CHAPTER EIGHT

EVALUATING FAD EFFECTIVENESS IN DEVELOPMENT PROJECTS: THEORY AND PRAXIS

Jeffrey C. Johnson and James D. Murray

INTRODUCTION

Fish aggregation devices (FADs) have been viewed by some as an important technological innovation for increasing productivity among small-scale artisanal fishermen in developing countries (Pollnac, this volume).[1] Development agencies, such as AID and international commissions, such as the Indo-Pacific Fishery Commission, have seen FADs as potentially useful for increasing productivity in underdeveloped fisheries and as a tool for the management of small-scale fisheries. In addition, FADs have been viewed as potentially useful in increasing the safety of small-scale fishers by reducing the distances traveled offshore and range of search for less seaworthy indigenous watercraft. One concern of these agencies and commissions has been evaluating the effectiveness and efficiency of FADs in a given context, particularly for lesser developed countries.

Evaluation of FAD projects is not a simple matter of increased harvesting efficiency. FADs have been recognized as potentially problematic, since aggregation itself adds nothing to the productivity of an ecosystem nor does it increase the carrying capacity of any given system. Thus, depending on a number of biological, economic, social, and technological factors, FADs may actually prove to be more detrimental than helpful in the long run. Increasing the ability to harvest at ever increasing efficiency, without increasing productivity, will ultimately lead to non-sustainable yields. Samples and Spraul (1985), for example, do not portray the use of FADs in commercial fisheries in a good light in terms of increasing sustainable yields, increased profits in the long run, and expanding demand for labor under open access. They suggest control of harvest levels and fishing effort as a potential solution to these problems.

The question of sustainability aside, there is still a question as to the extent to which FADs increase efficiency in harvests under various economic, social, biological, and technological conditions. In developing countries, and particularly with respect to artisanal fisheries, FADs may provide a cost effective, simple solution to increasing harvests where there are sufficient harvestable populations. As a technology for improving efficiency in the harvest of marine species, FADs are no more problematic than any other technological means for improving efficiency such as the introduction of or improvements in net technology, fish finders, etc. Thus, FADs should be evaluated no differently than any other technological improvement, particularly in terms of their use in developing fisheries. With this in mind, this paper examines different concerns in evaluating the effectiveness of FADs under a variety of conditions. The first portion of the paper focuses on issues of research design, taking into account the practical aspects of design in developing fisheries. This is followed by a discussion of problems inherent in the statistical analysis of data typically collected in the course of evaluating FAD effectiveness.

EXPERIMENTAL DESIGN

In evaluating the use of FADs there is the simple question of "How effective are FADs?" But this question is a relative one and can only be answered in terms of relative effectiveness (i.e., as compared to alternatives). Depending on the particular context of the projects, this problem presents itself as a classic experimental design issue in terms of treatments and controls. By using different types of FADs, different in relation to both themselves and, possibly, some control (traditional fishing practices without FADs) and keeping fishing methods either constant or varying them by site or some other condition (e.g., depth), one can examine factors affecting variations in catch. Ultimately, there is an interest in statistically assessing differences among both treatments and controls and coming up with estimates of catch per unit effort (or any other estimate of interest). The former allows for assessments of relative effectiveness while the latter will help in assessing sustainability of these fishing practices under different bioeconomic or sociological conditions.

There are a variety of references on experimental design that are certainly appropriate for the problem at hand. Optimally, a project might best be evaluated using some randomized balanced block design which in turn requires relatively more sophisticated parametric analytical tests. Practically, however, the human component, particularly in developing fisheries, will limit the complexity of experimental design and the nature of catch data will call into question the use of standard parametric statistical tests. In actual field experiments, random is not always random and attempts to control for extraneous or confounding influences are not all that controllable. It is not highly trained technicians collecting samples, but usually local fishers who are a part of a fishing culture with norms and relations that ultimately will affect both the nature of design and the validity of both the data and the statistical analysis.

