AD-A058 809 CASE WESTERN RESERVE UNIV CLEVELAND OHIO DEPT OF MA--ETC F/G 12/1
SIMULTANEOUS CONFIDENCE INTERVALS FOR LINEAR COMBINATIONS OF TW--ETC(U)
AUG 78 S ZACKS N00014-75-C-0529

UNCLASSIFIED TR-32 NL

| OF |
AD
A058809

END
DATE
FILMED
11-78
DDC

LEVEL

SIMULTANEOUS CONFIDENCE INTERVALS FOR LINEAR COMBINATIONS OF TWO INVERSE LINEAR REGRESSION PARAMETERS.

(6) (12) B.S.

by

(10) S. Zacks

(9) Technical Report No. 32

(11) 21 Aug [redacted] 78

(12) 22 p.

(14) TR-32

PREPARED UNDER CONTRACT

(15) N[redacted]00014-75-C-0529, PROJECT NR 042-276

OFFICE OF NAVAL RESEARCH

Department of Mathematics and Statistics
CASE WESTERN RESERVE UNIVERSITY
Cleveland, Ohio

1. Introduction.

Consider the linear model $Y = \alpha + \beta x + e$, where x is a deterministic regressor, α and β unknown regression coefficients and e a normal random variable, with zero mean and variance σ^2. For a fixed value, η, $x(\eta) = (\eta - \alpha)/\beta$ is called the inverse (linear) regression parameter. Given n points $(x_1, Y_1), \ldots, (x_n, Y_n)$ one can determine exact γ-level confidence intervals to the inverse regression, on the basis of the least-squares estimators of α and β, by applying the celebrated Fieller's Theorem (Fieller, 1944). Many different applications and extensions of Fieller's Theorem are available in the literature. In a recent issue of the American Statistician, Zerbe (1978) applied Fieller's Theorem to obtain confidence intervals for the ratio of arbitrary linear combinations of the parameters of the general linear model. Zerbe's generalization yields, as a special case, confidence intervals for the reltive potency of dilution bioassays. These confidence limits are obtained from the confidence limits of the difference of two inverse linear regression parameters, which correspond to parallel regression lines. This however, was originally solved by Fieller, in his famous 1944 paper. In the case of non-parallel linear regressions one cannot obtain exact confidence intervals as a special case of Zerbe's method. Formulae available in the literature are generally based on asymptotic approximations (see Armitage (1971), pp. 446-447). In the present paper we develop a method for determining exactly confidence intervals for any linear combination of the corresponding inverse linear regression parameters of two non-parallel lines. The method is based on a simultaneous application of the basic statistic used in Fieller's Theorem to the bivariate case. We obtain

a joint confidence set, at level γ, for the two inverse linear regression parameters. The set is closed and convex, with continuous boundary, similar to an ellipse which has been asymmetrically extended. This simultaneous confidence region yields simultaneous confidence intervals for any linear combination of the parameters, by considering the proper tangential lines. The method is developed in Section 2 in a general manner and reduced by proper transformations to a canonical form. Fortran subroutine programs for determining the upper and lower confidence limits of the difference are provided in Appendix I. In Section 3 we compare the exact simultaneous confidence limits of the difference to the asymptotic limits. The comparison is done numerically by simulating samples of size $n = 50$ and $n = 100$; determining the least-squares estimators of the regression lines, the inverse regression estimates and the confidence limits for the difference of the inverse regression parameters. It is shown that the asymptotic approximations yield intervals which are too short compared to the exact method, even in samples of size $n = 100$. This illustrates that the commonly applied asymptotic formulae lead to intervals with smaller coverage probability than the nominal confidence level.

There is in the literature a considerable disagreement on the question whether the notion of relative-potency is meaningful in non-dilution assays. Cornfield (1964), in his famous paper on the role of parallelism in comparative bioassays, provided a long discussion of the theoretical aspects of this question. Cornfield developed in his paper a Bayesian procedure for determining (posterior) confidence intervals of the log-relative-potency, as functions of the inverse regression of the test

preparation, $x_t(\eta)$. We apply in Section 4 the method developed in the present paper to obtain classical confidence intervals of log-relative-potency of comparative (non-dilution) photodynamic bioassays. These assays were performed by Epstein et al (1965) to test the toxicity of benzen-soluble organic extracts from air pollution samples. The standard preparation consisted of factory made benzo (a) pyrene (BaP). Some description of the assays and the nature of the biological response is given in Section 4. We illustrate the method developed here on actual sample data from Boston, Mass. and Chattanooga, Tenn. For an extensive statistical analysis and new indices based on photodynamic bioassays see Bialik, Epstein and Zacks (1978) and Bialik (1978).

In conclusion, the method developed here of determining exact confidence intervals is shown to be easily applicable. Although it requires computer analysis, the program is simple and the results can be obtained in a matter of seconds on any time-sharing equipment. Programmable calculators can be used too for determining these confidence limits.

