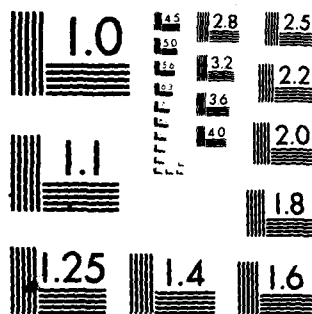




MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

12

AD A108332

ON TESTING MONOTONE TENDENCIES*

Richard L. Dykstra**

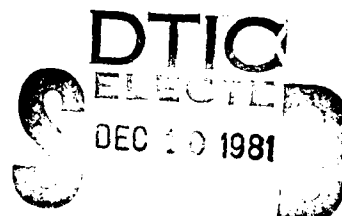
and

Tim Robertson

Department of Statistics
University of Missouri-Columbia
Columbia, Missouri

Department of Statistics
The University of Iowa
Iowa City, Iowa 52242

Technical Report 80
October 1981



*This research was sponsored by the Office of Naval Research under Contracts N00014-78-C-0655 and N00014-80-C-0321.

**A portion of the work done by this author was done on sabbatical leave at The University of Iowa.

DTIC FILE COPY

81 12 08 026

ON TESTING MONOTONE TENDENCIES

Richard L. Dykstra and Tim Robertson

SUMMARY

In certain problems, it may be expected that a regression function has a substantial overall tendency to be monotone and yet we may not be certain that all of the restrictions imposed by a simple order are satisfied. Distribution theory for likelihood ratio tests of homogeneity of a collection of normal means when the collection is "decreasing on the average" and for testing "decreasing on the average" as a null hypothesis, is presented. The restriction "decreasing on the average" is less restrictive than the usual monotone restriction and allows the data to give rise to "reversals" over short ranges of values of the parameter set. It is closely related to the "starshaped ordering" restriction discussed in Shaked (Ann. Statist. (1979)).

Accession For

WEIS GRAD	X
DEED CAR	
THIRD CAR	
THIRD CAR	

A

INTRODUCTION AND SUMMARY. The detection of a monotone relationship between two variables is an important problem in statistics. One approach to such problems is to base a conclusion about such a relationship upon an estimate of, or a test of a hypothesis about, a regression function. Procedures for making inferences about parameters which are known or suspected to satisfy a trend have received considerable attention in the statistical literature and a comprehensive treatment of much of the early work done on order restricted inference is given in Barlow, Bartholomew, Bremner and Brunk (1972).

In certain problems, it may be expected that the regression function has a substantial overall tendency to be monotone and yet we may not be certain that all of the restrictions imposed by a simple order are satisfied (this point is made on page 165 of Barlow et al. (1972)). For example, it is generally believed that mortality rates increase with age. The ages 15 through 35 crude mortality rates per 10,000 insured male lives are shown in Figure 1. This data is taken from the 1973 Reports of Mortality and Morbidity Experience of the Transactions of the Society of Actuaries. The rates are the 1965-70 ultimate experience and are based upon approximately 35,000 insured lives in each age group so that the standard error should be roughly 1.7 per 10,000 lives. The hypothesis that the mortality rates tend to increase with age seems to be confirmed by the data in Figure 1. However, it is not at all clear that the underlying regression function is strictly increasing over this range of ages. In fact, actuaries now believe that there is a "bump" in the mortality rate at about age 20 and that the mortality rate actually decreases from, roughly, ages 20 through 25. The 1965-70 graduations were among the first to reflect this "bump."

