

## On the $f$ Statistic and Comparable Measures in Linear Free-Energy Relationship Fittings<sup>1</sup>

S. Ehrenson

Department of Chemistry, Brookhaven National Laboratory, Upton, New York 11973

Received July 25, 1978

The properties and validity of a goodness-of-fit statistic,  $f$ , previously employed in dual-substituent linear free-energy equation fittings are carefully examined and verified. Confidence level tables are presented for  $f$  ratio comparisons and examples given for their use. These confidence level measures constitute an original (for linear free-energy applications) means of judging whether one system of parametrization ultimately based upon experimental observation is significantly (statistically) better than another similarly derived in given correlation applications. Brief analysis of relationships to other currently favored measures, most of which are not directly comparative, is presented. The value and means of testing independence among multiple variable sets are also given brief scrutiny.

Attempts at improved modeling in linear free energy relationship (LFER) applications over the past 15 years have been accompanied by searches for statistical fitting measures of greater resolution.<sup>2</sup> Among those adopted, one of a particularly simple and, it was felt, revealing nature, the  $f$  statistic, has been extensively used by the author and co-workers for dual substituent LFER fittings, particularly for alternative parameter set comparisons, without formal demonstration of its validity.<sup>3,4</sup> Recognizing the subjective nature of elements of applied statistics,<sup>5</sup> it came as no surprise thereafter to find the meaning and by implication the value of this measure questioned<sup>6</sup> in efforts to analyze advantages of those several in common use. The following should serve to answer such questions and through the confidence level tabulations better quantify its use for inter-set significance comparisons. Where it was felt to serve a useful purpose beyond review, the rather simple connections which exist with other measures are indicated. These are, of course, not unexpected since all ultimately refer to residuals minimized by linear or linearized least-squares procedures.

**Nature of and Confidence Levels for the  $f$ -Ratio Distribution.** There is in extensive use among crystallographers,<sup>7,8</sup> but otherwise not very well known, an agreement, or "R factor", statistic which is defined as,

$$R = \left\{ \frac{\sum_{i,j} (\phi_i^{\text{obsd}} - \phi_i^{\text{calcd}}) w_{ij} (\phi_j^{\text{obsd}} - \phi_j^{\text{calcd}})}{\sum_{i,j} \phi_i^{\text{obsd}} w_{ij} \phi_j^{\text{obsd}}} \right\}^{1/2} \quad (1)$$

where the  $\phi$ 's are a set of observed functions,  $n$  in number ( $i, j = 1, \dots, n$ ), or the values calculated for them in some parametrized model representation, depending upon the superscript as shown. The measure is best viewed as one of composite goodness of approximation of the functions relative to the observed composite magnitude of the functions themselves. Where the weight matrix, of which  $w_{ij}$  are elements, is diagonal and unitary, conforming to the implicit weighting scheme employed in linear or linearized LFER least-squares procedures,<sup>9,10</sup> this meaning becomes transparent as  $R$  is seen to reduce and in so doing become identical with  $f$ ,

$$R = \left\{ \frac{\sum_i (\phi_i^{\text{obsd}} - \phi_i^{\text{calcd}})^2}{\sum_i (\phi_i^{\text{obsd}})^2} \right\}^{1/2} \equiv f = \frac{\left\{ \sum_i \delta_i^2/n \right\}^{1/2}}{\left\{ \sum_i (P_i^{\text{obsd}})^2/n \right\}^{1/2}} \quad (2)$$

$P_i$  is  $\log k_i$ ,  $\log k_i/k^0$ , or otherwise, dependent upon the LFER form used, and  $\delta_i$  is the observed to calculated  $P_i$  value dif-

ference. The numerator of the rightmost side of eq 2 should be quickly recognized as the root-mean-square (rms) deviation of the fitted set and the denominator as the rms of observed values.<sup>11</sup>

