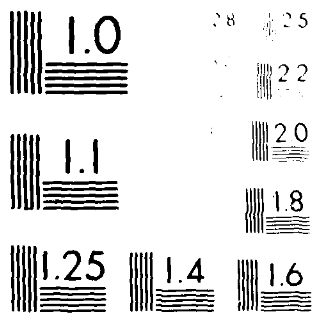
AD-A112 212 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
TESTING WHETHER ONE REGRESSION FUNCTION IS EVERYWHERE LARGER TH--ETC(U)
FEB 82 R L BERGER DAAG29-79-C-0158
UNCLASSIFIED FSU-STATISTICS-M607 ARO-16713.17-M NL

1 OF 1
AD A
112212

END
DATE
FILMED
04-82
DTIC


1.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

ARO 16713.17-M

(12)

AD A112212

TESTING WHETHER ONE REGRESSION FUNCTION IS EVERYWHERE LARGER THAN ANOTHER

by

Roger L. Berger

FSU Statistics Report M607
USARO Technical Report D-52

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

February, 1982

This research was supported in part by the U.S. Army Research Office Grant No. DAAG29-79-C-0158. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

SUMMARY

The problem of testing whether one regression function is larger than another on a specified compact set R is considered. The regression functions must be linear functions of the parameters but need not be linear functions of the independent variables. The proposed test statistic is compared to a standard t percentile. The test has an exactly specified size in typical situations. Properties of the power function of the test are investigated. The related question of comparing a regression function to a specified function is also considered.

1. Introduction

In this paper we consider testing whether the regression function from one population is everywhere above the regression function from another population. A medical researcher, for example, might be interested in testing whether the mean of a response variable in a diseased population is larger than the mean of a response variable in a healthy population for all possible values of an independent variable. Examples such as these have been discussed by Tsutakawa and Hewett (1978), Hewett and Lababidi (1980) and Spurrier, Hewett and Lababidi (1980) who have considered this testing problem.

The model considered herein generalizes the models of Tsutakawa and Hewett (1978) and Hewett and Lababidi (1980) in two ways. First the regression functions may be functions other than linear functions of the independent variables. For example, the regression functions may be quadratic or higher degree polynomials. Second, the two regression functions need not be of the same functional form. For example, one may be assumed to be a linear function and the other a quadratic fu[illegible] The test proposed herein reduces to the tests of Tsutakawa and Hewett (1978) and Hewett and Lababidi (1980) for the special models they consider. The proposed test requires no new tables for its implementation. Only a standard t table is needed.

The model and test are presented in Section 2.1. A numerical example using the organ weight data of normal and diabetic mice is presented in Section 2.2. Properties of the test, including its power, are discussed in Section 3. The related problem of testing whether a single regression function everywhere exceeds a specified function is discussed in Section 4.

2. Model, Test and Application

2.1. Model and Test

Let $\{(X_{1j}, Y_{1j}), j = 1, \ldots, n_1\}$ and $\{(X_{2j}, Y_{2j}); j = 1, \ldots, n_2\}$ denote two independent sets of observations where $X_{ij} = (X_{ij1}, \ldots, X_{ijk})$. The X_{ij} may be observed random vectors or design variables fixed by the experimenter. The entire analysis is conditioned on the observed values of X_{ij}. Let R denote a closed and bounded subset of k-dimensional Euclidean space. R is the set of possible values of the independent variables X_{ij}. We assume that given the X_{ij} the Y_{ij} are independent normal random variables with

$$E(Y_{ij} | X_{ij1} = x_{ij1}, \ldots, X_{ijk} = x_{ijk}) = \sum_{m=1}^{p_i} \beta_{im} f_{im}(x_{ij1}, \ldots, x_{ijk}) = f_i(x_{ij})\beta_i$$

and

$$\mathrm{var}(Y_{ij} | X_{ij1} = x_{ij1}, \ldots, X_{ijk} = x_{ijk}) = \sigma^2.$$

$\beta_i = (\beta_{i1}, \ldots, \beta_{ip_i})'$, $i = 1, 2$, and σ^2 are unknown parameters.

The $f_i(x) = (f_{i1}(x), \ldots, f_{ip_i}(x))$, $i = 1, 2$, are known vectors of functions which define the functional form of the regression functions. For example the f_{ij} might be polynomials such as 1, x_1, x_2^2, or x_1x_3. By allowing $p_1 \neq p_2$ and $f_{1m} \neq f_{2m}(x)$, this model allows the two regression functions to have different functional forms. The first might be a linear function and the second a quadratic function. But usually $p_1 = p_2$ and $f_1(x) = f_2(x)$ will be chosen so the regression functions have the same functional form.