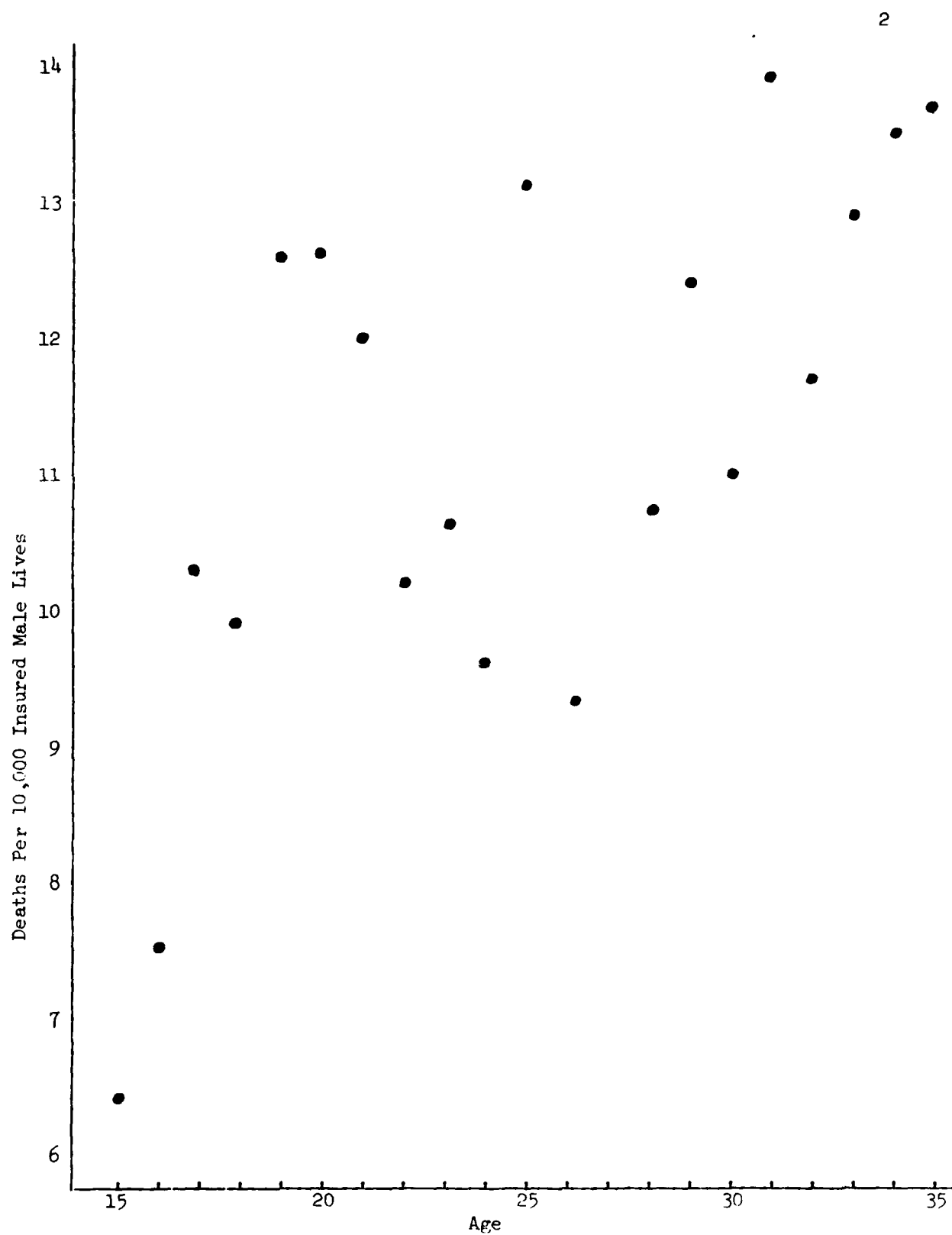


Figure 1. Crude Mortality Rates Per 10,000 Insured Male Lives.

The "usual" order restricted inference procedures do not allow the data to give rise to such "bumps" and the least squares order restricted estimate of the mortality rate based upon this data is actually constant from ages 19 through 28. These considerations suggest that inference procedures which account for a somewhat less restrictive monotone relationship could be of value. This paper is an account of our studies of one method of modeling a monotone relationship which allows the regression function to go counter to the overall trend over short ranges of values.

Def: Suppose $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ is a vector of parameters and $w = (w_1, w_2, \dots, w_k)$ is a vector of positive weights. We say that θ is "decreasing on the average from the left" (DAL) with respect to w provided

$$\bar{\theta}_1 \geq \bar{\theta}_2 \geq \dots \geq \bar{\theta}_k \quad \text{where} \quad \bar{\theta}_i = (\sum_{j=1}^i w_j)^{-1} \cdot (\sum_{j=1}^i \theta_j w_j); \quad i=1, 2, \dots, k. \quad (1.1)$$

Increasing on the average from the left (IAL), increasing on average from the right (IAR) and decreasing on the average from the right (DAR) are defined analogously.

