Predictor Sort Sampling, Tight T's, and the Analysis of Covariance

STEVE VERRILL*

In this article we revisit a method of sample allocation that has long been known to statisticians and has recently been "discovered" by wood strength researchers. The method allocates experimental units to blocks on the basis of the values of a variable, x, that is known to be correlated with the response, y. We call this allocation method "predictor sort sampling." We demonstrate that the associated paired T analysis recommended in statistical texts is deficient if the sample size is small and the correlation between x and y is high. We temper this criticism of standard statistical intuition with a proof that the approach is asymptotically correct. In a related development we show that a modified pooled T approach can be taken to this data with a resultant increase in power. We compare these approaches to an analysis of covariance approach and discuss the advantages of each. Finally, we warn against the intuitively attractive but incorrect power calculations that are likely to be performed in association with a predictor sort experiment.

KEY WORDS: Analysis of covariance; Asymptotics; Experimental design; Paired T; Pooled T; Power; Sample size; T distribution.

1. INTRODUCTION

In recent years wood strength researchers have begun to replace experimental unit allocation via random sampling with allocation via sorts based on nondestructive measurements of strength predictors, such as modulus of elasticity and specific gravity. Warren and Madsen (1977) described the procedure as follows:

> Specifically, then, all the boards in the experiment are ordered from weakest to strongest as nearly as can be judged from their moduli of elasticity, knot size, and slope of grain. To divide the material into J equivalent groups, the first J boards, after ordering, are taken and randomly allocated one to each group. This is repeated with the second, third, fourth, etc., sets of J boards. The strength distributions of the resulting groups should then be essentially the same.

This allocation procedure has come to be known as predictor sort sampling. Depending on the level of their statistical sophistication, wood strength researchers have analyzed the resultant data via blocked or unblocked analyses of variance.

As one would expect, predictor sort allocation has long been known to statisticians. Cox (1957) compared seven procedures that one might use given the availability of a correlated predictor. One of the procedures is a predictor sort coupled with a randomized block analysis of variance. Another approach is an analysis of covariance. Cox concluded that for $\rho < .6$, blocked ANOVA's are essentially as efficient as an analysis of covariance. He also noted that a blocked ANOVA can be superior to an analysis of covariance if the relationship between the covariate and the response is not adequately modeled.

Cox's calculations showed that the effective variance in both the blocking and analysis of covariance situations is $(1 - \rho^2)\sigma_y^2$, a fact also noted by Cochran (1957). (Here σ_y^2 is the variance of y and ρ is the correlation between the predictor x and the response y.)

When the number of treatments is two, a blocked analysis of variance amounts to a paired T test, and an unblocked ANOVA corresponds to a pooled T test. In Figure 1 we plot a histogram of the paired T statistic for the case in which $\rho = 1.0$ and $n = 2k = 16$. We overlay the histogram with the pdf of a T distribution with $k - 1 = 7$ degrees of freedom. In Figure 2 we plot a histogram of the pooled T statistic in the case in which $\rho = .7$ and $n = 2k = 40$. We overlay this second histogram with the pdf of a T distribution with $2k - 2 = 38$ degrees of freedom. These figures illustrate why we refer to paired and pooled T's associated with predictor sort sampling as "tight T's." (Each of the histograms is based on 10,000 simulation trials.)

On reflection, most professional statisticians would realize that a T distribution is only an approximation to the distribution of the tight paired T. It is likely, however, that statisticians, and certainly scientists, would neglect this discrepancy in the course of an analysis (see, for example, Cox 1958, ex. 3.3; Finney 1972, sec. 13.17; Ostle and Mensing 1975, ex. 11.3; Snedecor and Cochran 1967, ex. 4.11.1).

A priori, there is no reason to believe that this neglect is acceptable. However, we show that for $\rho < 1$ the approach is asymptotically justified, and that for small samples the approximation is adequate unless ρ is high and k is small.

Further, we show that it is possible to analyze data of this sort via a modified pooled T. As we will see, such an approach yields power that is superior to that achievable via a paired T. In fact, in some cases this power exceeds that available from an analysis of covariance.

In Section 2 of this article, we establish that, given a predictor sort allocation, for $0 \le \rho < 1$ the tight paired T and the tight pooled T (suitably scaled) are asymptotically $N(0, 1)$. This result is a special case of a more general theorem that covers n-way balanced ANOVA's. In Section 3 we present estimates of the small sample critical values of the tight T's, and identify the sample sizes needed to permit use of the asymptotic critical values. In Section 4 we discuss a power study that should be useful in selecting sample sizes, and we show that we need not greatly concern ourselves about entering the critical value tables via estimates of ρ. We include in this power study an analysis of covariance approach, and identify those cases in which it is inferior to one of the tight T approaches. In Section 5 we warn against intuitively attractive but incorrect power calculations.