We wish to compare the regression functions $f_1(x)\beta_1$ and $f_2(x)\beta_2$. In particular we are interested in whether $f_1(x)\beta_1$ is always greater than $f_2(x)\beta_2$. The test we will propose is a size α test of

$$H_0:\ f_1(x)\beta_1 \le f_2(x)\beta_2 \qquad \text{for at least one } x \in R$$

vs.

$$H_A:\ f_1(x)\beta_1 > f_2(x)\beta_2 \qquad \text{for every } x \in R.$$

Let b_1 and b_2 denote the least squares estimates of β_1 and β_2 and let

$$s^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (y_{ij} - f_i(x_{ij})b_i)^2/\nu$$

denote the pooled estimate of σ^2 where $\nu = n_1 - p_1 + n_2 - p_2$. The estimate b_i has a multivariate normal distribution with mean β_i and covariance matrix $\sigma^2 D_i^{-1}$ where the (m, n) element of D_i is $\sum_{j=1}^{n_i} f_{im}(x_{ij})f_{in}(x_{ij})$. Let $e(x) = f_1(x)D_1^{-1}f_1'(x) + f_2(x)D_2^{-1}f_2'(x)$.

Then the variance of $f_1(x)b_1 - f_2(x)b_2$ is $\sigma^2 e(x)$.

The test we will propose may be motivated in this way. Consider comparing the two regression functions only at a single point $x \in R$. Let $T_x = (f_1(x)b_1 - f_2(x)b_2)/s\sqrt{e(x)}$. The test which rejects H_{0x}: $f_1(x)\beta_1 \leq f_2(x)\beta_2$ in favor of H_{Ax}: $f_1(x)\beta_1 > f_2(x)\beta_2$ when $T_x > t_{1-\alpha}(\nu)$ is a size α test where $t_{1-\alpha}(\nu)$ is the $(1 - \alpha)$th percentile of a t distribution with ν degrees of freedom. The alternative of interest H_A is the intersection of all the H_{Ax} for all $x \in R$. It seems reasonable to decide in favor of H_A only if the test based on T_x decides in favor of H_{Ax} for every $x \in R$. This leads us to propose the following test.

Define the test statistic T by

$$T = \min_{x \in R} T_x. \tag{2.1}$$

Reject H_0 in favor of H_A if and only if $T > t_{1-\alpha}(\nu)$. This test is always a level α test in that the probability of a type one error is always less than or equal to α. This test has <u>size exactly equal to α</u> if the f_{ij} are continuous functions of x and there are values of β_1 and β_2 such that $f_1(x)\beta_1 = f_2(x)\beta_2$ for one value of $x \in R$ and $f_1(x)\beta_1 > f_2(x)\beta_2$ for all other $x \in R$. These facts are proven in the Appendix. Section 3 contails examples of when the condition is satisfied and the size is exactly α.

The test statistic T will usually have to be evaluated by numerical methods. This is discussed in Section 3. In some cases it can be evaluated explicitly. In particular it is shown in the Appendix that this test is equivalent to the tests proposed by Tsutakawa and Hewett (1978) and Hewett and Lababidi (1980) for the special models they consider if $\alpha \le .5$.

2.2. A Numerical Application

We consider the data relating body and kidney weight for diabetic and healthy mice which was presented by Bishop (1973). These data were analyzed by Tsutakawa and Hewett (1978). The data are plotted in Figure 1. The independent variable x is the body weight and the dependent variable y is the kidney weight. We consider modelling the mean of the kidney weight as a quadratic function of the body weight. We consider testing H_0: $\beta_{11} + \beta_{12}x + \beta_{13}x^2 \le \beta_{21} + \beta_{22}x + \beta_{23}x^2$ for some $x \in R$ versus H_A: $\beta_{11} + \beta_{12}x + \beta_{13}x^2 > \beta_{21} + \beta_{22}x + \beta_{23}x^2$ for all $x \in R$ where the first population is the diabetic population and $R = \{x: 26 \le x \le 52\}$. The least squares line for the 9 diabetic mice is given by $1821.14 - 49.92x + .77x^2$. The least squares line for the 25 healthy mice is given by $-434.68 + 62.70x - .85x^2$. $s^2 = (66052 + 196505)/(9 - 3 + 25 - 3) = 9377.0$. The matrices D_1^{-1} and D_2^{-1} are

$$D_1^{-1} = \begin{bmatrix} 408.0894 & -19.1506 & .2214 \\ -19.1506 & .9031 & -.0105 \\ .2214 & -.0105 & .0001 \end{bmatrix}$$

and

$$D_2^{-1} = \begin{bmatrix} 81.1483 & -4.6954 & .0666 \\ -4.6954 & .2735 & -.0039 \\ .0666 & -.0039 & .0001 \end{bmatrix}$$

Thus $T_x = a(x)/s\sqrt{e(x)}$ where $a(x) = 2255.82 - 112.62x + 1.62x^2$ and $e(x) = 489.2377 - 47.6920x + 1.7526x^2 - .0288x^3 + .0002x^4$. Numerical minimization of T_x for x between 26 and 52 yields the test statistic T = .4797 which corresponds to x = 40.10. This T is at approximately the 68th percentile of a t distribution with 28 degrees of freedom.

