Aquaculture Research, 1998, **29**, 373–379

Statistical aspects of aquaculture research: sample sizes for pond experiments

T S Smart & J Riley
IACR-Rothamsted, Harpenden, UK

P Edwards
Agricultural and Aquatic Systems Program, School of Environment, Resources and Development, Asian Institute of Technology, Bangkok, Thailand

Correspondence: Janet Riley, IACR-Rothamsted, Harpenden AL5 2JQ, UK

Abstract

The present paper illustrates the need for greater pond replication in pond experiments. It demonstrates that increasing the number of fish sampled in each pond will not reduce the number of ponds needed when treatment effects are estimated. Sampling bias, skew distributions and heterogeneous fish populations are discussed in relation to fish sample sizes. When it is impossible to increase the number of replicate ponds, several methods of accounting for the pond variability are suggested.

Introduction

Experiments on fish ponds are often used to compare treatments; however, the ability to detect treatment differences is often restricted by the number of ponds used. As few as three replicate ponds per treatment may be used in a typical experiment to investigate the effects of treatments on the growth of fish. Fish weight and length are measured at the start and end of the experiment, with further measurements possibly being taken during the experiment. The large variability between ponds combined with the fish variability result in imprecise estimates of treatment effects when so few replicates are used. Thus, only very large treatment differences are detected in such circumstances. The present paper considers how the precision of estimates is affected by the number of fish sampled per pond and the number of replicate ponds per treatment. The first of these can usually be determined by the researcher, but the availability of ponds is often limited. It is assumed that the primary objective of an experiment is to estimate treatment effects. The effect of skewness in the distribution of the fish on sample size is considered, and there is a brief discussion of sampling bias and the problems associated with sampling during the experiment.

Data from several pond aquaculture experiments completed at the Asian Institute of Technology (AIT), Thailand, and at the Institute of Aquaculture, Stirling, UK, are examined to show how fish sample size and pond replication can affect the precision of estimates and the sensitivity of tests for treatment differences (Table 1).

Method

For all of the experiments, the mean fish weights of each pond were calculated and the variance was estimated from their analysis of variance (ANOVA). These data were used to estimate the minimum relative treatment differences which would be detectable for different numbers of pond replicates. The variability at each level, between fish within ponds, between ponds with the same treatment and between ponds with different treatments, was investigated in more depth. These variabilities were estimated using residual maximum likelihood (REML) (Patterson & Thompson 1971), which provides estimates of the variance at each level, even if the number of fish sampled in each pond is not the same.

Table 1 A summary of the experimental design and the fish measurements of the experiments (standard errors are given in parentheses)

Experiment	Number of treatments	Number of pond replicates	Mean number of fish sampled per pond	Mean weight (g)	Mean length (mm)
1	6	3	47	121 (6.8)	188 (3.1)
2	5	3	503	56 (4.4)	142 (3.1)
3	5	3	418	65 (3.8)	152 (2.8)
4	5	3	388	88 (9.2)	166 (5.8)
5	5	3	409	101 (12.6)	174 (6.7)
6	1	4	29	178 (22.9)	198 (10.5)
7	2	2	66	581 (40.6)	250 (3.0)

A simulation was also used to consider the extreme case, where the distribution of fish within a pond was both heterogeneous and skew. The extreme values estimated from the experiments were combined in the simulation.

Results and discussion

Number of replicate ponds

Many pond aquaculture experiments have only a few replicates of treatment ponds, i.e. rarely more than three. With so few replicate ponds, only very large treatment differences will be detectable. When a *t*-test is used, the detectable difference is

$$\textit{detectable difference} = t_{(df,0.05)}\sqrt{\frac{2s^2}{n}} \qquad (1)$$

where $t_{(df,0.05)}$ is the *t*-value at the 5% level for *df* degrees of freedom, *n* is the number of pond replicates for each treatment and s^2 is the pond mean variance estimated from an ANOVA.

Different sized fish were used in each experiment, so a relative detectable difference was calculated to enable comparisons to be made between experiments. This was calculated as a percentage of the mean fish weight across all ponds in the experiment.

$$\textit{relative detectable difference} = t_{(df,0.05)}\sqrt{\frac{2s^2}{n}} \times \frac{100}{\textit{mean}} \qquad (2)$$

where *mean* is the mean fish weight across all ponds.