2. Derivations.

Consider two regression lines:

$$Y = \alpha_s + \beta_s x + e \qquad \text{(the "standard" line)}$$

and

$$Y = \alpha_t + \beta_t x + e \qquad \text{(the "test" line)}$$

where α_s, β_s, α_t and β_t are the regression coefficients and e is a random variable having a normal distribution, $N(0,\sigma^2)$. It is assumed that the variance, σ^2, around the two lines is the same. n_s independent trials are conducted to estimate the coefficients of the "standard"

line, and n_t independent trials are conducted to estimate the coefficients of the "test" lines. Let $\{x_{s,1},\ldots,x_{s,n_s}\}$ be the points at which the "standard" trials are performed and $\{x_{t,1},\ldots,x_{t,n_t}\}$ those for the "test" trials. We denote by $\bar{x}_s$ and $\bar{x}_t$ the averages of $x_{s,i}$ $(i=1,\ldots,n_s)$ and $x_{t,i}$ $(i=1,\ldots,n_t)$, respectively. Let $SD_s = \sum_{i=1}^{n_s}(x_{s,i} - \bar{x}_s)^2$ and define SD_t similarly. It is assumed that the "standard" and the "test" trials yield mutually <u>independent</u> random variables $Y_{s,i}$ and $Y_{t,j}$ $(i = 1,\ldots,n_s;\ j = 1,\ldots,n_t)$.

Denote by a_s, b_s, a_t and b_t be the least-squares estimators of α_s, β_s, α_t and β_t, respectively. Let $\hat{\sigma}^2$ be the pooled-variance around the least-squares regression lines (see, Graybill (1977) for the theory and formulae of least-squares estimation). Fix η and let $\hat{x}_s$ and $\hat{x}_t$ denote the estimators of the inverse regression parameters $\xi_s = (\eta-\alpha_s)/\beta_s$ and $\xi_t = (\eta-\alpha_t)/\beta_t$, in which the least-squares estimators a_s, b_s; a_t and b_t are substituted, respectively. We define the pivotal variables

$$u_s = (\eta - a_s) - \xi_s b_s = b_s(\hat{x}_s - \xi_s)\,,$$

$$u_t = (\eta - a_t) - \xi_t b_t = b_t(\hat{x}_t - \xi_t)\,. \tag{2.1}$$

It follows immediately from the least-squares theory that u_s and u_t are independent random variables, normally distributed with zero means and variances

$$\mathrm{Var}\{u_s\} = \sigma^2\left(\frac{1}{n_s} + \frac{(\bar{x}_s - \xi_s)^2}{SD_s}\right),$$

$$\mathrm{Var}\{u_t\} = \sigma^2\left(\frac{1}{n_b} + \frac{(\bar{x}_t - \xi_t)^2}{SD_t}\right). \tag{2.2}$$

It follows that,

$$(2.3) \qquad \frac{n_s b_s^2(\hat{x}_s - \xi_s)^2}{1 + \frac{n_s(\bar{x}_s - \xi_s)^2}{SD_s}} + \frac{n_t b_t^2(\hat{x}_t - \xi_t)^2}{1 + \frac{n_t}{SD_t}(\bar{x}_t - \xi_t)^2} \sim \sigma^2 \chi^2[2]$$

where $\chi^2[\nu]$ designates a chi-squared random variable with ν degrees of freedom. The pooled-variance estimator $\hat{\sigma}^2$ is independent of the random variable specified in (2.3) having a distribution like that of $\sigma^2\chi^2[\nu_p]/\nu_p$, where $\nu_p = n_s + n_t - 4$. Let $F_\gamma[2,\nu_p]$ denote the γ-fractile of a (central) F-distribution with 2 and ν_p degrees of freedom. It follows that

$$(2.4) \qquad \frac{b_s^2(\hat{x}_s - \xi_s)^2}{1 + \frac{n_s(\bar{x}_s - \xi_s)^2}{SD_s}} + \frac{\frac{n_t}{n_s} b_t^2(\hat{x}_t - \xi_t)^2}{1 + \frac{n_t(\bar{x}_t - \xi_t)^2}{SD_t}} \leq \frac{2\hat{\sigma}^2}{n_s} F_\gamma[2,\nu_p]$$

holds with probability γ for all $(\alpha_s, \beta_s, \alpha_t, \beta_t, \sigma)$. Inequality (2.4) specifies a γ-level simultaneous confidence region for (ξ_s, ξ_t). Define,

$$(2.5) \qquad \zeta_s = \sqrt{\frac{n_s}{SD_s}}(\bar{x}_s - \xi_s), \qquad \zeta_t = \sqrt{\frac{n_t}{SD_t}}(\bar{x}_t - \xi_t),$$

and

$$(2.6) \qquad \hat{z}_s = \sqrt{\frac{n_s}{SD_s}}(\hat{x}_s - \bar{x}_s), \qquad \hat{z}_t = \sqrt{\frac{n_t}{SD_t}}(\hat{x}_t - \bar{x}_t),$$

and

$$B_s^2 = \frac{SD_s}{n_s} b_s^2 ,$$

(2.7)

$$B_t^2 = \frac{SD_t}{n_s} b_t^2 ,$$

then inequality (2.4) is reduced to the canonical form

$$\frac{B_s^2(\hat{z}_s - \zeta_s)}{1 + \zeta_s^2} + \frac{B_t^2(\hat{z}_t - \zeta_t)^2}{1 + \zeta_t^2} \leq R^2 , \tag{2.8}$$

where $R^2 = 2\hat{\sigma}^2 F_\gamma[2, \nu_p]/n_s$. Since both summands on the LHS of (2.8) are non-negative, it follows that a necessary condition for (ζ_s, ζ_t) to satisfy (2.8) is that $\zeta_s^{(1)} \leq \zeta_s \leq \zeta_s^{(2)}$, where $\zeta_s^{(i)}$ $(i = 1,2)$ are the two real solutions of the quadratic equation

$$B_s^2(\hat{z}_s - \zeta_s)^2 = R^2(1 + \zeta_s^2) . \tag{2.9}$$

These roots, if exist, are given by

$$\zeta_s^{(i)} = \frac{B_s^2 \hat{z}_s}{B_s^2 - R^2} + (-1)^i \frac{R}{B_s^2 - R^2}[B_s^2(1 + \hat{z}_s^2) - R^2]^{\frac{1}{2}} . \tag{2.10}$$

