right) = \frac{1}{9^2}\sum_{i=1}^{9}\hat{\text{var}}(\hat{S}_i)\,,$$

or, here

$$\hat{\text{se}}\left(\frac{1}{9}\sum_{i=1}^{9}\hat{S}_i\right) = \frac{1}{9}\left(\sum_{i=1}^{9}[\hat{\text{se}}(\hat{S}_i)]^2\right)^{1/2}$$

$$= 0.0103\,.$$

Thus, for this example, pooling the nine lots and using the unknown capture history analysis gives virtually the same result as analyzing the unweighted means of the nine $\hat{S}_i$.

4.1.4.5. Variance estimation. – Either of the point estimates of $\overline{S}$ is reasonable; however, the theoretical standard error estimate of $\hat{\text{se}}(\hat{\overline{S}}) = 0.0103$ is not reliable and we recommend against it. A valid estimate of the variance that we should associate with $\hat{\overline{S}}$ in this example, under the assumption that the true $S_1, \ldots, S_9$ can themselves be considered a sample, is

$$\hat{\text{var}}(\hat{\overline{S}}) = \frac{\sum_{i=1}^{9}(\hat{S}_i - \hat{\overline{S}})^2}{(9)(8)}$$

$$= 0.0010844\,,$$

or $\hat{\text{se}}(\hat{\overline{S}}) = 0.0329$. The ratio of this empirical standard error (estimate) to the theoretical standard error (which assumes all $S_i \equiv S$) is

$$\frac{0.03293}{0.0102} = 3.228\,.$$

This empirical variance of $\hat{\overline{S}}$ is based on only 8 df; thus, one might want to test that this ratio is statistically significantly larger than one. A quick way to make this judgment is to treat $8(3.228)^2$ as a realization of a chi-square variable with 8 df. Symbolically, this test statistic is

$$\frac{[\text{empirical } \hat{\text{var}}(\hat{\overline{S}})]\ \text{df}}{[\text{theoretical } \hat{\text{var}}(\hat{\overline{S}})]},$$

where df is the degrees of freedom of the empirical estimator. Here, this test statistic has the value 83.4, which is highly significant ($P < 0.001$). We conclude that the theoretical standard error for $\hat{\overline{S}}$ should not be used; we will, therefore, use the empirical $\hat{\text{se}}(\hat{\overline{S}})$.

Given that the empirical $\hat{\text{se}}(\hat{\overline{S}})$ is to be used, the 95% CI for $E(S)$ is based on the t-distribution for 8 df. From Table 6.7, the appropriate t-value in $\hat{\overline{S}} \pm t\hat{\text{se}}(\hat{\overline{S}})$ is $t = 2.306$; hence, the 95% CI limits are 0.817 ± 0.076, or 0.741 to 0.893.

4.2. Empirical Variance Estimation

4.2.1. Estimation of the Variance-Inflation Factor

Here we consider a hypothetical example where practical constraints allowed only the release of 8,860 fish in one lot at dam 1. Only batch marks (one mark for treatment fish, one for control fish) were used and recaptured fish were removed at dams 2 and 3 ($k = 3$); thus, the first capture history protocol applied.

The biologists initially believed the treatment effect would be acute but quite severe at $S \sim 0.7$. From Part 6, the optimal number of treatment fish is $R_{t1} = 1.215\,R_{c1}$. Thus, $R_{t1} = 4{,}860$ and $R_{c1} = 4{,}000$. Approximately 60% of the recapture effort was planned at dam 2, and the remaining effort at dam 3. From previous experience, biologists believed $p_{t2} = p_{c2} = 0.12$ and $p_{t3} = p_{c3} = 0.08$ could be achieved. This planning, based on model $H_{1\phi}$, resulted in careful consideration of all of the practical details. The resulting data for the single lot are

Group	R_{v1}	m_{v12}	m_{v13}	r_{v1}
t	4,860	346	146	492
c	4,000	413	166	579

With $k = 3$, TEST 1.*T*2 can be computed to test (approximately) $p_{t2} = p_{c2}$. In the example, the result of this test yields $\chi^2 = 0.13$, 1 df, $P = 0.72$. Thus, model $H_{1\phi}$ is supported. TEST 1.*R*1 tests $\phi_{t1} = \phi_{c1}$, and this is strongly rejected ($\chi^2 = 39.10$, 1 df, $P = <0.001$). Thus, a

treatment effect on survival is shown.

The estimate of treatment survival under model $H_{1\phi}$ is 0.699 (se = 0.040). ($\hat{S}$ = 0.724, se = 0.081 under model H'_{2p}.) The model selected seems satisfactory and the point estimate also appears satisfactory. A valid criticism at this point is the measure of precision. The estimated theoretical standard error is based on multinomial sampling variation and may, indeed, be too low (i.e., the theoretical se may be underestimated). We have emphasized the importance of some form of replication so an empirical measure of precision can be obtained. Here, only one lot of fish was released; however, further analysis may yield insights into the adequacy of the variance estimate.

In the example, we find that the recapture data m_{v1j} were recorded for each gatewell in dams 2 and 3. These data are tabulated below, partitioned by dam and gatewell.

	Dam 2		Dam 3	
Gatewell	*t*	*c*	*t*	*c*
1	18	24	20	19
2	61	51	29	47
3	89	101	63	51
4	49	57	13	18
5	38	55	12	15
6	91	125	9	16
Total	346	413	146	166

These data can be used to estimate the variance-inflation factor c (as presented in Section 4.1.3). First, the two 2×6 contingency tables are analyzed by PROC CHISQ in program RELEASE.

Dam	χ^2	df	*P*
2	5.70	5	0.34
3	7.40	5	0.19
Total	13.10	10	0.22

The estimator of the variance-inflation factor c is $\hat{c} = \chi^2/\text{df} = 13.10/10 = 1.31$ (from Section 4.1.3). Thus, $\hat{\text{se}}(\hat{S}) = \sqrt{1.31} \times$ the theoretical standard error; or $1.144 \times 0.040 = 0.046$. An approximate 95% CI for the example is

$$\hat{S} \pm t_{\alpha/2,10}\hat{\text{se}}(\hat{S})$$

or

$$0.699 \pm 2.228(0.046)$$

or

$$0.699 \pm 0.102.$$

This interval compares with 0.699 ± 0.078 if the variance-inflation factor is not incorporated.