Clearly if $\theta_1 \geq \theta_2 \geq \dots \geq \theta_k$ then θ is DAL (also IAR). The order restriction, DAL, is closely related to the "starshaped ordering" discussed in Shaked (1979). A vector θ is said to be lower-starshaped if, in addition to (1.1), $\bar{\theta}_k \geq 0$. Shaked (1979) studies estimates of normal and Poisson means subject to the restriction that they are starshaped and gives examples from reliability theory and from branching processes in which such orderings are of interest.

In Section 2, assuming that θ is a vector of normal means, we derive the restricted maximum likelihood estimate of θ . Our restrictions are

somewhat more general than those considered by Shaked (1979), in that we allow " $\leq$ ", " $=$ ", and " $\geq$ " to be intermixed in the restriction (1.1). Our method of derivation is different from that of Shaked and leads to the distributions of likelihood ratio statistics for tests where either the null or alternative hypothesis requires that the parameter vector is monotone on the average. These hypothesis tests are discussed in Sections 3 and 4. The test statistics have null-hypothesis distributions which are mixtures of chi-square or beta distributions--the so-called chi-bar-square ($\bar{\chi}^2$) and E-bar-square ($\bar{E}^2$) distributions.

2. RESTRICTED MAXIMUM LIKELIHOOD ESTIMATES. Suppose $Y_1, Y_2, \dots, Y_k$ are independent random variables and that $Y_i \sim n(\theta_i, a_i \sigma^2)$ where $a_1, a_2, \dots, a_k$ are known positive constants. Let $w_i = (a_i \sigma^2)^{-1}$ and $W_i = \sum_{j=1}^i w_j$; $i = 1, 2, \dots, k$. We wish to find the maximum likelihood estimator (MLE) of $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ subject to the constraints

$$\bar{\theta}_{i_{l-1}} = \bar{\theta}_{i_l} \quad (\text{or equivalently } \bar{\theta}_{i_{l-1}} = \theta_{i_l}); \quad l = 1, 2, \dots, m, \quad (2.1)$$

where $\bar{\theta}_i = w_i^{-1} \cdot \sum_{j=1}^i w_j \theta_j$ and $i_1 < i_2 < \dots < i_m$ is an ordered subset of $\{1, 2, \dots, k\}$. Using the constraints (2.1) to write θ_{i_j} ; $j = 1, 2, \dots, m$ in terms of θ_l for $l \notin \{i_1, i_2, \dots, i_m\}$ we obtain

$$\theta_{i_j} = \sum_{\alpha=1}^j h_{j\alpha} \sum_{\beta=i_{\alpha-1}+1}^{i_{\alpha}-1} w_{\beta} \theta_{\beta} \quad (2.2)$$

where

$$h_{j\alpha} = w_{i_{j-1}}^{-1} \prod_{\gamma=\alpha}^{j-1} w_{i_{\gamma}} \cdot w_{i_{\gamma}-1}^{-1}.$$

(We adopt the convention that a product over the empty set equals one while a sum over the empty set equals zero.) Substituting (2.2) into the log-likelihood function and equating the derivatives to zero we obtain the equations

$$(\theta_j - y_j) - \sum_{\alpha=l(j)}^m (y_{i_\alpha} - \theta_{i_\alpha}) h_{\alpha, l(j)} w_{i_\alpha} = 0; \quad j \in \{i_1, i_2, \dots, i_m\} \quad (2.3)$$

where $l(j) = l$ iff $i_{l-1} < j < i_l$. If we define $\bar{y}_i = W_i^{-1} \sum_{j=1}^i w_j y_j$, then the equations (2.3) have a surprisingly tractable solution.

