PREDICTING EFFECTIVENESS OF WOOD PRESERVATIVES FROM SMALL SAMPLE FIELD TRIALS

Carol L. Link

Mathematical Statistician

and

Rodney C. DeGroot

Research Plant Pathologist
U.S. Department of Agriculture, Forest Service
Forest Products Laboratory[1]
One Gifford Pinchot Drive
Madison, WI 53705-2398

(Received January 1989)

ABSTRACT

Field tests of wood preservatives use groups of stakes treated at various retentions. Although only an average value is reported for a given group of stakes, the lifetime of individual stakes is quite variable. This paper explores presentations of data that reflect such variability. We also consider the feasibility of predicting the effectiveness of a preservative before all stakes fail. For sample sizes of ten replicate stakes, we suggest that reports include box plots of the actual failure times, and that studies use the sample median for the reported lifetime value rather than the sample mean and report the first quartile as a lower bound for the population average.

Keywords: Box plots, censored data, nonparametric statistics.

INTRODUCTION

Wood has considerable natural variability in its properties. Engineering design values account for this feature in different ways. For example, grading rules for identified wood species quote a strength value that is near a low percentile of the strength distribution, whereas the stiffness value is closer to the population average. However, current procedures for field testing preservative-treated wood do not include variability as a major component of the data interpretation.

The effectiveness of wood preservatives is usually summarized by sample mean values. For example, Gjovik and Gutzmer (1989) report sample lifetime mean values, and Colley (1970) and Hartford (1972) report mean stake ratings over time. Ideally, statistical statements should be made about the population rather than the sample mean. The analysis should also include a measure of variability. Consumers might be more interested in the minimal lifetime of a treated stake rather than the average lifetime. However, as a retention of a wood preservative is often tested with a series of ten replicate stakes, little information exists for the lower part of the distribution of stake lifetimes. Data from tests of different preservatives and retentions cannot be readily combined to give distributional

---

[1] The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin.

information that would be valid for an individual preservative and retention. With ten replicate stakes, only average trends are usually discussed. The questions are: Does the current method of reporting best describe the materials on test and Is this method an adequate predictor of the performance of treated wood products that might subsequently go into service?

The purpose of this paper is to explore alternative methods of presenting information about preservative-treated stakes in field trials to reflect data variability. Using parametric and nonparametric procedures, we will also explore the feasibility of using early failures in a group of replicate stakes to predict the sample median of that group. We define failure as a stake destroyed by decay and/or termites [rated 5 or E by the Forest Products Laboratory (FPL) rating scheme; Gjovik and Gutzmer 1989]. However, other definitions of failure could be used, such as time to significant decay and/or termite attack—a log score of 7 (as used by Colley 1970) or a decay index of 50 (as used by Hartford 1972).

DATA UNDER CONSIDERATION

This report considers data from stakes placed on test by FPL. During the past 50 years, FPL has tested the effectiveness of wood preservatives using field tests of 19,000 treated and untreated stakes in Florida, Louisiana, Maine, Mississippi, and Wisconsin. The vast majority of stakes are tested in groups of ten replicates at a given preservative and retention level. The progress of these field trials has been periodically reported in a series of research notes, FPL-02 (most recent report, Gjovik and Gutzmer 1989). Each group of replicate stakes is described in terms of the percentage of stakes that are good, serviceable but showing signs of decay and/or termites, or removed from test because of decay and/or termites. After all stakes have failed, the mean lifetime of a group of replicate stakes is reported.

Too many stakes exist to consider all of them at once. Therefore, in this report we initially confine our attention to groups of ten replicate 2- by 4- by 18-in. Southern Pine stakes placed on test in Mississippi. We will first analyze the variability of 510 control stakes. Then, we will consider methods for estimating the sample median lifetime before all stakes fail. Finally, we will consider confidence intervals for the population median.

