

POTENTIAL OF THE S_B AND S_{BB} DISTRIBUTIONS FOR DESCRIBING MECHANICAL PROPERTIES OF LUMBER¹

R. G. Pearson

Professor

Department of Wood and Paper Science, North Carolina State University
Raleigh, NC 27650

(Received 1 February 1980)

ABSTRACT

The S_B distribution can fit a much wider range of shapes of frequency distributions than the log-normal and Weibull, or even the beta, distributions. Maximum likelihood estimates of the four parameters can be obtained.

A bivariate distribution between correlated variables A and B, e.g. modulus of rupture and modulus of elasticity, is readily calculated if an S_B distribution is fitted to each variable, the only additional information needed being the value of the correlation coefficient between A and B. This distribution, known as the S_{BB} distribution, is based on a transformation of the original values into normal deviates and so its properties can be determined from normal distribution theory. The conditional distribution of variable A for a given value of B is itself an S_B distribution, the form of which may vary with the value of B to fit the test distribution of the A-values.

The S_{BB} distribution offers a possible means of obtaining a probabilistic relationship between the various strength properties of lumber such as modulus of rupture and maximum tensile strength. Means of estimating the correlation coefficient between different strength properties are suggested and the effect of error in the estimate of the correlation coefficient is discussed.

Keywords: Frequency distribution, lognormal distribution, Weibull distribution, lumber, mechanical properties, S_B distribution, S_{BB} distribution.

INTRODUCTION

It is often convenient to describe the results of mechanical tests on lumber in terms of some recognized distribution function. Unfortunately, the frequency distributions of the test results do not follow a consistent pattern. The normal distribution provides a reasonable fit in some cases, but the lognormal and the Weibull distributions have generally been found to give a better fit. The Weibull, which has been increasingly preferred in recent years, has the advantage of being able to represent both negatively and positively skewed data, whereas the lognormal is restricted to positively skewed data. One problem with these three distributions is that they are very limited in the values of skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) that they can accommodate. The normal distribution is theoretically appropriate only for zero skewness and for $\beta_2 = 3$, while for the other two distributions, β_1 is a near linear function of β_2 . This is illustrated in Fig. 1, which is a plot of β_2 against β_1 , the square of the skewness. The normal distribution is represented by a single point, the lognormal by a single line, and the Weibull by two lines, the shorter line being for negative skewness. There is a lower bound

This paper is Paper No. 6289 of the Journal Series of the N.C. Agricultural Research Service. The comments and helpful suggestions of Dr. W. L. Hafley, North Carolina State University, are gratefully acknowledged.

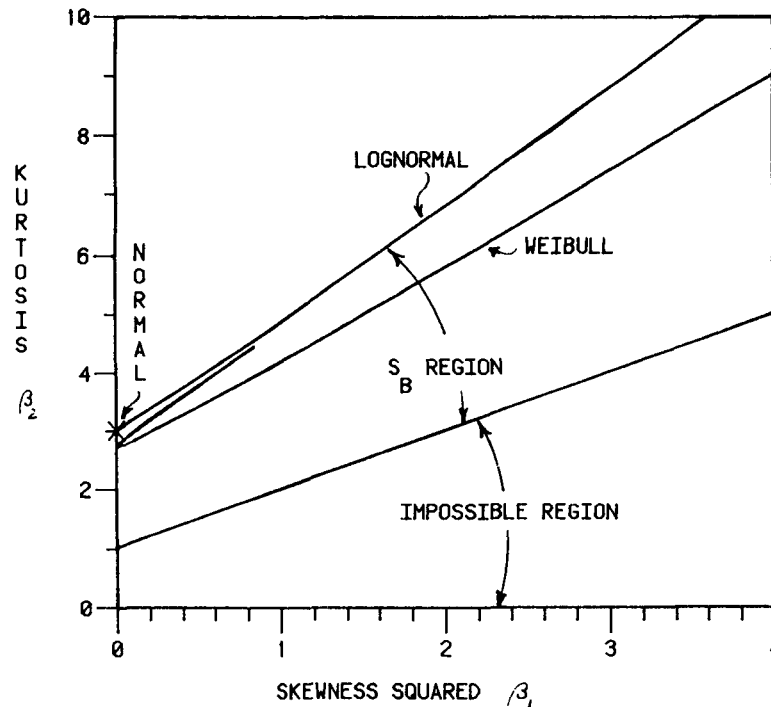


FIG. 1. β_1, β_2 space showing the values covered by the normal, lognormal, Weibull and S_B distributions.

