Biometrical Journal **44** (2002) 7, 801–812

Exact Multiple Comparisons of Three or More Regression Lines: Pairwise Comparisons and Comparisons with a Control

JOHN D. SPURRIER*

Department of Statistics, University of South Carolina, Columbia, SC 29208 U.S.A.

Summary

The problem of finding exact simultaneous confidence bounds for differences in regression models for k groups via the union-intersection method is considered. The error terms are taken to be iid normal random variables. Under an assumption slightly more general than having identical design matrices for each of the k groups, it is shown that an existing probability point for the multivariate studentized range can be used to find the necessary probability point for pairwise comparisons of regression models. The resulting methods can be used with simple or multiple regression. Under a weaker assumption on the k design matrices that allows more observations to be taken from the control group than from the $k-1$ treatment groups, a method is developed for computing exact probability points for comparing the simple linear regression models of the $k-1$ groups to that of the control. Within a class of designs, the optimal design for comparisons with a control takes the square root of $(k-1)$ times as many observations from the control than from each treatment group. The simultaneous confidence bounds for all pairwise differences and for comparisons with a control are much narrower than Spurrier's intervals for all contrasts of k regression lines.

Key words: Confidence bound; Multivariate studentized range; Simultaneous confidence; Simple linear regression; Union-intersection.

1. Introduction

There is a wealth of literature on comparing the means of 3 or more groups under the assumption of iid normal errors. HOCHBERG and TAMHANE (1987) and HSU (1996) give excellent summaries of this literature. At times, one wishes to compare groups based on some parametric function other than the mean. For example, MASUDA, SAITO, and INUI (1997) studied the effects of three treatments (indomethacin, ketoprofen, and a control) on the accumulation of the chemotherapy agent [^{3}H]methotrexate in rat renal tissue as a function of time. The data suggest a linear relationship between accumulation and time for each treatment in the time interval used in the study. Thus, one can compare the three treatments by performing multiple comparisons of the three regression models.

* Corresponding author: spurrier@stat.sc.edu

 0323-3847/02/0710-0801 $ 17.50+.50/0

Another example involves the degradation of drugs over time. It may be of interest to compare the relationship between drug content and storage time for different batches, different types of storage container, or different formulations of a drug. RUBERG and STEGEMAN (1991) discuss the need for comparing batches in the determination of drug shelf-life.

There is little literature on the multiple comparison of regression models. ROYEN (1990) used the example of all pairwise comparisons of $k \geq 3$ linear regression models, under the assumptions of iid normal errors and the use of the same design matrix for each group, as motivation for studying the distribution of the multivariate studentized range. SPURRIER (1999) used the union-intersection method to find exact simultaneous intervals for all contrasts of $k \geq 3$ simple linear regression models under the same assumptions. Spurrier also showed that the SCHEFFÉ (1959) method is conservative when comparing all contrasts of $k \geq 3$ regression lines.

LIU, JAMSHIDIAN, and ZHANG (2001) developed a simulation based method for approximating the probability point for simultaneous comparison of regression models. Their method allows one to consider a subset of contrasts and to restrict the predictor variable to a finite interval. These actions result in narrower confidence bounds. They also allow for more than one predictor variable and for different design matrices for each treatment. Although the LIU et al. method has many advantages, it has the disadvantages that the intervals are approximate rather than exact due to the simulation and that the simulation must be repeated for each application. In this article, I develop exact (up to the accuracy of the numerical methods) simultaneous bounds for all pairwise comparisons of regression lines and for comparisons with a control under certain restrictions on the design matrices and no restrictions on the predictor variable.

Denote the model for the n_i observations from the ith group by

$$Y_{ij} = \alpha_i + \beta_i x_{ij} + \varepsilon_{ij} \tag{1}$$

for $i = 1, \ldots, k$, $j = 1, \ldots, n_i$. Assume that the error terms are iid $N(0, \sigma^2)$ and that there are at least two distinct x_{ij} values for each group. The $n_i \times 2$ design matrix $\boldsymbol{X}_i$ for the ith group has the first column $(1, \ldots, 1)'$ and second column $(x_{i1}, \ldots, x_{in_i})'$, $i = 1, \ldots, k$. Let

$$\boldsymbol{A}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} = \left\{ \begin{matrix} a_{i11} & a_{i12} \\ a_{i12} & a_{i22} \end{matrix} \right\}, \tag{2}$$

$i = 1, \ldots, k$.