Research Design: Simple Example In this section we provide a brief example of a simple design for evaluating the effectiveness of FADs in a fixed structure recreational fishing context in Wrightsville Beach which is located near Wilmington, North Carolina (Figure 1). Although this is an example drawn from the United States, the pure and practical design issues encountered here will illustrate the actual problems of evaluation stemming from the human component.

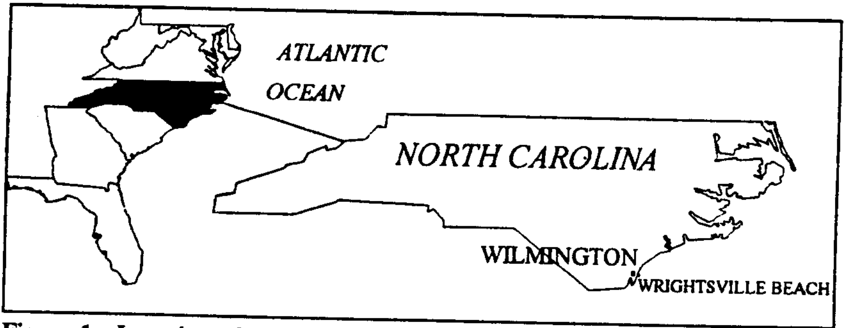

Figure 1. Location of Wrightsville Beach in the Southeastern United States.

The example comes from Murray et al. (1987) on a comparison of two piers, one using a FAD, the other not. Johnson served as a consultant on this project in terms of the overall research design. The objective was a simple one: to determine the effectiveness of FADs for fixed structure recreational fishing. Early design considerations were also influenced by the number, kind, and configuration of the FADs to be used. In this case, a single configuration of a specific kind (McCintosh) was to be explored. A single pier could suffice in an alternating placement and removal of the FAD units. However, there was always a chance that statistically significant differences in catches between periods could be a function of weather, water temperature, or a variety of ethological or ethnological factors. Thus, two piers that share a high degree of similarity were preferable.

Ultimately two piers were selected that were within proximity to one another (but not so close so as to worry about spatial autocorrelation) and had similar characteristics in terms of length, surrounding water depth, fishing restrictions, water temperature, seasonality, and sand substructure. But the concern for similarity in characteristics did not guarantee any "natural" difference between the two piers. The question then was, "Were there any differences in catch per unit effort between the two structures?"

In pursuit of this question, the basis for the experimental procedure was developed in which piers were to be sampled simultaneously three days per week, three hours per day with hours chosen on a random basis each day of the 5-to-6-week sampling period. In addition, the pier was divided into three sections of 76 meters each--nearshore, middle, and end. Pre-experimental assessment of any differences in piers was conducted for any indication of possible differences in CPUE that might be a function of existing differences between the piers. This was done 10 weeks prior to the start of the study.

After pre-experimental assessment, piers were sampled as described above in an alternating treatment/control sequence with sampling periods of 5-6 weeks in duration. Sampling days and hours were chosen randomly with days being divided into three six-hour periods containing each of six one-hour segments. Segments within periods were also chosen at random. Data collected included conditions (e.g., windspeed, water temperature, etc.), wet gear time, species caught, amount and type of gear used, and weight and length of fish caught. In addition, a similar design was used in the direct observation of FADs while in place (i.e., observation of control and treatment sites by divers).

This is a simple example of an attempt to control for extraneous variables in assessing the effectiveness of FADs. Concern for similarity of pier characteristics, the characteristics of the surrounding ocean environment, and the sampling design itself were all carefully considered so that any observed differences between the FAD and non-FAD sites could be attributed exclusively to our hypothesized effect (i.e., FAD pier will have higher catch rates), at least theoretically. Of course, even under the best of experimental conditions, we still must be cautious in assigning true cause and effect. Nevertheless, under the scenario described we have applied reasonable common sense in attempting to control for any confounding variables.