to the value of β_2 for any value of β_1 defined by the line $\beta_2 - \beta_1 - 1 = 0$, shown in Fig. 1 as the boundary of the "impossible region." The lognormal and Weibull lines are not widely separated, especially for small values of skewness, and so it is to be expected that either distribution would provide a moderate fit to data with β_1 and β_2 values that plot near those lines.

As an illustration of the scatter of test values in the β_1, β_2 space, Fig. 2 reproduces part of Fig. 1 with the plotted values of b_1 and b_2 (the estimates of β_1 and β_2) for the properties of the three samples of lumber mentioned later in this paper. It should be remembered, however, that large sampling errors are usually associated with the b estimates. Consequently, this diagram does not necessarily indicate, or rule out, any particular distribution as the "best" distribution.

The purpose of this paper is to discuss some features of a four-parameter, univariate distribution called the S_B distribution (Johnson 1949a), and the corresponding bivariate distribution known as the S_{BB} distribution (Johnson 1949b). In Fig. 1, the S_B distribution covers the β_1, β_2 space between the lognormal curve and the impossible region. Consequently, it is possible to fit it to data with a wider range of skewness and kurtosis values than even the beta distribution, which covers the region between the Weibull line and the impossible region.

The S_B and S_{BB} distributions have been applied by Schreuder and Hafley (1977) and by Hafley and Schreuder (1977) to tree heights and diameters with considerable success. As far as the author is aware, they have not been applied to the

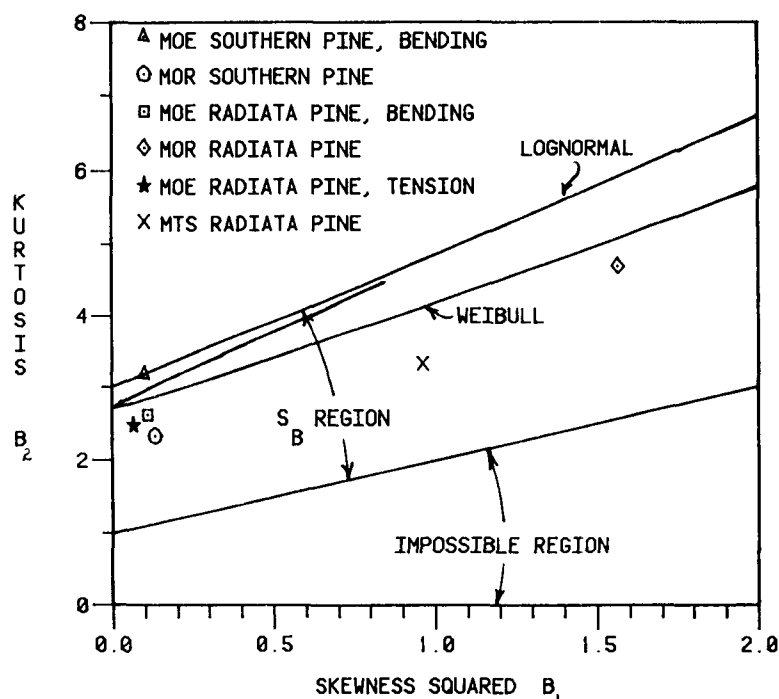


FIG. 2. Values of b_1 , b_2 for modulus of rupture (MOR), modulus of elasticity (MOE) and maximum tensile strength (MTS) of 2- × 8-inch southern pine and 2- × 4-inch radiata pine lumber.

mechanical properties of wood. Warren (1978) discussed Schreuder and Hafley's work, and commented briefly on the possible application of these distributions to the mechanical properties. In earlier work (Warren 1973), he found that the Weibull distribution fitted the lumber data he has been working with better than the four-parameter Pearson type 1 distribution, of which the beta distribution is a special case. The Weibull has the theoretical justification of being compatible with a weakest link theory of failure. Consequently, Warren appears to favor further exploration of the Weibull system in preference to the S_B distribution. However, the final sentence of his paper states "... very little work has been done on fitting the Weibull bivariate (or any other bivariate distribution) to modulus of rupture and modulus of elasticity data, so that a verdict on its applicability or otherwise must be deferred until the necessary evidence has been accumulated."