* Steve Verrill is Mathematical Statistician, U.S. Department of Agriculture Forest Products Laboratory, Madison, WI 53705.

Journal of the American Statistical Association
March 1993, Vol. 88, No. 421, Theory and Methods

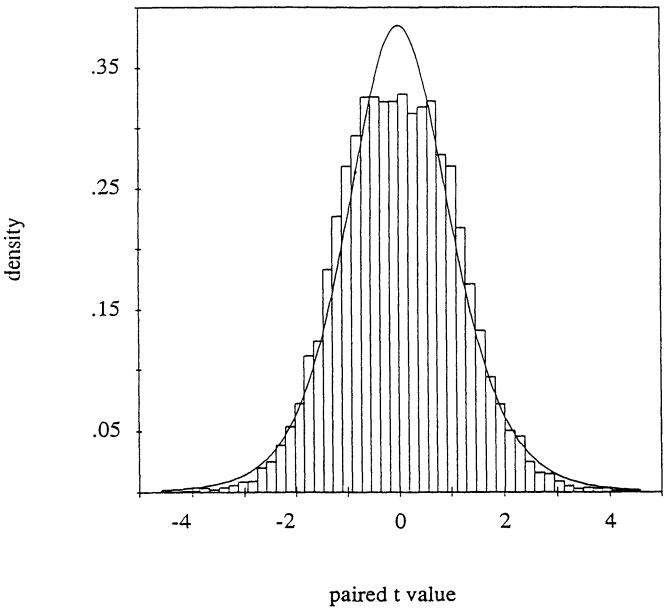

Figure 1. Histogram of the Paired T Statistic for $\rho = 1.0$ and $k = 8$ Overlayed With a T_7 Density Function.

2. ASYMPTOTIC DISTRIBUTIONS

Theorem. Assume that the predictor variable and the variable of interest have a joint bivariate normal distribution with correlation ρ. Let the allocation of samples be as described in Section 1. (For the n-way case, enough adjacent experimental units are chosen at a time to provide one additional observation for each cell.) Then for $0 \leq \rho < 1$, the asymptotic distribution of the statistic that treats the groups of "equivalent" experimental units as a block (the "paired approach") is χ^2_{J-1}. The asymptotic distribution of the statistic that ignores the block structure generated by these groups (the "pooled approach") is $(1 - \rho^2)\chi^2_{J-1}$.

Proof. The proof appears in the Appendix.

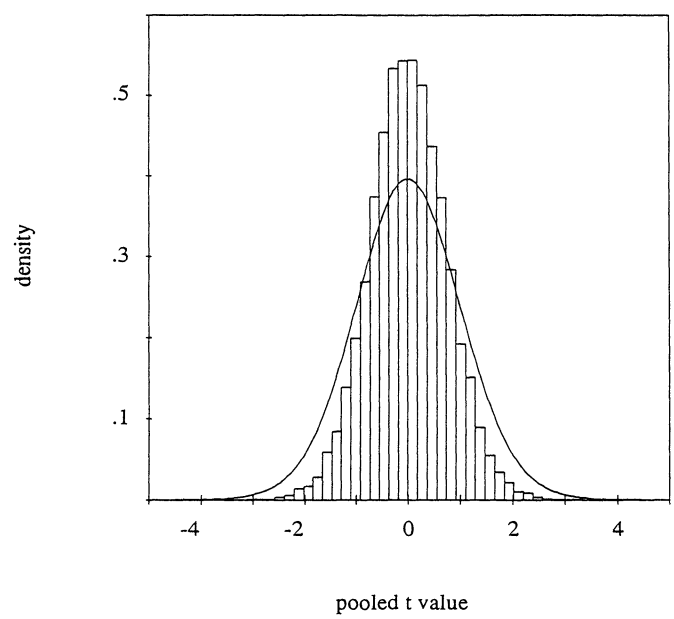

Figure 2. Histogram of the Pooled T Statistic for $\rho = .70$ and $k = 20$ Overlayed With a T_{38} Density Function.