Let $\hat{\alpha}_i$ and $\hat{\beta}_i$ denote the least squares estimators of α_i and β_i, $i = 1, \ldots, k$, and let $\hat{\sigma}^2$ denote the pooled error mean square with $\nu = n_1 + \ldots + n_k - 2k$ degrees of freedom. The vectors $(\hat{\alpha}_i - \alpha_i, \hat{\beta}_i - \beta_i)'$ are independent $N(\boldsymbol{0}, \sigma^2\boldsymbol{A}_i)$, $i = 1, \ldots, k$. The variable $\nu\hat{\sigma}^2/\sigma^2$ has a chi−squared distribution with ν degrees of freedom and is independent of the least squares estimators.

The exact simultaneous confidence intervals for comparing pairs of regression lines are derived using the pivotal variables

$$F_{ij} = \{[(\hat{\alpha}_i - \hat{\alpha}_j) - (\alpha_i - \alpha_j), (\hat{\beta}_i - \hat{\beta}_j) - (\beta_i - \beta_j)]\,(\boldsymbol{A}_i + \boldsymbol{A}_j)^{-1} \\ \times [(\hat{\alpha}_i - \hat{\alpha}_j) - (\alpha_i - \alpha_j), (\hat{\beta}_i - \hat{\beta}_j) - (\beta_i - \beta_j)]'\}/(2\hat{\sigma}^2) \tag{3}$$

for $1 \leq i < j \leq k$. Let C denote the set of (i, j) pairs for which it is desired to compare the regression line of group i to that of group j. For all pairwise comparisons, $C = \{(i, j), 1 \leq i < j \leq k\}$ and for comparisons with a control (group k), $C = \{(i, j), i = 1, \ldots, k-1, j = k\}$.

For fixed $0 < \alpha < 1$, let $c > 0$ be a constant such that

$$P[F_{ij} \leq c^2/2, \quad \text{for all} \quad (i, j) \in C] = P(F^* \leq c^2/2) = 1 - \alpha\,, \tag{4}$$

where F^* denotes the maximum F_{ij} over all $(i, j) \in C$. By inverting equation (4) using methods analogous to WORKING and HOTELLING (1929) and SCHEFFÉ (1959), we get the $100(1-\alpha)\%$ simultaneous confidence intervals

$$(\alpha_i + \beta_i x) - (\alpha_j + \beta_j x) \in (\hat{\alpha}_i + \hat{\beta}_i x) - (\hat{\alpha}_j + \hat{\beta}_j x) \pm c\hat{\sigma}[(1, x)\,(\boldsymbol{A}_i + \boldsymbol{A}_j)\,(1, x)']^{1/2} \tag{5}$$

for all $-\infty < x < \infty$ and all $(i, j) \in C$. It remains to determine the value of c which depends on $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k$, C, ν, and α. Dividing the interval endpoints in (5) by x and taking limits as $x \to \infty$ yields confidence intervals for the difference of slope parameters, $\beta_i - \beta_j$, $(i, j) \in C$.

In Section 2, existing probability points for the multivariate studentized range are used to form simultaneous confidence bounds for all pairwise comparisons of linear regression models under the assumption of $\boldsymbol{A}_1 = \ldots = \boldsymbol{A}_k$. These results also hold for multiple regression. This is the confidence interval dual of the simultaneous testing problem considered by ROYEN (1990). In Section 3, I find exact probability points for comparing $k - 1$ simple linear regressions to a control (group k) regression under the assumption $\boldsymbol{A}_1 = \ldots = \boldsymbol{A}_{k-1} = b\boldsymbol{A}_k$, where $b > 0$ is a constant. In Section 4, a computational method is sketched for pairwise comparisons and for comparisons with a control under no restrictions on the $\boldsymbol{A}_i$ matrices when $k = 3$. A comparison of the bounds in Sections 2 and 3 with the bounds of SPURRIER (1999) and a discussion of optimal designs for comparison with a control are given in Section 5. The Masuda et al. data are analyzed in Section 6. Concluding remarks are given in Section 7.

2. All Pairwise Comparison of Regression Lines

In this section, I will show that if we consider all pairwise comparisons of regression lines and if

$$\boldsymbol{A}_1 = \ldots = \boldsymbol{A}_k, \tag{6}$$

then the probability point c can be determined from existing tables for the multivariate studentized range. The result will be shown for the simple linear regression model in (1) but is also true for comparing multiple regressions models.

An equivalent formulation of condition (6) is

$$n_1 = \ldots = n_k\,, \qquad \sum_{j=1}^{n_1} x_{1j} = \ldots = \sum_{j=1}^{n_k} x_{kj}\,, \quad \text{and} \quad \sum_{j=1}^{n_1} x_{1j}^2 = \ldots = \sum_{j=1}^{n_k} x_{kj}^2\,. \tag{7}$$

One way to satisfy these conditions is to use the same design matrix for each group.