The minimum relative differences that would be detectable for different numbers of pond replicates are given in Table 2. In a typical experiment with three replicates, the minimum relative detectable differences observed range from 37% to 80%. Thus, such experiments will be able to detect only very large treatment differences. Even with 10 replicate ponds the minimum detectable differences are still very large.

Number of fish sampled per pond

In general, if more fish are sampled per pond, this will not lead to much improvement in the precision of estimates of treatment differences. With an increase in the number of sampled fish, a very precise estimate for any particular pond may be obtained, but this does not necessarily give more precise information about treatment effects. The treatments are applied to individual ponds, and thus, treatment comparisons should be made at a pond level and not at a fish level. Therefore, the precision of any treatment effects is determined by both pond and fish variability. Yates & Zacopanay (1935) discussed how to combine such variance components. The estimate of the variance used to estimate treatment differences is given by:

$$v_t = v_p + \frac{v_f}{n} \qquad (3)$$

where v_t is the estimated variance used in any treatment comparisons, v_p is the between-pond variance, v_f is the variance between fish within the same pond and *n* is the number of fish sampled per

Table 2 Detectable difference as a percentage of the mean fish weight (g) for five experiments with different means and variances, and different numbers of replicate ponds[1]

			Number of replicate ponds					
Experiment	**Mean**	**Variance**	**2**	**3**	**4**	**5**	**6**	**10**
1	121	830	48	39	34	30	27	21
2	56	290	61	50	43	38	35	27
3	65	220	46	37	32	29	26	20
4	88	1270	81	66	57	51	48	36
5	100	2380	98	80	69	62	56	44

[1]The estimates are conservative because a *t*-value of 2 was used, but the degrees of freedom from which the variance is estimated are often small and a larger *t*-value is needed. Thus, an experiment with 12 ponds and four treatments each replicated three times has eight residual degrees of freedom if the treatments are allocated at random. With eight degrees of freedom, $t_{(8,0.05)} = 2.31$, and the relative detectable difference would increase by a factor of $2.31/2 = 1.16$.

pond. If the same number of fish are sampled in each pond, then $v_t = s^2$.

When each treatment is replicated in r ponds, the standard error of a treatment effect (i.e. the estimated mean response under the treatment) is given by:

$$se_t = \sqrt{\frac{v_p}{r} + \frac{v_f}{rn}} \qquad (4)$$

It can be seen that, as n becomes larger, se_t becomes smaller, but se_t will never be smaller than $\sqrt{(v_p/r)}$. Therefore, the major influence on se_t is replication number and not the number of fish sampled. Hence, it is essential either to increase the number of ponds available or to find methods of reducing the between pond variability.

The effects of increased sample size and increased numbers of replicate ponds are shown in Fig. 1. From this, it can be seen clearly that a sample of more than 10 fish has little effect on the precision of the treatment estimates, but an increase in the number of replications has a very large effect.

Is this always true or is it specific to the example used? What happens if the pond and fish variabilities are very different? Equation 4 can be split into two parts: (a) a scaling factor dependent on the number of replicates and the variability between ponds, and (b) a term purely dependent on the ratio of fish variability to pond variability, $k = v_f / v_p$:

$$se_t = \sqrt{\frac{v_p}{r}} \times \sqrt{1 + \frac{k}{n}} \qquad (5)$$

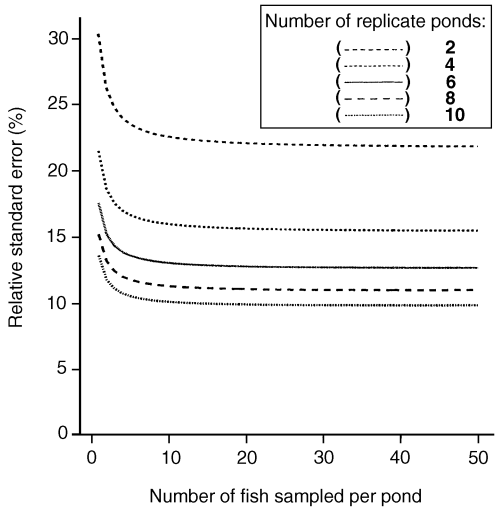

Figure 1 Relative standard error as a percentage of the mean fish weight for different sampling strategies showing how increasing the number of fish sampled and the number of replicate ponds affect the precision of the estimates. The variances used were those estimated from experiment 2.