3. MONTE CARLO ESTIMATES OF CRITICAL VALUES OF THE T'S

We ran 10,000 trials for each combination of $n = 2k = 4, 6, 8(4)40(8)120, 160, 200, 300$ and $\rho = .40(.05).95, .99, 1.0$. (Here k is the number of observations per treatment and ρ is the correlation between the predictor x and the response y.) To do so we used the uniform (UNI) and normal (RNOR) random number generators developed by Kahaner and Marsaglia. The absolute values of the scaled pooled statistic (the usual pooled statistic divided by $\sqrt{1 - \rho^2}$) and the usual paired statistic were ordered, and order statistics 8001, 9001, 9501, 9801, and 9901 were used as estimates of the two-sided .20, .10, .05, .02, and .01 critical values. Using the techniques described by Verrill and Johnson (1988), one can see that this approach yields the following .999 probability intervals on the true sizes: [.187, .213], [.090, .110], [.043, .057], [.015, .025], and [.0066, .0134] (e.g., $\Pr(\xi_{.090} \leq .10\ cv \text{ estimate} \leq \xi_{.110}) = .999$).

For each ρ size combination, the critical values for $n = 8(4)40(8)120, 160, 200, 300$ were smoothed via the equation $cv = a_0 + a_1/n^{1/2} + a_2/n + a_3/n^{3/2}$. For $\rho < 1.0$, a_0 was fixed at the appropriate asymptotic critical value.

For $\rho = 1$ (not covered by the Theorem) there are heuristic reasons for believing that $\sqrt{kln(k)}$ times the usual pooled statistic converges in distribution to something approximating a linear combination of independent double exponentials. Thus in this case we calculated the smoother of the Monte Carlo critical values of the usual paired statistic, but we calculated the smoother of the Monte Carlo critical values of $\sqrt{kln(k)}$ (rather than $1/\sqrt{1 - \rho^2}$) times the usual pooled statistic. To perform the pooled $\rho = 1.0$ calculations, we fixed a_0 at the average of the estimated critical values for $n = 104$, 112, 120, 160, 200, and 300. (Plots indicated that the critical values seemed to have leveled off at an "asymptotic value" by $n = 104$.) In the paired $\rho = 1.0$ case, a_0 was a free parameter estimated in the smoothing process.

The coefficients for the pooled statistics and size .05 are presented in Table 1. Those for the paired statistics and size .05 are presented in Table 2. Because we found it necessary to exclude the $n = 4, 6$ results from the smoothing process, the critical values for these two n's and size .05 are presented in Table 3. (Again the critical values reported in Table 3 for $\rho = 1$ are the critical values of the usual paired statistic,

Table 1. Smoothing Curve Coefficients, Pooled T, Two-Sided Size = .05

ρ	a_0	a_1	a_2	a_3
.40	1.960	−.2577	3.683	2.803
.45	1.960	.04724	.9800	8.326
.50	1.960	.1034	.5330	9.370
.55	1.960	.4342	−2.825	16.62
.60	1.960	.03690	.9038	8.550
.65	1.960	−.3309	5.768	−3.737
.70	1.960	−.02812	2.638	4.050
.75	1.960	−.01677	2.068	6.746
.80	1.960	.5183	−2.803	18.36
.85	1.960	.07835	.9213	14.07
.90	1.960	−.08358	5.158	6.189
.95	1.960	−.01327	8.132	9.684
.99	1.960	−1.726	56.24	−52.81
1.0	2.315	.3015	−2.284	−6.309

Table 2. Smoothing Curve Coefficients, Paired T, Two-Sided Size = .05

ρ	a_0	a_1	a_2	a_3
.40	1.960	.2104	−.7322	27.61
.45	1.960	.6684	−4.402	34.34
.50	1.960	.7335	−5.161	36.02
.55	1.960	.6474	−4.129	32.81
.60	1.960	.2880	−.7502	25.99
.65	1.960	.6407	−4.543	33.54
.70	1.960	.5610	−3.486	31.37
.75	1.960	.6104	−4.190	32.83
.80	1.960	.6726	−3.872	31.15
.85	1.960	.3756	−2.169	28.55
.90	1.960	.3746	−.8042	22.85
.95	1.960	.8558	−5.915	34.22
.99	1.960	.8197	−7.140	34.78
1.0	1.681	2.847	−13.91	39.97

Table 4. n Required for "Good" Asymptotics, Pooled Statistic

	Two-sided size				
ρ	.20	.10	.05	.02	.01
.40	20	35	50	39	50
.45	20	38	52	46	71
.50	20	35	59	39	52
.55	20	31	63	42	66
.60	22	32	50	39	52
.65	21	32	62	43	59
.70	25	36	53	46	60
.75	24	36	56	46	59
.80	29	43	90	50	60
.85	36	60	80	55	76
.90	48	74	116	77	98
.95	88	126	227	140	231
.99	309	381	430	264	281
1.0	—	—	—	—	—

but the critical values of $\sqrt{kln(k)}$ times the usual pooled statistic.)