Experimental design will naturally be determined by the problem at hand. For example, Rountree (1989) designed a study that was interested in understanding the effect of FAD structure size on extent of aggregation for what he calls his protection hypothesis. He used a randomized block design in order to control for both location effects and temporal effects (e.g., tidal effects, various behavioral effects). Cayre et al. (1991) reports on a study that, unlike Rountree's observational approach, actually attempts to determine the effectiveness of FADs in a more real world setting involving an artisanal fishery in the Comoros Islands. In this study there were open sea conditions with no FADs (control) and fishing near FADs (treatment). In addition, motorized vessels with troll lines and artificial lures and non-motorized canoes with handlines and natural bait were compared under control and treatment conditions. Data were collected by technicians at each of 17 artisanal fishing landing sites.

Despite careful concern for specific design elements, it is often the case that humans will inevitably complicate, and sometimes compromise, our ability to test the hypothesized effects of interest. In the Murray et al. (1987) study, for example, pier owners, concerned over the possible interference of FAD units with normal fishing (i.e., getting lines tangled), requested the units be placed no closer than 229 meters from the pier. Referring to this as "unnecessarily restrictive," Murray et al. (ibid:145) recognized this distance as having a potential impact on the results of the study. In developing countries, FAD placement may be subject to not just environmental concerns (e.g., depth, water temperature), but social ones as well, such as fishing territoriality, share systems, and so on (Pollnac 1988). Thus, design issues must consider both natural and human factors as potential confounding variables and in the possible presence of any possible measurement error.

Another important consideration in the overall design of the research involves the types of statistical analysis to be used. Complicated block designs, for example, often require complex parametric analyses, which in turn require that certain assumptions be met (e.g., normality, random) for testing hypotheses. The next section discusses potential analytical problems as they relate to the type of fisheries data normally found in assessing differences between treatment and controls based in terms of catch per unit effort (CPUE).

Potential Analytical Pitfalls Ultimately any elevation will be determined by some standard in order to determine what constitutes a "true" difference between alternatives. The gold standard used by most is the idea of statistical significance. A detailed discussion of significance is beyond the scope of this paper, but as we shall see, means for meeting the assumptions of both commonly used test statistics (e.g., t-test) and means for assessing significance (e.g., probability distribution) must always be called into question, particularly with regard to the types of data routinely collected in evaluating FADs. Instead of wading through a terse discussion of assumptions surrounding different statistical approaches, in this section and the sections to follow we will explore the problems inherent in FAD evaluation data and provide an example an discussion of potential analytical pitfalls.

First, data used to evaluate FAD effectiveness is almost universally messy. That is to say, catch or observational data typically found in FAD evaluation research is often highly skewed, heavily tied, sometimes sparse, particularly on a species-by-species basis, many involve small sample sizes, comparison groups (samples) of different sizes, and may not truly constitute a random sample. Two examples of the kind of data typically found in research of this kind can be found in Table 1. In both cases (one an example of catch per unit effort data, the other representative of observational data collected by divers) standard deviations are approximately two to three times that of the mean, reflecting highly skewed data.

Table 1. Means and standard deviations for catch data from studies evaluating FAD effectiveness: catch or observation data for treatments by species.

	Species	Mean	SD
Murray et al. (1987)	*Pomatomus saltatrix*	.006	.016
Rountree (1989)	*Decoptdrus punctatus*	171	325

Much of this skewness is due to zero catch rates or no observations leading to a large number of ties at one end of the distribution. The presence of these ties, particularly at the zero end, have implications for choosing appropriate methods for both transforming or reexpressing the raw data and the eventual selection of proper analytical techniques that minimize any potential violations in assumptions.

Table 2. Means and standard deviations for CPUE for species combined for both piers.

	Mean	SD
Pomatomus saltatrix	0.325	0.802
Leiostomus xanthurus	0.187	0.651
Trachinotus spp.	0.065	0.375
Scomberomous maculatus	0.104	0.718

In order to understand the nature of these pitfalls and any potential solutions, we analyze a portion of data from the overall study by Murray et al. (1987). Two sampling periods from the month of September were selected on the basis of the presence of relatively large numbers of observations for each of the piers in the study. For this period, pier 1 is the treatment (i.e., FAD pier), while pier 2 is the control. Data to be compared is catch per unit effort data defined as the catch per minute per unit of gear (i.e., rod). Total number of observations for pier 1 are 21 and for pier 2, 44, providing a good example of data containing samples of unequal size. This will provide a simple example of a situation involving typical data based on a relatively modest sample size.