Table 3 shows the estimated values of v_f, v_p and k for various datasets. The ratio of v_f and v_p ranges from about 1, when fish within a pond are very similar, to a value of the order 10, when a non-uniform stock of fish was used (experiments 6 and 7). Figure 2 shows how this affects the standard error for increased fish samples for k in the range observed in the experiment. Unless the pond variability can be reduced, there is no need to sample more than 10 fish per pond. The fish should be

Table 3 Estimates of fish variability, v_f, pond variability, v_p, and the ratio, k, for various pond aquaculture experiments

Experiment	Measurement	v_f	v_p	k
1	Weight (g)	690	810	0.9
	Length (mm)	200	170	1.2
2	Weight (g)	290	290	1.0
	Length (mm)	270	150	1.8
3	Weight (g)	240	210	1.2
	Length (mm)	130	120	1.1
4	Weight (g)	530	1270	0.4
	Length (mm)	220	510	0.4
5	Weight (g)	530	2380	0.2
	Length (mm)	170	670	0.3
6	Weight (g)	10 836	1686	6.4
	Length (mm)	1839	368	5.0
7	Weight (g)	57 400	5700	10.1
	Length (mm)	1050	19	55.4*

*In this experiment, the fish were very different at the start of the experiment, but were distributed such that there was a similar distribution in each pond. By the end of the experiment, the between-pond variability was still small in comparison with the between-fish variability.

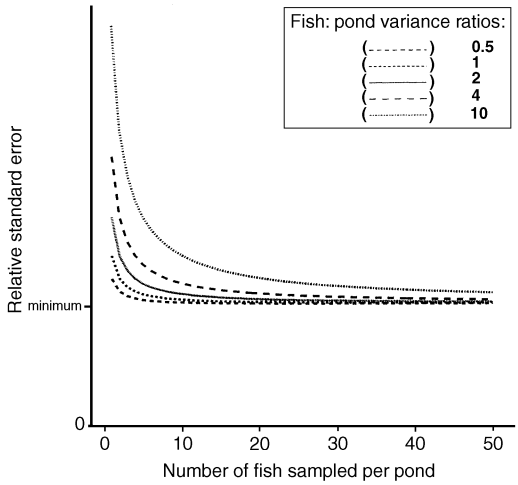

Figure 2 Relative standard error for different fish:pond variance ratios and different numbers of fish sampled per pond.

stocked at whatever density is appropriate for the experiment, but only 10 fish need to be sampled per pond. The fish variability to pond variability ratio remains of the order of between 1 and 10 throughout the experiment. Therefore, provided that the objective is to estimate treatment effects, this sampling strategy would be appropriate whenever sampling is necessary.

Skewed distributions

These examples have relied on Eqns 4 and 5, which are based on the assumption that the fish measurements come from underlying normal distributions. Additional work by one of the authors on the distribution of fish measurements within and between ponds has shown that fish weights within a pond are often skewed by either a few very large fish, or if breeding has occurred during the experiment, a few very small fish. Even if the distribution of measurements is skewed, the distribution of the sample mean for each pond tends towards a normal distribution. Therefore, a slightly larger fish sample size per pond than that previously stated may be required, but with the levels of skewness observed (Table 4), even with a sample size of 10, much of the skewness is removed. For example, the skewness of the distribution of the mean weight when 10 fish are sampled from the most extreme pond in experiment 7 is reduced from 1.47 to 0.40. For any distribution, the sample mean tends to a normal distribution as the sample size increases (for an example, see Mead, Curnow & Hasted 1993; for a proof, see Hogg & Tanis 1993). The large negatives values of skewness in length in a few ponds were caused by the inclusion of the offspring produced during the experiment. If these are removed from the analysis, then the skewness is reduced considerably. For the pond with a