The quality of the smoothed critical values was tested by performing an additional 10,000 trials for each n, ρ combination ($n \geq 8$). These tests demonstrated that if one uses the smoothed critical values, then the two-sided size will not be off by more than .01 for sizes .20 and .10, and by more than .005 for size .05. In fact these figures are fairly conservative.

In the course of these tests we also investigated the acceptability of replacing the smoothed critical values with tabled T critical values. In the pooled case the tabled T values yield good approximations to the critical values of the scaled statistic for $\rho \leq .80$. In the paired case the T approximation is satisfactory for ρ as large as .90. For low n, however, as ρ increases beyond .90, the actual size falls below the nominal size and a small amount of power ($\leq$.15) is lost.

To get an idea of the sample sizes needed for the asymptotic critical values to be satisfactory, for each ρ size combination we fit a curve of the form $a_0 + a_1/n^{1/2} + a_2/n + a_3/n^{3/2}$ to the counts of the times that the asymptotic critical values were exceeded in the 10,000 trials. (a_0 was fixed at 2,000, 1,000, 500, 200, or 100 depending on the size in question.) We then found the n's at which these curves descended below 2,200, 1,100, 550, 250, or 125. These values are reported in Table 4 for the pooled T and in Table 5 for the paired T. Note that as the size decreases or ρ increases (for the pooled case), the n needed to achieve good performance of the asymptotic values also increases. On the other hand, for all but the highest ρ good performance is achieved for fairly small n (<100 in the pooled case).

4. POWER STUDY

We performed a power study that covered the cases in which ρ = .4(.1).8, .85(.05).95, .99, $n = 2k$ = 12(12)48, 72, 96; Δ/σ_y = .0(.25)1.5, and ρ = .4(.1).8, .85(.05).95, .99, $n = 2k$ = 4(2)14, Δ/σ_y = .0(.5)3.0. Here Δ is the mean difference between treatment 1 and treatment 2, and σ_y^2 is the variability of the y's. The study can be summarized as follows:

- A predictor sort followed by a standard analysis yields poor power properties. (Because the unscaled statistic is "tight," for smaller Δ the power associated with it is actually *lower* than what one could obtain from standard random sampling.) Unfortunately, this is the approach currently taken by many wood strength researchers.
- The gain in efficiency of the tight T (pooled or paired)

Table 3. Monte Carlo Critical Values, Two-Sided Size = .05

	Pooled		Paired	
ρ	n = 4	n = 6	n = 4	n = 6
T value	4.30	2.78	12.7	4.30
.40	4.19	2.87	12.4	4.42
.45	4.45	2.81	13.5	4.20
.50	4.31	2.74	11.5	4.36
.55	4.59	2.86	12.7	4.37
.60	4.40	2.84	12.3	4.22
.65	4.26	2.83	12.2	4.27
.70	4.51	2.90	12.8	4.39
.75	4.55	2.84	12.4	4.11
.80	4.56	2.97	12.7	4.24
.85	4.66	3.09	12.1	4.06
.90	4.84	3.29	12.6	4.15
.95	5.41	3.91	12.8	3.97
.99	9.08	7.09	11.8	3.69
1.0	1.40	1.69	11.1	3.58

Table 5. n Required for "Good" Asymptotics, Paired Statistic

	Two-sided size				
ρ	.20	.10	.05	.02	.01
.40	34	56	94	71	85
.45	35	67	140	92	143
.50	32	73	134	87	113
.55	28	63	144	90	122
.60	31	59	122	74	109
.65	28	59	113	81	105
.70	31	57	120	78	118
.75	32	67	125	73	112
.80	31	69	159	98	112
.85	31	63	94	83	159
.90	34	69	111	91	123
.95	36	66	150	97	192
.99	36	64	88	70	106
1.0	—	—	—	—	—

predictor sort approach over a standard random sampling approach ranges from 33% to 800% as ρ increases from .5 to .99.