In summary, the recapture data m_{v1j} were partitioned by gatewell to allow a type of empirical variance to be computed using quasi-likelihood methods. Model $H_{1\phi}$ was selected, and the estimated treatment survival rate was 0.699 (se = 0.046). We believe the quasi-likelihood approach to variance estimation deserves full consideration. However, the use of replication or multiple lots is the superior procedure to be followed.

4.2.2. Replication Only

In the case of true replication, the underlying parameters (in particular S) are the same for each replicate and we infer that this is so from the design of the study. If parameters do not vary over replicates, this simplifies obtaining an optimal point estimate of S and an empirical variance estimator of $\hat{S}$. The closest we can come to replicates is the use of sublots based on different (or unique) within-lot marks.

For the purpose of discussion, we assume that the data have been thoroughly analyzed to determine a model, and that, on the basis of that model, $\hat{S}_1, \ldots, \hat{S}_n$ have been computed for the n replicates. By assumption, these estimated effects are independent. Also available are the corresponding estimated theoretical sampling variances $\hat{\text{var}}(\hat{S}_i \mid S_i)$. We make particular reference here to the simple estimator (valid under $H_{1\phi}$)

$$\hat{S}_i = \frac{r_{ti1}/R_{ti1}}{r_{ci1}/R_{ci1}},$$

with sampling variance

$$\operatorname{var}(\hat{S}_i \mid S_i) = c(S)^2\left[\frac{1}{E(r_{ti1})} - \frac{1}{R_{ti1}} + \frac{1}{E(r_{ci1})} - \frac{1}{R_{ci1}}\right].$$

Here, r_{ti1} and r_{ci1} are the total recaptures (r_{t1} and r_{c1}) partitioned by sublot i. The partitioned releases at dam 1 are $R_{v11}, \ldots, R_{vn1}$; $v = t$ or c.

To get a best estimate of $\hat{S}$, pool the data over replicates and analyze the pooled data under the chosen model. In particular, if model $H_{1\phi}$ is used, the optimal estimator of S is

$$\hat{S} = \frac{r_{t1}/R_{t1}}{r_{c1}/R_{c1}}.$$

General statistical theory also tells us that the above pooled estimator is (asymptotically) equivalent to the optimal weighted average of the separate $\hat{S}_1, \ldots, \hat{S}_n$:

$$\hat{S} = \frac{\sum_{i=1}^{n} w_i \hat{S}_i}{\sum_{i=1}^{n} w_i},$$

where

$$w_i = \frac{1}{\operatorname{var}(\hat{S}_i \mid S_i)}.$$

It is partly because these weights are not known, rather they have to be estimated, that we recommend pooling the replicates to get $\hat{S}$. The rest of the reason is that a nonlinear pooling is often superior to even the optimal linear weighted average.

The formula for the empirical variance of $\hat{S}$ is

$$\operatorname{var}(\hat{S}) = \frac{\sum_{i=1}^{n} w_i(\hat{S}_i - \hat{S})^2}{\left(\sum_{i=1}^{n} w_i\right)(n-1)}.$$

To compute an estimate, $\hat{\text{var}}(\hat{S})$, from this formula, we must know, or estimate, either the weights, or the relative weights,

$$\frac{w_i}{\sum_{j=1}^{n} w_j}, \quad i = 1, ..., n .$$

Usually, one thinks of replicates as all having the same sample size. If that is true, then all R_{ti1} are the same, as are all R_{ci1}, and then (whether or not the $R_{ti1} = R_{ci1}$) all the weights are the same, $w_i = w$, so they drop out giving us the usual formula

$$\hat{\text{var}}(\hat{S}) = \frac{\sum_{i=1}^{n} (\hat{S}_i - \hat{S})^2}{n(n-1)} .$$

We recommend equal sizes for all replicates because the relative weights are then known making empirical variance estimation easy.

In general, under model $H_{1\phi}$, the weight is given by

$$\frac{1}{w_i} = \frac{1}{R_{.i1}} \left[\frac{1}{\gamma_i} \left(\frac{1}{S\lambda_{c1}} - 1 \right) + \frac{1}{1-\gamma_i} \left(\frac{1}{\lambda_{c1}} - 1 \right) \right],$$

where

$$R_{.i1} = R_{ti1} + R_{ci1} ,$$

and

$$\gamma_i = \frac{R_{ti1}}{R_{.i1}} .$$

If the release ratios are kept constant (e.g., at $R_{ti1} = R_{ci1}$), then all $\gamma_i = \gamma$ are the same and we can take the relative weights as $w_i = R_{.i1}$. In this case, the formula for the empirical variance of $\hat{S}$ is

$$\hat{\text{var}}(\hat{S}) = \frac{\sum_{i=1}^{n} R_{i1}(\hat{S}_i - \hat{S})^2}{\left(\sum_{i=1}^{n} R_{i1}\right)(n-1)}.$$

The optimal release ratios of treatment to control are the same for all replicates because the true parameters are the same over all n replicates. It should not arise, except by accident, that the ratios γ_i differ. Slight differences in the γ_i have little effect and it is then better to use the weight as $w_i = R_{i1}$ than to estimate w_i.

4.2.3 Random Multiple Lots Only

Replication, as above, is difficult to achieve in these fisheries experiments. The more likely design is multiple lots. We treat here the case where it is assumed that the $S_1, \ldots, S_n$ are a random sample, with $\overline{S} = \hat{E}(S)$ and population variance σ^2.

From general linear least-squares statistical theory, the optimal linear combination of the $\hat{S}_i$ is

$$\hat{\overline{S}} = \frac{\sum_{i=1}^{n} w_i \hat{S}_i}{\sum w_i},$$

where now

$$w_i = \frac{1}{\sigma^2 + \text{var}(\hat{S}_i \mid S_i)}.$$

Both σ^2 and the true sampling variance $\text{var}(\hat{S}_i \mid S_i)$ are unknown. The notation $\text{var}(\hat{S}_i)$ and $\text{var}(\hat{S}_i \mid S_i)$ mean the same thing; we merely wish to emphasize here the conditional nature of the within-lot sampling variation of $\hat{S}_i$. Note also that only the relative values of these weights need to be known to compute $\hat{\overline{S}}$.