It is proposed in this paper to advance some arguments for studying the applicability of S_B and S_{BB} distributions to the results of mechanical tests on lumber.

THE UNIVARIATE S_B DISTRIBUTION

Like the beta distribution, the S_B distribution is defined only for the range (0,1), so the test values must be transformed to bring them within that range. To do this, the lower and upper limits of the population values must be known or estimated. If x is the test value, ξ is the lower bound and λ is the range of the population, then the variate u lies within the range (0,1), where

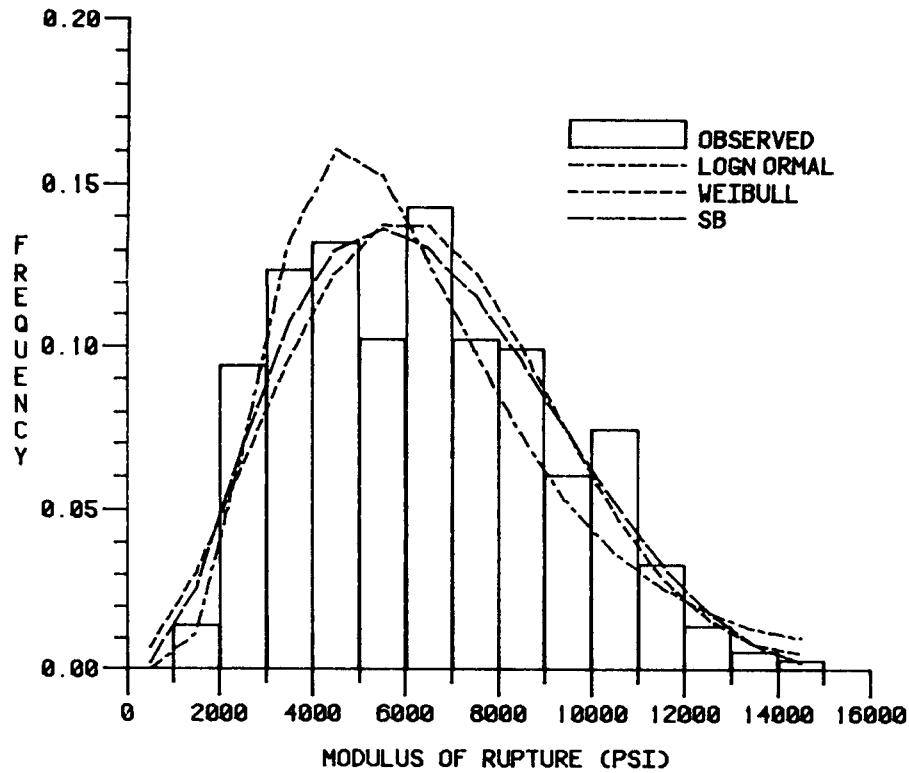


FIG. 3. Fit of the lognormal, Weibull and S_B distributions to modulus of rupture of 363 pieces of mixed grades of 2- x 8-inch southern pine lumber.

$$u = (x - \xi)/\lambda \quad (1)$$

The probability distribution function for u is given by

$$f(u) = (\delta/\lambda \sqrt{2\pi}) \{u(1-u)\}^{-1} \exp[-0.5\{\gamma + \delta \ln u/(1-u)\}^2] \quad (2)$$

where δ and γ are shape parameters.

If both the lower and upper bounds of the distribution are specified, then closed maximum likelihood solutions of δ and γ exist. Often, only the lower bound can be reliably specified, in which case the upper bound may be estimated by an iterative procedure. It is also possible to obtain maximum likelihood estimates of both bounds by an iterative procedure if neither can be reliably estimated beforehand. A suitable computer program enables ready calculation of the parameters for any of these cases. Although values of the population bounds for the mechanical properties are unknown, informed estimates can be made. If the bounds obtained by the mathematical search routines appear unrealistic, then the researcher can make some adjustments as indicated later.