Let $\varrho = a_{k12}/[a_{k11}a_{k22}]^{1/2}$, and for $i = 1, \ldots, k$, let

$$Z_{i1} = \frac{\hat{\alpha}_i - \alpha_i}{\sigma(a_{k11})^{1/2}}$$

and

$$Z_{i2} = \frac{-\varrho(\hat{\alpha}_i - \alpha_i)}{(1-\varrho^2)^{1/2}\,\sigma(a_{k11})^{1/2}} + \frac{\hat{\beta}_i - \beta_i}{(1-\varrho^2)^{1/2}\,\sigma(a_{k22})^{1/2}}\,. \tag{8}$$

It follows that $Z_{11}, Z_{12}, \ldots, Z_{k1}, Z_{k2}$ are iid $N(0,1)$.

Define the 2×2 matrix

$$B = \begin{Bmatrix} (a_{k11})^{1/2} & 0 \\ \varrho(a_{k22})^{1/2} & [(1-\varrho^2)\,a_{k22}]^{1/2} \end{Bmatrix}. \tag{9}$$

Under condition (6),

$$\begin{aligned} F_{ij} &= (Z_{i1} - Z_{j1}, Z_{i2} - Z_{j2})\,\boldsymbol{B}'(2\boldsymbol{A}_k)^{-1}\,\boldsymbol{B}(Z_{i1} - Z_{j1}, Z_{i2} - Z_{j2})'/(2\hat{\sigma}^2/\sigma^2) \\ &= (Z_{i1} - Z_{j1}, Z_{i2} - Z_{j2})\,(Z_{i1} - Z_{j1}, Z_{i2} - Z_{j2})'/(4\hat{\sigma}^2/\sigma^2)\,. \end{aligned} \tag{10}$$

Thus, for the class of all pairwise comparisons,

$$F^* = \max_{i \neq j}\,[|(Z_{i1}, Z_{i2})' - (Z_{j1}, Z_{j2})'|^2]/(4\hat{\sigma}^2/\sigma^2) = (Q_{2,k,\nu})^2/4\,, \tag{11}$$

where $Q_{2,k,\nu}$ is the bivariate studentized range. See ROYEN (1991) for a discussion of the multivariate studentized range.

ROYEN (1990) performed a massive simulation study using 10^8 replications to approximate $q(\alpha; d; k; \nu)$, the upper α probability point of the multivariate studentized range where d is the dimension of the normal vectors. In our case $d = 2$. It follows from equations (4) and (11) that the probability point

$$c = q(\alpha;\, 2;\, k;\, \nu)/(2)^{1/2}. \tag{12}$$

An analogous development follows for multiple regression with no restriction on the values of the predictor variables. In this case, the analog of model (1) involves

d regression coefficients. The $c^2/2$ term in equation (4) becomes c^2/d. If the analog of equation (6) holds, for example if the same design matrix is used for each group, then the analog of equation (11) becomes $F^* = (Q_{d,k,\nu})^2/(2d)$, where $Q_{d,k,\nu}$ is the d-variate studentized range. The $100(1-\alpha)\%$ simultaneous confidence bounds for all pairwise differences of the k regression models are the analog of display (5) with

$$c = q(\alpha; d; k; \nu)/(2)^{1/2}. \tag{13}$$

If there are functional relationships among the predictor variables, as for example in polynomial regression, then these simultaneous intervals are conservative.

It follows from ROYEN (1984, 1990, 1991) that the distribution function of $Q_{d,k,\nu}$ can be written as a multi-dimensional integral when $k = 3$. The integral is difficult to evaluate numerically if $d > 3$.

3. Comparison of $k-1$ Regression Lines with a Control

Let us now find an algorithm for computing c for comparing the regression models in (1) for groups $i = 1, \ldots k-1$ to the model for group k, the control, under the assumption

$$\boldsymbol{A}_1 = \ldots = \boldsymbol{A}_{k-1} = b\boldsymbol{A}_k, \tag{14}$$

where $b > 0$ is a constant. Condition (14) is equivalent to

$$bn_1 = \ldots = bn_{k-1} = n_k,\ b\sum_{j=1}^{n_1} x_{1j} = \ldots = b\sum_{j=1}^{n_{k-1}} x_{(k-1)j} = \sum_{j=1}^{n_k} x_{kj}\,,$$
$$\text{and}\quad b\sum_{j=1}^{n_1} x_{1j}^2 = \ldots = b\sum_{j=1}^{n_{k-1}} x_{(k-1)j}^2 = \sum_{j=1}^{n_k} x_{kj}^2\,. \tag{15}$$

One way to satisfy condition (14) is to use the same design matrix for groups 1 to $k-1$ and have the group k design matrix contain b copies of each row of the design matrix used for the other groups. If $b > 1$, one takes more observations from the control group.