3. Properties of the Test

3.1 Power Function Properties

Tho test we propose has size exactly α if the f_{ij} are continuous functions and there are values of β_1 and β_2 such that $f_1(x)\beta_1 = f_2(x)\beta_2$ for one value of $x \in R$ and $f_1(x)\beta_1 > f_2(x)\beta_2$ for all other $x \in R$. This condition is satisfied if the $f_{ij}(x)$ include the constant 1, the linear functions x_i, i = 1, ..., k, and the quadratic functions x_ix_j, i = 1, ..., k, j = 1, ..., i.

Then β_1 and β_2 can be chosen so that $f_1(x)\beta_1 - f_2(x)\beta_2 = (x - x_0)(x - x_0)'$, the square of the distance from x to x_0, where x_0 is a fixed element of R. For this choice, $f_1(x)\beta_1 - f_2(x)\beta_2$ is zero for $x = x_0$ and is positive for all other x. Another situation in which the condition is satisfied is if $f_i(x)\beta_i = \beta_{i(k+1)} + \sum_{j=1}^{k} \beta_{ij}x_j$ and $R = \{x: x_{j*} \le x_j \le x_j^*, j = 1, \ldots, k\}$, the model considered by Tsutakawa and Hewett (1978) and Hewett and Lababidi (1980). The choice of $\beta_2 = 0$, $\beta_{11} = \cdots = \beta_{1k} = 1$ and $\beta_{1(k+1)} = -\sum_{j=1}^{k} x_{j*}$ yields $f_1(x)\beta_1 - f_2(x)\beta_2 = \sum_{j=1}^{k} (x_j - x_{j*})$ which is zero for $x = (x_{1*}, \ldots, x_{k*})$ and positive for all other $x \in R$. A numerical corroboration of the fact that the size is exactly α is found in the following simulation study.

A simulation study was conducted to investigate the power function of the test. In this study the regression functions were $f_i(x)\beta_i = \beta_{i1} + \beta_{i2}x + \beta_{i3}x^2$, $i = 1, 2$. The variance σ^2 was set equal to one. The sample sizes n_1 and n_2 were both 10 with 3 observations at each of $x = 1$ and $x = -1$ and 4 observations at $x = 0$. $R = \{x: -1 \le x \le 1\}$. The size of the test was fixed at $\alpha = .05$ by using $t_{.95}(14) = 1.761$. The International Mathematics and Statistics Library programs GGNSM and GGCHS were used to generate the random vector $b_1 - b_2$ and the random variable S^2. A total of 3000 repetitions were used to obtain each of the estimates in Tables 1 and 2.

The maximum probability of a Type I error takes place when $f_1(x)\beta_1 = f_2(x)\beta_2$ for one x and $f_1(x)\beta_1 - f_2(x)\beta_2$ becomes large for all other x. This can be observed in Table 1 where the probability of a Type I error is given for various values of β_1 and β_2 in the null hypothesis. As one proceeds down Columns II, III, IV or V of Table 1, $f_1(x)\beta_1 = f_2(x)\beta_2$ for one value of x(x = -1 for Columns III, IV and V and x = 0 for Columns II) and $f_1(x)\beta_1 - f_2(x)\beta_2$ is becoming large for all other values of x. The probability of a Type I error increases to $\alpha = .05$ as one proceeds down any column. The estimates slightly exceed .05 in a few cases due to sampling error.

The power function of the proposed test exhibits the following monotonicity property. If (β_1, β_2) and (β_1^*, β_2^*) are two parameter vectors which satisfy

$$(3.1) \qquad f_1(x)\beta_1^* - f_2(x)\beta_2^* \geq f_1(x)\beta_1 - f_2(x)\beta_2$$

for every x with strict inequality for some x, then the power at (β_1^*, β_2^*) is greater than the power at (β_1, β_2). This property is apparent as one proceeds down any column in Tables 1 or 2, across the first three columns in any row of Table 2, or across the last three columns in any row of Table 1. It is also apparent if one compares any two corresponding entries in Table 2a and 2b since the regression functions are farther apart in 2b than in 2a.

The power of the test is near one only if $f_1(x)\beta_1 - f_2(x)\beta_2$ is large for all x. This is the case for the lower entries in Table 2. In Table 2, the minimum distance between the regression functions is c. The power nears one only as the minimum distance c becomes large.

The test we propose is biased in that the probability of rejecting H_0 is less than α for some (β_1, β_2) in H_A. The feature was noted by Tsutakawa and Hewett (1978) for the special model they considered and it continues to exist for the more general models we consider. This biasedness can be observed in the entries for c = .5 which are less than .05 in Table 2. But as noted by Tsutakawa and Hewett (1978) for their special case, the test we propose is consistent in that, for any fixed point (β_1, β_2) in H_A, the power can be made arbitrarily near one by choosing the sample sizes sufficiently large. Although we do not feel this bias is serious, it should be noted that the power of the test we propose may be small if $f_1(x)\beta_1$ exceeds $f_2(x)\beta_2$ by only a small amount over most of R.

The power function properties we have described in this section are true in general, not just for the case of quadratic regression we considered in the simulation experiment. The proofs of these facts can be accomplished using the methods employed in proofs in the Appendix.