Figures 3 and 4 show the lognormal, Weibull and S_B distributions fitted to the values of modulus of rupture (MOR) and flatwise modulus of elasticity (MOE), respectively, given by Doyle and Markwardt (1966) in their Fig. 9 for four grades of 363 pieces of 2- x 8-inch southern pine lumber.

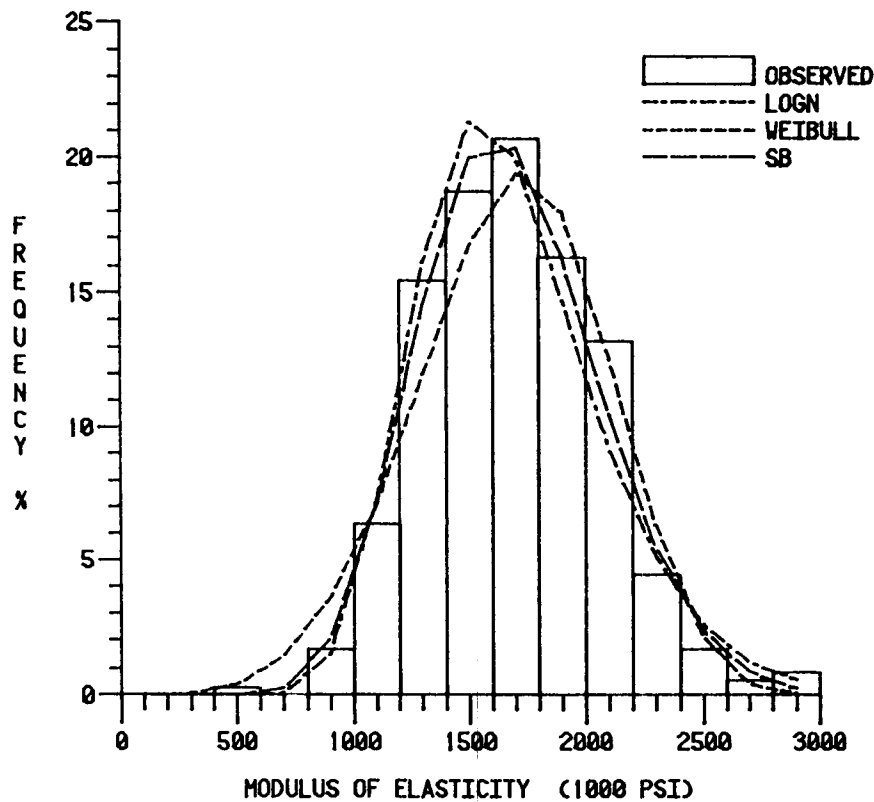


FIG. 4. Fit of the lognormal, Weibull and S_B distributions to modulus of elasticity of 363 pieces of mixed grades of 2- x 8-inch southern pine lumber.

For both properties, the S_B distribution gave the best fit according to log likelihood and Kolmogorov-Smirnov statistics. The log likelihood estimates of the goodness of fit for the three distributions are given in Table 1, the smaller the value the better the fit. The rank of the distributions is given in parentheses. That the S_B distribution should give the best fit for modulus of rupture is to be expected because the plotted point in Fig. 2 for skewness and kurtosis lies well below the Weibull line. However, the corresponding point for modulus of elasticity lies on the lognormal line, but Fig. 4 shows that the shape of the lognormal distribution does not fit the pattern of the experimental values quite as well as does the S_B distribution.

THE BIVARIATE S_{BB} DISTRIBUTION

The bivariate distribution can be readily found for correlated variables x_1 , x_2 , for example, modulus of rupture and modulus of elasticity, if an S_B distribution is fitted to the test results for each variable. Apart from the parameters of these marginal distributions, the only additional information required is the correlation coefficient for the two variables.

The test values, after being transformed according to equation (1), are further transformed to yield normal variates as under

TABLE 1. Comparison of distributions for 363 pieces of 2- × 8-inch pine lumber.