Define the variables $U = \hat{\sigma}^2/\sigma^2$ and

$$W_{i1} = \frac{\hat{\alpha}_i - \alpha_i}{\sigma(a_{i11})^{1/2}}\,,$$
$$W_{i2} = \frac{-\varrho(\hat{\alpha}_i - \alpha_i)}{(1-\varrho^2)^{1/2}\,\sigma(a_{i11})^{1/2}} + \frac{\hat{\beta}_i - \beta_i}{(1-\varrho^2)^{1/2}\,\sigma(a_{i22})^{1/2}}\,, \qquad i = 1, \ldots, k\,. \tag{16}$$

The variables $W_{11}, W_{12}, \ldots, W_{k2}$ are iid $N(0,1)$. The variable U is distributed as a chi-squared variable with ν degrees of freedom divided by ν and is independent of the W variables. Under condition (14), for $i = 1, \ldots, k-1$ we have from equations (3) and (16)

$$\begin{aligned} F_{ik} &= \{[(\hat{\alpha}_i - \hat{\alpha}_k) - (\alpha_i - a_k),\, (\hat{\beta}_i - \hat{\beta}_k) - (\beta_i - \beta_k)]\,[(b+1)\,\boldsymbol{A}_k]^{-1} \\ &\qquad \times [(\hat{\alpha}_i - \hat{\alpha}_k) - (\alpha_i - \alpha_k),\, (\hat{\beta}_i - \hat{\beta}_k) - (\beta_i - \beta_k)]'\}/(2\hat{\sigma}^2) \\ &= [b^{1/2}W_{i1} - W_{k1},\, b^{1/2}W_{i2} - W_{k2}]\,\boldsymbol{B}'(\boldsymbol{A}_k)^{-1}\,\boldsymbol{B} \\ &\qquad \times [b^{1/2}W_{i1} - W_{k1},\, b^{1/2}W_{i2} - W_{k2}]'/[2(b+1)\,\hat{\sigma}^2/\sigma^2] \\ &= \frac{(W_{i1} - W_{k1}/b^{1/2})^2 + (W_{i2} - W_{k2}/b^{1/2})^2}{2(b+1)\,U/b}\,. \end{aligned}$$

Substituting equation (17) into equation (4) yields

$$\begin{aligned} 1 - \alpha &= P(F^* \le c^2/2) \\ &= P[(W_{i1} - W_{k1}/b^{1/2})^2 + (W_{i2} - W_{k2}/b^{1/2})^2 \le (b+1)\,Uc^2/b, \\ &\qquad i = 1, \ldots, k-1]\,. \end{aligned} \tag{18}$$

Conditioning on $W_{k1} = w_{k1}$, $W_{k2} = w_{k2}$, and $U = u$ yields

$$\begin{aligned} 1 - \alpha = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty & P[(W_{i1} - w_{k1}/b^{1/2})^2 + (W_{i2} - w_{k2}/b^{1/2})^2 \\ & \qquad \le (b+1)\,uc^2/b,\ i = 1, \ldots, k-1] \\ & \times \varphi(w_{k1})\,\varphi(w_{k2})\,g(u)\,\mathrm{d}w_{k1}\,\mathrm{d}w_{k2}\,\mathrm{d}u\,, \end{aligned} \tag{19}$$

where φ is the standard normal density function and g is the density function of U.

Let h_λ and H_λ denote the density and distribution function of a noncentral chi-squared random variable with two degrees of freedom and noncentrality parameter λ. Let

$$R = W_{k1}^2 + W_{k2}^2 \quad \text{and} \quad r = w_{k1}^2 + w_{k2}^2\,. \tag{20}$$

Using the independence of the W variables and relationships between normal and chi-squared variables, we have

$$\begin{aligned} 1 - \alpha &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \{H_{r/b}[(b+1)\,uc^2/b]\}^{k-1}\,\varphi(w_{k1})\,\varphi(w_{k2})\,g(u)\,\mathrm{d}w_{k1}\,\mathrm{d}w_{k2}\,\mathrm{d}u \\ &= \int_0^\infty \int_0^\infty \{H_{r/b}[(b+1)\,uc^2/b]\}^{k-1}\,h_0(r)\,g(u)\,\mathrm{d}r\,\mathrm{d}u\,. \end{aligned} \tag{21}$$

Making the transformation $T = \exp(-R/2)$ yields

$$1 - \alpha = \int_0^\infty \int_0^1 \{H_{-2\ln(t)/b}[(b+1)\,uc^2/b]\}^{k-1}\,g(u)\,\mathrm{d}t\,\mathrm{d}u\,. \tag{22}$$