3.2. Computational Shortcuts

In (2.1) the test statistic T was defined as the minimum of T_x over the set R. Typically the computation of the test statistic will be accomplished by a numerical minimization of T_x. But to perform the test the actual value of T need not be computed. One only needs to know whether $T > t_{1-\alpha}(\nu)$ or $T \le t_{1-\alpha}(\nu)$. In this section we describe two shortcuts which allow the determination of whether $T > t_{1-\alpha}(\nu)$ or $T \le t_{1-\alpha}(\nu)$ without the actual computation of T.

3.2.1. Shortcut for determining if H_0 is accepted.

Let X* denote an arbitrary finite subset of R. For example, if $R = \{x: x_{i*} \le x_i \le x_i^*, i = 1, \ldots, k\}$, X* might be the set of 2^k extreme points $(x_1, \ldots, x_k)$ where $x_i = x_{i*}$ or $x_i = x_i^*$. Let $T' = \min_{x \in X^*} T_x$. Since $T \le T'$, if $T' \le t_{1-\alpha}(\nu)$ accept H_0. Furthermore if $T' = t'$ the significance probability associated with T is at least $P(T_0 > t')$ where T_0 has a central t distribution with ν degrees of freedom.

For the body kidney weight data of Section 2.2, $R = \{x: 26 \le x \le 52\}$. T_x for $x = 26$ is .9962 and T_x for $x = 52$ is .6347. Either one of these points is less than $T_{.95}(28) = 1.7011$ so the test accepts H_0 at level .05. Furthermore the significance probability associated with T is at least $P(T_0 \ge .6347) = .27$. The exact significance probability computed in Section 2.2 was .32.

In Tables 1, and 2 the second (middle) number for each entry is the proportion of the acceptances which were determined by the shortcut method. These values indicate that the usefulness of this shortcut depends on the actual regression functions. But, in many cases, a large proportion of the acceptances were determined by this shortcut. In 18 out of the 62 cases in which there were some acceptances, all of the acceptances were determined by the shortcut.

3.2.2. Shortcut for determining if H_0 is rejected.

For this shortcut to be valid, α must be no more than .5. Since α usually satisfies $\alpha \le .1$, this restriction is not practically important. Let m denote the number of distinct nonconstant functions in $\{f_{ij}(x)\colon\ i = 1, 2;\ j = 1, \ldots, p_i\}$ Let Z^* denote the set of 2^m points $(z_1^*, z_2^*) = ((z_{11}, \ldots, z_{1p_1}), (z_{21}, \ldots, z_{2p_2}))$ formed by replacing $f_{ij}(x)$ by either $\max_{x\in R} f_{ij}(x)$ or $\min_{x\in R} f_{ij}(x)$ in $(f_1(x), f_2(x))$. Note that if $f_{1r}(x) = f_{2s}(x)$ then $z_{1r} = z_{2s}$, i.e., the maximum or minimum is used on both z_1^* and z_2^*. Let $T^* = \min_{z^*\in Z^*} T_{z^*}$ where

$T_{z^*} = (z_1^* b_1 - z_2^* b_2)/s \sqrt{z_1^* D_1^{-1} z_1^{*\prime} + z_2^* D_2^{-1} z_2^{*\prime}}$. If $T^* > t_{1-\alpha}(\nu)$ then $T > t_{1-\alpha}(\nu)$ and H_0 can be rejected. Furthermore, if $T^* = t^*$, the significance probability associated with T is less than or equal to $P(T_0 > t^*)$ where T_0 has a central T distribution with ν degrees of freedom. If $T^* = T_{z^*}$ where $z^* = (f_1(x), f_2(x))$

for some $x \in R$ then, in fact, the significance probability equals $P(T_0 > t^*)$.

For example, suppose $f_1(x) = (1, x, x^2)$, and $f_2(x) = (1, x)$, and $R = \{x: -1 \le x \le 2\}$. There are two distinct nonconstant functions, x and x^2, and the $2^2 = 4$ values of (z_1^*, z_2^*) in Z^* are $((1, -1, 0), (1, -1))$, $((1, -1, 4), (1, -1))$, $((1, 2, 0), (1, 2))$ and $((1, 2, 4), (1, 2))$. Points like $((1, -1, 0), (1, 2))$ where x has been replaced by its minimum in z_1^* and its maximum in z_2^* are not in Z^*.

The validity of this shortcut is based on two facts, the fact about functions which are the ratios of linear functions and square roots of positive quadratic functions mentioned in the *proof of Theorem 3 and the fact that* $A = \{(f_1(x), f_2(x)): x \in R\}$ is a subset of

$$B = \{(z_1, z_2): \min_{x\in R} f_{ij}(x) \le z_{ij} \le \max_{x\in R} f_{ij}(x) \text{ and } z_{1r} = z_{2s} \text{ if } f_{1r}(x) = f_{2s}(x)\}$$

so a minimum over A is not less than a minimum over B.

This shortcut was used in the simulation study of Section 3.1. In this case $R = \{x: -1 \le x \le 1\}$. $f_1(x) = f_2(x) = (1, x, x^2)$ so there are $m = 2$ distinct nonconstant functions. The $2^m = 4$ points in Z^* are $((1, -1, 0), (1, -1, 0))$, $((1, -1, 1), (1, -1, 1))$, $((1, 1, 0), (1, 1, 0))$ and $((1, 1, 1), (1, 1, 1))$.