Property	Values of log likelihood and rank of distribution		
	Lognormal	Weibull	S _B
Modulus of rupture	-881.5 (3)	-872.1 (2)	-864.9 (1)
Modulus of elasticity	-159.2 (2)	-166.4 (3)	-154.8 (1)

$$\begin{aligned} z_1 &= \gamma_1 + \delta_1 \ln\{u_1/(1 - u_1)\} \\ z_2 &= \gamma_2 + \delta_2 \ln\{u_2/(1 - u_2)\} \end{aligned} \quad (3)$$

The correlation coefficient ρ between z_1 and z_2 may then be calculated as follows:

$$\rho = \sum_{i=1}^n z_{1i}z_{2i}/n \quad (4)$$

where n = total number of paired values.

The probability distribution function for the bivariate distribution of z_1 and z_2 is then given by

$$f(z_1, z_2; \rho) = \{2\pi\sqrt{1 - \rho^2}\}^{-1} \exp\{-0.5(1 - \rho^2)^{-1}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\} \quad (5)$$

The conditional probability of z_2 for any given z_1 is also an S_B distribution with parameters given by

$$\begin{aligned} \gamma' &= (\gamma_2 - \rho z_1)/\sqrt{1 - \rho^2} \\ \delta' &= \delta_2/(1 - \rho^2) \end{aligned} \quad (6)$$

Because the z 's are normal variates, the percentile limits for z_2 for given z_1 can be readily computed. If k is the normal deviate corresponding to a particular percentile α , e.g. $k = -1.645$ for the 5th percentile, then the α -percentile for u_2 for given u_1 is

$$u_2 = \Psi/(1 + \Psi) \quad (7)$$

where

$$\Psi = \theta\{u_1/(1 - u_1)\}^\phi \exp\{k\sqrt{1 - \rho^2}/\delta_2\} \quad (8)$$

$$\theta = \exp\{(\rho\gamma_1 - \gamma_2)/\delta_2\} \quad (9)$$

$$\phi = \rho\delta_1/\delta_2 \quad (10)$$

The mean regression is too complicated to be readily manipulated. However, the median regression is more amenable. It is obtained from Eq. (7) by putting $k = 0$ in Eq. (8) to give

$$u_{2m} = \theta u_{1m}^\phi / [(1 - u_{1m})^\phi + \theta u_{1m}^\phi] \quad (11)$$

the subscript "m" indicating the median value.

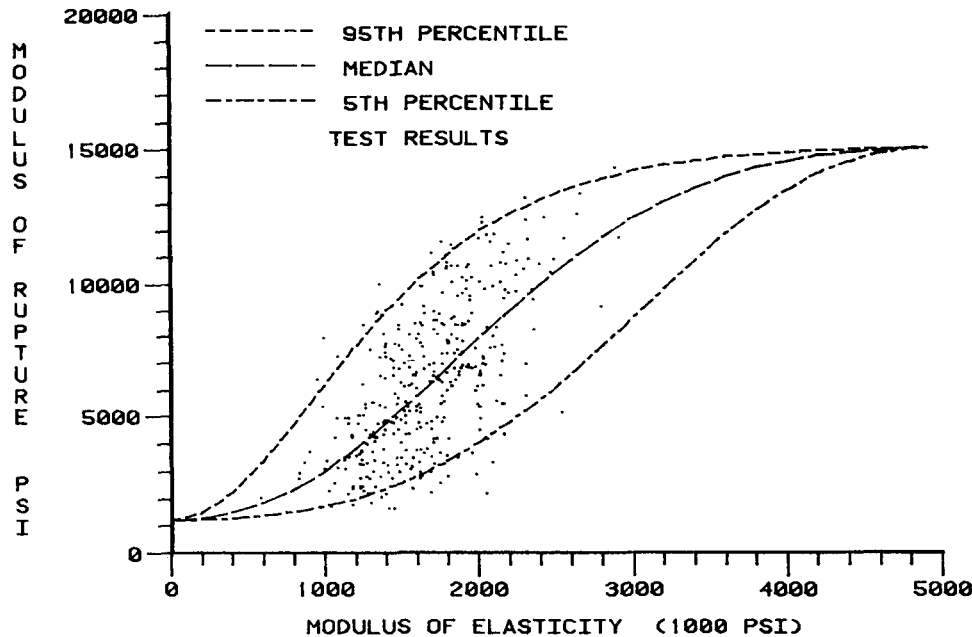


FIG. 5. Regression lines for the median and 5th and 95th percentiles for the bivariate distribution of modulus of rupture on modulus of elasticity of 2- x 8-inch southern pine lumber.