The theoretical variance of $\hat{\bar{S}}$ is

$$\text{var}(\hat{\bar{S}}) = \frac{1}{\sum_{i=1}^{n} w_i}.$$

This requires the absolute weights. Although $\text{var}(\hat{S}_i \mid S_i)$ is estimable if we assume $c = 1$, it is still necessary to know σ^2 to use the above formula.

An alternative, which leads to an empirical variance estimator, is

$$\text{var}(\hat{\bar{S}}) = \frac{\sum_{i=1}^{n} w_i(\hat{S}_i - \hat{\bar{S}})^2}{\left(\sum_{i=1}^{n} w_i\right)(n-1)}.$$

When $\sigma^2 = 0$, the above reduces to the case of n replicates. Often, it will be better to take these weights, w_i, as equal rather than trying to estimate them (because the estimates are subject to uncertainty). Thus, a practical formula will often be the simple one of

$$\text{var}(\hat{\bar{S}}) = \frac{\sum_{i=1}^{n} (\hat{S}_i - \hat{\bar{S}})^2}{n(n-1)},$$

with

$$\hat{\bar{S}} = \frac{1}{n}\sum_{i=1}^{n} \hat{S}_i.$$

This simple formula is reasonable if the $\text{var}(\hat{S}_i \mid S_i)$ are all nearly equal, or if they are all small relative to σ^2 (for example, if σ^2 is an order of magnitude greater than any $\text{var}(S_i \mid S_i)$). In theory, when the S_i vary, the $\text{var}(\hat{S}_i \mid S_i)$ will not be all equal. However, there is so much computational difficulty and statistical variation associated with estimating both σ^2 and $\text{var}(S_i \mid S_i)$ that treating these weights as all equal will often give better results than estimating them.

With low sampling effort (so $\text{var}(\hat{S}_i \mid S_i)$ dominates σ^2), it suffices to use the approximation

$$w_i = \frac{1}{\text{var}(\hat{S}_i \mid S_i)};$$

that is, we treat the n lots as if they were replicates because most of the variation in the $\hat{S}_i$ is sampling variation. Now, again, only the relative values of the weights are needed (so the value of c is not needed). In this case we recommend the weights be estimated as

$$\hat{w}_i = \frac{1}{\hat{\text{var}}(\hat{\bar{S}} \mid S_i)}.$$

That is, replace $\hat{S}_i$ in $\hat{\text{var}}(\hat{S}_i \mid S_i)$ by $\hat{\bar{S}}$. For model $H_{1\phi}$, this corresponds to taking the relative weights as

$$\hat{w}_i = \left[\frac{1}{r_{ti1}} - \frac{1}{R_{ti1}} + \frac{1}{r_{ci1}} - \frac{1}{R_{ti1}}\right]^{-1},$$

then computing the weighted mean and empirical $\hat{\text{var}}(\hat{\bar{S}})$. In general, this procedure corresponds to using the weights as

$$\hat{w}_i = (\text{cv}(\hat{S}_i))^{-2},$$

which we recommend if $\text{var}(\hat{S}_i \mid S_i)$ dominates σ^2.

If one is unwilling to settle for approximate relative weights, then σ^2 must be estimated. To estimate σ^2 requires a reliable (unbiased) estimate of the $\text{var}(\hat{S}_i \mid S_i)$, $i = 1, \ldots, n$ in order to separate σ^2 from sampling variation. Given such estimators, $\hat{\text{var}}(\hat{S}_i \mid S_i)$, the estimate of σ^2 requires iterative solution of a complicated equation. Because this is a matter of estimating variance components, we defer its discussion until Chapter 4.3.

4.2.4. Treatment Effect as a Relative Risk

Inference about the ratio of two proportions arises in many subject areas. In medical contexts, this problem is referred to as relative risk. There is an extensive literature on relative

risk (Gart 1985). In our notation, the general statistical model assumed for relative risk is

$$r_{ti1} \mid R_{ti1} \sim \text{bin}(R_{ti1}, S_i\lambda_{ci1}) ,$$

$$r_{ci1} \mid R_{ci1} \sim \text{bin}(R_{ci1}, \lambda_{ci1}) ,$$

$i = 1, ..., n$. These n independent, paired data sets are replicates if $\lambda_{ci1} = \lambda_{c1}$ and $S_i = S$ for all $i = 1, ..., n$. We have dealt with this case. If the λ_{ci1} and S_i vary, then the MLE of S_i is

$$\hat{S}_i = \frac{r_{ti1}/R_{ti1}}{r_{ci1}/R_{ci1}} , \quad i = 1, ..., n .$$

We referred to this case as random multiple lots.

There is an intermediate case: all $S_i = S$ but the λ_{ci1} vary. Gart (1985) gives the ML solution for this intermediate case. The ML estimator of S is not closed-form; rather, it must be found by numerical iterative methods (program SURVIV can be used to do this computation). However, the pooled estimator has comparable efficiency:

$$\hat{S} = \frac{\sum_{i=1}^{n} r_{ti1} / \sum_{i=1}^{n} R_{ti1}}{\sum_{i=1}^{n} r_{ci1} / \sum_{i=1}^{n} R_{ci1}} .$$

In medical and pharmacological applications, the sample sizes R_{ti1} and R_{ci1} are often <100 (and often like the data of Stier and Kynard 1986); certainly such data are one to three orders of magnitude less than in many fisheries experiments. With such small sample sizes, it may be worth the effort to find the exact MLE. At small sample sizes, however, there is no guarantee that the MLE is optimal.

4.3. Estimation of Variance Components

4.3.1. Some Theory

Two conceptually distinct types of variation constitute the essence of "variance components": sampling variance and parameter variance across samples. These topics were

introduced in Section 1.2.3. Let S_i represent the treatment effect for lot i, with n random lots in a study. Data analysis results in the estimates, $\hat{S}_i$, $i = 1, ..., n$. The uncertainty associated with $\hat{S}_i$ as an estimator of S_i is $\text{var}(\hat{S}_i \mid S_i)$, the sampling variance of $\hat{S}_i$. Usually we have denoted this sampling variance as $\text{var}(\hat{S}_i)$; the extended notation here is to emphasize the conditional nature of this sampling variance.