In Tables 1 and 2, the third (bottom) number for each entry is the proportion of the rejections which were detected by this shortcut. The usefulness of this shortcut is seen to depend on the actual value of the regression function but in many cases the proportion is fairly high. In 17 of 71 cases all of the rejections were detected by this method, avoiding numerical minimization of T_x.

3.2.3. Numerical minimization of T_x.

If neither of the shortcuts presented in Sections 3.2.1 or 3.2.2 determine whether H_0 is to be accepted or rejected or if the exact value of T is desired to compute an exact significance probability, then the function T_x must be minimized by numerical methods to determine the value of the test statistic T. The problem of minimizing a function such as T_x which is the ratio of two functions of x has been studied extensively in the mathematical programming literature by Charnes and Cooper (1962), Swarup (1965) Sharma (1967) and Craven and Mond (1973, 1975a, and 1975b). These authors have found that this non-linear programming problem is equivalent to other nonlinear programming problems which do not involve fractions. These results could simplify the numerical minimization of T_x.

4. Comparison of a Regression Function with a Specified Function.

The hypothesis that the regression function from a population is everywhere above a specified function can be tested with a test similar to the one described in Section 2. For this problem we only have the sample $\{(X_{1j}, Y_{1j});\ j = 1, \ldots, n_1\}$ from the first population. We wish to test

$$H_0: \ f(x)\beta_1 \le g(x) \quad \text{for at least one } x \in R$$

vs

$$H_A: \ f(x)\beta_1 > g(x) \quad \text{for all } x \in R$$

where $g(x)$ is a specified function. The function $g(x)$ may be a constant c in which case we are testing whether the mean of the response is greater than c for all possible values of the independent variable x. Let

$$s_1^2 = \sum_{j=1}^{n_1} (y_{1j} - f_1(x_{1j})b_1)^2/\nu_1$$

where $\nu_1 = n_1 - p_1$. Let $e_1(x) = f_1(x)D_1^{-1}f_1'(x)$ and $T_{1x} = (f_1(x)b_1 - g(x))/s_1\sqrt{e_1(x)}$. Define the test statistic T_1 by

$$T_1(x) = \min_{x \in R} T_{1x}.$$

A level α test of H_0 verses H_A is given by reject H_0 if $T_1 > t_{1-\alpha}(\nu_1)$.

This test enjoys all the same properties as the test based on T described in Sections 2 and 3. For example, the test has size exactly α if $f_{11}, \ldots, f_{1p_1}$ and g are all continuous functions and there is a value of β_1 such that $f(x)\beta_1 = g(x)$ for one value of $x \in R$ and $f_1(x)\beta_1 > g(x)$ for all other $x \in R$. The proofs of these properties are analogous to those for T with the function $f_1(x)\beta_1 - f_2(x)\beta_2$ replaced by $f_1(x)\beta_1 - g(x)$. These proofs are not given herein.

Table 1

Power of the Test and Percentage of Acceptances and Rejections by Shortcuts[1] for Selected Points in H_0[2].

c	I	II	III	IV	V
0	.0003	.0003	.0003	.0003	.0003
	99	99	99	99	99
	100	100	100	100	100
	.0027	.0080	.0037	.0067	.0087
1	100	90	98	98	99
	50	33	27	50	50
	.0053	.0287	.0110	.0237	.0327
2	100	40	97	98	99
	75	22	30	39	63
	.0060	.0483	.0267	.0493	.0523
5	100	0	97	100	100
	100	16	16	32	78
	.0060	.0510	.0460	.0523	.0523
25	100	0	99	100	100
	100	15	9	30	100
	.0060	.0510	.0523	.0523	.0523
1000	100	0	100	100	100
	100	15	8	30	100

[1] First (top) entry: estimated power of the test
Second (middle) entry: percentage of acceptances detected by shortcut in Section 3.2.1
Third (bottom) entry : percentage of rejections detected by shortcut in Section 3.2.2

[2] Column I : $f_1(x)\beta_1 - f_2(x)\beta_2 = c(1 - x^2)$
Column II : $f_1(x)\beta_1 - f_2(x)\beta_2 = cx^2$
Column III: $f_1(x)\beta_1 - f_2(x)\beta_2 = c(x + 1)^2/4$
Column IV : $f_1(x)\beta_1 - f_2(x)\beta_2 = c(x + 1)/2$
Column V : $f_1(x)\beta_1 - f_2(x)\beta_2 = c(-x^2 + 2x + 3)/4$

Table 2a

Power of the Test and Percentage of Acceptances and Rejections by Shortcuts[1] for Selected Points in H_A^2

c	I	II	III	IV	V
	.0053	.0290	.0547	.0180	.0560
.5	98	94	96	99	73
	38	33	51	76	21
	.0510	.1460	.2023	.0990	.2327
1	94	87	94	99	51
	45	44	63	78	36
	.5197	.6897	.7357	.6030	.7967
2	82	81	96	99	30
	65	65	81	92	59
	.9323	.9647	.9667	.9387	.9907
3	91	93	99	100	29
	90	90	97	100	88
	1.000	1.000	1.000	1.000	1.000
5	--	--	--	--	--
	100	100	100	100	100