Theorem 2.1. The solutions to the equations, (2.3), are given by

$$\hat{\theta}_j = y_j + \sum_{b=l(j)}^m (y_{i_b} - \bar{y}_{i_b-1}) w_{i_b} W_{i_b}^{-1}; \quad j \in \{i_1, i_2, \dots, i_m\} \quad (2.4)$$

which implies that

$$\hat{\theta}_{i_b} = \bar{y}_{i_b} + \sum_{a=b+1}^m (y_{i_a} - \bar{y}_{i_a-1}) w_{i_a} W_{i_a}^{-1}; \quad b=1, 2, \dots, m. \quad (2.5)$$

Proof: We begin by arguing that if the vector $\hat{\theta}$ satisfies (2.4) and the restrictions, (2.1), then its values for $j \in \{i_1, i_2, \dots, i_m\}$ must be given by (2.5). First, note that

$$\begin{aligned} \hat{\theta}_{i_1} &= W_{i_1-1}^{-1} \sum_{j=1}^{i_1-1} w_j \hat{\theta}_j \\ &= \bar{y}_{i_1-1} + \sum_{b=1}^m (y_{i_b} - \bar{y}_{i_b-1}) w_{i_b} W_{i_b}^{-1} \\ &= \bar{y}_{i_1-1} [1 - w_{i_1} W_{i_1}^{-1}] + y_{i_1} w_{i_1} W_{i_1}^{-1} + \sum_{b=2}^m (y_{i_b} - \bar{y}_{i_b}) w_{i_b} W_{i_b}^{-1} \\ &= \bar{y}_{i_1} + \sum_{b=2}^m (y_{i_b} - \bar{y}_{i_b}) w_{i_b} W_{i_b}^{-1}. \end{aligned}$$

We proceed by induction. Assuming (2.5) holds for b and noting that

$$\sum_{j=1}^{i_b} w_j \hat{\theta}_j = w_{i_b} \hat{\theta}_{i_b}, \text{ we consider}$$

$$\begin{aligned} \hat{\theta}_{i_{b+1}} &= w_{i_{b+1}}^{-1} \left[w_{i_b} \hat{\theta}_{i_b} + \sum_{j=i_b+1}^{i_{b+1}-1} w_j \hat{\theta}_j \right] \\ &= w_{i_{b+1}}^{-1} \left[w_{i_b} \bar{y}_{i_b} + w_{i_b} \sum_{a=b+1}^m (y_{i_a} - \bar{y}_{i_a-1}) w_{i_a} w_{i_a}^{-1} + w_{i_{b+1}}^{-1} \bar{y}_{i_{b+1}-1} - w_{i_b} \bar{y}_{i_b} \right. \\ &\quad \left. + (w_{i_{b+1}}^{-1} - w_{i_b}) \sum_{a=b+1}^m (y_{i_a} - \bar{y}_{i_a-1}) w_{i_a} w_{i_a}^{-1} \right] \\ &= w_{i_{b+1}}^{-1} \left[w_{i_{b+1}}^{-1} \bar{y}_{i_{b+1}-1} + w_{i_{b+1}} \sum_{a=b+1}^m (y_{i_a} - \bar{y}_{i_a-1}) w_{i_a} w_{i_a}^{-1} \right] \\ &= \bar{y}_{i_{b+1}} + \sum_{a=b+2}^m (y_{i_a} - \bar{y}_{i_a-1}) w_{i_a} w_{i_a}^{-1}. \end{aligned}$$

Thus, by induction, (2.5) holds for $b = 1, 2, \dots, m$.

Comparing (2.4) and (2.3) it suffices to show that

$$\sum_{b=\ell}^m (y_{i_b} - \bar{y}_{i_b-1}) w_{i_b} w_{i_b}^{-1} = \sum_{b=\ell}^m (y_{i_b} - \hat{\theta}_{i_b}) h_{b,\ell} w_{i_b} \quad (2.6)$$

for $\ell = 1, 2, \dots, m$. The equation, (2.6), holds for $\ell = m$ since

$$\hat{\theta}_{i_m} = \bar{y}_{i_m}, \quad h_{mm} = w_{i_m}^{-1} \quad \text{and}$$

$$(y_{i_b} - \bar{y}_{i_b-1}) w_{i_b} w_{i_b}^{-1} = (y_{i_b} - \bar{y}_{i_b}) w_{i_b} w_{i_b}^{-1}; \quad b = 1, 2, \dots, m. \quad (2.7)$$