If one knows the values of $S_1, ..., S_n$, inference is simple:

$$\overline{S} = \frac{1}{n} \sum_{i=1}^{n} S_i ,$$

(i.e., $\hat{E}(S) = \overline{S}$),

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (S_i - \overline{S})^2 ,$$

and

$$\hat{\text{se}}(\overline{S}) = \frac{\hat{\sigma}}{\sqrt{n}} .$$

Unfortunately, the true S_i for lot i is never known. We have only estimators $\hat{S}_1, ..., \hat{S}_n$, each subject to a possibly different sampling variance. Within the context of the entire study, the total variation of $\hat{S}_i$ is

$$\text{var}(\hat{S}_i) = \sigma^2 + \text{var}(\hat{S}_i \mid S_i) .$$

Often, the sampling variance $\text{var}(\hat{S}_i \mid S_i)$ is larger than σ^2. However, $\text{var}(\hat{S}_i \mid S_i)$ depends on the size of lot i and recapture rates, whereas σ^2 is independent of the sample sizes and parameters that affect $\text{var}(\hat{S}_i \mid S_i)$; σ^2 depends on differences among lots with respect to fish characteristics and environmental factors (some controllable, some not). If, during a 1-month study, 10 lots are released, one every 3 days, then it is likely there will be differences in the parameters $S_1, ..., S_{10}$. The fact that fish in the 10th lot were larger and older at release than those in lot 1 could have an effect.

In Section 4.2.3, we discussed estimation of $E(S)$ in simple situations when we can avoid estimating σ^2. In particular, if all true (as opposed to estimated) sampling variances are the same then use

$$\hat{\bar{S}} = \frac{1}{n}\sum_{i=1}^{n}\hat{S}_i ,$$

with theoretical (total) variance

$$\text{var}(\hat{\bar{S}}) = \frac{\sigma^2 + E[\text{var}(\hat{S} \mid S)]}{n}$$

and the unbiased estimator

$$\hat{\text{var}}(\hat{\bar{S}}) = \frac{\sum_{i=1}^{n}(\hat{S}_i - \hat{\bar{S}})^2}{n(n-1)} .$$

Even though $\text{var}(\hat{\bar{S}})$ is the sum of two conceptually distinct components of variation, we can, in this simple case, estimate $\text{var}(\hat{\bar{S}})$ without having separate estimators of σ^2 and $E(\text{var}(\hat{S} \mid S))$. This is not always true. In more general situations, it becomes necessary to estimate σ^2 in the processes of computing $\hat{\bar{S}}$ and $\text{var}(\hat{\bar{S}})$. For this separation of variance components to be valid, one must first have a valid estimator of the sampling variance, $\text{var}(\hat{S}_i \mid S_i)$.

From the above, when we can assume the $\text{var}(\hat{S}_i \mid S_i)$ are all equal, then

$$\hat{\sigma}^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\hat{S}_i - \hat{\bar{S}})^2\right] - \frac{1}{n}\left[\sum_{i=1}^{n}\hat{\text{var}}(\hat{S}_i \mid S_i)\right]. \qquad (4.1)$$

This derives from

$$E[\hat{\text{var}}(\hat{\bar{S}})] = \frac{\sigma^2 + E[\text{var}(\hat{S} \mid S)]}{n}$$

and

$$\hat{E}[\text{var}(\hat{S} \mid S)] = \frac{1}{n}\sum_{i=1}^{n}\hat{\text{var}}(\hat{S}_i \mid S_i) .$$

Even when the sampling variances are not all equal, the above provides a (non-optimal) estimator of σ^2, which can be useful as a starting value for iterative solution of the better estimator below.

The general theory, as introduced in Section 4.2.3, is to use a weighted procedure, with weights equal to the reciprocals of the total variance of $\hat{S}_i$,

$$w_i = \frac{1}{\sigma^2 + \text{var}(\hat{S}_i \mid S_i)},$$

$$\hat{\bar{S}} = \frac{\sum_{i=1}^{n} w_i \hat{S}_i}{\sum_{i=1}^{n} w_i},$$

with theoretical variance

$$\text{var}(\hat{\bar{S}}) = \frac{1}{\sum_{i=1}^{n} w_i}$$

and empirical variance estimator

$$\hat{\text{var}}(\hat{\bar{S}}) = \frac{\sum_{i=1}^{n} w_i (\hat{S}_i - \hat{\bar{S}})^2}{\left(\sum_{i=1}^{n} w_i\right)(n-1)}.$$

When the w_i are the true (unknown) weights, then

$$E[\hat{\text{var}}(\hat{\bar{S}})] = \text{var}(\hat{\bar{S}}).$$

Therefore, if we have reliable estimators of sampling variance, we can solve the following equation for σ^2:

$$\frac{1}{n-1} \sum_{i=1}^{n} \hat{w}_i (\hat{S}_i - \hat{\bar{S}})^2 = 1, \tag{4.2}$$

where $\hat{\bar{S}}$ is the weighted average

$$\hat{\bar{S}} = \frac{\sum_{i=1}^{n} \hat{w}_i \hat{S}_i}{\sum_{i=1}^{n} \hat{w}_i} \tag{4.3}$$

and

$$\hat{w}_i = \frac{1}{\hat{\sigma}^2 + \hat{\text{var}}(\hat{S}_i \mid S_i)} . \tag{4.4}$$

Because there is only one unknown here, $\hat{\sigma}^2$, numerical solution of this equation is straightforward. If the best estimator of each $\text{var}(\hat{S}_i \mid S_i)$ is the average

$$\hat{\text{var}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\text{var}}(\hat{S}_i \mid S_i) , \tag{4.5}$$

then the solution is the simple one, $\tilde{\sigma}^2$, given in equation (4.1).

A useful approximate test exists for the null hypothesis H_0: $\sigma^2 = 0$. Compute

$$\chi^2 = \frac{(n - 1)\hat{\text{var}}(\hat{\bar{S}})}{\hat{\text{var}}(\hat{\bar{S}} \mid \sigma^2 = 0)} , \tag{4.6}$$

where

$$\hat{\text{var}}(\hat{\bar{S}} \mid \sigma^2 = 0) = \frac{1}{\sum_{i=1}^{n} \frac{1}{\hat{\text{var}}(\hat{S}_i \mid S_i)}} .$$

Under H_0: $\sigma^2 = 0$, χ^2 is distributed as chi-square with $n - 1$ df.