[1]See Table 1 footnote

[2]Column I : $f_1(x)\beta_1 - f_2(x)\beta_2 = c$

Column II : $f_1(x)\beta_1 - f_2(x)\beta_2 = (x + 1)^2/4 + c$

Column III : $f_1(x)\beta_1 - f_2(x)\beta_2 = (-x^2 + 2x + 3)/4 + c$

Column IV : $f_1(x)\beta_1 - f_2(x)\beta_2 = -x^2 + 1 + c$

Column V : $f_1(x)\beta_1 - f_2(x)\beta_2 = x^2 + c$

Table 2b

Power of the Test and Percentage of Acceptances and Rejections by Shortcuts[1] for Selected Points in H_A[2].

c	I	II	III	IV	V
	.0053	.0870	.1453	.0250	.1510
.5	98	94	100	100	0
	38	11	79	100	17
	.0510	.2543	.3167	.1107	.3680
1	94	92	100	100	0
	45	15	87	100	28
	.5197	.7557	.7700	.6077	.8460
2	82	94	100	100	0
	65	32	96	100	57
	.9323	.9677	.9680	.9387	.9933
3	91	99	100	100	0
	90	68	100	100	88
	1.0000	1.0000	1.0000	1.0000	1.0000
5	--	--	--	--	--
	100	100	100	100	100

[1]See Table 1 footnote.

[2]Column I : $f_1(x)\beta_1 - f_2(x)\beta_2 = c$
Column II : $f_1(x)\beta_1 - f_2(x)\beta_2 = (x+1)^2 + c$
Column III: $f_1(x)\beta_1 - f_2(x)\beta_2 = -x^2 + 2x + 3 + c$
Column IV : $f_1(x)\beta_1 - f_2(x)\beta_2 = -4x^2 + 4 + c$
Column V : $f_1(x)\beta_1 - f_2(x)\beta_2 = 4x^2 + c$