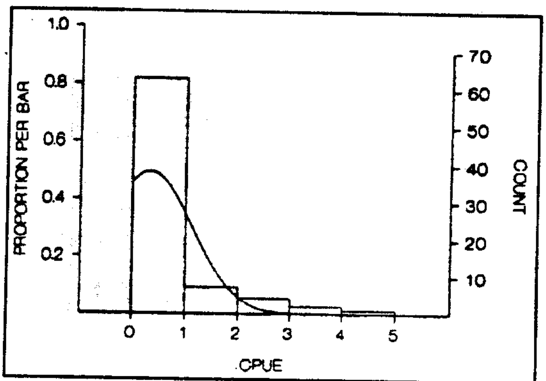

Figure 2. Density plot of CPUE for *Pomatomus saltatrix*.

Means and standard deviation by species are provided in Table 2. As can be seen, the data are highly skewed. Figure 2 is a density plot for CPUE for *Pomatomus saltatrix* for both piers, including a normal smoothing. Again, it is readily apparent that the data are highly skewed and heavily tied. The question is: What kinds of problems do data of this form present in terms of appropriate selection of analysis and assessment of statistical significance.

Statistical Tests and Probability Distributions One's initial reaction may be to rely on a test statistic familiar to all introductory statistics students, the t-test. After all, the test is familiar to most, used by many, and gets at what we want, namely whether there are any differences between the mean catch rates of pier 1 and pier 2. We are really interested in whether the FAD pier is catching better than the non-FAD pier.

Although we "secretly" seek to prove that the FAD pier catches better than the non-FAD pier, in a statistical test we actually test the null hypothesis (H_0) that there is no difference between the mean catch rate of pier 1 and pier 2 ($H_0: \mu_1 - \mu_2 = 0$). The t-test provides us with a single value that helps in assessing departure from the null hypothesis, the t-value. But this single value alone is not sufficient to tell us whether the null hypothesis can be rejected. We need some probability distribution for the test statistic under the assumption that the null hypothesis is true. In parametric statistics such probability distributions are derived by putting conditions or structure on the null hypothesis that makes it theoretically possible to calculate a distribution. But in order to use these distributions in evaluating the null hypothesis, we are required to make rather restrictive assumptions about our sample data and the population from which it was drawn. In parametric approaches these assumptions typically require us to consider whether our data are from a random sample from a population that has particular characteristics (e.g., variables are independently and normally distributed with constant means and variance). Thus, theoretically at least, a test of the null hypothesis is valid if and only if these conditions are met. Now the question becomes: Given the types of data we find in the course of evaluating FAD projects, do we generally meet these rather restrictive assumptions?

The answer to this question is typically "no." Putting aside for a moment the condition of independence, we may often call into question the conditions of randomness or normality, let alone other special conditions (e.g., regarding means and variance). However, in all fairness to parametric tests, such as the t-test, there is evidence (primarily from extensive simulation research) that the sample data need only be of sufficiently large size and approximately normal in tests to be valid. In addition, there are ways for dealing with other problems as, for example, unequal variances (e.g., independent samples, separate variance t-test).

We will return to the issue of randomness shortly. For now, let us return to a discussion of the data found inTable 2.

As pointed out, the data in this table are highly skewed, possibly not lending themselves very easily to parametric tests. There are transformations of the raw data that will sometimes improve upon normality problems within the samples and include log and rank reexpressions. However, data are sometimes so skewed or messy that transformations are largely unsatisfactory. Figures 3 and 4 show the raw and transformed density plots for the CPUE for all species combined for both piers. Although there is some improvement in normality as we move from Figure 3 to Figure 4, the reexpression of the raw data is still far from normal or approximate normal.