Table 1

Values of c for 95% Simultaneous Confidence Bounds for Comparing $k-1$ Regression Lines to a Control Under Condition (14) with $b = 1$, 1.5, 2, and 3

b	n_1	n_k	k					
			3	4	5	6	7	8
1	3	3	5.136	4.648	4.380	4.214	4.103	4.024
	4	4	3.657	3.594	3.557	3.535	3.523	3.517
	5	5	3.289	3.310	3.325	3.338	3.351	3.364
	6	6	3.123	3.179	3.215	3.244	3.269	3.290
	7	7	3.029	3.103	3.152	3.189	3.220	3.247
	8	8	2.969	3.054	3.111	3.154	3.189	3.218
	9	9	2.927	3.020	3.081	3.128	3.166	3.198
	10	10	2.896	2.994	3.060	3.109	3.149	3.183
	11	11	2.872	2.975	3.043	3.095	3.136	3.171
	12	12	2.853	2.959	3.030	3.083	3.126	3.162
	13	13	2.838	2.946	3.019	3.074	3.118	3.154
	14	14	2.825	2.936	3.010	3.066	3.111	3.148
	15	15	2.815	2.927	3.002	3.059	3.105	3.143
	16	16	2.805	2.919	2.996	3.054	3.100	3.138
	17	17	2.798	2.913	2.990	3.049	3.095	3.134
	18	18	2.791	2.907	2.986	3.044	3.092	3.131
	19	19	2.785	2.902	2.981	3.041	3.088	3.128
	20	20	2.780	2.898	2.977	3.037	3.085	3.125
	25	25	2.760	2.882	2.963	3.025	3.074	3.115
	30	30	2.748	2.871	2.954	3.017	3.067	3.109
	35	35	2.739	2.864	2.948	3.012	3.062	3.104
	40	40	2.733	2.859	2.944	3.008	3.059	3.101
	45	45	2.728	2.854	2.940	3.005	3.056	3.099
	50	50	2.724	2.851	2.937	3.002	3.054	3.097
	55	55	2.721	2.849	2.935	3.000	3.052	3.095
	60	60	2.718	2.846	2.933	2.998	3.051	3.094
1.5	4	6	3.394	3.447	3.470	3.485	3.495	3.505
	6	9	3.043	3.139	3.198	3.241	3.275	3.304
	8	12	2.927	3.038	3.108	3.161	3.202	3.237
	10	15	2.870	2.987	3.064	3.121	3.166	3.203
	12	18	2.835	2.957	3.037	3.096	3.144	3.183
	14	21	2.813	2.937	3.019	3.080	3.129	3.170
	16	24	2.796	2.922	3.006	3.069	3.119	3.160
	18	27	2.784	2.912	2.997	3.060	3.111	3.153
	20	30	2.775	2.903	2.989	3.054	3.105	3.148
	30	45	2.748	2.880	2.968	3.035	3.088	3.132
	40	60	2.735	2.868	2.958	3.026	3.080	3.124
	50	75	2.728	2.862	2.952	3.021	3.075	3.120
2	3	6	3.690	3.775	3.805	3.814	3.816	3.814
	4	8	3.245	3.347	3.402	3.437	3.462	3.482
	5	10	3.076	3.188	3.254	3.300	3.335	3.364
	6	12	2.988	3.105	3.178	3.230	3.270	3.303
	7	14	2.933	3.054	3.131	3.187	3.230	3.266
	8	16	2.896	3.020	3.100	3.158	3.204	3.241