- Because of the difference in degrees of freedom, the pooled approach yields greater power than the paired approach. For small n, the gain can be substantial. For example, for $\alpha = .01$, $n = 8$, $\rho = .7$, and $\Delta/\sigma_y = 2.5$, the paired tight T yields power .42 and the pooled tight T yields power .81.
- For $\rho \leq .85$, the pooled tight T performs as well as an analysis of covariance. For small n (say, $n \leq 14$) the pooled tight T actually performs better than an analysis of covariance. For example, for $\alpha = .01$, $n = 6$, $\rho = .4$, and $\Delta/\sigma_y = 3.0$, a predictor sort followed by an analysis of covariance yields a power of .29, and the pooled tight T yields power .45—small n's and large Δ/σ_y's are common in certain areas of wood strength research.
- In the pooled case, for $n \geq 12$, $\rho \leq .95$, entering the critical value tables via estimated ρ's causes no problems. In our simulations, for each trial $\hat{\rho}$ was calculated as the average of the correlation between the predictor and the response for the treatment 1 sample and the correlation between the predictor and the response for the treatment 2 sample. For $\hat{\rho}$ less than .40, standard T tables were used to obtain critical values. For $\hat{\rho}$ between .40 and .90, the interpolation was linear between the two nearest bracketing ρ's. For $\hat{\rho}$ greater than .90, the critical value was a quadratic interpolation/extrapolation of the critical values for $\rho = .90$, .95, and .99.

  For $n < 12$, $\rho \leq .95$, entering the tables via $\hat{\rho}$ can yield inflated sizes, but if ρ is known from experience to within .05–.10, nominal and actual sizes match well. In our simulations, for each trial "ρ from experience" was drawn from a $N(\rho, .05^2)$ distribution for $\rho \leq .90$, from a $N(\rho, .025^2)$ distribution for $.90 < \rho \leq .95$, and from a $N(\rho, .005^2)$ distribution for $\rho = .99$.

  For $\rho = .99$, entering the tables via $\hat{\rho}$ yields inflated sizes, but the inflation decreases to acceptable levels as n increases. For $\rho = .99$, entering the tables via a ρ "known from experience" is unacceptable.
- Entering the critical value tables via estimated ρ's causes no problems in the paired case.
- For the pooled tight T, power results can be approximated well by taking a noncentral T approach with noncentrality parameter equal to

$$(\Delta/\sigma_y)/\sqrt{2(1-\rho^2)/k}$$

  and $2k - 2$ degrees of freedom. Alternatively, power can be estimated by assuming that $(\bar{Y}_2 - \bar{Y}_1 - \Delta)/(s_{\text{pooled}}\sqrt{1-\rho^2}\sqrt{2/k}) \approx N(0, 1)$. The noncentral T approach yields perfectly adequate approximations to the true power for $\rho \leq .80$. For higher ρ it tends to overestimate power, but the overestimation decreases as n increases. The normal approach yields a more significant overestimate of power, and its use should probably be restricted to lower ρ and higher n.
- Similarly, for the paired tight T, power results can be approximated by taking a noncentral T approach with the same noncentrality parameter as in the pooled tight T case, but with $k - 1$ degrees of freedom. Also, power can be estimated by the same normal approximation used in the pooled tight T case. Again, the noncentral T approach yields good power approximations for $\rho \leq .80$ and for higher ρ, n combinations. The power overestimation associated with the normal approach is, of course, even more pronounced in the paired case.
- The noncentrality parameter associated with a two-treatment analysis of covariance is

$$(\Delta/\sigma_y\sqrt{1-\rho^2}) \div \sqrt{\frac{2}{k} + \frac{(\bar{X}_{.2} - \bar{X}_{.1})^2}{\sum (X_{i1} - \bar{X}_{.1})^2 + \sum (X_{i2} - \bar{X}_{.2})^2}}.$$

  Thus it pays to minimize $\bar{X}_{.2} - \bar{X}_{.1}$. For this reason an analysis of covariance associated with predictor sort sampling performs slightly better (a .01–.10 increase in power) than an analysis of covariance associated with a standard random allocation.
- If the relationship between the predictor and the response is misspecified, then, even for very high ρ, the tight T's can yield better power than an analysis of covariance. For moderate n, the misspecification can be difficult to detect. For example, for $\rho = .95$, $n = 24$, and $\Delta/\sigma_y = .0(.1).8$, we list the powers associated with tight T analyses and an analysis of covariance in Table 6. The data sets for this analysis were generated by drawing 24 x values from a $N(20, 7^2)$ distribution and obtaining y's via $y = x^3 + \varepsilon$, where the ε's were $N(0, \sigma^2)$ with σ chosen so that the sample correlation between the x's and y's was approximately .95. A (fairly) typical example of such a data set is presented in Figure 3.

Full details of the simulation studies used to estimate critical values, to test these critical values, and to estimate power properties may be found in Verrill and Green (1993).