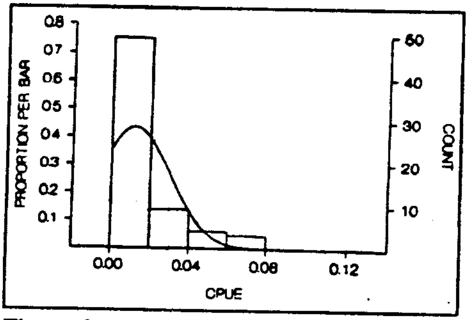

Figure 3. Density plot of the raw data for total CPUE.

A test statistic that has no distributional assumptions and is related to the t-test is the non-parametric Wilcoxon rank sum test. This test compares the sum of the ranks between the two samples and is therefore less affected by distributional factors such as outliers or skewness. Although tied ranks can affect the outcome, there are means for correcting for ties, but except for under any but the most extreme conditions the correction alters the outcome little. As we shall see, this is an excellent test for the analysis of the kind of data we are interested in here, but

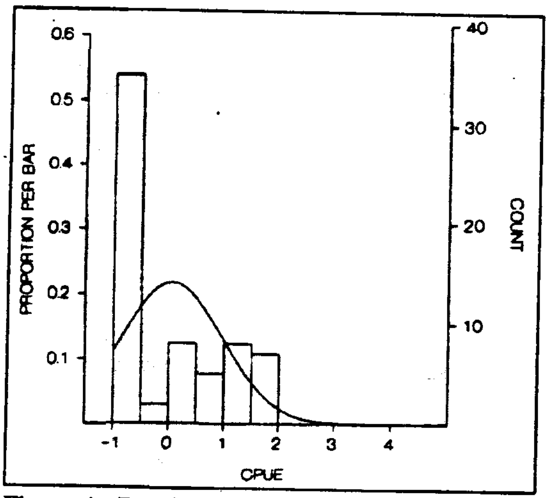

Figure 4. Density plot of log transformed data for total CPUE.

is limited to test of the difference between two samples. In addition, traditional means for assessing significance (e.g., asymptotic p-values) may be problematic under certain conditions (e.g., small sample sizes).

Any potential problems in assessing significance even for nonparametric tests can be eliminated through the use of exact permutational or randomization procedures. The probability distribution to be used is specifically related to the problem at hand; it is not theoretical but empirical. Thus, instead of relying on a theoretical distribution requiring restrictive assumptions we can utilize exact permutational, Monte Carlo estimates of exact permutations, or randomized distributions that require little in the way of assumptions. Similar to theoretical probability distributions, the probability distribution for exact or randomized tests represent the distribution for the test statistic under the assumption the null hypothesis is true. However, because these distributions are empirically based, generated with the use of computer intensive methods, assumptions about the nature of the population or the sample are unnecessary.

The ability to empirically generate probability distributions for the problem at hand allows us to move beyond the rather limited number of traditional test statistics (and their corresponding theoretical distributions) and develop test statistics that are more relevant to a specific research problem. Randomization tests more generally are computer intensive methods for testing hypothesis that are essentially assumption-free. The general principal involves the calculation of the observed test statistics from the sample data. The empirical probability distribution is generated by randomizing or shuffling the sample data within categories, experimental groups, treatments/controls, etc. Following each shuffle, the test statistic is calculated and this is repeated for an adequate number of times (e.g.,1,000), resulting in a distribution of randomized test statistics. The observed test statistic is compared to the distribution of empirically derived test statistics under the assumption the null hypothesis is true. If the observed value is extreme in relation to the empirical values, then the null will be rejected (see Noreen 1989 for a good review).

In the following section we provide three examples of the analysis of data from a FAD evaluation project described earlier. Each provides an example of the type of statistical tests discussed above. The first example uses a conventional parametric test, the second an exact parametric test, and the third a randomization test of our own design.