It has been shown<sup>12</sup> that if a least-squares refinement (determination) of  $p$  parameters  $\mathbf{X}$  (here usually the  $\rho$ 's but on occasion otherwise, with or without intercepts) has resulted in the best set for which  $f_0$  pertains and there is another set  $\mathbf{X}_1$  for which the  $f$  value is  $f_1$ , then

$$(f_1^2 - f_0^2)/f_0^2 \quad (3)$$

is distributed as  $[p/(n-p)]F_{p,n-p}$ , where  $F$  is the well-known Fisher (Snedecor) distribution,<sup>13,14</sup> assuming approximate linearity in the hypotheses. As in the crystallographic problems cited,<sup>7,12</sup> this is expected to pertain here in that all parameter sets are ultimately derived from and applied to the same or similar subsets of a defining data base similarly viewed, i.e., on  $P$  values for aromatic reactivities, spectral intensities, NMR shifts, etc., fitted to  $\rho\sigma$  products and sums thereof. Hence to test the hypothesis

$$H_0: \mathbf{X} = \mathbf{X}_1 \quad (4)$$

the  $f$  ratio,  $\mathcal{R} = f_1/f_0$ , is compared with

$$\mathcal{R}_{p,n-p,\alpha} \equiv \left\{ \frac{p}{n-p} F_{p,n-p,\alpha} + 1 \right\}^{1/2} \quad (5)$$

If  $\mathcal{R} > \mathcal{R}_{p,n-p,\alpha}$ , the hypothesis, to wit, parameter set  $\mathbf{X}_1$  provides as good fitting as  $\mathbf{X}$ , may be rejected on the  $100\alpha\%$  significance (confidence) level. Values for the  $\mathcal{R}$  statistic can readily be derived from tables of  $F$  commonly available: Hamilton's text<sup>12</sup> has tables most useful for the large sample cases ( $n$  large) generally encountered in crystallographic structure refinements. Tables I-III presented here cover the small  $n$  range over what is probably the full useful confidence level (CL) span for the special cases,  $p$  equal 1, 2, and 3, of most interest in current LFER applications.

What we have here, then, to recapitulate, is a compact, easily used, and, it would appear in the LFER context, a heretofore unexploited means by which the often qualitatively made statement that one data-derived parameter set is significantly better at correlating a given series of related observations than another may be quantitatively tested. The implications of such tests to further proliferation of parameter sets, and in the development of new equation forms, should be profound.

**Discussion of Meaning and Uses in LFER Regression Analysis.** Examination of these CL compilations, particularly in Table II, reveals the  $f$ -ratio estimate factor of 2 for the lower limit of significance between alternative dual substituent parameter sets given in ref 3 to have been sound. The choice

Table I.  $\mathcal{R}_{1,n-p,\alpha}$  Values

$n - p$	$\alpha$							
	0.50	0.75	0.90	0.95	0.975	0.990	0.995	0.999
1	1.414	2.613	6.392	12.75	25.47	63.66	127.3	636.6
2	1.155	1.512	2.294	3.203	4.500	7.089	10.01	22.37
3	1.093	1.294	1.687	2.092	2.610	3.517	4.418	7.528
4	1.066	1.205	1.462	1.711	2.014	2.510	2.972	4.420
5	1.051	1.157	1.346	1.524	1.732	2.062	2.357	3.230
6	1.042	1.127	1.276	1.413	1.571	1.814	2.026	2.630
7	1.035	1.107	1.230	1.341	1.467	1.658	1.822	2.276
8	1.031	1.092	1.197	1.290	1.395	1.552	1.684	2.044
9	1.027	1.081	1.172	1.252	1.342	1.474	1.585	1.881
10	1.024	1.072	1.153	1.223	1.301	1.416	1.511	1.762
11	1.022	1.065	1.137	1.200	1.269	1.370	1.453	1.670
12	1.020	1.059	1.125	1.181	1.243	1.333	1.407	1.598
13	1.018	1.054	1.114	1.166	1.222	1.303	1.369	1.540
14	1.017	1.050	1.105	1.153	1.204	1.278	1.338	1.491
15	1.016	1.047	1.098	1.141	1.189	1.257	1.311	1.451
16	1.015	1.044	1.091	1.132	1.176	1.238	1.289	1.417
17	1.014	1.041	1.085	1.123	1.164	1.222	1.269	1.387
18	1.013	1.039	1.080	1.116	1.154	1.208	1.252	1.362
19	1.012	1.036	1.076	1.109	1.145	1.196	1.237	1.339
20	1.012	1.034	1.072	1.103	1.137	1.185	1.224	1.319
22	1.011	1.031	1.065	1.093	1.124	1.167	1.201	1.286
24	1.010	1.029	1.059	1.085	1.113	1.152	1.182	1.259
26	1.009	1.026	1.054	1.078	1.103	1.139	1.167	1.236
28	1.008	1.024	1.050	1.072	1.096	1.128	1.154	1.217
30	1.008	1.023	1.047	1.067	1.089	1.119	1.143	1.201
40	1.006	1.017	1.035	1.050	1.066	1.088	1.105	1.147
60	1.004	1.011	1.023	1.033	1.043	1.057	1.068	1.095

there was entirely subjective. Only for rather meager sets ( $n < 7$ ) does such a ratio suggest less than 95% confidence in distinguishability; where  $n$  is 10 or more the CL is  $>99.5\%$ .