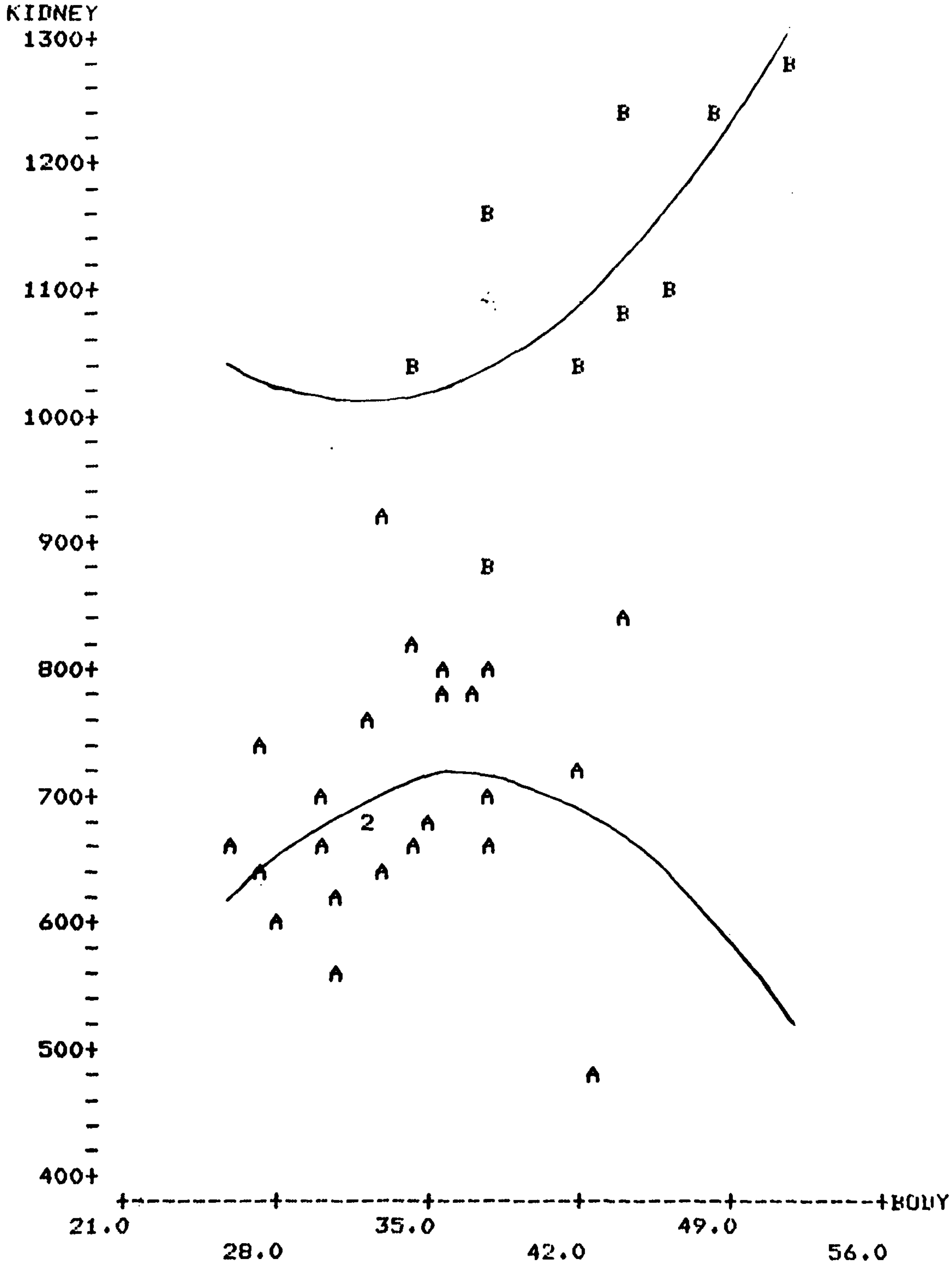


Figure 1
Body and Kidney Weight for
Healthy (A) and Diabetic (B) Mice

REFERENCES

Charnes, A. and Cooper, W.W. (1962). Programming with linear fractional functionals. Naval Research Logistics Quarterly 9, 181-186.

Craven, B.D. and Mond, B. (1973). A note on mathematical programming with fractional objective functions. Naval Research Logistics Quarterly 20, 577-581.

Craven, B.D. and Mond, B. (1975a). Nonlinear fractional programming. Bulletin of the Australian Mathematical Society 12, 391-397.

Craven, B.D. and Mond, B. (1975b). On fractional programming and equivalence. Naval Research Logistics Quarterly 22, 405-410.

Hewett, J.E. and Lababidi, Z. (1980). Comparison of two populations with multivariate data. Biometrics 36, 671-675.

Hewett, J.E., Lababidi, Z. and Spurrier, J.D. (1980). Comparison of two regression lines over a finite interval. University of South Carolina Statistics Technical Report.

Hewett, J.E. and Tsutakawa, R.K. (1978). Comparison of two regression lines over a finite interval. Biometrics 34, 391-398.

Sharma, I.C. (1967). Feasible direction approach to fractional programming problems. Opsearch 4, 61-72.

Swarup, K. (1965). Programming with quadratic fractional functions. Opsearch 2, 23-30.

Appendix

Proofs regarding the size of the test and the equivalence of the test to the test proposed by Tsutakawa and Hewett (1978) and Hewett and Lababidi (1980) are given in this appendix.

The size of the test.

Theorem 1: Under the assumptions of our model the test has level α, i.e.,

$$\text{A.1} \qquad \sup_{(\beta_1,\beta_2)\in H_0} P_{\beta_1,\beta_2}(T > t_{1-\alpha}(\nu)) \le \alpha.$$

Proof: Fix $(\beta_1, \beta_2) \in H_0$. There is an $x_0 \in R$ such that $f_1(x_0)\beta_1 \le f_2(x_0)\beta_2$. Then, with probability one, $T \le T_{x_0} \le Q$ where

$$Q = \frac{f_1(x_0)b_1 - f_2(x_0)b_2 - (f_1(x_0)\beta_1 - f_2(x_0)\beta_2)}{S\sqrt{c(x_0)}}$$

Q has a t distribution with ν degrees of freedom. So

$$P_{\beta_1,\beta_2}(T > t_{1-\alpha}(\nu)) \le P(Q > t_{1-\alpha}(\nu)) = \alpha.$$

Since β_1 and β_2 were arbitrary, A.1 is true. ||

Theorem 2: Suppose that all the $f_{ij}(x)$, $i = 1, 2$, $j = 1, \ldots, p_i$, are continuous on R. If there exist β_1^* and β_2^* such that $f_1(x_0)\beta_1^* = f_2(x_0)\beta_2^*$ for one $x_0 \in R$ and $f_1(x)\beta_1^* > f_2(x)\beta_2^*$ for all other $x \in R$ then the test has size exactly α, i.e.,

$$\sup_{(\beta_1,\beta_2)\in H_0} P_{\beta_1,\beta_2}(T > t_{1-\alpha}(\nu)) = \alpha.$$

The proof of Theorem 2 will use Lemma 1 which can be proved using standard analysis methods.

Lemma 1: Let $g_n(x)$, $n = 1, 2, \ldots$, be continuous functions on a compact set R. Suppose there exists an $x_0 \in R$ such that $g_n(x_0)$ is constant (say c) for all n. Suppose $g_n(x)$ increases to infinity as $n \to \infty$ for all $x \neq x_0$. Then

(A.2) $$\lim_{n\to\infty} \min_{x\in R} g_n(x) = c.$$

Proof of Theorem 2: By Theorem 1 it suffices to show there exists a sequence (β_1^n, β_2^n), $n = 1, 2, \ldots$, such that, $(\beta_1^n, \beta_1^n) \in H_0$ for $n = 1, 2, \ldots$, and

(A.4) $$\lim_{n\to\infty} P_{\beta_1^n,\beta_2^n}(T > t_{1-\alpha}(\nu)) \geq \alpha .$$