Figure 5 shows the regression lines corresponding to the 5th, 50th (median), and 95th percentiles for an S_{BB} distribution fitted to the same results for modulus of rupture and flatwise modulus of elasticity as in Figs. 3 and 4. The figure illustrates how the distribution accommodates the change in skewness of the modulus of rupture results as the modulus of elasticity changes. For low values of modulus of elasticity, the median is closer to the 5th percentile line than to the 95th line, but for high values it is closer to the 95th percentile line.

The lower and upper limits were determined for the distribution by iteration but are not realistic in that horizontal asymptotes are implied for the modulus of rupture. The inference may be drawn that the data do not contain sufficient information to provide good estimates of the limits. One can specify certain relationships or values for the limits that will lead to more realistic shapes for the regression lines in the extrapolated regions. An illustration of such a modification is given in Fig. 6 for the S_{BB} distribution for maximum tensile strength (MTS) and modulus of elasticity of 74 pieces of 2- x 4-inch radiata pine lumber from data supplied by Leicester (1979). A substantial change from the original lower and upper bounds made little difference to the regression lines in the region occupied by the test results. There are, of course, very significant changes in the regression lines outside that region. As usual, extrapolation beyond the scope of the data is fraught with uncertainty. Exploration of procedures for obtaining better estimates of the bounds may be a worthwhile future study.

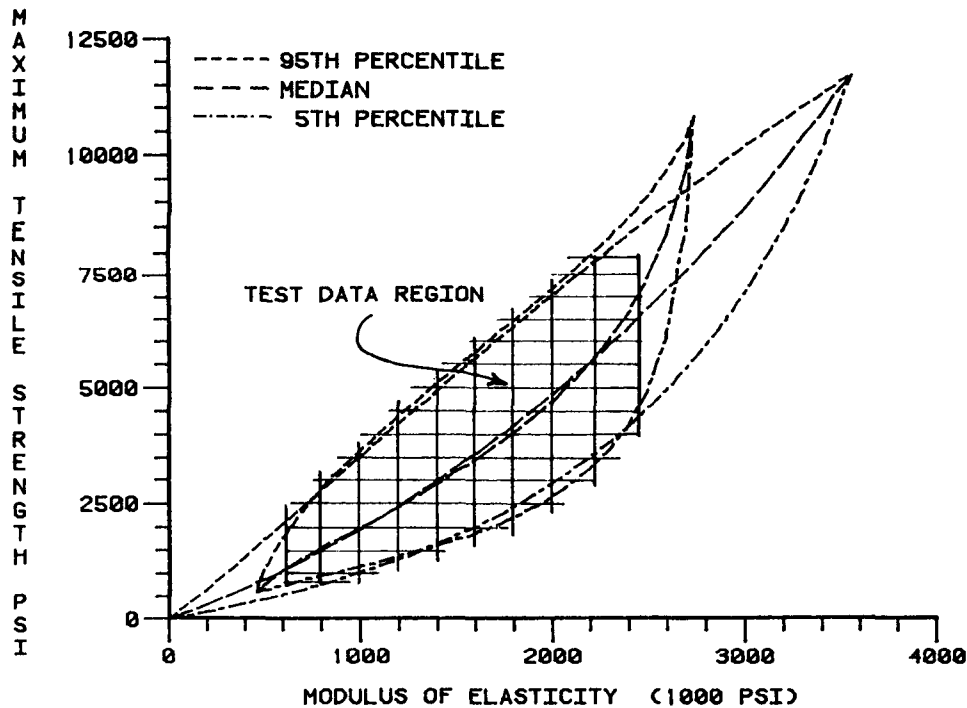


FIG. 6. Effect of change in the estimated values of the upper and lower bounds of the population values of maximum tensile strength and modulus of elasticity in tension of 2- x 4-inch radiata pine lumber.