Analytical Example Murray et al. (1987) used a parametric t-test in this analysis of CPUE data between control and treatment piers. Frusher (1986) employed a conventional t-test to test differences in the size of fish caught at FAD and non-FAD sites. Rountree (1989) used two- and three-way ANOVA models in comparing FADs of different size. Cayre et al. (1991) employed the Wilcoxon rank sum test using conventional asymptotic p-values, a nonparametric analogue to the t-test, in testing for differences in FAD and non-FAD sites. In this section, we will compare the results of three different approaches to the analysis of the subset of data from Murray et al . (1987). Particular attention will be focused on the structure of the null hypotheses in relation to the theoretical or empirical probability distribution, related assumptions, and interpretations and comparisons of resulting p-values.

Table 3 shows the output of a t-test analysis using the statistical package SYSTAT (Wilkinson 1990). Because of the sparseness of data for individual species, we will look at differences in catch per unit effort for all species combined (total CPUE). As mentioned earlier, skewed data of this kind is typically reexpressed in order to address problems of normality. In this case, analysis of the raw and log transformed data will be compared. Figures 3 and 4 show the raw and log transformed density plots respectively. As is evident, even the log transformed data is highly skewed.[2]

Table 3. SYSTAT results of t-tests for raw and log-transformed data.

Independent samples t-test on log of total CPUE grouped by pier

Pier	N	Mean	SD
1	21	0.303	0.945
2	44	-0.065	0.894

Separate variances t = 1.493
DF = 37.5 P = 0.144 (.072)
Pooled variances t = 1.523
DF = 63.0 P = 0.133 (.067)

Independent samples t-test on total CPUE grouped by pier

Pier	N	Mean	SD
1	21	0.016	0.020
2	44	0.010	0.017

Separate variances t = 1.133
DF = 33.7 P = 0.265 (.132)
Pooled variances t = 1.209
DF = 63.0 P = 0.231 (.110)

We are interested in testing the mean difference between the treatment and control piers with expectations that the treatment pier will have a higher CPUE than will the control pier. However, we actually test the null hypothesis as follows:

$$H_0: \mu_1 - \mu_2 = 0$$

The null hypothesis is simply that there is no difference between the piers. In this case, we are interested in a one-tailed test of the following type:

$$H_a: \mu_1 - \mu_2 > 0$$

The alternative hypothesis reflects our expectations that the treatment pier will have a higher CPUE than the control pier. Because this is a parametric approach, the test is only valid if the two groups are from random samples from the same normal population. Depending on whether this is a pooled or separate variance t-test, we must also be concerned about whether the two samples have equal variance.

The results in Table 3 show that in either the raw or log transformed case the null hypothesis cannot be rejected at the "gold standard" 0.05 level. The reported probabilities are two-tailed and should be divided by 2 (a/2). One-tailed probabilities are reported in parentheses. It should be pointed out that the real probabilities should be reported independent of whether they meet or exceed a threshold significance value or not (e.g., 0.05). There can be vast differences between analyses where probabilities miss by a mile or barely make it under the wire (Noreen 1989).

We next examine an analysis of the data using a nonparametric test with both a p-value estimate on the basis of a theoretical distribution (asymptotic p-value) and one based on an exact estimate of the p-value. The Wilcoxon rank sum test, described earlier, is a test statistic based on a comparison of the ranked sums between two samples. The null hypothesis is stated as follows:

H_0: the distribution of differences is symmetrical around 0

This null hypothesis is similar to that for the parametric t-test in that it assumes no differences between control and treatment. The alternative hypothesis could be stated as follows:

H_a: differences will be larger than 0

That is to say one distribution is stochastically larger than the other. This is again a one-tailed probability problem.

Results of an analysis of the data using the computer pack-

Table 4. Output from STATXACT for Wilcoxon rank sum test showing the asymptotic and exact estimate for p.