Not all available data were used for modeling purposes. Some data were used to create or choose a model, and other data were used to validate the model. The data used for modeling purposes were from 79 groups of Southern Pine 2- by 4- by 18-in. stakes on test in Mississippi, which were treated with chromated zinc chloride, fluor chrome arsenate phenol (type A), copper naphthenate, coal-tar creosote, and pentachlorophenol, from plots 2, 5, 6, 20, 24, 38, 41, 48, 55, 59, and 67. These treatments were chosen because they represent a variety of inorganic and organic preservatives in water- or oil-based formulations. Historical data were also chosen so that adequate failure information was available for the statistical analysis. The data sets used to validate the models were from other preservatives and types of wood products on test in Mississippi and Wisconsin.

Variability of data

The failure times of any group of replicate stakes vary. The difference between the first and last failure times of a particular group of ten stakes ranges from years to decades. Stake plots are inspected at discrete intervals, usually ranging from every 6 months after the plot is first installed to every 2 years after that plot has been on test for a decade or more. Failure time is currently defined as the length of time between the installation date and the inspection date when the stake is observed to have failed. In reality, the stake failed some time between the last inspection date before failure and the inspection date of failure. Because of the discrete nature of the inspection times, there are often multiple failures at a given inspection date. To better reflect the fact that the deterioration process is continuous, we adjusted the recorded failure times so that only one failure occurs at any specific time. For example, if four failures were discovered at the year 2 inspection, and the prior inspection was at year 1, then the failures are assumed to have occurred at 1.25, 1.5, 1.75, and 2 years. This conservative approach did not account for seasonal influence on rate of deterioration and had the most impact on data for untreated control stakes that tended to deteriorate over a relatively short time span. Compared to failure times for untreated stakes, failure times for treated stakes were spread out over a greater number of years and were less clustered; hence, the data for treated stakes were less influenced by this procedure. Unique failure times were necessary for some statistical procedures used in this paper. The slight adjustment of failure times did not significantly affect any results, given the large variability in failure times.

The variability of individual lifetimes is due to the inherent variability in wood stakes, in the actual preservative retention, and in the presence of decay fungi and termites over time in a particular location of the test plot. However, there are no quantitative measurements of any of these factors other than individual stake preservative retention values.

Variability of controls

To explore the variability of stake data, let us first consider the controls. When each plot is established, a set of untreated stakes is included to serve as controls, or a reference baseline. If the underlying nature of the wood or the experimental environment has not changed over time, similar failure patterns should occur for control stakes of similar species. Therefore, the variability observed within or between control stakes in the plots is likely to serve as a lower bound for the variability expected in field tests of wood preservatives.

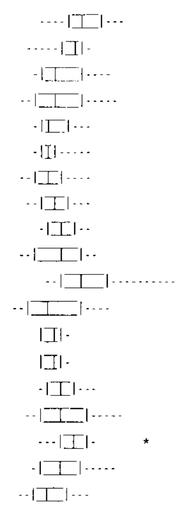
We will consider 51 groups (1 per plot) of Southern Pine sapwood 2- by 4- by 18-in. control stakes installed between December 1938 and May 1982 at the Harrison Experimental Forest in Saucier, Mississippi. The variability of the stake lifetimes within a given plot can be presented in different ways. In Table 1, failure time data are represented as a histogram and a box plot. In the modified histogram or grouped frequency distribution, time is broken into a ¼-year grid with the number of failures listed in each cell. In the box plot (Velleman and Hoaglin 1981), a box surrounds the center 50% of the data, the median is represented by a perpendicular bar, and "whiskers" extend to the smallest and largest observations, except for outliers, which are denoted by an asterisk.

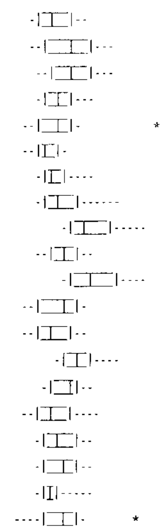
Actual failure times range from a few months to 6 years. Neither representation indicates sample means. However, the box plots show the median lifetime as well as some indication of the variability in each group. The sample mean and median lifetimes are similar within a plot and range from 1.4 to 3.6 years. Because the

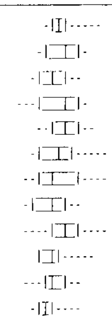
TABLE 1. *Presentation of failure times of Southern Pine 2- by 4- by 18-in. controls.*