Notice that $R^2 = O(\frac{1}{n_s})$ and therefore, if n_s is sufficiently large, $B_s^2 > R^2$ with high probability. Indeed, $SD_s/n_s = O(1)$ and $b_s^2 \to \beta_s^2 > 0$, with probability one. If the real roots (2.10) do not exist, (2.8) does not hold and one cannot obtain exact confidence limits. The probability of such an event is, however, smaller than $1-\gamma$. For every ζ_s in the interval $(\zeta_s^{(1)}, \zeta_s^{(2)})$, define

$$\Psi_R(\zeta_s; \hat{z}_s) = R^2 - B_s^2(\zeta_s - \hat{z}_s)^2/(1 + \zeta_s^2). \tag{2.11}$$

It follows from (2.8) that the corresponding value of ζ_t, for a point within the confidence region, is between the two limits

$$\zeta_t^{(i)}(\zeta_s) = \frac{B_t^2 \hat{z}_t}{B_t^2 - \Psi_R(\zeta_s;\hat{z}_s)} + (-1)^i[\Psi_R(\zeta_s;\hat{z}_s)\cdot B_t^2 \cdot (1 + \hat{z}_t^2)$$

(2.12)

$$- \Psi_R^2(\zeta_s;\hat{z}_s)]^{\frac{1}{2}} / (B_t^2 - \Psi_R(\zeta_s;\hat{z}_s)), \quad i = 1,2.$$

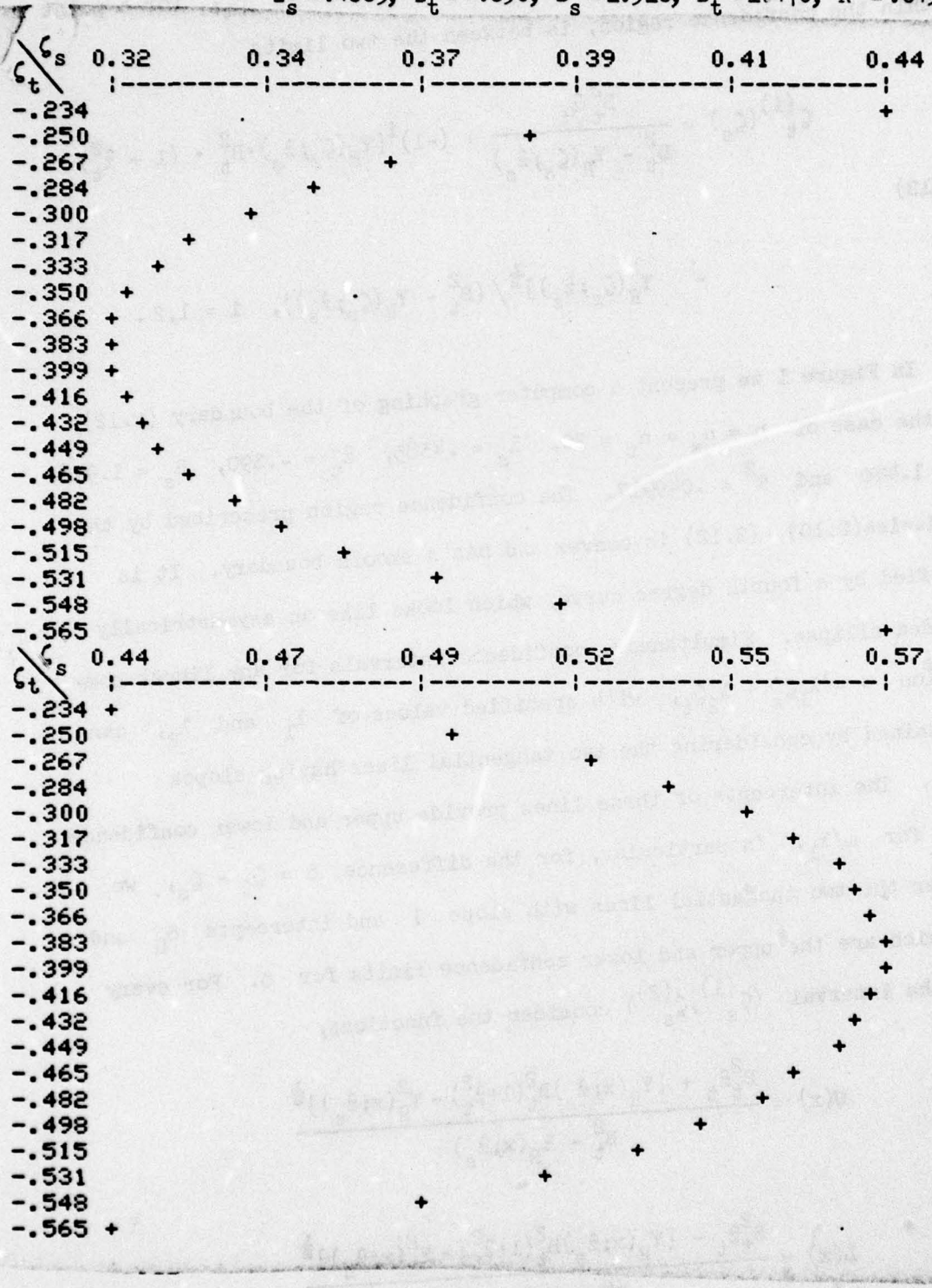
In Figure 1 we present a computer graphing of the boundary (2.12), for the case of $n = n_s = n_t = 20$, $\hat{z}_s = .4385$, $\hat{z}_t = -.390$, $B_s = 1.918$, $B_t = 1.470$ and $R^2 = .049917$. The confidence region prescribed by the boundaries(2.10) - (2.12) is convex and has a smooth boundary. It is specified by a fourth degree curve, which looks like an asymmetrically extended ellipse. Simultaneous confidence intervals for any linear combination $\omega = \lambda_1\zeta_s + \lambda_2\zeta_t$, with specified values of λ_1 and λ_2, can be obtained by considering the two tangential lines having slopes $-\lambda_1/\lambda_2$. The intercepts of these lines provide upper and lower confidence limits for ω/λ_1. In particular, for the difference $\delta = \zeta_t - \zeta_s$, we consider the two tangential lines with slope 1 and intercepts δ_U and δ_L, which are the upper and lower confidence limits for δ. For every x in the interval $(\zeta_s^{(1)},\zeta_s^{(2)})$ consider the functions,