One example should suffice to illustrate this most direct use of the  $f$ -ratio test. In ref 4 it was reported that equilibrium constants for ionization of para-substituted phenols in acetone (reaction series 14, Tables VI and VII with  $n = 10$ ) are correlated with  $f$  values of 0.093, 0.173, 0.151, 0.256, and 0.146, respectively, employing the dual substituent sets  $\sigma_{R(A)^-}$ ,  $\sigma_{R(BA)}$ ,  $\sigma_{R^0}$ , and  $\sigma_{R^-}$  (all with  $\sigma_I$ ) and the Swain and Lupton  $FM$  set.<sup>15</sup> Let us proceed to test the hypothesis that the set with  $\sigma_{R(A)^-}$  values is not distinguishably better at correlating the effects of substitution on these reactivities. The  $\mathcal{R}$  values pertaining to the sets following  $\sigma_{R(A)^-}$  in the order given above, and relative thereto, are 1.86, 1.62, 2.75, and 1.57, respectively. Hence, the CL (percentages) at which the individual hypotheses of equality of representation can be rejected are 99, 97.5, 99.9, and 95. In all cases then we may reasonably conclude that the set with  $\sigma_{R(A)^-}$  is superior in fitting these data. Note well, only pairwise comparisons can be and are drawn here with the statistic as presented. As well, the reaction series tested was *not* part of the data base for determination of the  $\sigma_{R(A)^-}$  set. (Being so it should be recognized gives any parameter set an advantage.)

Comparison of the fitting possible with the "best" single substituent set (i.e., the  $\sigma_p^-$  set) serves to further illustrate the utility of the  $f$  ratio measure. From the same source,<sup>4</sup> we see the  $f$  value for the  $\sigma_p^-$  fitting is 0.133, yielding an  $\mathcal{R}$  of 1.43 against the dual substituent  $\sigma_{R(A)^-}$  set result. The linear hypothesis is here of order one, that is, in the removal of one parameter (one of the two  $\rho$ 's), and the distribution for reference, generally

$$\mathcal{R}_{b,n-p,\alpha} = \left\{ \frac{b}{n-p} F_{b,n-p,\alpha} + 1 \right\}^{1/2} \quad (6)$$

is now  $\mathcal{R}_{1,8,\alpha}$ . We quickly find from Table I that the null hypothesis, to wit, representation by the single parameter equation with  $\sigma_p^-$  is indistinguishable from that by the best

dual substituent set, may be rejected at the 97.5% level. It should be noted that the test as described is strictly appropriate only if the second parameter set is derived through restriction of the first set (often but not necessarily upon complete removal of  $b$  of the original  $p$  parameters; see, e.g., example 6 of ref 7). This is not exactly the case here, but closely so, in that  $\sigma_p^-$  is to a good approximation an equally weighted linear combination of  $\sigma_I$  and  $\sigma_{R(A)^-}$ . Again, then, we may conclude with considerable confidence that the  $\sigma_I$  and  $\sigma_{R(A)^-}$  fitting is significantly better than  $\sigma_p^-$ . An even more favorable appraisal would of course accompany comparison against either of the individual ( $\sigma_I$  or  $\sigma_{R(A)^-}$ ) sets alone.

**Relationship to Other Measures and Tests.** Several observations concerning these and related measures and the conclusions which may be drawn from them are usefully considered at this point. The first is: unlike most conventional uses of confidence level tests for significance of correlation in the LFER literature, the tests here are for one representation (parameter set) compared to another and not whether the correlation with any one is "good" (meaningful). In the latter case the null hypothesis is essentially that the relationship between dependent and independent variables(s) is random, not really a very useful line of division for alternatives, particularly upon refutation. This doubtlessly accounts for the huge  $F$  values and high CL's generally reported and, in retrospect, to have been expected. Figure 1 reveals just how poor these measures are in distinguishing correlation quality by comparison with  $f$ ; recall  $f \leq 0.1$  was chosen,<sup>3,4</sup> essentially arbitrarily, as the range of good quality correlation.  $F_{2,n-2,\alpha}$ , appropriate for the dual substituent equation with no intercept, may be shown equal to  $(n-2)(f^{-2}-1)/2$  and the CL's (100 $\alpha\%$ ), generally obtainable upon series expansions in terms of  $F_{b,n-p,\alpha}$ ,<sup>14c</sup> are here particularly simple; i.e.,  $\alpha = 1 - f^{n-2}$ .