Installation		Modified histogram[a]
Date	Plot	Failure time (yr) 0 1 2 3 4 5 6
12/38	2	1 1 11121 1 1
1/40	3	1 1 2222
6/40	4	1212 1 1 1 1
2/41	5	1111112 1 1
2/42	7	231 111 1
2/42	9	12321 1
2/42	10	1 21211 1 1
12/43	12	1 212111 1
12/43	13	21211111
12/44	14	11 111 11111
12/45	15	11 2111 1 1 1
4/48	20	1111111 1 1 1
12/48	24	32221
12/48	25	32221
12/49	26	2121111 1
12/49	27	1 1111111 1 1
12/49	28	1 1 21211 1
3/52	32	111111111 1
3/52	33	1 2111111 1
3/52	34	211111111
3/52	35	111 111 111 1
3/52	36	11 1111111 1
12/52	37	1221111 1
4/53	38	111111111 1
12/54	40	1 22221
12/54	41	22211 1 1
12/54	42	21211 1 1 1
12/56	47	2111111 1 1
12/57	48	111121111
12/58	53	11111111 1 1
12/59	54	1 11111121
12/59	55	1 21111111
12/60	56	121211 1 1
12/60	57	11121211
7/61	59	1111121 1 1
12/63	61	121111111
12/63	62	211111111
12/63	63	123111 1
1/67	66	1 1 1 12111 1
11/67	67	2232 1
12/71	68	11211121
5/76	74	21121111
12/76	75	1 11 1 2121
11/78	78	111112111
4/80	80	11211111 1
6/80	81	1111111 1 1 1
6/80	82	111121111
5/81	83	1 1 11211 1 1
12/81	85	22221 1
5/82	86	1 1212111
5/82	87	1331 1 1

Box plot[b] — Failure time (yr) 0 1 2 3 4 5 6







[a]Time broken into a 1/4-year grid with number of failures.

[b]A box runs from the first to the third quartile, the median is represented by a perpendicular bar within the box, and "whiskers" extend to the smallest and largest observations, except for outliers denoted by *. Outliers are those data points that are far away from the center of the data, compared to the interquartile range (Velleman and Hoaglin 1981).

median lifetimes do not show a pattern over time, no evidence exists that the lifetime of the control stakes has increased or decreased over time. It is important to note the variability of failure times within each plot, which is not summarized in an average lifetime.