$$U(x) = \frac{B_t^2\hat{z}_t + [\Psi_R(x;\hat{z}_s)B_t^2(1+\hat{z}_t^2) - \Psi_R^2(x;\hat{z}_s)]^{\frac{1}{2}}}{B_t^2 - \Psi_R(x;\hat{z}_s)} \tag{2.13}$$

and

$$L(x) = \frac{B_t^2\hat{z}_t - [\Psi_R(x;\hat{z}_s)B_t^2(1+\hat{z}_t^2) - \Psi_R^2(x;\hat{z}_s)]^{\frac{1}{2}}}{B_t^2 - \Psi_R(x;\hat{z}_s)} . \tag{2.14}$$

FIGURE 1. Computer Graphing Of The Boundary (2.12) For $n_s = n_t = 20$, $\hat{z}_s = .4385$, $\hat{z}_t = -.390$, $B_s = 1.918$, $B_t = 1.470$, $R^2 = .049917$.



These two functions consist of the upper and lower boundary points of the confidence region defined by (2.12). The function $U(x)$ is concave, having a maximum at $x = \hat{z}_s$. Let x_u be the (unique) point at which $\frac{\partial}{\partial x} U(x) = 1$, and x_L the point at which $\frac{\partial}{\partial x} L(x) = 1$. It is easy to verify that

$$\frac{\partial}{\partial x} U(x) = \left\{ \frac{B_t^2 \hat{z}_t}{(B_t^2 - \Psi_R(x;\hat{z}_s))^2} + \right.$$

$$\tfrac{1}{2}[\Psi_R(x;\hat{z}_s)B_t^2(1 + \hat{z}_t^2) - \Psi_R^2(x;\hat{z}_s)]^{-\frac{1}{2}} \cdot$$

$$\cdot \frac{B_t^2(1 + \hat{z}_t^2) - 2\Psi_R(x;\hat{z}_s)}{B_t^2 - \Psi_R(x;\hat{z}_s)} + [\Psi_R(x;\hat{z}_s)B_t^2(1 + \hat{z}_t^2)$$

$$\left. - \Psi_R^2(x;\hat{z}_s)]^{\frac{1}{2}} / (B_t^2 - \Psi_R(x;\hat{z}_s))^2 \right\} \cdot \frac{\partial}{\partial x} \Psi_R(x;\hat{z}_s) \,, \tag{2.15}$$

where

$$\frac{\partial}{\partial x} \Psi_R(x;\hat{z}_s) = \frac{2B_s^2(\hat{z}_s - x)}{(1 + x^2)^2} \,. \tag{2.16}$$

The derivative of $L(x)$ can be expressed similarly, with minus signs before the second and third terms in the curly brackets of (2.15). The solution of the equations $\frac{\partial}{\partial x} U(x) = 1$ and $\frac{\partial}{\partial x} L(x) = 1$ is performed numerically. The upper and lower confidence limits for δ are given by

$$\delta_u = U(x_u) - x_u \,,$$

$$\delta_L = U(x_L) - x_L \,. \tag{2.17}$$

In Appendix I we provide the subroutine functions which determine numerically the values of δ_u and δ_L.

We conclude the present section with some comments about the asymptotic (large sample) confidence intervals. It can be shown (see Zacks (1971), pp. 241) that the asymptotic distribution of $\hat{z}_s$ and $\hat{z}_t$ are normal with means ζ_s and ζ_t, respectively; and asymptotic variances

$$(2.18) \qquad AV\{\hat{z}_s\} = \frac{\sigma^2}{B_s^2 n_s}(1 + \zeta_s^2),$$

and

$$AV\{\hat{z}_t\} = \frac{\sigma^2}{B_t^2 n_s}(1 + \zeta_t^2).$$

Accordingly, the asymptotic confidence limits of $\delta = \zeta_t - \zeta_s$, at nominal level γ, are given by

$$(2.19) \qquad \hat{\delta}_i = (\hat{z}_t - \hat{z}_s) + (-1)^i u_{\frac{1+\gamma}{2}} \frac{\hat{\sigma}}{\sqrt{n_s}} \left[\frac{1 + \hat{z}_s^2}{B_s^2} + \frac{1 + \hat{z}_t^2}{B_t^2} \right]^{\frac{1}{2}}, \quad i = 1, 2$$

where $u_{\frac{1+\gamma}{2}}$ is the $(\frac{1+\gamma}{2})$th fractile of the standard normal distribution. In the following section we compare the behavior of the asymptotic limits with those of the exact ones.