Summary of exact distribution of Wilcoxon rank sum statistic:

Min.	Max.	Mean	St.dev.	Observed	Standardized
378.0	1,154	693	65.48	800	1.634

Mann-Whitney Statistic = 569.0

Asymptotic Inference:
One-sided p-value: Pr {Test Statistic .GE. Observed} = .0511
Two-sided p-value: 2 * one-sided = .1022

Monte Carlo p-value estimates at 99.00% level of confidence:
1-sided: Pr { Test Statistic .GE. Observed } = .0540 q .0130
2-sided: Pr {|Test Statistic - Mean| .GE. |Observed - Mean|} = .1105 q .0180

age for exact nonparametric inference called STATXACT (Mehta and Patel 1991) as well as the results from the Wilcoxon rank sum test are presented in Table 4. In this case both the asymptotic p-values and the Monte Carlo estimate of p meets the "gold" standard for rejection of the null hypothesis that the two distributions are stochastically the same. However, the Monte Carlo estimate of p and the 99 percent confidence interval yields the most potentially valid estimate even when sample sizes are small and the data messy (i.e., tied, sparse, skewed; see Mehta and Patel 1991).

Finally we present an analysis of the data using a simple randomization test based on differences in means between the control and treatment groups. As described above, means are subtracted representing the observed difference between pier 1 and pier 2. The data is then randomized within categories (piers) and a difference in means is calculated. This is repeated for a large number of tests (e.g., 2,000), creating a distribution of mean differences under the null hypothesis. The null hypothesis in this case is:

$$H_0: X_1 - X_2 = 0$$

Once again, since we are interested in whether the FAD pier is catching better than the non-FAD pier, the alternative hypothesis is:

$$H_a: X_1 - X_2 > 0$$

Noreen (1989) provides code in several different programming languages for carrying out this and other related hypothesis tests. Here we use a randomization statistical program called RESAMPLING STATS (Simon 1991). The program uses a simple programming language for customizing your own randomization or bootstrap tests. Our

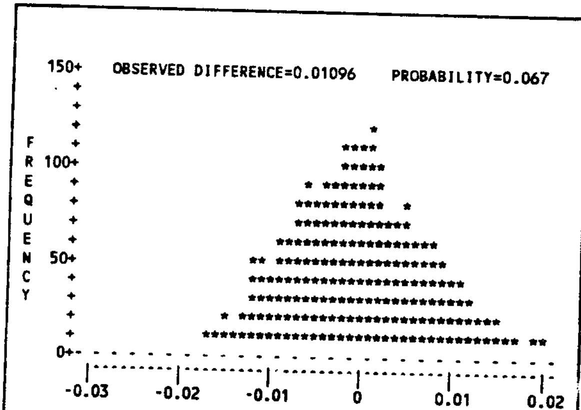

Figure 5. Output from RESAMPLING STATS.

test is based on a comparison of the observed difference in means to a distribution of 2,000 randomized differences. Figure 5 is output from RESAMPLING STATS for a histogram of the empirical differences in means. The observed difference is 0.01096 and lies to the far right in the empirical distribution. The area to the right of the observed difference is 0.067 (the one-tailed probability). Although this does not meet the 0.05 rejection criterion, the observed difference in means is still a relatively rare event and yields results similar to those of the t-test for the log transformed data. Like the previous example, this hypothesis test requires no assumptions about the sample, the population from which it was drawn, or, for that matter, whether it was even a random sample. We care only if the catch rates of the two piers are different and no inference is made concerning any population.

Table 5. Comparison of statistical tests and assumptions.

TEST	ASSUMPTIONS	COMMENTS
Parametric t-test independent samples equal variance (pooled variance)	two samples are independent, random samples from a normal population, two population variances are equal	sizes of samples should be approximately equal; sample sufficiently large and approximately normal; Outliers should be dealt with.
Parametric t-test independent samples unequal variance (separate variance)	two samples are independent, random samples from normal population	sample sufficiently large and approximately normal; outliers should be dealth with.
Wilcoxon rank sum test using exact or Monte Carlo estimate of exact probability	two samples are independent (random)	exact nature of sample data less problematic
Randomization of test for differences in means	independence	need for random sample relaxed but independence is still important

The assumptions surrounding the probability distribution for each of the hypothesis tests discussed in this section are summarized in Table 5. In all cases, the independence of observation is still required for the analysis to be valid (Edgington 1987). However, as we move down through the various tests, the number of required assumptions to ensure validity of the p-values declines dramatically. Also, we observed differences in the magnitudes of the p-values between the different approaches. The Wilcoxon rank sum test using an exact estimate of p yielded statistically significant and valid results as compared to conventional parametric tests.