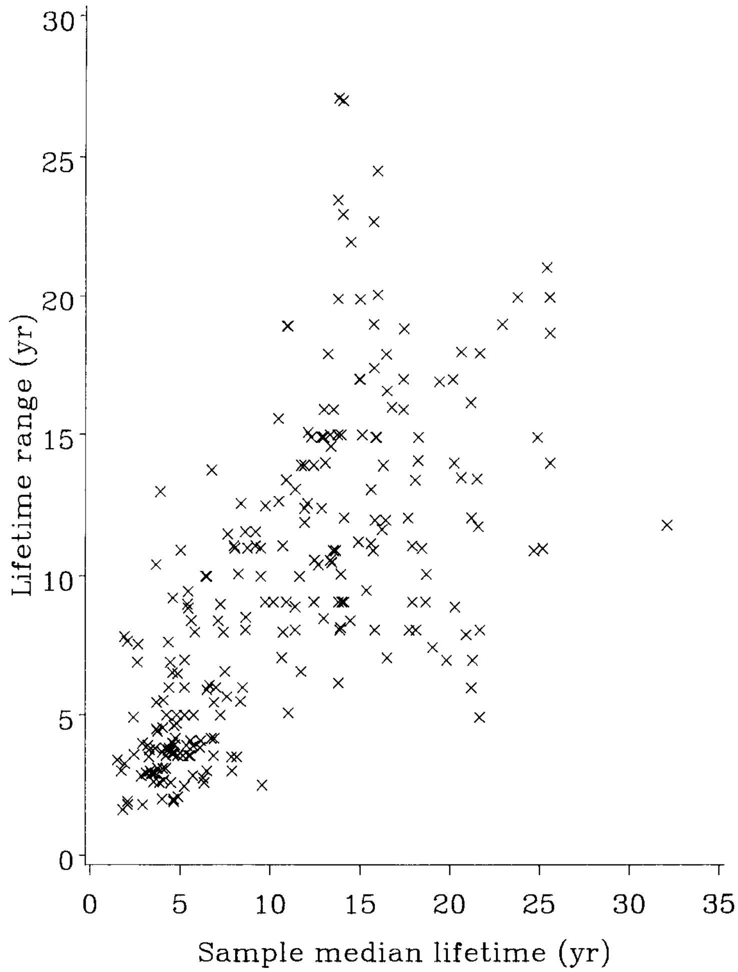


FIG. 1. Median lifetime compared to lifetime range for a group of ten failed replicate stakes treated with different retentions of chromated zinc chloride, fluor chrome arsenate phenol, copper naphthenate, coal-tar creosote, or pentachlorophenol.

Variability of treated stakes

Like the controls, the individual failure times vary widely within a group of ten replicate stakes. For groups of ten replicate stakes treated with either chromated zinc chloride, fluor chrome arsenate phenol (type A), copper naphthenate, coal-tar creosote, or pentachlorophenol in which all stakes have failed, variability in failure time is demonstrated by plotting the sample range (maximum–minimum) against the sample median lifetime (Fig. 1). We detected no difference in distribution patterns of individual stake failure times about the median values for each treatment group. An average lifetime, whether sample mean or median, by itself conceals the large underlying variability of the data.

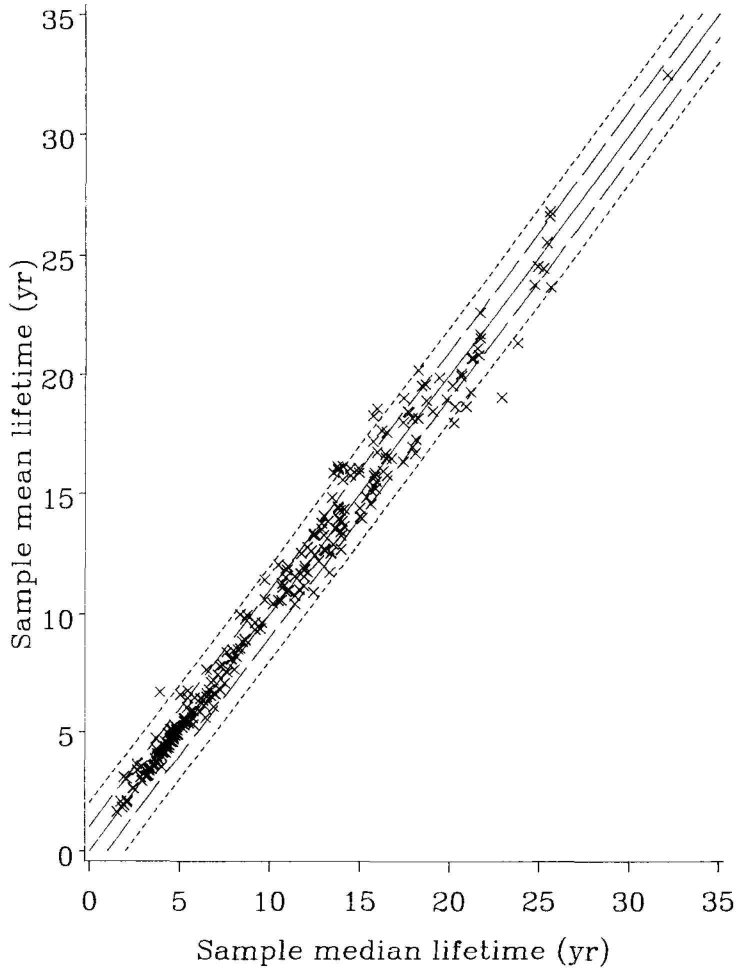


FIG. 2. Mean compared to median stake lifetime. Solid line indicates identical mean and median values; long-dashed lines include points where mean and median differ by at most 1 year; short-dashed lines include points where mean and median differ by at most 2 years.