5. TIGHT T's AS "PARTIALLY PAIRED" T's

It is edifying to see how the noncentrality parameter described earlier,

$$(\Delta/\sigma_y)/\sqrt{2(1-\rho^2)/k},$$

differs from those in the "pure" pooled and paired cases. (Bear in mind, however, that as ρ increases to 1, the noncentral T approach to estimating power loses validity.) Con-

Table 6. Power, $\rho = .95$, $n = 24$, Nominal Size = .01, Misspecified Model

	Δ/σ								
Approach	.0	.1	.2	.3	.4	.5	.6	.7	.8
Tight paired T	.01	.02	.09	.24	.45	.65	.78	.87	.91
Tight pooled T	.00	.01	.05	.15	.33	.55	.73	.85	.92
ANCOVA (Predictor Sort)	.00	.00	.03	.12	.30	.53	.73	.86	.93

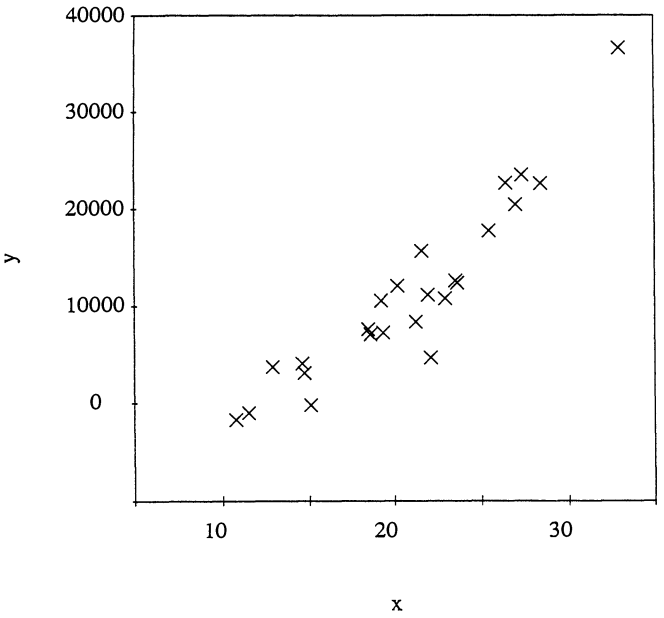

Figure 3. $y = x^3 + \varepsilon$. 24 (x, y) pairs where x is drawn from a N(20, 7^2) distribution, $y = x^3 + \varepsilon$, ε is drawn from a N(0, σ^2) distribution, and σ is chosen so that the sample correlation between x and y is approximately .95.

sider the following error model:

$$(Y - \mu_y)/\sigma_y = \delta + \gamma_y + \varepsilon_y$$

and

$$(X - \mu_x)/\sigma_x = \delta + \gamma_x + \varepsilon_x,$$

where Y is the property of interest, X is the predictor, and δ, γ_y, γ_x, ε_y, and ε_x are independent random variables with means equal to 0 and variances equal to σ_δ^2, σ_γ^2, σ_γ^2, σ_ε^2, and σ_ε^2. (Thus $\sigma_\delta^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2 = 1$.) Here δ represents the "natural variation" shared by Y and X, γ_y and γ_x represent the natural variation unique to Y and X, and ε_y and ε_x are the measurement errors. In this case the noncentrality parameter appropriate to a pure pooled T analysis would be

$$(\Delta/\sqrt{\sigma_y^2(\sigma_\delta^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2)})(\sqrt{k}/\sqrt{2}).$$

The noncentrality parameter that a statistician might recommend and that a scientist might use (incorrectly) to calculate the sample sizes needed for a pure paired T analysis would be

$$(\Delta/(\sigma_y\sigma_\varepsilon))(\sqrt{k}/\sqrt{2}).$$

Because $\rho = \sigma_\delta^2/(\sigma_\delta^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2) = \sigma_\delta^2$, the noncentrality parameter that one would use (correctly) to calculate the sample sizes needed for a predictor sort experiment would be

$$(\Delta/\sqrt{\sigma_y^2(\sigma_\gamma^2 + \sigma_\varepsilon^2)(1 + \sigma_\delta^2)})(\sqrt{k}/\sqrt{2}).$$

We see from this that the predictor sort approach succeeds in partially blocking out natural variation (σ_δ^2). If σ_δ^2 is large in comparison to σ_γ^2 and σ_ε^2 (ρ is high), then the tight T noncentrality parameter is much larger than the pure pooled noncentrality parameter—the tight T yields a large increase in power. But the predictor sort approach does not succeed in blocking out all natural variation (the γ's remain), and if σ_γ^2 is not small in comparison to σ_ε^2, "standard" paired T power calculations can seriously underestimate the sample sizes needed.