A confidence interval can be constructed for σ^2 by solving two modified versions of equation (4.2). Assume that we want a $(1 - \alpha)100\%$ CI, where $\alpha = \alpha_L + \alpha_U$ and L and U stand for lower and upper, respectively. Usually we will take $\alpha_L = \alpha_U$; hence, for a 95% CI, $\alpha_L = \alpha_U = 0.025$. One first looks up the percentile (critical) values for the central chi-square distribution corresponding to α_L and $1 - \alpha_U$, i.e., find $\chi^2_{n-1,\,\alpha_L}$ and $\chi^2_{n-1,1-\alpha_U}$. For example, for 10 df, $\chi^2_{10,0.025} = 3.25$ and $\chi^2_{10,0.975} = 20.5$. To find the upper limit, $\hat{\sigma}_U^2$, on σ^2, solve the equation

(for σ_U^2)

$$\frac{1}{n-1}\sum_{i=1}^{n}\hat{w}_i\left(\hat{S}_i-\hat{\bar{S}}\right)^2=\frac{\chi^2_{n-1,\,\alpha_L}}{n-1}, \tag{4.7}$$

where $\hat{\bar{S}}$ is the weighted average, as in equation (4.3), and the weights are as in equation (4.4). To find the lower limit, $\hat{\sigma}_L^2$, solve the equation (for σ_L^2)

$$\frac{1}{n-1}\sum_{i=1}^{n}\hat{w}_i\left(\hat{S}_i-\hat{\bar{S}}\right)^2=\frac{\chi^2_{n-1,1-\alpha_U}}{n-1}. \tag{4.8}$$

If equation (4.8) does not have a positive solution for σ^2, then set $\hat{\sigma}_L^2 = 0.0$ and adjust to a one-sided $(1-\alpha)100\%$ CI by redefining $\alpha_U = \alpha$.

Sometimes it suffices to have all $\hat{\text{var}}(\hat{S}_i \mid S_i)$ the same. Thus, one replaces each $\hat{\text{var}}(\hat{S}_i \mid S_i)$ by $\hat{\text{var}}$ of equation (4.5). Equation (4.1) is then the solution to equation (4.2). The confidence limits now also have explicit solutions:

$$\hat{\sigma}_L^2=\frac{\sum_{i=1}^{n}\left(\hat{S}_i-\hat{\bar{S}}\right)^2}{\chi^2_{n-1,1-\alpha_U}}-\hat{\text{var}} \tag{4.9}$$

and

$$\hat{\sigma}_U^2=\frac{\sum_{i=1}^{n}\left(\hat{S}_i-\hat{\bar{S}}\right)^2}{\chi^2_{n-1,\,\alpha_L}}-\hat{\text{var}}, \tag{4.10}$$

where, for these two equations,

$$\hat{\bar{S}}=\frac{1}{n}\sum_{i=1}^{n}\hat{S}_i.$$

The estimator of σ^2 obtained by equations (4.2)-(4.4) is not the MLE. The equations for the MLE are easily derived; they also require iterative solution in general. In the special case of all $\text{var}(S_i \mid S_i)$ being the same, the MLE exists in closed form; it is then given by equation

(4.1) *modified* to have n, not n - 1, in the denominator of the first term. When the number of lots is small, then the MLE of σ^2 is substantially biased. Of course, this bias can be substantially reduced by a simple adjustment to the likelihood equations. This bias-adjusted MLE would be acceptable for small n, and is preferred asymptotically as n gets large. However, n (lots) is unlikely to be large in these release-recapture experiments. Consequently, there is no compelling reason to use the MLE of σ^2 here. We would still use the MLE if we did not want confidence intervals on σ^2. Large-sample likelihood theory provides for likelihood intervals, but those intervals are not reliable for small sample sizes (n). Instead, we believe it is better, when n is small, to base a CI for σ^2 on the result that $\Sigma w_i(\hat{S}_i - \overline{\hat{S}})^2$ has approximately a central chi-square distribution.

Using asymptotic theory, the relative efficiency of the chi-square based estimator to the MLE is

$$\left[1 + \frac{\sum_{i=1}^{n} (w_i - \overline{w})^2}{n(\overline{w})^2}\right]^{-1};$$

here

$$\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i,$$

$$w_i = \frac{1}{\sigma^2 + \text{var}(\hat{S}_i \mid S_i)}.$$

Notice that this is just $(1 + (\text{cv})^2)^{-1}$, cv being the coefficient of variation among the true $w_1, \ldots, w_n$. This relative efficiency is usually quite high; however, we again emphasize that our reason for recommending the moment estimator for small n is the more reliable confidence interval that can then be computed.

4.3.2. Simple Example

Section 1.2.2 introduced variance components and gave a numerical example. The parameter there is a proportion; the five true p_i are 0.13, 0.17, 0.16, 0.13, and 0.16; $\overline{p} = 0.015$; and the estimate of σ based directly on these p_i is $\tilde{\sigma} = 0.0187$:

$$\tilde{\sigma}^2 = \frac{1}{4}\sum_{i=1}^{5} (p_i - \overline{p})^2.$$

In this example, the true sampling variances are the same for all 5 years. Consequently, the simple method works, equation (4.1):

$$\hat{\overline{p}} = \frac{1}{5}\sum_{i=1}^{5}\hat{p}_i$$

$$= 0.16$$

and

$$\hat{\text{var}}(\hat{\overline{p}}) = \frac{\sum_{i=1}^{5}(\hat{p}_i - \hat{\overline{p}})^2}{(5)(4)}$$

$$= 0.00034969.$$

Also,

$$\hat{\text{var}}(\hat{\overline{p}} \mid \sigma^2 = 0) = \frac{1}{5^2}\sum_{i=1}^{n}\hat{\text{var}}(\hat{p}_i)$$

$$= 0.000300675.$$

The estimate of σ^2 is thus

$$\hat{\sigma}^2 = 5\left[\hat{\text{var}}(\hat{\overline{p}}) - \hat{\text{var}}(\hat{\overline{p}} \mid \sigma^2 = 0)\right]$$

$$= 0.00024508,$$

or, $\hat{\sigma} = 0.0157$.