3. Simulation Comparison of the Exact and Asymptotic Confidence Intervals for Differences.

In the present section we illustrate the exact and the asymptotic confidence intervals by performing simulation runs and computing the confidence limits for the simulated values. For the sake of simplicity we consider two-point designs, in which both the "standard" and the "test" are performed at the points $z = -1$ and $z = 1$. Altogether there are $n = 2m$ observations around each regression line, m observations at $z = -1$ and m at $z = 1$. When there are only two regression points the least-squares re-

gression line passes through the sample means corresponding to these points. Thus, we have generated in each case two independent standard normal variates U_{-1}, U_1 and simulated the sample means by

$$\bar{Y}_{-1} = \mu_{-1} + \frac{\sigma}{\sqrt{m}} U_{-1}$$

$$\bar{Y}_1 = \mu_1 + \frac{\sigma}{\sqrt{m}} U_1 \quad .$$

The slope of the regression line is then $b = (\bar{Y}_1 - \bar{Y}_{-1})/2$ and its intercept is $a = (\bar{Y}_1 + \bar{Y}_{-1})/2$.

After computing these estimates for the "standard" and the "test" regression lines we determined $\hat{z}_s = (\eta - a_s)/b_s$, $\hat{z}_t = (\eta - a_t)/b_t$ and the exact confidence intervals, at level $\gamma = .95$, according to the subroutine functions of Appendix I. In our computations we assumed that σ^2 is known and applied the formula $R^2 = \sigma^2 \chi^2_\gamma[2]/n$. Asymptotic confidence limits were computed in each run according to formula (2.19) using the time value of σ instead of the estimator $\hat{\sigma}$. In Tables 1 and 2 we present the results of 50 such simulation runs. Table 2 presents cases with sample sizes twice as large ($n = 100$) as those of Table 1 ($n = 50$). We see that the asymptotic limits yield intervals which are too small, even when $n = 100$. Thus, the asymptotic formulae yield intervals whose coverage probabilities are smaller than the nominal ones. We remark that the computation of 50 runs took about six seconds on a time-sharing (Honeywell GE-430) computer. The determination of the exact confidence limits in each case is a matter of a split of a second.

TABLE 1. .95-Confidence Limits for δ; n = 50; $\alpha_s = 3.5$, $\alpha_t = 3.5$, $\beta_s = 2.5$, $\beta_t = 1.75$, $\eta = 4.0$, $\sigma = .5$.

Run No.	Point Estimates			Exact Limits		Asy. Limits	
	$\hat{z}_s$	$\hat{z}_t$	$\hat{\delta}$	$\hat{\delta}_L$	$\hat{\delta}_u$	$\hat{\hat{\delta}}_L$	$\hat{\hat{\delta}}_u$
1	0.141	0.102	-.039	-.199	0.123	-.168	0.090
2	0.275	0.209	-.066	-.228	0.099	-.197	0.064
3	0.185	0.146	-.039	-.202	0.127	-.170	0.092
4	0.178	0.088	-.090	-.245	0.067	-.214	0.035
5	0.179	0.143	-.036	-.191	0.120	-.160	0.088
6	0.240	0.215	-.025	-.187	0.140	-.155	0.106
7	0.255	0.121	-.134	-.286	0.019	-.255	-.012
8	0.214	0.204	-.009	-.172	0.157	-.141	0.122
9	0.220	0.186	-.034	-.191	0.127	-.161	0.093
10	0.214	0.213	-.002	-.156	0.156	-.126	0.123
11	0.150	0.148	-.002	-.153	0.150	-.123	0.118
12	0.203	0.145	-.058	-.214	0.101	-.183	0.068
13	0.167	0.130	-.037	-.188	0.116	-.158	0.084
14	0.215	0.169	-.046	-.199	0.110	-.169	0.078
15	0.194	0.188	-.005	-.165	0.157	-.134	0.123
16	0.218	0.114	-.103	-.255	0.049	-.225	0.018
17	0.241	0.079	-.162	-.338	0.015	-.303	-.022
18	0.191	0.094	-.097	-.252	0.058	-.221	0.026
19	0.248	0.069	-.178	-.335	-.022	-.303	-.053
20	0.188	0.186	-.002	-.171	0.172	-.139	0.134
21	0.264	0.178	-.086	-.249	0.080	-.217	0.045
22	0.153	0.203	0.051	-.111	0.218	-.081	0.182
23	0.216	0.134	-.083	-.236	0.072	-.205	0.040
24	0.154	0.103	-.051	-.212	0.114	-.181	0.080
25	0.152	0.062	-.090	-.240	0.060	-.210	0.029
26	0.212	0.241	0.029	-.125	0.187	-.095	0.154
27	0.183	0.060	-.123	-.287	0.041	-.254	0.008
28	0.177	0.169	-.008	-.173	0.161	-.141	0.126
29	0.182	0.125	-.057	-.212	0.099	-.182	0.067
30	0.160	0.093	-.068	-.225	0.091	-.194	0.059
31	0.193	0.161	-.032	-.181	0.119	-.152	0.088
32	0.183	0.073	-.110	-.276	0.057	-.242	0.023
33	0.239	0.133	-.106	-.265	0.055	-.234	0.022
34	0.208	0.204	-.004	-.161	0.158	-.131	0.124
35	0.233	0.232	-.001	-.172	0.177	-.140	0.138
36	0.228	0.173	-.055	-.215	0.108	-.183	0.074
37	0.224	0.070	-.154	-.313	0.004	-.281	-.028
38	0.235	0.199	-.037	-.204	0.135	-.172	0.098
39	0.235	0.071	-.165	-.323	-.006	-.291	-.038
40	0.149	0.197	0.048	-.121	0.222	-.088	0.184
41	0.217	0.196	-.021	-.185	0.148	-.153	0.112
42	0.247	0.122	-.124	-.291	0.045	-.258	0.010
43	0.220	0.061	-.160	-.318	-.000	-.287	-.032
44	0.224	0.110	-.114	-.273	0.047	-.241	0.014
45	0.204	0.304	0.100	-.068	0.274	-.036	0.236
46	0.184	0.153	-.030	-.191	0.133	-.160	0.099
47	0.256	0.181	-.076	-.235	0.087	-.204	0.053
48	0.266	0.181	-.085	-.241	0.072	-.210	0.040
49	0.269	0.145	-.124	-.287	0.040	-.255	0.006
50	0.223	0.217	-.006	-.178	0.171	-.145	0.133