ESTIMATION OF THE CORRELATION COEFFICIENT

Often it is not possible to obtain paired x_1, x_2 values to enable the correlation coefficient to be calculated. For example, modulus of rupture and maximum tensile strength cannot be determined for the same piece of lumber because each involves a destructive test. S_B distributions can be fitted to available data on the separate properties, but some estimate of the correlation coefficient for the two properties is needed to obtain the bivariate distribution. One is tempted to base the estimate on the correlation coefficient obtained for each of the x_1, x_2 variables with some other common property. For example, the correlation coefficient between tensile strength and modulus of elasticity can be obtained, as can that between bending strength and modulus of elasticity. Unfortunately, no theory exists for estimating the x_1, x_2 correlation coefficient from the correlation between x_1 and x_2 separately with some other property. However, one intuitively expects that knowledge of the latter correlations should provide an indication of the x_1, x_2 correlation coefficient.

Figure 7 illustrates the effect on the percentile regressions of using two estimates of the correlation coefficient relating maximum tensile strength and modulus of rupture. The data were from tests on 74 pieces of radiata pine (Leicester

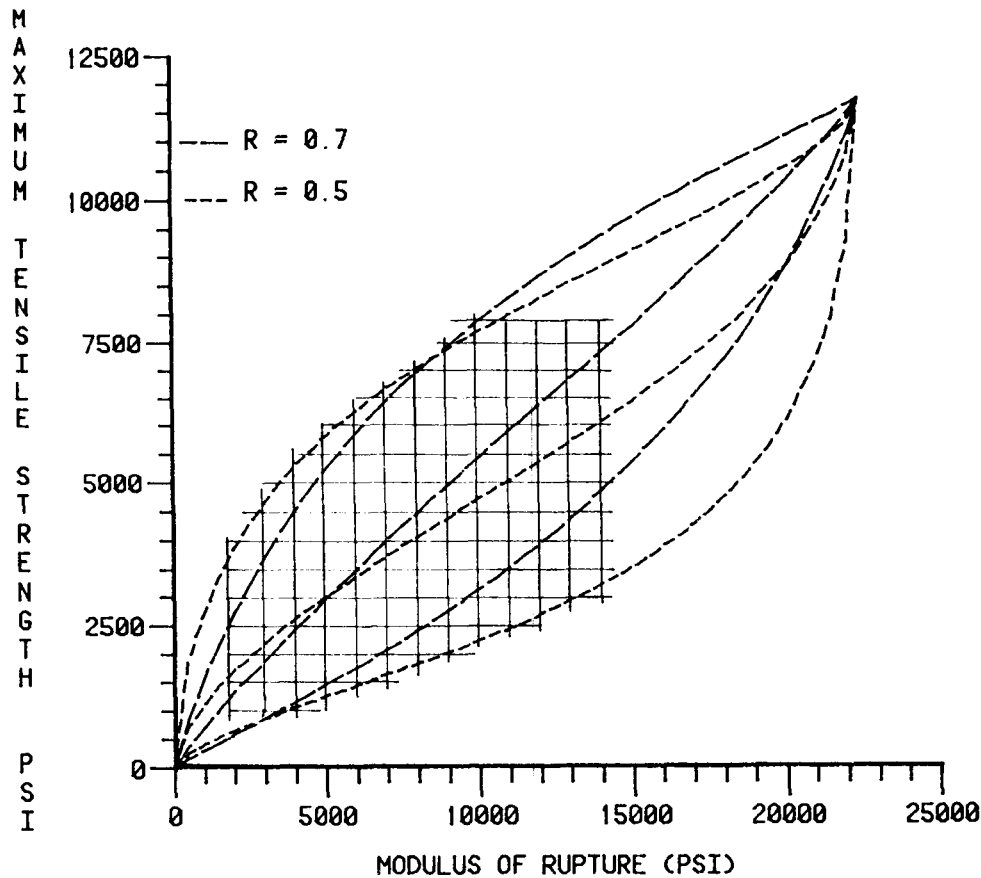


FIG. 7. Effect of change in the assumed correlation coefficient between maximum tensile strength and modulus of rupture of 2- x 4-inch radiata pine lumber.

1979), the correlation coefficient between tensile strength and modulus of elasticity being 0.78 and that between modulus of rupture and modulus of elasticity being 0.75. In Fig. 7, regression lines are drawn for correlation coefficients of 0.7 and 0.5. As expected, the smaller coefficient gave a wider spread for the theoretical distribution. However, it should be noted that the percentile lines based on the two correlation coefficients cross over each other. There is, therefore, only a small difference in the values near those intersections.