6. SUMMARY

Cox (1957) showed that, applied to predictor sort data, blocked analysis of variance methods can be competitive with analysis of covariance provided that $\rho \leq .60$. We have noted that although the assumptions of a blocked analysis of variance are not strictly met in a predictor sort situation, the approach is asymptotically justified. Also, our simulations indicate that (neglecting a minor loss of power in high ρ, low n cases) the approach works well even for small samples. We have also suggested the use of a pooled tight T that is competitive with analysis of covariance for $\rho \leq .85$ and that performs better than an analysis of covariance in low n situations. The tight T statistics can also perform better than an analysis of covariance when the relationship between the predictor and the response is misspecified. Finally, we have warned against careless power calculations in predictor sort situations.

APPENDIX:

Let H denote the inverse of the $N(0, 1)$ distribution function.

Lemma. Let U_{1n} denote the first order statistic from a sample of n uniform (0, 1)'s. Then $H(U_{1n})/\sqrt{n}$ converges in probability to 0.

Proof. Because $(H(U_{1n}) - H(1/n))\sqrt{2ln(n)}$ converges in distribution to an extreme value distribution, and $-H(1/n) \sim \sqrt{2ln(n)}$ (see, for example, David 1981, sec. 9.3), the lemma follows.

A.1 Proof of the Main Result

For ease of exposition, we will present the proof for the one-way case. The extension to a proof of the n-way case is straightforward.

Let $\{X_i, i = 1, \ldots, n\}, \{Z_i, i = 1, \ldots, n\}$ be iid $N(0, 1)$ random variables. define $Y_i \equiv \rho X_i + \sqrt{1 - \rho^2}Z_i$. Then the Y_i's are iid $N(0, 1)$ and

$$\begin{aligned} \operatorname{corr}(X_i, Y_j) &= \rho \quad \text{if } i = j, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Because the ANOVA F statistics are invariant under changes in location and scale, it is clear that we can obtain statistics that have the relevant distributions by ordering the X's, bringing along the Y's, and randomly dividing $Y_{(i-1)J+1,n}, \ldots, Y_{iJ,n}$ among the J treatments. (Here $Y_{l,n}$ is the lth order statistic among the Y's.)

Let W_{ij} denote the ith Y that is assigned to treatment j. Then $W_{ij} = \rho X_{k(i,j),n} + \sqrt{1 - \rho^2}P_{ij}$, where $k(i, j) \in \{(i - 1)J + 1, \ldots, iJ\}$ and the P_{ij} are iid $N(0, 1)$ and are independent of the X's. (Here $X_{l,n}$ is the lth order statistic among the X's.)

A.1.1 The Numerator of the F Statistics. The numerator of both the blocked and unblocked F statistics equals $\sum_{j=1}^{J} I(\bar{W}_{\cdot j} - \bar{W}_{\cdot\cdot})^2$, where $I = n/J$. This equals

$$\begin{aligned} \sum_{j=1}^{J} I(&\rho^2(\bar{X}_{k(\cdot,j),n} - \bar{X}_{k(\cdot,\cdot),n})^2 \\ &+ 2\rho\sqrt{1 - \rho^2}(\bar{X}_{k(\cdot,j),n} - \bar{X}_{k(\cdot,\cdot),n})(\bar{P}_{\cdot j} - \bar{P}_{\cdot\cdot}) \\ &+ (1 - \rho^2)(\bar{P}_{\cdot j} - \bar{P}_{\cdot\cdot})^2). \end{aligned}$$

Now

$$\sum_{j=1}^{J} I(\bar{X}_{k(\cdot,j),n} - \bar{X}_{k(\cdot,\cdot),n})^2 = \sum_{j=1}^{J} I\left(\sum_{i=1}^{I} (X_{k(i,j),n} - \bar{X}_{k(i,\cdot),n})\Big/ I\right)^2$$

$$\le \sum_{j=1}^{J} \left(\sum_{i=1}^{I} (X_{iJ,n} - X_{(i-1)J+1,n})\right)^2 \Big/ I \le \sum_{j=1}^{J} (X_{n,n} - X_{1,n})^2 \Big/ I$$

which converges in probability to 0 by the Lemma. Also, it is clear that $\sum_{j=1}^{J} I(\bar{P}_{\cdot j} - \bar{P}_{\cdot\cdot})^2 \sim \chi^2_{J-1}$. These results, together with the Cauchy–Schwarz inequality, imply that

$$\sum_{j=1}^{J} I(\bar{W}_{\cdot j} - \bar{W}_{\cdot\cdot})^2 \xrightarrow{D} (1 - \rho^2)\chi^2_{J-1}. \tag{A.1}$$