Table 1. Continued

b	n_1	n_k	k					
			3	4	5	6	7	8
	9	18	2.869	2.995	3.077	3.137	3.184	3.224
	10	20	2.849	2.977	3.060	3.122	3.170	3.210
	11	22	2.833	2.962	3.047	3.109	3.159	3.200
	12	24	2.820	2.950	3.036	3.099	3.150	3.191
	13	26	2.810	2.941	3.027	3.091	3.142	3.184
	14	28	2.801	2.933	3.020	3.085	3.136	3.179
	15	30	2.794	2.926	3.014	3.079	3.131	3.174
	16	32	2.787	2.920	3.008	3.074	3.126	3.170
	17	34	2.782	2.915	3.004	3.070	3.122	3.166
	18	36	2.777	2.911	3.000	3.066	3.119	3.163
	19	38	2.773	2.907	2.996	3.063	3.116	3.160
	20	40	2.769	2.903	2.993	3.060	3.113	3.158
	25	50	2.755	2.890	2.981	3.049	3.103	3.148
	30	60	2.746	2.882	2.973	3.042	3.097	3.142
	35	70	2.739	2.876	2.968	3.037	3.092	3.138
	40	80	2.735	2.872	2.964	3.034	3.089	3.135
	50	100	2.728	2.866	2.959	3.029	3.085	3.131
3	3	9	3.322	3.474	3.559	3.613	3.649	3.675
	4	12	3.083	3.222	3.305	3.361	3.403	3.436
	5	15	2.977	3.114	3.199	3.259	3.305	3.341
	6	18	2.918	3.055	3.142	3.204	3.252	3.291
	7	21	2.880	3.017	3.106	3.169	3.219	3.260
	8	24	2.854	2.991	3.081	3.146	3.197	3.238
	9	27	2.834	2.972	3.062	3.128	3.180	3.223
	10	30	2.819	2.958	3.048	3.115	3.168	3.211
	11	33	2.808	2.946	3.037	3.105	3.158	3.202
	12	36	2.798	2.937	3.029	3.097	3.150	3.194
	13	39	2.790	2.929	3.021	3.090	3.144	3.188
	14	42	2.784	2.923	3.015	3.084	3.138	3.183
	15	45	2.778	2.917	3.010	3.079	3.134	3.179
	16	48	2.773	2.913	3.006	3.075	3.130	3.175
	17	51	2.769	2.908	3.002	3.071	3.126	3.172
	18	54	2.765	2.905	2.998	3.068	3.123	3.169
	19	57	2.762	2.902	2.995	3.065	3.120	3.166
	20	60	2.759	2.899	2.993	3.062	3.118	3.164
	25	75	2.748	2.888	2.982	3.053	3.109	3.156
	30	90	2.741	2.881	2.976	3.047	3.103	3.150
	35	105	2.736	2.876	2.971	3.043	3.099	3.147
	40	120	2.732	2.873	2.968	3.040	3.097	3.144

Let Z_1 and Z_2 be iid $N(0,1)$ random variables. The function

$$H_\lambda(s) = P[Z_1^2 + (Z_2 - \lambda^{1/2})^2 \leq s]$$

$$= 2 \int_0^{s^{1/2}} \{\Phi[\lambda^{1/2} + (s - z_1^2)^{1/2}] - \Phi[\lambda^{1/2} - (s - z_1^2)^{1/2}\} \varphi(z_1)\, dz_1 , \qquad (23)$$

where Φ is the standard normal distribution function.

Equation (22) can be solved numerically for c using numerical quadrature and an iterative root finding technique. Note that under condition (14) the value of c depends on α, k, b, and ν. Table 1 gives values of c for comparisons with a control for $\alpha = 0.05$, $k = 3, \ldots, 7$, $b = 1, 1.5, 2$, and 3, and selected values of $n_k = bn_1$.

The following steps were taken using double precision FORTRAN to find the c values in Table 1. The integral in equation (23) was evaluated using 64-point Gaussian quadrature. The value of Φ was determined using the error function. The inner integral in (22) was evaluated using 64-point Gaussian quadrature. The outer integral in (22) was represented as the sum of integrals over the intervals $[0, u_1]$, $[u_1, u_2]$, and $[u_2, u_3]$, where

$$\begin{aligned} u_1 &= \{1 - 2/(9\nu) - 2.326[2/(9\nu)]^{1/2}\}^3\,, \\ u_2 &= \{1 - 2/(9\nu) + 2.326[2/(9\nu)]^{1/2}\}^3\,, \\ u_3 &= \{1 - 2/(9\nu) + 10[2/(9\nu)]^{1/2}\}^3\,. \end{aligned} \tag{24}$$

Sixty-four point Gaussian quadrature was used on each of the integrals. The points in (24) are good approximations of the .01, .99, and a most extreme point of the distribution of U (see equation 26.4.17 of ZELEN and SEVERO (1964)). The secant method was used to iterate to the solution.

4. The Case of Three Groups

For $k = 3$ it is possible to calculate c for comparisons with a control and for all pairwise differences for unrestricted $\boldsymbol{A}_i$'s. The approach is to rewrite the F_{ij} variables in terms of a 4-dimensional t vector. The components of the t vector are constructed from

$$\begin{aligned} Y_1 &= (\hat{\alpha}_1 - \hat{\alpha}_3) - (\alpha_1 - \alpha_3)\,, \qquad & Y_2 &= (\hat{\beta}_1 - \hat{\beta}_3) - (\beta_1 - \beta_3)\,, \\ Y_3 &= (\hat{\alpha}_2 - \hat{\alpha}_3) - (\alpha_2 - \alpha_3)\,, \qquad & Y_4 &= (\hat{\beta}_2 - \hat{\beta}_3) - (\beta_2 - \beta_3)\,, \end{aligned} \tag{25}$$

and $\hat{\sigma}$. One then numerically integrates the t vector over the region specified in equation (4) and solves for c. This method is analogous to the SPURRIER and ISHAM (1985) technique for the comparison of three normal means. More details are given in SPURRIER (2002).