Table 4 Skewness at harvest of fish weight (g) and length (mm) for individual ponds. The mean skewness and the maximum in absolute value are given

	Weight		Length	
Experiment	**Mean**	**Maximum**	**Mean**	**Maximum**
1	0.19	0.75	–0.46	–2.85
2	–0.23	–1.07	–1.02	–2.82
3	–0.03	1.35	–0.39	–0.88
4	0.10	0.87	–0.46	–1.04
5	–0.13	0.89	–0.63	–1.2
6	1.02	1.56	0.29	0.99
7	1.02	1.47	0.35	0.49

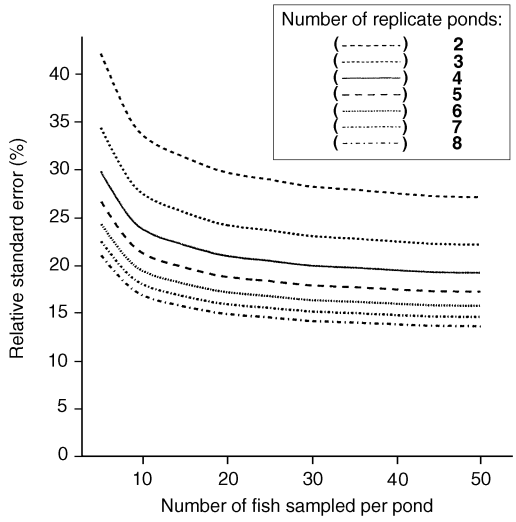

Figure 3 Relative standard error as a percentage of the mean fish weight for the simulation of the extreme case, where skewness = 3 and $v_f / v_p = 10$, with different numbers of fish sampled and replicate ponds.

skewness of –2.82, if fish less than 8 cm are removed, then the skewness becomes –0.702.

If it is suspected that the distribution is very skewed or that the fish were heterogeneous at the start of the experiment, then the sample size should be increased. The present authors simulated the extreme situation of the maximum observed skewness, 3, and a $v_f : v_p$ ratio of 10, and a sample size of 20 fish per pond seemed appropriate here (Fig. 3). This is an extreme case, and is unlikely to arise if similar fish are used at the start of the experiment. In experiments 6 and 7, where the $v_f : v_p$ ratios were large, fish of different ages were used with initial weights ranging from 107 to 913 g. Even in such an extreme case, there is no need to measure as many as 50 or 100 fish per pond when estimating treatment effects between ponds.

Sampling bias

One reason for sampling a large number of fish per pond is to reduce the effect of sampling bias. Most sampling techniques introduce bias because fish cannot easily be selected at random; for example, in an experiment where the first 160 out of 300 fish removed from holding nets were tagged. On average, the tagged fish were larger than those not tagged; the mean difference in weight at the start was 53 g (SE 16 g on 298 d.f.). This difference became greater as the fish grew, and at the end of the experiment, the mean difference was 95 g (SE 29.7 g on 266 d.f.). Various sampling techniques are possible, but those which introduce less bias, such as draining the ponds, are likely to be more stressful to the fish, and hence, less representative of what happens in ponds on farms. A compromise is needed. If the same sampling technique is used for each pond, then the positive bias should be similar for each pond. This will only affect treatment estimates and not estimates of treatment differences.

Other improvements

There are some improvements which can be made in the design and analysis of pond experiments even if the number of replicate ponds cannot be increased. These include: blocking (Cochran & Cox 1957); use

of covariates; regression analysis; non-parametric tests (Siegel & Castellan 1988), where an analysis of variance is inappropriate; and REML (Robinson 1987; Payne, Lane, Digby, Harding, Leech, Morgan, Todd, Thompson, Tunnicliffe Wilson, Welham & White 1993). Simple linear regression, multiple regression and analysis of covariance are discussed in most statistical texts; for example, Snedecor & Cochran (1989) and Mead *et al.* (1993). Under certain circumstances, all of these methods can improve the precision with which estimates are made or the sensitivity of the tests applied. However, their improvement is limited in comparison to the use of an increased number of ponds. Any known differences between ponds prior to the experiment should be incorporated into the design. This naturally leads to blocking of designs, where similar ponds are grouped together in blocks. Treatment comparisons can then be made between similar ponds within blocks, removing some of the between-block variability.