In order to establish (2.6) in general, we use induction. Assume that (2.6)

holds for $\ell = c+1$. Then

$$\begin{aligned}
\sum_{b=c}^m (y_{i_b} - \bar{y}_{i_b-1}) w_{i_b} w_{i_b}^{-1} &= (y_{i_c} - \bar{y}_{i_c}) w_{i_c} w_{i_c}^{-1} + \sum_{b=c+1}^m (y_{i_b} - \bar{y}_{i_b}) w_{i_b} w_{i_b-1}^{-1} \\
&= w_{i_c-1}^{-1} \left\{ y_{i_c} - \left[\bar{y}_{i_c} + \sum_{b=c+1}^m (y_{i_b} - \bar{y}_{i_b}) w_{i_b} w_{i_b-1}^{-1} \right] \right\} w_{i_c} \\
&\quad + \sum_{b=c+1}^m (y_{i_b} - \bar{y}_{i_b}) w_{i_b} w_{i_b-1}^{-1} \cdot \left[1 + w_{i_c} w_{i_c-1}^{-1} \right] \\
&= w_{i_c-1}^{-1} \left\{ y_{i_c} - \hat{\theta}_{i_c} \right\} w_{i_c} + \sum_{b=c+1}^m (y_{i_b} - \hat{\theta}_{i_b}) h_{b,c+1} w_{i_b} \\
&\quad \cdot \left[1 + w_{i_c} w_{i_c-1}^{-1} \right] \\
&= [y_{i_c} - \hat{\theta}_{i_c}] h_{cc} w_{i_c} + w_{i_c} \cdot w_{i_c-1}^{-1} \cdot \sum_{b=c+1}^m (y_{i_b} - \hat{\theta}_{i_b}) h_{b,c+1} w_{i_b} \\
&= \sum_{b=c}^m (y_{i_b} - \hat{\theta}_{i_b}) h_{b,c} w_{i_b}
\end{aligned}$$

using (2.7), (2.5), the induction hypothesis, $w_{i_c-1}^{-1} = h_{cc}$ and the fact that $h_{b,c+1} \cdot w_{i_c} \cdot w_{i_c-1}^{-1} = h_{b,c}$. Since (2.6) holds for all l , the theorem is established.

We now show a remarkable property of the solutions given by (2.4) and (2.5). Since $\hat{\theta}_{i_c} = w_{i_c-1}^{-1} \cdot \sum_{j=1}^{i_c-1} w_j \hat{\theta}_j$,

$$\begin{aligned}
\sum_{j=1}^{i_c} \hat{\theta}_j w_j &= w_{i_c-1}^{-1} \hat{\theta}_{i_c} + w_{i_c} \hat{\theta}_{i_c} \\
&= w_{i_c} \hat{\theta}_{i_c} \\
&= w_{i_c} \left[\bar{y}_{i_c} + \sum_{b=c+1}^m (y_{i_b} - \bar{y}_{i_b-1}) w_{i_b} w_{i_b}^{-1} \right]
\end{aligned} \tag{2.8}$$

by (2.5). Suppose $i \notin \{i_1, i_2, \dots, i_m\}$ and also that i_c is the largest element of $\{i_1, i_2, \dots, i_m\}$ such that $i_c < i$. Then, using (2.8) and (2.4),

$$\begin{aligned}
 \bar{\theta}_{i-1} &= W_{i-1}^{-1} \sum_{j=1}^{i-1} \hat{\theta}_j w_j \\
 &= W_{i-1}^{-1} \left[\sum_{j=1}^{i_c} \hat{\theta}_j w_j + \sum_{j=i_c+1}^{i-1} \hat{\theta}_j w_j \right] \\
 &= W_{i-1}^{-1} \left\{ W_{i_c} \left[\bar{y}_{i_c} + \sum_{b=c+1}^m (y_{i_b} - \bar{y}_{i_b-1}) w_{i_b} W_{i_b}^{-1} \right] + \sum_{j=i_c+1}^{i-1} w_j y_j \right. \\
 &\quad \left. + \sum_{b=c+1}^m (y_{i_b} - \bar{y}_{i_b-1}) w_{i_b} W_{i_b}^{-1} \cdot [W_{i-1} - W_{i_c}] \right\} \\
 &= \bar{y}_{i-1} + \sum_{b=c+1}^m (y_{i_b} - \bar{y}_{i_b-1}) w_{i_b} W_{i_b}^{-1}.
 \end{aligned} \tag{2.9}$$