A.1.2 The Unblocked F Denominator. We have $\sum_{j=1}^{J} \sum_{i=1}^{I} \times (W_{ij} - \bar{W}_{\cdot j})^2/J(I-1) = \sum_{j=1}^{J} \sum_{i=1}^{I} (W_{ij} - \bar{W}_{\cdot\cdot} + \bar{W}_{\cdot\cdot} - \bar{W}_{\cdot j})^2/J(I-1) = \sum_{j=1}^{J} \sum_{i=1}^{I} (W_{ij} - \bar{W}_{\cdot\cdot})^2/J(I-1) - \sum_{j=1}^{J} I \times (\bar{W}_{\cdot j} - \bar{W}_{\cdot\cdot})^2/J(I-1)$. It is clear that the first term in the last sum converges to 1 in probability as $I = n/J$ goes to infinity. By (A.1), the second term in the last sum converges in probability to 0, so the unblocked denominator converges in probability to 1.

A.1.3 The Blocked F Denominator. We have

$$\sum_{j=1}^{J} \sum_{i=1}^{I} (W_{ij} - \bar{W}_{i\cdot} - \bar{W}_{\cdot j} + \bar{W}_{\cdot\cdot})^2/(I-1)(J-1)$$

$$= \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{(W_{ij} - \bar{W}_{i\cdot})^2 - 2(W_{ij} - \bar{W}_{i\cdot})(\bar{W}_{\cdot j} - \bar{W}_{\cdot\cdot}) + (\bar{W}_{\cdot j} - \bar{W}_{\cdot\cdot})^2}{(I-1)(J-1)}. \tag{A.2}$$

By (A.1), the last term in (A.2) converges in probability to 0 as $I = n/J$ goes to infinity. The first term equals $\sum_{j=1}^{J} \sum_{i=1}^{I} (\rho(X_{k(i,j),n} - \bar{X}_{k(i,\cdot)}) + \sqrt{1-\rho^2}(P_{ij} - \bar{P}_{i\cdot}))^2/(I-1)(J-1)$. Clearly,

$$\sum_{j=1}^{J} \sum_{i=1}^{I} (X_{k(i,j),n} - \bar{X}_{k(i,\cdot)})^2/(I-1)(J-1) \le \sum_{j=1}^{J} \sum_{i=1}^{I} (X_{iJ,n} - X_{(i-1)J+1,n})^2/(I-1)(J-1).$$

Then, because $\sum |a| \xrightarrow{p} 0$ implies that $\sum a^2 \xrightarrow{p} 0$, and $\sum_{i=1}^{I} (X_{iJ,n} - X_{(i-1)J+1,n}) \le X_{n,n} - X_{1,n}$, the Lemma, together with the Cauchy–Schwarz inequality, implies that the first term in (A.2) converges in probability to $1 - \rho^2$. Making one last use of the Cauchy–Schwarz inequality (to show that the second term in (A.2) converges in probability to 0), we see that the blocked denominator converges in probability to $1 - \rho^2$.

[*Received January 1991, Revised February 1992.*]

REFERENCES

Cochran, W. G. (1957), "Analysis of Covariance: Its Nature and Uses," *Biometrics,* 13, 261–281.

Cox, D. R. (1957), "The Use of a Concomitant Variable in Selecting an Experimental Design," *Biometrika,* 44, 150–158.

——— (1958), *Planning of Experiments,* New York: John Wiley.

David, H. A. (1981), *Order Statistics* (2nd ed.), New York: John Wiley.

Finney, D. J. (1972), *An Introduction to Statistical Science in Agriculture,* New York: John Wiley.

Ostle, B., and Mensing, R. W. (1975), *Statistics in Research,* Ames, IA: Iowa State University Press.

Snedecor, G. W., and Cochran, W. G. (1967), *Statistical Methods,* Ames, IA: Iowa State University Press.

Verrill, S. P., and Green, D. W. (1993), "Predictor Sort Sampling, Tight *T*'s, and the Analysis of Covariance: Theory, Tables, and Examples," technical report, Madison, WI: USDA Forest Products Laboratory.

Verrill, S. P., and Johnson, R. A. (1988), "Tables and Large-Sample Distribution Theory for Censored-Data Correlation Statistics for Testing Normality," *Journal of the American Statistical Association,* 83, 1192–1197.

Warren, W. G., and Madsen, B. (1977), "Computer-Assisted Experimental Design in Forest Products Research: A Case Study Based on Testing the Duration-of-Load Effect," *Forest Products Journal,* 27, 45–50.