For parametric statistics hypotheses are concerned with characteristics of a population and as such randomness is an important assumption. However, even though samples may not be random or randomness may be in question, we may be simply interested in whether variables are related in some way or whether groups differ in some particular manner without any concern for generalizing to a population. Randomization and exact tests allow for valid means to test hypotheses of this kind (Noreen 1989).

SUMMARY

Careful consideration of both research design and analytical issues will help to ensure valid assessment of the effectiveness of a given FAD project. An important concern analytically is the degree to which the data collected meet the assumptions of probability distributions used for testing the null hypothesis. Given the rather messy data generally found in studies on FADs, and considering the sample data may not be from a random sample, randomization and exact tests provide an assumption-free alternative to traditional statistical approaches.

As we have seen from the examples, there is currently software available for personal computers to carry out analysis of this type. In addition, Edgington (1987) and Noreen (1989) provide discussions of the use of randomization tests for more complex research designs such as randomized block designs with repeated measures. Given the potential problems associated with evaluating the effectiveness of FAD projects in a development context, researchers need the latest tools at their disposal in order to make the most valid assessments possible.

Finally, the design of experiments for evaluating FAD effectiveness needs to be aware of sociological factors that may impact valid assessment. This is of particular concern in a developing country context where such things as fishing territoriality or secrecy may impact FAD placement and the reporting of catch data. Even the best of experimental designs can be undermined by the cultural context of the evaluation study itself.

NOTES

1. This work was funded in part from Sea Grant development funds. We would like to thank Richard Pollnac for his recognition and concern for design and analytical issues of the type reviewed in this paper. Thanks also to Patricia for providing some extra work space and motivation.

2. A constant of .5 was added to the raw data in order to perform a log transformation.

REFERENCES CITED

Cayre, P., P. Letouz, D. Anungee, and J. Williams (1991) Artisanal fishing of tuna around fish aggregating devices in Comoros Islands: Preliminary estimate of FAD efficiency. *In* Symposium on Artificial Reefs and Fish Aggregation Devices as Tools for the Management and Enhancement of Marine Fisheries Resources. Pp. 61-74. Bangkok: Regional Office for Asia and the Pacific/Food and Agricultural Organization of the United Nations.

Edgington, E.S. (1987) Randomization Tests. Marcel Dekker: New York.

Frusher, S.D. (1986) Utilization of small-scale fish aggregation devices by Papua New Guinea's artisanal fishermen. Proceedings of the First Asian Fisheries Forum, pp. 371-374. Asian Fisheries Society: Manila.

Mehta, C. and N. Patel (1991) STATXACT. Cytel Software Corporation: Cambridge, MA.

Murray, J.D., J.C. Howe, D.G. Lindquist, and D.C.Griffith (1987) Using FADs to attract fish at coastal fishing piers. Marine Fisheries Review 49(2):143-154.

Noreen, E.W. (1989) Computer Intensive Methods for Testing Hypotheses. John Wiley & Sons: New York.

Pollnac, R. B. (1988) Social and Cultural Characteristics of Fishing Peoples. Marine Behavior and Physiology 14:23-39.

Rountree, R.A. (1989) Association of fishes with fish aggregation devices: Effects of structure size on fish abundance. Bulletin of Marine Science 44(2):960-972.

Samples, K.C. and J.T. Spraul (1985) Fish aggregating devices and open-access commercial fisheries: A theoretical inquiry. Bulletin of Marine Science 37(1):305-317.

Simon, J.L. (1991) RESAMPLING: Probability and Statistics a Radically Different Way. Copyright Julian L. Simon.

Wilkinson, L. (1990) SYSTAT: The System for Statistics. Evanston, IL: SYSTAT, Inc.