Thus, comparing (2.4) and (2.9), we see that

$$\bar{\theta}_{i-1} > (<) \hat{\theta}_i \quad \text{if and only if} \quad \bar{y}_{i-1} > (<) y_i,$$

regardless of $i_1, i_2, \dots, i_m$. It follows that if we wish to find the MLE of the vector θ subject to the constraints

$$\bar{\theta}_{i-1} \geq (\leq) \theta_i \quad (\text{equivalently } \bar{\theta}_{i-1} \geq (\leq) \bar{\theta}_i); \quad i=1,2,\dots,k$$

then we know that the i^{th} constraint $\bar{\theta}_{i-1} = \theta_i$ needs to be imposed if and only if $\bar{y}_{i-1} < (>) y_i$ (see Barlow et al. (1972), page 89).

Compared to other restricted optimization problems, this is a noteworthy property and it allows us to write the solutions in a concise form.

Adopting the standard notation, $a^+ = \max\{a, 0\}$ and $a^- = \min\{a, 0\}$ we have established the following theorem which generalizes the work of Shaked (1979).

Theorem 2.2. The MLE of the vector $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ subject to the constraints $\bar{\theta}_{i-1} (\geq) (=) (\leq) \bar{\theta}_i$; $i = 1, 2, \dots, k$ is given by

$$\hat{\theta}_i = \psi_i + \sum_{j=i+1}^m (y_j - \bar{y}_{j-1})^{\nu_i} w_j w_j^{-1}; \quad i = 1, 2, \dots, k$$

where

$$(\psi_i, \nu_i) = \begin{cases} (\max(\bar{y}_i, y_i), +) \\ (\bar{y}_i, 1) \\ (\min(\bar{y}_i, y_i), -) \end{cases} \quad \text{if the } i^{\text{th}} \text{ constraint is } \begin{cases} \geq \\ = \\ \leq \end{cases}.$$

For example, if the restrictions require that θ is DAL, then

$$\hat{\theta}_i = \max(\bar{y}_i, y_i) + \sum_{j=i+1}^k (y_j - \bar{y}_{j-1})^+ w_j w_j^{-1}. \quad (2.10)$$

If h is some integer ($1 < h < k$) and our restrictions require that $\bar{\theta}_1 \leq \bar{\theta}_2 \leq \dots \leq \bar{\theta}_h \geq \bar{\theta}_{h+1} \geq \dots \geq \bar{\theta}_k$, then

$$\hat{\theta}_i = \min(\bar{y}_i, y_i) + \sum_{j=i+1}^k (y_j - \bar{y}_{j-1})^- w_j w_j^{-1} \quad \text{for } i \leq h \quad \text{while}$$

$$\hat{\theta}_i = \max(\bar{y}_i, y_i) + \sum_{j=i+1}^k (y_j - \bar{y}_{j-1})^+ w_j w_j^{-1} \quad \text{for } h < j \leq k.$$

The MLE of θ subject to the requirement that θ is lower starshaped (θ is DAL and $\bar{\theta}_k \geq 0$), say $\tilde{\theta}$, is easily expressed in terms of the $\hat{\theta}$ given in (2.10). In particular,

$$\tilde{\theta}_i = \begin{cases} \hat{\theta}_i & \text{if } \bar{y}_k \geq 0 \\ \hat{\theta}_i - \bar{y}_k & \text{if } \bar{y}_k < 0. \end{cases}$$

The proof of this is a straightforward verification of the properties characterizing projections in Theorem 7.8 of Barlow et al. (1972).

3. RESTRICTED HYPOTHESIS TESTS. Consider the problem of testing the null hypothesis $H_0: \theta_1 = \theta_2 = \dots = \theta_k$ when the parameter vector θ is known to be DAL. In other words, test H_0 against the alternative $H_1 - H_0$ (H_1 but not H_0) where $H_1: \bar{\theta}_1 \geq \bar{\theta}_2 \geq \dots \geq \bar{\theta}_k$. We consider a likelihood ratio statistic which has a surprisingly tractable distribution.