TABLE 2. .95-Confidence Limits for δ; $n = 100$, $\alpha_s = 3.5$, $\alpha_t = 3.5$, $\beta_s = 2.5$, $\beta_t = 1.75$, $\eta = 4.0$, $\sigma = .5$.

Run No.	Point Estimates			Exact Limits		Asy. Limits	
	$\hat{z}_s$	$\hat{z}_t$	$\hat{\delta}$	$\hat{\delta}_L$	$\hat{\delta}_u$	$\hat{\delta}_L$	$\hat{\delta}_u$
1	0.168	0.121	-.048	-.134	0.040	-.117	0.022
2	0.241	0.179	-.062	-.150	0.026	-.133	0.008
3	0.192	0.144	-.047	-.135	0.041	-.118	0.024
4	0.188	0.113	-.075	-.160	0.011	-.144	-.006
5	0.189	0.143	-.046	-.131	0.040	-.114	0.023
6	0.221	0.183	-.039	-.126	0.049	-.109	0.032
7	0.230	0.131	-.100	-.184	-.015	-.167	-.032
8	0.207	0.177	-.031	-.118	0.058	-.101	0.040
9	0.211	0.166	-.044	-.131	0.042	-.114	0.025
10	0.208	0.182	-.026	-.111	0.060	-.095	0.043
11	0.173	0.146	-.027	-.110	0.058	-.094	0.041
12	0.201	0.144	-.057	-.143	0.029	-.126	0.012
13	0.181	0.136	-.046	-.129	0.039	-.113	0.022
14	0.208	0.157	-.051	-.135	0.035	-.119	0.018
15	0.197	0.168	-.029	-.115	0.058	-.099	0.041
16	0.210	0.127	-.083	-.167	0.002	-.151	-.015
17	0.222	0.110	-.112	-.203	-.020	-.185	-.039
18	0.195	0.116	-.079	-.164	0.006	-.148	-.011
19	0.225	0.102	-.124	-.209	-.038	-.192	-.055
20	0.193	0.165	-.028	-.118	0.062	-.100	0.044
21	0.234	0.162	-.072	-.160	0.016	-.143	-.002
22	0.174	0.175	0.001	-.086	0.090	-.069	0.072
23	0.209	0.138	-.071	-.156	0.014	-.139	-.003
24	0.174	0.122	-.053	-.140	0.035	-.123	0.018
25	0.174	0.098	-.076	-.159	0.008	-.143	-.009
26	0.206	0.198	-.009	-.094	0.077	-.078	0.060
27	0.191	0.099	-.092	-.179	-.004	-.162	-.021
28	0.188	0.157	-.031	-.119	0.059	-.102	0.041
29	0.190	0.133	-.057	-.143	0.028	-.126	0.011
30	0.178	0.116	-.062	-.148	0.024	-.131	0.007
31	0.196	0.153	-.043	-.126	0.041	-.110	0.024
32	0.190	0.106	-.085	-.173	0.004	-.155	-.014
33	0.222	0.138	-.084	-.170	0.003	-.154	-.014
34	0.204	0.176	-.028	-.114	0.059	-.097	0.042
35	0.218	0.189	-.029	-.119	0.062	-.102	0.044
36	0.215	0.159	-.056	-.143	0.032	-.126	0.014
37	0.213	0.103	-.110	-.196	-.024	-.180	-.041
38	0.219	0.172	-.047	-.136	0.043	-.119	0.025
39	0.219	0.103	-.116	-.202	-.030	-.185	-.047
40	0.172	0.171	-.001	-.090	0.089	-.073	0.071
41	0.209	0.171	-.038	-.126	0.051	-.109	0.033
42	0.226	0.132	-.094	-.182	-.004	-.165	-.022
43	0.211	0.098	-.113	-.199	-.026	-.182	-.043
44	0.213	0.125	-.088	-.174	-.001	-.158	-.018
45	0.202	0.230	0.028	-.061	0.118	-.044	0.100
46	0.191	0.149	-.043	-.130	0.045	-.113	0.027
47	0.231	0.163	-.068	-.154	0.020	-.137	0.002
48	0.237	0.164	-.072	-.157	0.014	-.141	-.004
49	0.238	0.144	-.094	-.181	-.006	-.164	-.024
50	0.213	0.181	-.031	-.121	0.060	-.104	0.041