In point of fact, it is just because of similar limitations in the sample correlation coefficient, i.e., separating randomness from dependence on a poorly resolving and hence potentially easily misinterpretable scale, that it was abandoned for use with the relatively precise fittings now obtained in LFER studies. Exner<sup>16,17</sup> has alluded to this deficiency of  $r$  in the

Table II.  $R_{2,n-p,\alpha}$  Values

$n - p$	$\alpha$							
	0.50	0.75	0.90	0.95	0.975	0.990	0.995	0.999
1	2.000	4.000	10.00	20.00	40.00	100.0	200.0	1000.
2	1.414	2.000	3.162	4.472	6.325	10.00	14.14	31.62
3	1.260	1.587	2.154	2.714	3.420	4.642	5.848	10.00
4	1.189	1.414	1.778	2.115	2.515	3.162	3.761	5.624
5	1.149	1.320	1.585	1.821	2.091	2.512	2.885	3.981
6	1.122	1.260	1.468	1.648	1.849	2.154	2.418	3.162
7	1.104	1.219	1.389	1.534	1.694	1.931	2.132	2.683
8	1.091	1.189	1.334	1.454	1.586	1.778	1.939	2.371
9	1.080	1.167	1.292	1.395	1.507	1.668	1.802	2.155
10	1.072	1.149	1.259	1.349	1.446	1.585	1.699	1.995
11	1.065	1.134	1.233	1.313	1.398	1.520	1.619	1.874
12	1.059	1.122	1.212	1.284	1.360	1.468	1.555	1.778
13	1.055	1.113	1.194	1.259	1.328	1.425	1.503	1.701
14	1.051	1.104	1.179	1.239	1.301	1.389	1.460	1.638
15	1.047	1.097	1.166	1.221	1.279	1.359	1.424	1.585
16	1.044	1.091	1.155	1.206	1.259	1.334	1.393	1.540
17	1.042	1.085	1.145	1.193	1.242	1.311	1.366	1.501
18	1.039	1.080	1.136	1.181	1.227	1.292	1.342	1.468
19	1.037	1.076	1.129	1.171	1.214	1.274	1.322	1.439
20	1.035	1.072	1.122	1.162	1.203	1.259	1.303	1.412
22	1.032	1.065	1.110	1.146	1.183	1.233	1.272	1.369
24	1.029	1.059	1.101	1.133	1.166	1.212	1.247	1.334
26	1.027	1.055	1.093	1.122	1.152	1.194	1.226	1.304
28	1.025	1.051	1.086	1.113	1.141	1.179	1.208	1.280
30	1.023	1.047	1.080	1.105	1.131	1.166	1.193	1.259
40	1.017	1.035	1.059	1.078	1.097	1.122	1.142	1.188
60	1.012	1.023	1.039	1.051	1.063	1.080	1.092	1.122

course of development of a statistical measure  $\psi$ , which bears some resemblance to  $f$ . The principal difference lies in normalization of variance by the rms of the observed values about the mean, which, as previously mentioned, is consistent with the greater statistical freedom accompanying intercept determination. As has been variously noted, however,  $\psi$  is just an expanded scale complement of  $r$ , or more precisely a complement of what Swain and Lupton<sup>15</sup> call the coefficient of determination, " $C^2$ ", if the degrees of freedom in the sample populations are explicitly and correctly considered. The ratio  $f$ , on the other hand, adopted for forms where explicit standard substituent reference is assumed to preclude intercepts other than the residual  $\epsilon$  (which in the simplest situation distributes as  $N(0, \sigma^2)$ ),<sup>18</sup> is consistent with normalization by the absolute rms observed value.<sup>19</sup> It is nevertheless usefully recognized that  $\psi$  ratios comparing alternative set fittings would distribute identically with  $R_{b,n-p}$  as the  $f$  ratios (and for that matter as would the relative standard errors) for static populations.