Equations (4.9) and (4.10) can be used to compute the 95% CI on σ^2. Required quantities are

$$\hat{\text{var}} = \frac{1}{5}\sum_{i=1}^{n}\hat{\text{var}}(\hat{p}_i \mid p_i)$$

$$= 0.0015034,$$

$$\sum_{i=1}^{5} \left(\hat{p}_i - \hat{\overline{p}}\right)^2 = 0.0069938,$$

$$\chi^2_{4,0.025} = 0.484,$$

$$\chi^2_{4,0.975} = 11.1.$$

Thus, we find

$$\hat{\sigma}_L^{\,2} = \frac{0.0069938}{11.1} - 0.0015034$$

$$= -0.000873$$

and

$$\hat{\sigma}_U^{\,2} = \frac{0.0069938}{0.484} - 0.0015034$$

$$= 0.01295.$$

That the lower bound is negative shows we cannot be certain σ^2 is greater than zero here. It suffices to take the interval on σ as 0 to 0.114, with $\hat{\sigma} = 0.0157$.

The formal test of H_0: $\sigma^2 = 0$ is

$$\chi^2 = \frac{4\,\hat{\text{var}}(\hat{\overline{p}})}{\hat{\text{var}}(\hat{\overline{p}} \mid \sigma^2 = 0)}$$

$$= \frac{4\,(0.00034969)}{0.000300675}$$

$$= 4.652$$

(P = 0.325). We cannot reject that σ^2 might be zero. The power of this test is low here because the number of years sampled is small (n = 5) and there is high within-year sampling variance relative to σ^2. The ratio of $\hat{\sigma}^2$ to the average var($\hat{p}_i \mid p_i$) is

$$\frac{0.00024508}{0.0015034} = 0.163.$$

Thus, sampling variation far exceeds year-to-year variation in the true p_i. This makes it very difficult to detect changes in the p_i over time. Environmental studies commonly suffer from large sampling variation, making it difficult to detect effects of interest over space or time.

4.3.3. More Complex Example

We generated a large set of data under model $H_{1\phi}$ for six lots, to further illustrate inference about S with multiple random lots. Data for two groups (treatment and control) were available from six occasions (k = 6) and are summarized in Table 4.4. Ideally, we could imagine that these data were from uniquely marked releases; sublots would then be possible, each based on the last digit of the tag number. There would then be 10 tables, each similar to Table 4.4, and each with sample sizes about one-tenth of those now shown. Such replication would allow variances to be estimated empirically and additional tests would be possible to enable a full scrutiny of the results. We will ignore the possible sublots here and consider only the data set shown in Table 4.4.

Consider, for this illustration, that these data were collected on the Columbia River from releases over six biweekly periods. The objective of the experiment is to estimate the "total project" survival of salmon smolts for a particular dam (here called dam 1) and the pool upstream from the dam. Total project mortality then includes all deaths of fish in the pool and deaths associated with the spillways, various turbine units, deflecting barriers, screens, etc. Therefore, treatment fish are released at a point upstream from the pool and control fish are released at a later appropriate time just below dam 1 such that control and treatment fish begin migrating downstream at approximately the same time between dams 1 and 2. Perhaps the treatment fish in lot 1 were released above the pool on May 1 and 2 and their movements were monitored at dam 1. When they arrive at dam 1, the control fish are released, perhaps over a period of 3-8 days, depending on the continued arrival rate of the treatment fish.

Table 4.4. – Summary of synthetic data generated under model $H_{1\phi}$ of the first capture history protocol. Data were generated for six lots where the parameters are $\phi_{ti} = 0.95$, $i = 2, ..., 6$; $\phi_{c1} = 0.95$, $i = 1, ..., 6$; and $\phi_{t1} = 0.85, 0.80, 0.70, 0.65, 0.75$, and 0.80 for lots 1, ..., 6, respectively; and $p_{ti} = p_{ci} = 0.10, 0.17, 0.12, 0.08, 0.07$, and 0.05 for lots 1, ..., 6, respectively.

			Number recaptured and removed at dam j, m_{rij}				
Lot	Group	Releases R_{ri}	$j = 2$	3	4	5	6
1	t	20,000	1,752	1,445	1,256	1,013	918
	c	20,000	1,922	1,660	1,461	1,135	1,045
2	t	12,000	1,636	1,296	1,021	849	616
	c	12,000	2,014	1,549	1,198	936	786
3	t	27,000	2,302	1,936	1,591	1,241	1,117
	c	27,000	3,034	2,547	2,177	1,847	1,468
4	t	20,000	1,050	883	804	698	592
	c	13,000	986	874	734	670	555
5	t	14,000	682	617	573	506	440
	c	20,000	1,309	1,173	1,066	912	844
6	t	7,000	293	261	220	191	185
	c	7,000	307	288	277	252	197

If the six treatment releases are made at 2-week intervals, it is reasonable to think that the true treatment survival might vary over the course of the overall experiment. This variation among the parameters (σ^2) might be caused by flow and other variables affecting the river, as well as by dam variables such as head, spill, and wicket gate and turbine blade settings.

In this example, we know the true treatment effect (parameter) for each lot ($S = 0.895$, 0.842, 0.737, 0.684, 0.789, and 0.842 for lots 1 to 6, respectively). Consequently, the best possible estimates of $E(S)$ and σ are

$$\overline{S} = \frac{1}{6}\sum_{i=1}^{6} S_i = 0.798$$

and

$$\hat{\sigma}^2 = \frac{1}{5}\sum_{i=1}^{6} (S_i - \overline{S})^2 = 0.006016$$

or

$$\hat{\sigma} = 0.0776 .$$

From these results, we would have

$$\hat{\text{se}}(\bar{\hat{S}}) = \frac{0.0776}{\sqrt{6}} = 0.0317 ,$$

and, using equations (4.9) and (4.10), with $\hat{\text{var}} \equiv 0$ (and $\chi^2_{5,0.025} = 0.831$, $\chi^2_{5,0.975} = 12.8$), a 95% CI on σ is 0.048 to 0.190. These results would arise only if all sampling variances $\text{var}(\hat{S}_i \mid S_i)$ were zero.

Using model $H_{1\phi}$ on the data for all six lots, we used RELEASE to compute $\hat{S}_i$ and $\hat{\text{se}}(\hat{S}_i \mid S_i)$. Those results are shown in Table 4.5.

Table 4.5. – Summary of total releases, estimates of treatment effects, and their standard errors under model $H_{1\phi}$ for the synthetic data in Table 4.4.