4. Confidence Intervals for the Relative-Potency in Comparative Bioassays.

In the present section we illustrate the application of our method to the determination of exact confidence limits of the log-relative-potency in comparative bioassays. The experiments described here were performed in 1963 and 1964 (see Epstein et al (1965)). They consisted of a series of photodynamic bioassays in which air-pollution samples from various sites in the U.S.A. were chemically analyzed. The benzo-soluble organic extracts were then diluted in aceton and the substances applied on a Paramecium Caudatum under ultraviolate radiation. The measurement of response was the time required (in minutes) to kill (or deactivate) 90% of 30 cells in the suspension. This observed random variable is called the LT90. Three dosages of the organic extracts were applied (10^{-6}, 10^{-5}, and 10^{-4} [g/ml]). It was empirically shown that $Y = \ell n\,(LT90)$ is approximately normally distributed, and that the expectation of Y is related to the log-dose linearly. Each assay was repeated $m = 4$ times independently. A standard assay was also performed with similar doses of factory made Benzene-a-Peryne (BaP). Also in the standard preparations $Y = \ell n\,(LT90)$ was linearly related to the log-dose of the BaP. The slopes of the regression lines were, however, significantly different. The log-relative-potency at level η is defined as the difference δ between the inverse regression parameters ξ_s and ξ_t. In Table 3 we provide the LT90 values and the least-squares estimates of the regression parameters. We include the "test" preparations from assays performed with the 1963-air-pollution data from Boston, Mass. and Chattanooga, Tenn., and corrsponding "standard" preparations. In Table 4 we provide the exact .95-confidence limits for the log-relative-potency, corresponding to LT90 = 30(2) 50. These confidence

limits were computed according to the subroutine functions of Appendix I, after reducing the statistics to the canonical ones. We see in Table 4 that corresponding to LT90 of 40 minutes, the standard preparation of BaP is at least a hundred thousands times more potent that the organic extracts from Boston and at least two million times more potent than those from Chattanooga. The same method can be used to compute the confidence limits of the relative potency of the organic extracts from Chattanooga compared to Boston. These confidence limits are presented in Table 5.

TABLE 3. The LT90 Values and the ℓn (LT90) to log-dose Regression Estimates, for the Photodynamic Bioassays.

	Boston, Mass.			Chattanooga, Tn.			BaP		
dose [g/ml]	10^{-4}	10^{-5}	10^{-6}	10^{-4}	10^{-5}	10^{-6}	10^{-4}	10^{-5}	10^{-6}
LT90 (min)	10.91	25.86	85.62	14.93	37.42	85.23	5.97	7.41	9.37
	14.38	24.59	83.21	15.23	38.21	82.41	5.68	6.83	8.97
	14.22	28.53	86.50	15.68	38.78	83.16	6.22	7.38	9.64
	14.68	28.81	87.04	15.79	39.01	84.02	7.04	7.64	9.80
a	3.4468 ± .06324			3.6029 ± .03162			2.0199 ± .03162		
b	.9251 ± .07746			.8463 ± .03873			.2096 ± .03873		
$\hat{\sigma}^2$	.012			.003			*) -		

*) $\hat{\sigma}^2$ is the pooled estimate of the variance around the "standard" and the "test" regression lines.

TABLE 4. .95-Confidence Intervals of the Log-Relative Potency of the Standard (BaP) Against the Organic Extracts.

LT90	Boston		Chattanooga	
	Lower	Upper	Lower	Upper
30.0000	4.3784	13.5967	5.5451	8.9085
32.0000	4.5138	14.1597	5.7176	9.2310
34.0000	4.6398	14.6871	5.8798	9.5413
36.0000	4.7588	15.1839	6.0328	9.8334
38.0000	4.8720	15.6530	6.1778	10.1086
40.0000	4.9799	16.0960	6.3157	10.3688
42.0000	5.0804	16.5135	6.4473	10.6153
44.0000	5.1776	16.9221	6.5735	10.8492
46.0000	5.2708	17.3150	6.6906	11.0749
48.0000	5.3586	17.6859	6.8028	11.2952
50.0000	5.4442	18.0355	6.9111	11.5060

TABLE 5. .95-Confidence Limits to the Relative Potency of the Organic Extracts from Chattanooga Compared to Boston.

LT90	Lower Limit	Upper Limit
30	.471	.883
36	.494	.918
42	.507	.963
50	.513	1.032

Appendix I

We present here subroutine functions, in FORTRAN, for the computation of the upper and lower confidence limits of the difference $\delta = \zeta_t - \zeta_s$. The input variables are: $A \leftarrow B_s^2$, $B \leftarrow B_t^2$, $x \leftarrow \hat{z}_s$, $Y \leftarrow \hat{z}_t$ and $R \leftarrow R$. The confidence level, sample size, and the value of $\hat{\sigma}^2$ are all determining the value of R according ot the formula given in Section 2.

SUBROUTINE FUNCTION FOR THE UPPER CONFIDENCE LIMIT OF $\delta = \zeta_t - \zeta_s$.

```
100         FUNCTION WU(A,B,X,Y,R)
110         C=A
120         E=B
130         U=X
140         V=Y
150         Q=R
160         QS=Q*Q
170         DS=QS*(C*(1.+U*U)-QS)
180         HL=(C*U-SQRT(DS))/(C-QS)
190         HU=(C*U+SQRT(DS))/(C-QS)
200         D=.01
210         X1=HL+D
220       1 Y1=DU(X1,C,E,U,V,Q)
230         IF(1.-Y1) 2,3,3
240       2 X1=X1+D
250         GO TO 1
260       3 XU=X1-D/2.
270         FR=QS-C*(U-XU)*(U-XU)/(1.+XU*XU)
280         FXU=E*V/(E-FR)
290         GXU=FR*E*(1.+V*V)-FR*FR
300         GXU=SQRT(GXU)/(E-FR)
310         FXU=FXU+GXU
320         WU=FXU-XU
330         RETURN
340         END
350         FUNCTION DU(H,A,B,X,Y,R)
360         W=H
370         C=A
380         E=B
390         U=X
400         V=Y
410         Q=R
420         FR=Q*Q-C*(U-W)*(U-W)/(1.+W*W)
430         EFR=2.*C*(U-W)*(1.+W*U)/((1.+W*W)**2)
440         GFR1=E*V/((E-FR)**2)
450         SFR=FR*E*(1.+V*V)-FR*FR
460         GFR2=1./SQRT(SFR)
470         GFR2=(E*(1.+V*V)-2.*FR)*GFR2/(2.*(E-FR))
480         GFR3=SQRT(SFR)/((E-FR)**2)
490         DU=(GFR1+GFR2+GFR3)*EFR
500         RETURN
510         END
```