An alternative means of estimating the correlation coefficient may be available. As shown in the Appendix, Eq. (8) may be rewritten in the form

$$\rho = z_{2m}/z_{1m} \quad (12)$$

where z_{1m} , z_{2m} are paired median values of the transformed normal deviates.

Consequently, if estimates of paired median values can be obtained, the correlation coefficient can be estimated from Eq. (12). One possible way of doing this will now be described.

Suppose, for example, that the median regressions for modulus of rupture on modulus of elasticity in bending, and for maximum tensile strength on modulus of elasticity in tension have been determined from the results of tests on lumber from the same population. The distribution of the modulus of elasticity in bending usually differs from that in tension, so it will also be necessary to know the median regression for these two properties. From this regression, we can obtain for any set of median values of modulus of elasticity in bending a matched set of median values for modulus of elasticity in tension. Hence for this median set of modulus of elasticity values in bending, we can obtain the corresponding set of modulus of rupture and maximum tensile strength values from the appropriate median regressions. In this way, we can obtain nominally matched pairs of values of modulus of rupture and maximum tensile strength from which the correlation coefficient can be computed. Unfortunately, the author does not have access at this time to suitable test data to enable this procedure to be tried out. To obtain such data would require the tension specimens to be tested first in bending to obtain the modulus of elasticity value to be paired with the modulus of elasticity value from the subsequent tension test.

CONCLUSION

The S_B and S_{BB} distributions offer a degree of flexibility which should enable them to be of value in describing lumber data. They appear to have considerable potential where knowledge of the bivariate relationships between correlated properties is required. They may also provide a means of estimating the correlation coefficient between strength properties when a direct determination of that correlation is impossible.

REFERENCES

- DOYLE, D. V., AND L. J. MARKWARDT. 1966. Properties of southern pine in relation to strength grading of dimension lumber. U.S. Forest Service Research Paper FPL-64.
- HAFLEY, W. L., AND H. T. SCHREUDER. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. *Can. J. For. Res.* 7(3):481-487.
- JOHNSON, N. L. 1949a. Systems of frequency curves generated by means of translation. *Biometrika* 36:149-176.
- . 1949b. Bivariate distributions based on simple translation systems. *Biometrika* 36:297-304.
- LEICESTER, R. L. 1979. Division of Building Research, Commonwealth Scientific and Industrial Research Organization, Australia. Private communication.
- SCHREUDER, H. T., AND W. L. HAFLEY. 1977. A useful bivariate distribution for describing stand structure of tree heights and diameters. *Biometrics* 33(3):471-478.
- WARREN, W. G. 1973. Estimation of the exclusion limit for dimension lumber. *Proc. Meeting Division V, IUFRO, South Africa, September and October 1973*. Vol. 2:1148-1154.
- . 1979. Some recent developments relating to statistical distributions in forestry and forest products research. In J. K. Ord, G. P. Patl, and C. Taillie, eds. *Statistical distributions in ecological work*. International Co-operative Publishing House, Fairland, MD. 1979.

APPENDIX

Equation (11) may be written in the form

$$\theta \{u_{1m}/(1 - u_{1m})\}^\phi = u_{2m}/(1 - u_{2m})$$

$$\therefore \ln \theta + \phi \ln \{u_{1m}/(1 - u_{1m})\} = \ln \{u_{2m}/(1 - u_{2m})\}$$

Substitution of the values for θ and ϕ from equations (9) and (10) leads to

$$(\rho\gamma_1 - \gamma_2)/\delta_2 + \rho(\delta_1/\delta_2)\ln \{u_{1m}/(1 - u_{1m})\} = \ln \{u_{2m}/(1 - u_{2m})\}$$

$$\text{i.e. } \rho\gamma_1 + \rho\delta_1\ln \{u_{1m}/(1 - u_{1m})\} = \gamma_2 + \delta_2\ln \{u_{2m}/(1 - u_{2m})\}$$

$$\text{i.e. } \rho \cdot z_{1m} = z_{2m}$$

$$\therefore \rho = z_{2m}/z_{1m}$$