It is well known that the MLE's, under H_0 , are given by

$$\tilde{\theta}_i = \bar{Y} = W_k^{-1} \cdot \sum_{i=1}^k w_i Y_i \quad (\text{free of } \sigma^2); i=1,2,\dots,k.$$

and

$$\tilde{\sigma}^2 = k^{-1} \sum_{i=1}^k \frac{(Y_i - \bar{Y})^2}{a_i}.$$

Since the MLE of the vector θ which satisfies H_1 does not depend on σ^2 , it follows that the MLE's under H_1 are $\hat{\theta}$, as derived in Section 2 and

$$\hat{\sigma}^2 = k^{-1} \sum_{i=1}^k \frac{(Y_i - \hat{\theta}_i)^2}{a_i}.$$

The likelihood ratio, Λ , is then given by

$$\Lambda = \left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \right)^{k/2}.$$

Now

$$\begin{aligned} \frac{\tilde{\sigma}^2}{\sigma^2} &= \sum_{i=1}^k (Y_i - \bar{Y})^2 w_i \\ &= \sum_{i=1}^k (Y_i - \hat{\theta}_i)^2 w_i + \sum_{i=1}^k (\hat{\theta}_i - \bar{Y})^2 w_i + 2 \sum_{i=1}^k (Y_i - \hat{\theta}_i)(\hat{\theta}_i - \bar{Y}) w_i. \end{aligned} \quad (3.1)$$

The class of vectors which satisfy H_1 form a closed convex cone which contains all the constant functions. The vector $\hat{\theta}$ is the weighted least squares projection of the vector Y onto this cone so that

$$\sum_{i=1}^k (Y_i - \hat{\theta}_i) \hat{\theta}_i w_i = \sum_{i=1}^k (Y_i - \hat{\theta}_i) \bar{Y} w_i = 0$$

(see Barlow et al. (1972), page 318). Thus, the last term in (3.1) is zero and a likelihood ratio test rejects H_0 in favor of H_1 for large values of

$$Q = 1 - \Lambda^{2/k} = \frac{\sum_{i=1}^k (\hat{\theta}_i - \bar{Y})^2 w_i}{\sum_{i=1}^k (Y_i - \bar{Y})^2 w_i}. \quad (3.2)$$

Suppose that $i_1, i_2, \dots, i_m$ are those indices where $\bar{\hat{\theta}}_{i-1} = \bar{\hat{\theta}}_i$ when $\hat{\theta}$ is obtained (i.e., those indices for which $\bar{Y}_{i-1} \leq \bar{Y}_i$). Using (2.4) and (2.5) we can write

$$\begin{aligned} \sum_{i=1}^k (Y_i - \hat{\theta}_i)^2 w_i &= \sum_{a=1}^m \left[Y_{i_a} - \bar{Y}_{i_a} - \sum_{b=a+1}^m (Y_{i_b} - \bar{Y}_{i_b-1}) \cdot w_{i_b} w_{i_b}^{-1} \right]^2 w_{i_a} \\ &\quad + \sum_{a=1}^m \left[\sum_{b=a}^m (Y_{i_b} - \bar{Y}_{i_b-1}) w_{i_b} w_{i_b}^{-1} \right]^2 (w_{i_a-1} - w_{i_a-1}). \\ &= \sum_{a=1}^m \left[Y_{i_a} - \bar{Y}_{i_a} - S_{a+1} \right]^2 w_{i_a} + \sum_{a=1}^m S_a^2 (w_{i_a-1} - w_{i_a-1}) \end{aligned}$$

where $S_a = \sum_{b=a}^m (Y_{i_b} - \bar{Y}_{i_b-1}) w_{i_b} w_{i_b}^{-1}$. Adding the first term from each of the above sums and using (2.7), we obtain

$$\begin{aligned}
& [(Y_{i_1} - \bar{Y}_{i_1-1})W_{i_1-1}W_{i_1}^{-1}S_2]^2 W_{i_1} + [(Y_{i_1} - \bar{Y}_{i_1-1})W_{i_1}W_{i_1-1}^{-1}S_2]^2 W_{i_1-1} \\
& = (Y_{i_1} - \bar{Y}_{i_1-1})^2 W_{i_1} W_{i_1-1} W_{i_1}^{-1} + W_{i_1} S_2^2.
\end{aligned}$$