SUBROUTINE FUNCTION FOR THE LOWER CONFIDENCE LIMIT OF $\delta = \zeta_t - \zeta_s$.

```
100         FUNCTION WL(A,B,X,Y,R)
110         C=A
120         E=B
130         U=X
140         V=Y
150         Q=R
160         QS=Q*Q
170         DS=QS*(C*(1.+U*U)-QS)
180         HL=(C*U-SQRT(DS))/(C-QS)
190         HU=(C*U+SQRT(DS))/(C-QS)
200         D=.01
210         X1=HU-D
220       1 Y1=DL(X1,C,E,U,V,Q)
230         IF(1.-Y1) 2,3,3
240       2 X1=X1-D
250         GO TO 1
260       3 XU=X1+D/2.
270         FR=QS-C*(U-XU)*(U-XU)/(1.+XU*XU)
280         FXU=E*V/(E-FR)
290         GXU=FR*E*(1.+V*V)-FR*FR
300         GXU=SQRT(GXU)/(E-FR)
310         FXU=FXU-GXU
320         WL=FXU-XU
330         RETURN
340         END
350         FUNCTION DL(H,A,B,X,Y,R)
360         W=H
370         C=A
380         E=B
390         U=X
400         V=Y
410         Q=R
420         FR=Q*Q-C*(U-W)*(U-W)/(1.+W*W)
430         EFR=2.*C*(U-W)*(1.+W*U)/((1.+W*W)**2)
440         GFR1=E*V/((E-FR)**2)
450         SFR=FR*E*(1.+V*V)-FR*FR
460         GFR2=1./SQRT(SFR)
470         GFR2=(E*(1.+V*V)-2.*FR)*GFR2/(2.*(E-FR))
480         GFR2=-GFR2
490         GFR3=SQRT(SFR)/((E-FR)**2)
500         GFR3=-GFR3
510         DL=(GFR1+GFR2+GFR3)*EFR
520         RETURN
530         END
```

References.

[1] Armitage, P., Statistical Methods in Medical Research, Wiley Interscience, New York, 1971.

[2] Bialik, O., Statistical Methods for the Analysis of Photodynamic Bioassay and Indexing Airpollution Toxocity. Ph.D. Dissertation, Department of Mathematics and Statistics, Case Western Reserve University, 1978.

[3] Bialik, O., Epstein, S.S. and Zacks, S., Indexing of Air Pollution Based on Photodynamic Activity of Organic Pollutants, 1978 (forthcoming).

[4] Cornfield, J., Comparative Bioassay and the Role of Parallelism, Jour. Pharmacology and Experimental Therapeutics, 25, No. 4, (1964), 617-657.

[5] Epstein, S.S, Saroicki, E. and Falk, H.L., Photodynamic Bioassay of Polycyclic Atmospheric Pollutants, Jour. Air Pollution Control, 15, No. 4, (1965), 174-176.

[6] Fieller, E.C., A Fundamental Formula in the Statistics of Biological Assay and Some Applications, Quarterly Jour. of Pharmacy and Pharmacology, 17, (1944), 117-213.

[7] Graybill, F.A., Theory and Applications of the Linear Model, Duxbury Press, Massachusetts, 1976.

[8] Zacks, S., The Theory of Statistical Inference, John Wiley, New York, 1971.

[9] Zerbe, G.O., On Fieller's Theorem and the General Linear Model, The American Statistician, 32, No. 3, (1978), 103-105.

exact confidence limits are compared numerically to asymptotic confidence limits in order to illustrate the deficiency of the method based on asymptotic formula. An application to comparative (non-dilution) bio-assays is shown too.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #32	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SIMULTANEOUS CONFIDENCE INTERVALS FOR LINEAR COMBINATIONS OF TWO INVERSE REGRESSION PARAMETERS		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) S. Zacks		8. CONTRACT OR GRANT NUMBER(s) NR 00014-75-C-0529 PROJECT NR 042-276
9. PERFORMING ORGANIZATION NAME AND ADDRESS DEPARTMENT OF MATHEMATICS AND STATISTICS CASE WESTERN RESERVE UNIVERSITY		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS OFFICE OF NAVAL RESEARCH ARLINGTON, VIRGINIA 22217		12. REPORT DATE August 21, 1978
		13. NUMBER OF PAGES 20
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION OF DOCUMENT UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Fieller's Theorem, inverse regressions, simultaneous confidence intervals, compurative bioassays, relative-potency, air pollution, toxicity of pollutants.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Confidence intervals for the inverse linear regression parameter can be obtained from the least-squares estimators of the regression coefficients by applying the celebrated Fieller Theorem. In the present paper we develop a method based on the simultaneous application of Fieller's Theorem to obtain simultaneous confidence intervals to any linear combination of the inverse regressions of two independent regression lines, which are not parallel. Computer subroutine functions are provided to assist in actual computations. The

(continued on the back)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)