5. Comparison of Methods and Designs

The confidence intervals in (5) are identical to those presented by SPURRIER (1999) for all contrasts of k simple regression lines except for the value of c. As one reduces the set of comparisons of simple regression lines from all contrasts to all

Table 2
Values of c for 95% Simultaneous Confidence Bounds for All Contrasts, All Pairwise Differences and All Comparisons with a Control for k Simple Linear Regressions Under Condition (6)

k	$n_1 = \ldots = n_k$	ν	All Contrasts	All Pairwise Comparisons	Comparisons with Control
3	7	15	3.321	3.193	3.029
	12	30	3.118	3.000	2.853
5	6	20	3.984	3.564	3.215
	12	50	3.725	3.336	3.030

pairwise comparisons to all comparisons with a control, fewer statements are being made and the probability point c decreases. Decreasing c leads to narrower confidence bounds. The values of c for 95% simultaneous confidence bounds for all contrasts, all pairwise comparisons, and comparisons with a control are given in Table 2 for some values of k and $n_1 = \ldots = n_k$ under condition (6). The amount of improvement offered by the intervals from Sections 2 and 3 relative to those in SPURRIER (1999) increases as k goes from 3 to 5. The amount of improvement appears to decrease slightly as ν increases. The results in Table 2 are consistent with simulation results in LIU et al. (2001).

We now study the effect of b on confidence bound width for comparisons with a control. Assume that the design matrix for the ith group, $\boldsymbol{X}_{i,}$ is comprised of r_i copies of an $n_0 \times 2$ design matrix $\boldsymbol{X}_0$, $i = 1, \ldots, k$. Thus,

$$\boldsymbol{A}_i = (1/r_i)\,(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\,, \qquad i = 1, \ldots, k\,. \tag{26}$$

We will assume $r_1 = \ldots = r_{k-1}$. Thus, condition (14) holds with $b = r_k/r_1$. Under these assumptions, the confidence bounds for comparisons with a control in (5) become

$$\begin{aligned}&(\alpha_i + \beta_i x) - (\alpha_k + \beta_k x) \in (\hat{\alpha}_i + \hat{\beta}_i x) - (\hat{\alpha}_k + \hat{\beta}_k x)\\ &\pm c\hat{\sigma}\{(1, x)\left[\left(\frac{1}{r_1} + \frac{1}{r_k}\right)(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\right](1, x)'\}^{1/2}\,.\end{aligned} \tag{27}$$

It follows from logic similar to that of Dunnett (1964) that the width of the confidence bounds are minimized for a fixed total sample size by taking $b \approx (k-1)^{1/2}$.

For $k = 3$ and $n_0 = 5$, the 95% simultaneous confidence bounds for all comparisons with a control using $r_1 = r_2 = r_3 = 7$, $(b = 1)$ are 1.1% wider that the bounds using $r_1 = r_2 = 6$ and $r_3 = 9$, $(b = 1.5)$. For $k = 5$ and $n_0 = 5$, the 95% bounds using $r_1 = \ldots = r_5 = 6$, $(b = 1)$ are 4.5% wider that the bounds using $r_1 = \ldots = r_4 = 5$ and $r_5 = 10$, $(b = 2)$.

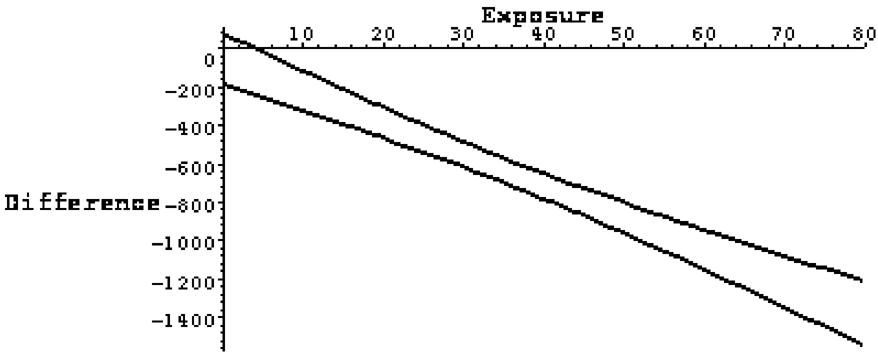

Fig. 1. Confidence Bound for Difference (Indomethacin-Control) in Expected Accumulation (fmol/mg Protein) as a Function of Exposure Time in Minutes.