Sampling during the experiment

If fish are sampled during the experiment, the effect of the sampling process on the fish is important. When fish are caught and measured, the animals become very stressed and may refrain from eating, sometimes with fatal consequences. In an experiment, it is important that the fish behaviour is representative of fish in farmers' ponds. If the fish stop feeding, then this will affect their growth. Hence, any treatment effects on the growth of the fish may be affected by the sampling procedure used. Therefore, the number of times the fish are sampled and the number sampled should be kept to a minimum. If the primary objective of the experiment is to estimate treatment effects on the weight or length of the fish at the end of the experiment, little extra information can be gained from sampling during the experiment. However, if the experiment is investigating changes over time, such as growth curves, then sampling during the experiment is necessary, but the effect on the fish must be considered.

Sample size and statistical power

The present paper has concentrated on the effects of sample size on detectable difference. It is also important to consider the power of the studies to detect differences when they exist. In order to increase the power (i.e. reduce the type II error, the probability of no difference being detected when there is a real difference), the sample sizes need to increase further. Again, this will require greater pond replication, rather than an increase in the number of fish sampled per pond. A true difference equal to the detectable difference will be detected 50% of the time, since the estimated difference will be less then the true difference in 50% of experiments (assuming a symmetrical distribution). To increase this rate to being 80% sure of detecting the same difference, the number of replicate ponds should be about double those needed when only considering the detectable difference with reference to a type I error (the probability of identifying a difference when no difference exists). Searcybernal (1994) gives a more detailed discussion on the power of tests and their interpretation in aquaculture experiments.

Conclusion

The major influence on the sensitivity of pond experiments is the number of ponds, and not the number of fish sampled within each pond. This is still true, even with the levels of heterogeneity and skewness observed, and with some limited positive bias in all sampling. Some improvements on the efficiency of the experiment are possible when the number of ponds cannot be increased, but these are very limited. By far the most effective way of increasing the efficiency of the experiment is to increase the number of pond replicates. With as few as three replicates, it may only be possible to detect differences in treatment effects relative to the overall mean of 40% or more.

Acknowledgments

This work was funded by the UK Overseas Development Administration although it claims no responsibility for the information provided or the views expressed. The IACR receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the UK. We are also grateful to Dr D. Little of the Asian Institute of Technology for the use of his data.

References

Cochran W.G. & Cox G.M. (1957) *Experimental Design*, 2nd edn. John Wiley & Sons, Chichester.

Hogg R.V. & Tanis E.A. (1993) *Probability and Statistical Inference*, 4th edn. Macmillan Publishing Company, New York, NY.

Mead R., Curnow R.N. & Hasted A.M. (1993) *Statistical Methods in Agriculture and Experimental Biology*, 2nd edn. Chapman and Hall, London.

Patterson H.D. & Thompson R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.

Payne R.W., Lane P.W., Digby P.G.N., Harding S.A., Leech P.K., Morgan G.W., Todd A.D., Thompson R., Tunnicliffe Wilson G., Welham S.J. & White R.P. (1993) *Genstat 5 Release 3 Reference Manual*. Oxford University Press, Oxford.

Robinson D.L. (1987) Estimation and use of variance components. *The Statistician* **36**, 3–14.

Searcybernal R. (1994) Statistical power and aquaculture research. *Aquaculture* **127**, 371–388.

Siegel S. & Castellan NJ., Jr (1988) *Nonparametric Statistics for the Behavioral Sciences*, 2nd edn. McGraw-Hill, New York, NY.

Snedecor G.W. & Cochran W.G. (1989) *Statistical Methods*, 8th edn. The Iowa State University Press, Ames, IA.

Yates F. & Zacopanay I. (1935) The estimation of the efficiency of sampling, with special reference to sampling for yields in cereal experiments. *Journal of Agricultural Science, Cambridge* **25**, 545–577.