Another useful comparison may be drawn from the second illustrated application, that is, for relative fitting to a more restricted equation form. It can be shown in general that the  $R$  tests consistent with the imposed constraint of complete parameter removal are identical, as regards confidence levels, to the  $t$  tests for significance usually reported for individual parameters.<sup>20</sup> The latter, cast as the null hypothesis, i.e., that the value of the parameter  $C_k$  is zero, is

$$t_k = C_k/s_k \quad (7)$$

and

$$(f_1^2 - f_0^2)/f_0^2 = t_k^2/(n - p) \quad (8)$$

where  $f_1$  is for the set with parameter  $k$  freedom removed and  $s_k$  is the standard error computed for the parameter in the original fitting from whence its  $C_k$  and  $f_0$  values obtain. Recognition of this similarity, besides providing an alternative way of computing comparative set fitting statistics, should

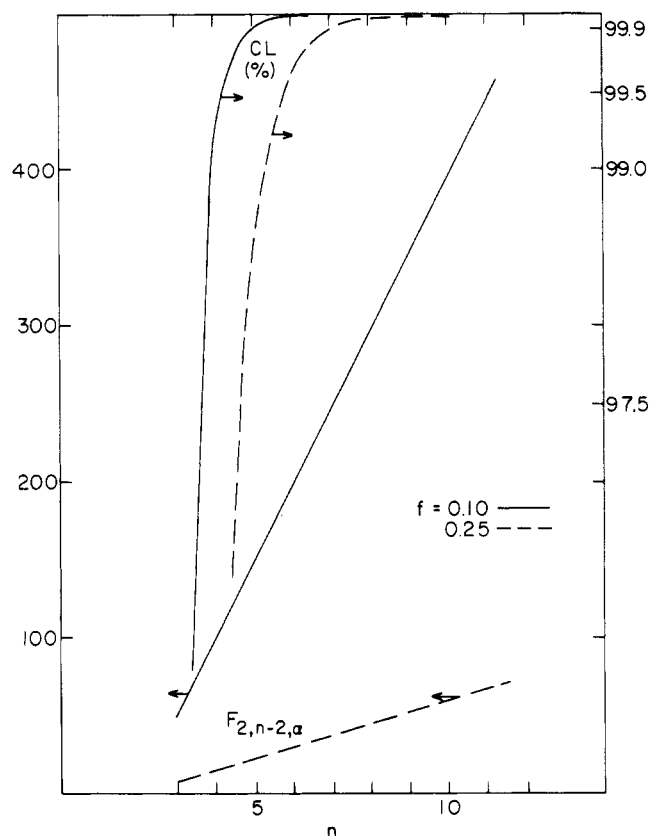


Figure 1. Fisher (Snedecor)  $F$  values and associated confidence levels corresponding to two values of  $f$  as functions of sample size.

tend to discourage misinterpretation of parameter confidence levels as routinely derived.<sup>21</sup>

The  $C_p$  statistic of Wold and Sjöström is pertinent to the present discussion.<sup>22,23</sup> Developed for use with the original

Table III.  $R_{3,n-p,\alpha}$  Values

$n - p$	$\alpha$							
	0.50	0.75	0.90	0.95	0.975	0.990	0.995	0.999
1	2.475	5.060	12.72	25.46	50.93	127.3	254.6	1273.
2	1.644	2.394	3.840	5.454	7.730	12.24	17.31	38.73
3	1.414	1.832	2.528	3.206	4.055	5.519	6.962	11.92
4	1.306	1.592	2.035	2.438	2.913	3.677	4.381	6.568
5	1.243	1.460	1.781	2.061	2.379	2.870	3.304	4.574
6	1.201	1.376	1.626	1.838	2.073	2.427	2.731	3.585
7	1.172	1.318	1.522	1.692	1.877	2.150	2.380	3.007
8	1.150	1.275	1.448	1.589	1.741	1.961	2.144	2.634
9	1.133	1.243	1.392	1.512	1.641	1.825	1.976	2.373
10	1.120	1.217	1.348	1.453	1.565	1.722	1.850	2.183
11	1.109	1.196	1.314	1.407	1.504	1.642	1.753	2.038
12	1.099	1.179	1.285	1.368	1.456	1.577	1.675	1.924
13	1.092	1.165	1.261	1.337	1.415	1.525	1.612	1.832
14	1.085	1.152	1.241	1.310	1.382	1.481	1.559	1.756
15	1.079	1.142	1.224	1.287	1.353	1.443	1.515	1.694
16	1.074	1.133	1.209	1.268	1.328	1.411	1.477	1.639
17	1.070	1.125	1.196	1.251	1.307	1.384	1.444	1.594
18	1.066	1.118	1.184	1.236	1.288	1.360	1.416	1.554
19	1.063	1.111	1.174	1.222	1.271	1.338	1.391	1.519
20	1.059	1.105	1.165	1.210	1.257	1.319	1.368	1.488
22	1.054	1.096	1.149	1.190	1.231	1.287	1.331	1.437
24	1.050	1.088	1.136	1.173	1.210	1.261	1.300	1.394
26	1.046	1.081	1.125	1.159	1.193	1.239	1.274	1.360
28	1.042	1.075	1.116	1.147	1.178	1.220	1.253	1.331
30	1.040	1.070	1.108	1.137	1.166	1.205	1.234	1.306
40	1.030	1.052	1.080	1.101	1.122	1.150	1.172	1.223
60	1.020	1.035	1.053	1.067	1.080	1.098	1.112	1.144