PREDICTING PRESERVATIVE EFFECTIVENESS

This section deals with predicting the sample average lifetime. Currently, the effectiveness of a wood preservative at a given retention level as reported in the FPL research notes is measured by the sample mean lifetime of a group of stakes treated with a certain preservative to a specified retention level. A lower percentile estimate might be a more appropriate descriptor if one is concerned with the lower portion of the lifetime distribution. However, the sample sizes currently used in preservative field trials allow meaningful comparisons of only an average value. Moreover, a confidence interval for the population average lifetime would indicate the variability associated with the estimate. Given the sample sizes currently under test, recommended confidence intervals (or lower bounds of confidence intervals) for the population average lifetime are based on nonparametric

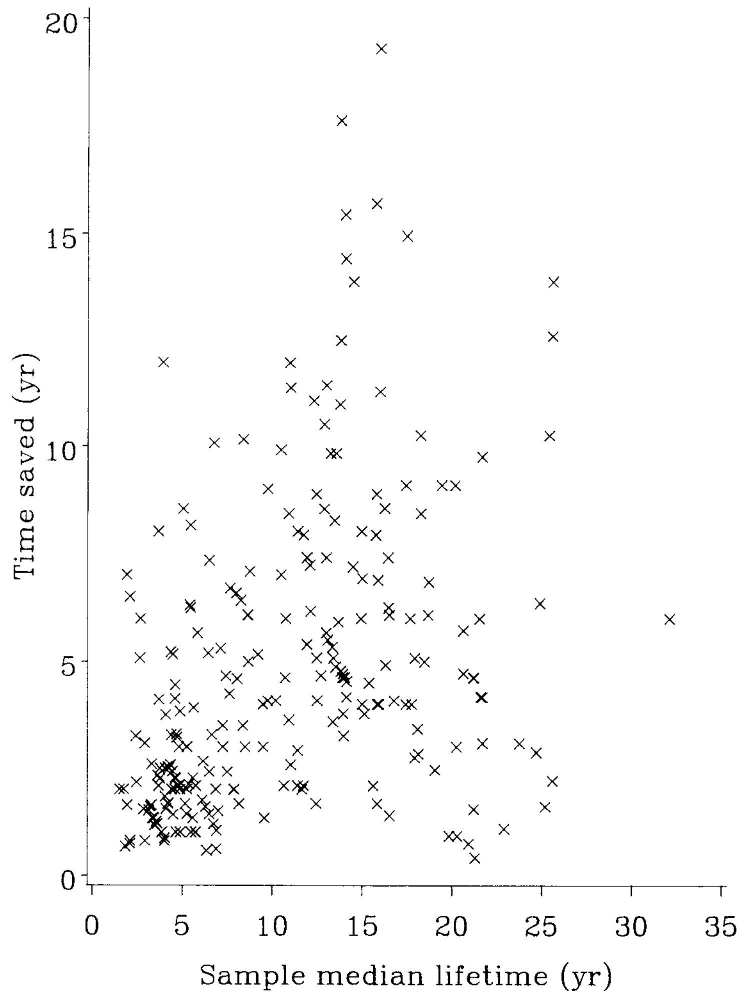


FIG. 3. Time saved by using median instead of mean lifetime to determine average lifetime of groups of ten replicate stakes. Time saved represents difference between inspection date when all stakes had failed and date when sixth stake had failed.

statistics. A good reference for nonparametric statistics is Conover (1980). The nonparametric statistics are presented in the following sections.

Median and mean values

Mean and median values are both estimates of the sample average lifetime. However, the mean can be computed only when all stakes have failed. The median can be calculated when the sixth stake fails (out of ten replicate stakes). Because the sample mean and median are both measures of average lifetimes, it is not surprising that the values are similar in any one group of stakes. In Fig. 2, the difference between the sample mean and median values is less than 1 year, 80% of the time. Similarly, the difference between the sample mean and median value is less than 1.5 (2) years, 90 (95)% of the time.