Lot i	R_1	$\hat{S}_i$	$\hat{\text{se}}(\hat{S}_i \mid S_i)$	True S_i
1	40,000	0.884	0.0123	0.895
2	24,000	0.836	0.0110	0.842
3	54,000	0.739	0.0087	0.737
4	33,000	0.685	0.0134	0.684
5	34,000	0.759	0.0156	0.789
6	14,000	0.871	0.0319	0.842
Unweighted average		0.796		0.798

The simplest way to compute an empirical estimate of $\text{var}(\hat{\bar{S}})$ from these six estimates of treatment survival is to use

$$\hat{\text{var}}(\hat{\bar{S}}) = \frac{\sum_{i=1}^{6} (\hat{S}_i - \hat{\bar{S}})^2}{(5)(6)} = 0.0010635 ,$$

where

$$\hat{\bar{S}} = \frac{1}{6} \sum_{i=1}^{6} \hat{S}_i = 0.796 ;$$

hence,

$$\hat{\text{se}}(\hat{\bar{S}}) = 0.0326 .$$

Computing

$$\hat{\text{var}} = \frac{1}{6} \sum_{i=1}^{6} \hat{\text{var}}(\hat{S}_i \mid S_i)$$

$$= 0.00298$$

and using equation (4.1), the simple estimate of σ^2 is

$$\tilde{\sigma}^2 = (6)(0.0010635) - 0.000298$$

$$= 0.006083 ,$$

or $\tilde{\sigma} = 0.078$. Equations (4.9) and (4.10) lead to the 95% CI on σ as 0.050 to 0.196. Because the sampling variances are so small here, these results are almost the same as if the six true S_i were known.

Corresponding to these unweighted procedures, there is a simple version of the test of H_0: $\sigma^2 = 0$: use equation (4.6) with

$$\hat{\text{var}}(\hat{\bar{S}}) = \frac{1}{30}\sum_{i=1}^{6}\left(\hat{S}_i - \hat{\bar{S}}\right)^2 = 0.0010635,$$

and

$$\hat{\text{var}}(\hat{\bar{S}} \mid \sigma^2 = 0) = \left(\frac{1}{6}\right)^2 \sum_{i=1}^{6} \hat{\text{var}}(\hat{S}_i \mid S_i)$$

$$= 0.0000496.$$

Thus, the test statistic is

$$\chi^2 = \frac{5(.0010635)}{0.0000496}$$

$$= 107.1,$$

which is highly significant. We conclude that there is substantial variation among the population parameters in addition to the sampling variation. Often, the estimation of σ^2 is as important as the separate point estimates of treatment effects.

The theoretically more efficient inference procedures require weighted analyses, as per equations (4.2), (4.3), (4.4), and (4.7) and (4.8), which themselves require iterative numerical solution. Solving the relevant equations, we find the weighted analysis gives

$$\hat{\sigma}^2 = 0.00638;$$

$$\hat{\sigma} = 0.080.$$

Then we can find the weighted results below:

$$\hat{\bar{S}} = 0.796,$$

$$\hat{\text{var}}(\hat{\bar{S}}) = 0.0010304,$$

and

$$\hat{\text{var}}(\hat{\overline{S}} \mid \sigma^2 = 0) = \frac{1}{\sum_{i=1}^{6} \frac{1}{\hat{\text{var}}(\hat{S}_i \mid S_i)}}$$

$$= 0.0000258 .$$

Hence, the weighted test of H_0: $\sigma^2 = 0$ is given by

$$\chi^2 = \frac{(5)(0.00103)}{0.0000258}$$

$$= 199.6 .$$

This weighted chi-square test statistic is much bigger than the unweighted one (of 107.1), but either one clearly rejects H_0: $\sigma^2 = 0$.

From the weighted analysis, $\hat{\text{se}}(\hat{\overline{S}}) = 0.0321$ and a 95% CI on $E(S)$ is $0.796 \pm (2.571)(0.0321)$ or 0.713 to 0.879. Also, from solving equations (4.7) and (4.8), we get the 95% CI on σ as 0.047 to 0.196 ($\hat{\sigma} = 0.080$). We see that there was no real advantage here to doing the weighted analysis. Unfortunately, that is not always true, and the only way to be sure is to do the weighted analysis.

To explore this example further, we generated another simulated six lots where the released numbers were one-tenth those shown in Table 4.4. Thus, for example, for lot 1, $R_{t1} = R_{c1} = 2{,}000$. All parameters were left unchanged. Thus, the only effect was to increase the within-lot sampling variances by a factor of 10. The results are shown in Table 4.6. It is fortuitous that the unweighted average of the $\hat{S}_i$ is so close to $\overline{S} = 0.798$. Note that the individual $\hat{S}_i$ are not as close to S_i now as in the original case (Table 4.5).

Table 4.6. – Summary of total releases and some results from the second simulated six lots under model $H_{1\phi}$; the same parameters apply as for Table 4.4.

Lot i	R_1	$\hat{S}_i$	$\hat{\text{se}}(\hat{S}_i \mid S_i)$	True S_i
1	4,000	0.880	0.0391	0.895
2	2,400	0.774	0.0321	0.842
3	5,400	0.743	0.0273	0.737
4	3,300	0.677	0.0418	0.684
5	3,400	0.739	0.0476	0.789
6	1,400	0.984	0.1145	0.842
Unweighted average		0.800		0.798

Table 4.7 summarizes the various results (such as $\hat{\overline{S}}$, $\hat{\sigma}$), unweighted and weighted (and the case of all $\text{var}(\hat{S}_i \mid S_i) = 0$), computed based on Table 4.6. The main efficiency of the weighted analysis is to achieve smaller $\hat{\text{se}}(\hat{\overline{S}})$ and a more powerful test of the null hypothesis H_0: $\sigma^2 = 0$ (i.e., of H_0: $S_1 = S_2 = S_3 = S_4 = S_5 = S_6$). These results illustrate that if the true $\text{var}(\hat{S}_i \mid S_i)$ vary substantially, the weighted analysis should be used.

If we further reduced the release numbers, a point would be reached where the sampling variation would exceed σ^2 and the test of H_0: $\sigma^2 = 0$ would have low power. One could easily get data so poor that there would be no evidence that the treatment effects varied. A tradeoff faces us in environmental studies: having both multiple lots and yet sufficient within-lot sample sizes to allow demonstrating any important lot-to-lot variation in the treatment effect, or other parameters (see, for example, Eberhardt 1978; Armour et al. 1983).