6. Example

Masuda et al. (1997) studied the effects of $k = 3$ treatments, indomethacin, ketoprofen, and a control, on the accumulation of the chemotherapy agent [^{3}H]methotrexate in rat renal tissue. They measured the accumulation (fmol/mg protein) in three independent replicates for each treatment after exposure time, x = 15, 30, and 60 minutes. Thus, the same design involving 9 observations was used for all three treatments. The sample regressions lines are

$$\begin{aligned} Y &= 310.92 + 4.60x \text{ for treatment 1 (indomethacin)} \\ Y &= 390.57 + 5.89x \text{ for treatment 2 (ketoprofen)} \\ Y &= 358.89 + 21.15x \text{ for treatment 3 (control)} \end{aligned} \tag{28}$$

The pooled estimate of σ is $\hat{\sigma} = 45.58$.

For 95% simultaneous comparisons with the control, we have $c = 2.927$ and the confidence bounds

$$\begin{aligned} &(\alpha_i + \beta_i x) - (\alpha_3 + \beta_3 x) \in (\hat{\alpha}_i + \hat{\beta}_i x) - (\hat{\alpha}_3 + \hat{\beta}_3 x) \\ &\pm 2.927(45.58)\,[1 - (2x/45) + (x^2/1575)]^{1/2} \end{aligned} \tag{29}$$

for $i = 1, 2$ and all x. The confidence bounds for comparing indomethacin to the control are shown in Figure 1. Note that the entire bounds are below zero for exposure times greater than 4.4 minutes. Thus, indomethacin is significantly inhibiting the uptake of the chemotherapy agent after 4.4 minutes. Dividing the endpoints by x and taking limits as $x \to \infty$ yields the confidence interval $(-19.93, -13.17)$ for $\beta_1 - \beta_3$.

7. Conclusions

It has been shown how to find exact simultaneous confidence intervals for all pairwise differences of k linear regression models under the assumption of iid normal errors and condition (6). The results hold for simple and multiple regression. It has also been shown how to find exact simultaneous confidence intervals for comparing each of $k-1$ simple regression lines to that of a control under the assumption of iid normal errors and condition (14). The resulting intervals are much narrower than those developed by SPURRIER (1999) for all contrasts of k regression lines. For comparisons with a control, it is more efficient to take more observations from the control than from each of the other groups.

Acknowledgements

The author thanks the referees for their helpful comments and Prof. Ken-ichi Inui for providing the example data.

References

DUNNETT, C. W., 1964: New Tables for Multiple Comparisons with a Control. *Biometrics* **20**, 482–491.

HOCHBERG, Y., and TAMHANE, A. C., 1987: *Multiple Comparison Procedures*. Wiley, New York.

HSU, J. C., 1996: *Multiple Comparisons: Theory and Methods*. Chapman and Hall, London.

LIU, W., JAMSHIDIAN, M., and ZHANG, Y., 2001: *Multiple Comparison of Several Linear Regression Models,* submitted.

MASUDA, S., SAITO, H., and INUI, K.-I., 1997: Interactions of Nonsteroidal Anti-inflammatory Drugs With Rat Renal Organic Anion Transporter, OAT-K1. *The Journal of Pharmacology and Experimental Therapeutics* **283**, 1039–1042.

ROYEN, T., 1984: Multiple Comparisons of Polynomial Distributions. *Biometrical Journal* **26**, 319–332.

ROYEN, T., 1990: Tables for Studentized Multivariate Maximum Ranges and their Application by Maximum Range Tests. *Biometrical Journal* **32**, 643–680.

ROYEN, T., 1991: Multivariate Gamma Distributions with One-Factorial Accompanying Correlation Matrices and Applications to the Distribution of the Multivariate Range. *Metrika* **38**, 299–315.

RUBERG, S. J., and STEGEMAN, J. W., 1991: Pooling Data for Stability Studies: Testing the Equality of Batch Degradation Slopes. *Biometrics* **47**, 1059–1069.

SCHEFFÉ, H., 1959: *The Analysis of Variance.* Wiley, New York.

SPURRIER, J. D., 1999: Exact Confidence Bounds for All Contrasts of Three or More Regression Lines. *Journal of the American Statistical Association* **94**, 483–488.

SPURRIER, J. D., 2002: Exact Pairwise Comparisons of Three or More Regression Lines: Pairwise Comparisons and Comparisons with a Control. *University of South Carolina Department of Statistics Technical Report Number* 210.

SPURRIER, J. D. and ISHAM, S. P., 1985: Exact Simultaneous Confidence Intervals for Pairwise Comparisons of Three Normal Means. *Journal of the American Statistical Association* **80**, 438–442.

WORKING, H., and HOTELLING, H., 1929: Application of the Theory of Error to the Interpretation of Trends. *Journal of the American Statistical Association*, Suppl. (Proc.) **24**, 73–85.

ZELEN, M., and SEVERO, N., 1964: Probability Functions. In M. Abramowitz and I. A. Stegun (eds.): *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables.* U.S. Government Printing Office, Washington, 925–995.

Received, February 2002
Revised, June 2002
Accepted, July 2002