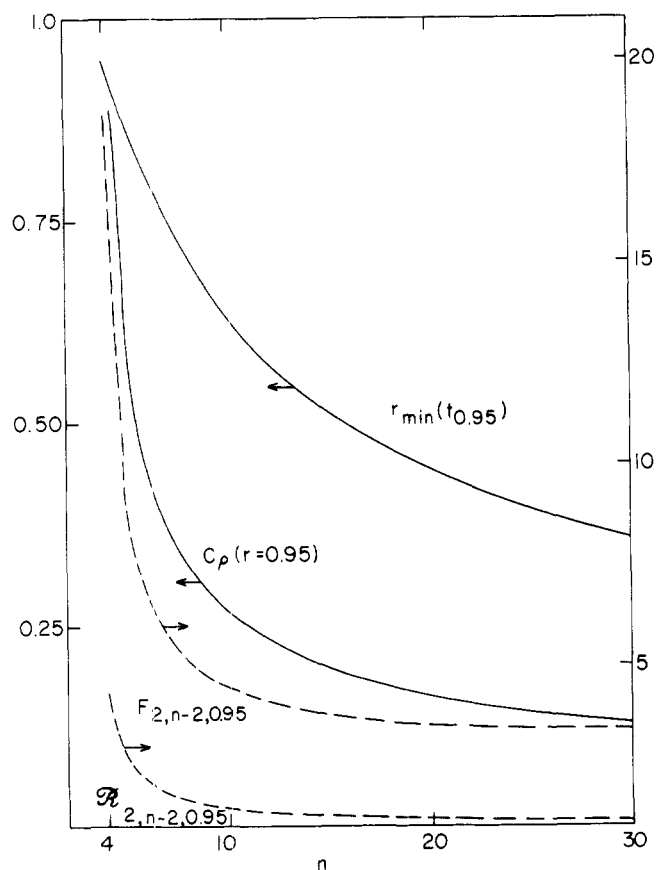


Figure 2. Comparison of the variation of several measures of sample-fitting significance with the number of data points.

Hammett form in yet another attempt to improve resolution among relatively good fittings, this statistic upon close examination can be recognized as a reciprocal "Student"  $t$ -measure for  $\rho$ , normalized to the 95% CL for the appropriate

number of degrees of freedom.

$$C_\rho = t(n - 2; 0.95)s_\rho/\rho \quad (9)$$

$$\rho = t_{ref}/t_\rho \quad (10)$$

Besides magnifying the region of significant correlation, which is the primary reason cited for its adoption, this measure is purported to decrease the strong sample size dependence exhibited by other measures such as  $r$  and  $\psi$  for given confidence levels. Whether it accomplishes this or not is moot since in the illustration provided (Figure 3 in ref 22)  $C_\rho$  clearly exhibits an increasing and by no means inconsequential variation with  $n$  for fixed values of  $r$  at small, but often just those encountered, sample sizes.

In fact, as it may readily be shown, the least significant value of  $r$  ( $r_{min.}$ ) at a given sample size  $n$ , obtained from the null hypothesis  $t$  test for  $r$ ,

$$t = \{(n - 2)r^2/(1 - r^2)\}^{1/2} \quad (11)$$

varies more slowly with change in  $n$  than does  $C_\rho$  for small  $n$  for all but the most highly correlated cases ( $r \geq 0.999$ ).<sup>24</sup> (Likewise,  $\psi$  measures of borderline significance of correlation may also be shown to vary more rapidly with  $n$  than  $r$  and hence to also be strongly sample size dependent.)