Table 4.7. – Comparisons of weighted and unweighted results based on the Monte Carlo study results for $\hat{S}_i$ in Table 4.6; the values when all $\text{var}(\hat{S}_i \mid S_i) = 0$ are the best possible (hence, "true" values) for these six lots.

Entity computed	If $\text{var}(\hat{S}_i \mid S_i) = 0$ (i.e., "truth")	Weighted analysis	Unweighted analysis
$\hat{\sigma}$	0.0776	0.0812	0.0959
$\hat{\overline{S}}$	0.798	0.780	0.800
$\hat{\text{se}}(\hat{\overline{S}})$	0.0317	0.0385	0.0458
$\hat{\sigma}_L$	0.048	0.028	0.039
$\hat{\sigma}_U$	0.190	0.261	0.269
χ^2	∞	30.1	18.5

4.4. Example with Four Groups and 10 Lots

In this chapter we will extend the synthetic example presented in Chapter 3.10. As before, we use data simulated under scheme B, $V = 4$ groups, $k = 6$ occasions, and $R_{v1} = 1{,}000$ fish released. Here, we extend the example with the inclusion of data for nine other lots. The parameters used are those in Table 3.5. Summaries of the m_{vij} arrays will not be given, but the data in Table 3.6 are representative. The data for each lot were generated under the same parameters. Therefore, there is no variance component issue because $\sigma^2 = 0$. Such data might arise from a large study using PIT tags and the lots correspond to the last digit of the tag numbers.

In this study, there are four groups. The first three are treatment groups ($v = 1, 2, 3$); each is then compared to the control group ($v = 4$). Thus, primary interest lies in the treatment effects S_{14}, S_{24}, and S_{34} (true values are 0.90, 0.75, and 0.70, respectively).

The results of various hypothesis tests are given in Table 4.8. The results of TEST 1.*R*1 convincingly show a strong treatment effect (i.e., $\phi_{t1} \neq \phi_{c1}$) for all 10 lots. If the test results for individual lots were less convincing, one might compute $\sum_{i=1}^{10} \chi^2_i = 897.5$, with 30 df. The results for TEST 1.*T*2 for each lot do not reject that $p_{t2} = p_{c2}$, which suggests model H_{2p} (which we know to be the correct model).

The results of TEST 1.*R*2 provide no evidence that ϕ_{t2} and ϕ_{c2} differ, tending to confirm that model H_{2p} is appropriate. The overall test for treatment effect (TEST 1) is highly significant. Although R_1 is only 1,000 individuals released, the model selection capability in this example is quite good because of reasonable recapture rates. We leave it to the reader to analyze this example with as few as, perhaps, 300 releases in each lot.

The 10 estimates of treatment effect are presented in Table 4.9. Within a lot, the estimates of S are somewhat variable; however, the mean of the 10 estimates is fairly precise for each of the three treatments. Because these 10 $\hat{S}_i$ are from replication with equal release numbers, the computing formulae used were

$$\hat{\bar{S}} = \frac{1}{10} \sum_{i=1}^{10} \hat{S}_i$$

and

$$\hat{\text{se}}(\hat{\bar{S}}) = \left(\frac{1}{90} \sum_{i=1}^{10} (\hat{S}_i - \hat{\bar{S}})^2 \right)^{1/2}.$$

Table 4.8. – Summary of the results of modeling selection tests for the simulated data: scheme B, $k = 6$, four groups, 10 lots.

	TEST 1.*R*1 H_0 vs. $H_{1\phi}$ (3 df)		TEST 1.*T*2 $H_{1\phi}$ vs. H_{2p} (3 df)		TEST 1.*R*2 H_{2p} vs. $H_{2\phi}$ (3 df)		TEST 1 H_0 vs. $H_{5\phi}$ (18 df)	
Lot	χ^2	P	χ^2	P	χ^2	P	χ^2	P
1	71.4	<0.001	52.6	<0.001	1.2	0.75	128.4	<0.001
2	93.9	<0.001	12.0	<0.001	4.7	0.20	125.9	<0.001
3	86.0	<0.001	54.6	<0.001	1.6	0.66	148.1	<0.001
4	92.0	<0.001	35.2	<0.001	1.5	0.68	137.5	<0.001
5	91.9	<0.001	26.1	<0.001	1.3	0.73	131.0	<0.001
6	80.2	<0.001	28.8	<0.001	0.5	0.92	127.8	<0.001
7	104.6	<0.001	60.2	<0.001	1.7	0.63	175.3	<0.001
8	90.3	<0.001	58.5	<0.001	4.3	0.23	164.3	<0.001
9	87.6	<0.001	47.6	<0.001	2.8	0.43	150.3	<0.001
10	99.6	<0.001	39.7	<0.001	3.1	0.37	149.9	<0.001

Table 4.9. – Summary of the estimates of treatment survival rate, under analysis model H_{2p}, for the simulated data: scheme B, $k = 6$, four groups, 10 lots. True parameter values are shown at the bottom of table.

Lot	$\hat{S}_{14}$	$\hat{\text{se}}(\hat{S}_{14})$	$\hat{S}_{24}$	$\hat{\text{se}}(\hat{S}_{24})$	$\hat{S}_{34}$	$\hat{\text{se}}(\hat{S}_{34})$
1	0.8889	0.0561	0.8303	0.0542	0.7315	0.0502
2	0.8597	0.0512	0.6927	0.0453	0.6151	0.0422
3	0.8906	0.0554	0.7339	0.0498	0.7125	0.0489
4	0.9442	0.0563	0.7333	0.0484	0.6881	0.0468
5	0.8226	0.0487	0.6944	0.0441	0.6793	0.0438
6	0.8331	0.0511	0.7204	0.0471	0.7075	0.0466
7	0.9031	0.0531	0.7040	0.0465	0.6981	0.0458
8	0.8746	0.0530	0.7801	0.0500	0.6788	0.0460
9	0.8856	0.0537	0.7169	0.0473	0.6858	0.0461
10	0.9112	0.0536	0.6722	0.0577	0.7113	0.0469
$\bar{\hat{S}}$	0.8814	0.0362	0.7278	0.0465	0.6908	0.0315
S	0.90		0.75		0.70	

RELEASE uses these simple unweighted formulae to compute averages and standard errors over multiple lots. Often, these will not be the appropriate formulae. This summarization capability is in RELEASE primarily as part of the simulation options. Proper weighted averages, standard errors, and estimates of variance components for real data must be done separately by the investigator.