The estimates b_i, $i = 1, 2$, can be written as $b_i = Z_i + \beta_i$ where Z_1, Z_2 and S are independent, and Z_i has an p_i-variate normal distribution with mean 0 and variance-covariance matrix $\sigma^2 D_i^{-1}$. In terms of these quantities, the statistics T and T_x can be written as

$$T = T(Z_1, Z_2, S, \beta_1, \beta_2) = \min_{x\in R} T_x(Z_1, Z_2, S, \beta_1, \beta_2)$$

and

$$T_x(Z_1, Z_2, S, \beta_1, \beta_2) = \frac{f_1(x)Z_1 - f_2(x)Z_2 + f_1(x)\beta_1 - f_2(x)\beta_2}{S\sqrt{e(x)}} .$$

Consider the sequence (β_1^n, β_2^n) defined by $\beta_i^n = n\beta_i^*$ where the β_i^* are defined in the statement of Theorem 2. For a fixed value of $z_1 \in R^{p_1}$, $z_2 \in R^{p_2}$ and $s > 0$, define

$$g_n(x) = T_x(z_1, z_2, s, \beta_1^n, \beta_2^n).$$

The $g_n(x)$ satisfy the conditions of Lemma 1 since 1) f_{ij} are continuous, 2) $s\sqrt{e(x)} > 0$, 3) $f_1(x_0)\beta_1^n = f_2(x_0)\beta_2^n$ and 4) $f_1(x)\beta_1^n - f_2(x)\beta_2^n$ increases to infinity as $n \to \infty$ for all $x \neq x_0$. By Lemma 2, $\lim_{n\to\infty} T(z_1, z_2, s, \beta_1^n, \beta_2^n) = T_{x_0}(z_1, z_2, s, \beta_1^*, \beta_2^*)$. Since z_1, z_2, and s were arbitrary, this implies that $T(Z_1, Z_2, S, \beta_1^n, \beta_2^n)$ converges to $T_{x_0}(Z_1, Z_2, S, \beta_1^*, \beta_2^*)$ with probability one and hence in distribution. Thus

$$\begin{aligned}\lim_{n\to\infty} P_{\beta_1^n,\beta_2^n}(T > t_{1-\alpha}(\nu)) &= \lim_{n\to\infty} P(T(Z_1, Z_2, S, \beta_1^n, \beta_2^n) > t_{1-\alpha}(\nu)) \\ &= P(T_{x_0}(Z_1, Z_2, S, \beta_1^*, \beta_2^*) > t_{1-\alpha}(\nu)) \\ &= \alpha. \quad ||\end{aligned}$$

<u>Equivalence with tests proposed by Tsutakawa and Hewett (1978) and Hewett and Lababidi (1980).</u>

<u>Theorem 3</u>: Suppose $f_i(x)\beta_i = \beta_{i0} + \sum_{j=1}^{k} \beta_{ij}x_j$ and R has the from $R = \{x\colon\ x_{j*} \le x_j \le x_j^*,\ j = 1, \ldots, k\}$. Consider the test which rejects H_0 if $T^* > t_{1-\alpha}(\nu)$ where $T^* = \min_{x\in X^*} T_x$ and X^* is the set of 2^k points for which x_j is either x_{j*} or x_j^*. Suppose $\alpha \le .5$. Then the tests based on T^* and T are equivalent.

Proof. For any k + 1 dimensional vectors b_1 and b_2 and $s > 0$, T_x is a linear function of $(x_1, \ldots, x_k)$ divided by the square root of a quadratic function of $(x_1, \ldots, x_k)$ which is positive for all $(x_1, \ldots, x_k) \in R^k$. Such a function has the property that $T^* = \min_{x \in X^*} T_x \geq 0$ implies $T^* = \min_{x \in R} T_x = T$. (This is easily proved for $k = 1$ and can be proven for general k by induction.). For any b_1, b_2 and $s > 0$, $T \leq T^*$ so if T rejects H_0, so does T^*. Suppose b_1, b_2 and s are such that T^* rejects H_0. Then $T^* > t_{1-\alpha}(\nu) \geq 0$, since $\alpha \leq .5$, so $T = T^*$ and T also rejects H_0. ||

NCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU Statistics M607 USARO D-52	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and subtitle) Testing Whether One Regression Function is Everywhere Larger Than Another		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Roger L. Berger		8. CONTRACT OR GRANT NUMBER(s) USARO Grant DAAG29-79-C-0158
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Florida State University Tallahassee, FL. 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P.O. Box 12211 Research Triangle Park, N.C. 27709		12. REPORT DATE February, 1982 13. NUMBER OF PAGES 23
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 unlimited

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

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS

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

The problem of testing whether one regression function is larger than another on a specified compact set R is considered. The regression functions must be linear functions of the parameters but need not be linear functions of the independent variables. The proposed test statistic is compared to a standard t percentile. The test has an exactly specified size in typical situations. Properties of the power function of the test are investigated. The related question of comparing a regression function to a specified function is also considered.

ATE

LMED

-8