Upon reflection, it seems undesirable to seek measures which obscure the greater uncertainties incumbent in parameter fitting to small rather than large samples, as well as the greater increment in information for the former upon equal increase in sample sizes. That this variation is in fact completely general and natural is seen in Tables I-III. The  $R$  measures for relative fittings vary similarly albeit considerably more gradually with sample size, as do the absolute measures for a given parameterization cited above, changing rapidly at small  $n$ , but at a decreasing rate and always toward conditions of more subtle distinction among alternatives with increasing sample size. Figure 2 contrasts the behavior of some of the measures considered with change in sample size.

As a final point, it may be useful to note that eq 11 is representative of null hypothesis tests which may be carried out

on significance of correlation among "independent" variables, as well as for sample correlations. The ramifications of such tests to multiparameter fittings should be clear. For example, the independence of  $\sigma_I$  and  $\sigma_R^0$ , claimed in ref 3 and presumably demonstrated by what appears to be a scatter plot and other arguments, conforms to the null hypothesis

$$H_0: r_{\sigma_I, \sigma_R^0} = 0 \quad (12)$$

From eq 11 we quickly find that this hypothesis may be rejected at the 80% CL ( $t = 1.53$  for 15 degrees of freedom). Adopting the conventional 95% CL criterion to discount correlations resulting from chance, it seems reasonable to conclude that these variables previously assumed linearly unrelated are indeed so.

A different conclusion might well be drawn for the variables  $\sigma^*$  and  $\nu$ , as employed in the amide hydrolysis correlations presented in ref 20. If all 11 alkyl groups, the nature of whose electrical and steric effects are matters of current controversy,<sup>20,25</sup> are considered,  $t$  is found to be 2.59 and the hypothesis that these parameter sets are uncorrelated can be rejected at somewhat more than the 95% level. Here, one might very reasonably conclude that the parameters are significantly correlated. Only upon removal of the *tert*-butyl group, as is generally done in the alternative fittings labeled A in the original correlations of these data<sup>26</sup> and carried over,<sup>20</sup> is the suggestion of interdependence at all countered (CL drops to 80%).<sup>27</sup>

The tests described for  $\sigma^*$  and  $\nu$  pertain to assumed linear dependence, i.e., for  $\sigma^* = a\nu + b$  with  $r_{\sigma^*, \nu} = 0$ , similar to that assumed between  $\sigma_I$  and  $\sigma_R^0$ . These tests have been generalized, however, for use on partial correlation coefficients and to multiple correlation (for dependent upon all independent variables) directly,<sup>12</sup> and their adoption for LFER fittings may be similarly useful.

### References and Notes

- (1) Research carried out at Brookhaven National Laboratory under contract with the U.S. Department of Energy and supported by its Office of Basic Energy Sciences.
- (2) Following the efforts of H. H. Jaffé, *Chem. Rev.*, **53**, 191 (1953), wherein extensive correlations employing the Hammett equation in its original form revealed the sample correlation coefficient,  $r$ , to be a poorly resolving and hence potentially deceptive statistic.
- (3) P. R. Wells, S. Ehrenson, and R. W. Taft, *Progr. Phys. Org. Chem.*, **6**, 147 (1968).
- (4) S. Ehrenson, R. T. C. Brownlee, and R. W. Taft, *Progr. Phys. Org. Chem.*, **10**, 1 (1973).
- (5) As judged by the fact that while most good texts include tables of the more common statistical distributions, they almost as often also contain at least one uncommon distribution toward which the author shows a predilection.
- (6) In personal inquiries mainly, but see also J. Shorter in "Correlation Analysis in Chemistry—Recent Advances", N. B. Chapman and J. Shorter, Eds.,

- Plenum Press, made available kindly by J. S. as a preprint in 1976.
- (7) W. C. Hamilton, *Acta Crystallogr.*, **18**, 502 (1965).
  - (8) A. J. C. Wilson, *Acta Crystallogr., Sect. A*, **32**, 994 (1976).
  - (9) S. Ehrenson, *Tetrahedron Lett.*, **7**, 351 (1964).
  - (10) S. Ehrenson, *Progr. Phys. Org. Chem.*, **2**, 195 (1964).
  - (11) Not, it should be noted, the rms of the observed values about their mean. The latter, often used in linear statistics where intercepts are determined, would be inconsistent upon removal of this degree of freedom and complicating as regards use of the generalized  $R$  factor. See, however, later where this normalization is further discussed.
  - (12) Cf., W. C. Hamilton, "Statistics in Physical Science", Ronald Press, New York, 1964, Chapter 5. Note, Hamilton's  $\alpha$  signifying the probability of error upon rejection is the complement of that employed here which signifies as a fraction the CL for hypothesis rejection.
  - (13) G. W. Snedecor and W. G. Cochran, "Statistical Methods", 6th ed., Iowa State University Press, Ames, Iowa, 1967.
  - (14) (a) W. H. Beyer, Ed., "Handbook of Tables for Probability and Statistics", The Chemical Rubber Publishing Co., Cleveland, Ohio, 1966, pp 19 and 240–246; (b) M. Abramowitz and I. Stegun, Eds., "Handbook of Mathematical Functions", Dover Publications, New York, 1965, Chapter 26; (c) *ibid.*, p 946, eq 26.6.4.
  - (15) C. G. Swain and E. C. Lupton, Jr., *J. Am. Chem. Soc.*, **90**, 4328 (1968).
  - (16) O. Exner, *Collect. Czech., Chem. Commun.*, **31**, 3222 (1966).
  - (17) O. Exner in "Advances in Linear Free Energy Relationships", N. B. Chapman and J. Shorter, Eds., Plenum Press, New York, 1972, Chapter 1.
  - (18) Reference 13, pp 166–170.
  - (19) Compare, e.g., with the accepted definition of residual mean square and standard error for the coefficient in the simplest nonintercept case,  $P_i = \rho\sigma_i$ ,

$$s_{P, \sigma}^2 = \frac{\sum P_i^2 - (\sum P_i \sigma_i)^2 / \sum \sigma_i^2}{(n-1)}$$

$$s_p = s_{P, \sigma} / (\sum \sigma_i^2)^{1/2}$$

- (20) Cf., M. Charton, *J. Am. Chem. Soc.*, **99**, 5687 (1977).
- (21) For example, reports of CL for the null hypothesis on parameters which are inconceivably zero (e.g., the intercepts in LFER where reference to a standard is not explicitly made) should be less likely if the broader implications of the  $t_k$  statistic discussed here are appreciated.
- (22) S. Wold and M. Sjöström, *Chem. Scr.*, **2**, 49 (1972).
- (23) M. Sjöström and S. Wold, *Chem. Scr.*, **6**, 114 (1974).
- (24) We explicitly contrast samples of size 5 and 20, as is done in ref 22, with a consistent definition of  $r_{\min}$ , i.e., at values of  $r < r_{\min}$ , the probability that the correlation is a result of chance is  $>0.05$ . For  $n = 5$ ,  $r_{\min} = 0.88$ , and at  $n = 20$   $r_{\min}$  is 0.44; more importantly, for the sake of the present argument  $(dr_{\min}/dn)_{r=0.95}$  is less than one-fifth the size for the latter compared to the former sample. See ref 12, p 183, for further details concerning eq 11.

Similar scrutiny may be given the individual series of  $f$  values. Consistent with the foregoing analysis, we may define, as a function of  $n$ , maximum values of  $f$ , above which the dual substituent equation correlation, for example, has a significant probability ( $>0.05$ ) of being the result of chance.

$$f_{\max} = \left( \frac{2F_{2, n-2, 0.95}}{n-2} + 1 \right)^{-1/2}$$

- One finds  $f_{\max}$  climbing sharply over the small sample range (0.05, 0.22, 0.37, 0.47 for  $n$  equal 3, 4, 5, and 6, respectively), to vary, obviously less sharply, between 0.69 and 0.85 for samples of 10 to 20 points, and thereafter to approach unity asymptotically. This appears to confirm earlier expectations that  $f \leq 0.1$  represents a meaningful level for correlation "goodness" for any reasonably populated reaction series.
- (25) A. J. MacPhee and J. E. Dubois, *Tetrahedron Lett.*, **28**, 2471 (1976).
  - (26) M. Charton, *J. Org. Chem.*, **41**, 2906 (1976).
  - (27) Compare, e.g., with the effect of removal of an extreme point, the  $\text{NO}_2$  group, for the  $\sigma_I, \sigma_R^0$  test. Here,  $t$  drops to 0.97 equivalent to 50% CL for rejection of the null hypothesis. Note, however, that  $\text{NO}_2$  does not enhance definition of this correlation as does *t*-Bu among the alkyls. The  $\sigma_I$  range is increased by  $\sim 20\%$  and  $\sigma_R^0$  not at all by  $\text{NO}_2$ , whereas *t*-Bu is seen to increase the range of  $\sigma^*$  and  $\nu$  both by more than 40%.