The term, $W_{i_1} S_2^2$, cancels part of the second term of the second sum. Proceeding, using similar reasoning we obtain

$$\sum_{i=1}^k (Y_i - \hat{\theta}_i)^2 W_i = \sum_{a=1}^m (Y_{i_a} - \bar{Y}_{i_a-1})^2 W_{i_a} W_{i_a-1} W_{i_a}^{-1}. \quad (3.3)$$

A fairly simple induction yields

$$\sum_{i=1}^k (Y_i - \bar{Y})^2 W_i = \sum_{i=2}^k (Y_i - \bar{Y}_{i-1})^2 W_i W_{i-1} W_i^{-1}. \quad (3.4)$$

Using (3.3) and (3.4) in our expression for Q we obtain

$$Q = \frac{\sum_{i \in I} Z_i^2}{\sum_i Z_i^2}$$

where $I = \{i_1, i_2, \dots, i_m\}$ and $Z_i = (Y_i - \bar{Y}_{i-1})W_i^{1/2} W_{i-1}^{1/2} W_i^{-1/2}$; $i = 2, 3, \dots, k$. Under H_0 , $Z_2, Z_3, \dots, Z_k$ are independent, standard normal random variables so that for a given set of indices, $I = \{i_1, i_2, \dots, i_m\}$, Q has a beta distribution with parameters $(k-m-1)/2$ and $m/2$.

Now, suppose $I \subset \{2, 3, \dots, k\} = I_0$ and that E_I is the event $\{Z_i \geq 0; i \in I \text{ and } Z_i < 0; i \notin I\}$. If we let $T_I = \sum_{i \in I} Z_i^2$, then we may write

$$\begin{aligned}
P[Q \geq t, E_I] &= P\left[\frac{T_{I_0-1}}{T_{I_0}} \geq t \mid E_I\right] P(E_I) \\
&= P[B_{(k-m-1)/2, m/2} \geq t] \cdot (1/2)^{k-1}
\end{aligned}$$

since T_{I_0-1}/T_{I_0} is independent of E_I .

If we partition the event $[Q \geq t]$ by intersecting it with all such events E_I and then collect terms, we obtain the following theorem.

Theorem 3.1. In testing $H_0: \theta_1 = \theta_2 = \dots = \theta_k$ against the alternative $H_1 - H_0$ where H_1 specifies that the parameter vector θ is DAL the likelihood ratio statistic $Q = 1 - \Lambda^{2/k} = (\sum_{i=1}^k (\hat{\theta}_i - \bar{Y})^2 w_i) / (\sum_{i=1}^k (Y_i - \bar{Y})^2 w_i)$ has a null hypothesis distribution given by

$$P[Q \geq t] = \sum_{m=0}^{k-1} \binom{k-1}{m} (1/2)^{k-1} P[B_{(k-m-1)/2, m/2} \geq t] \quad (3.5)$$

for all t , where $B_{\alpha, \beta}$ denotes a standard beta random variable with parameters α and β and $B_{0, \beta}$ ($B_{\alpha, 0}$) is taken to be degenerate at 0 (1) when $\beta > 0$ ($\alpha > 0$).

We note that the distribution of Q given in (3.5) is exactly the same as the distribution of

$$\frac{\sum_{i=1}^{k-1} (Z_i \wedge 0)^2}{\sum_{i=1}^{k-1} Z_i^2} \quad (3.6)$$

where $Z_1, Z_2, \dots, Z_{k-1}$ are independent standard normal random variables.

If, in fact, H_0 is not true then we can replace Y_i by $Y_i - \theta_i$ in our expression for Q and say that Q is distributed as

$$\frac{\sum_{i=1}^{k-1} [(Z_i + \delta_i) \wedge 0]^2}{\sum_{i=1}^{k-1} (Z_i + \delta_i)^2} = \frac{\sum_{i=1}^{k-1} [(Z_i + \delta_i) \wedge 0]^2}{\sum_{i=1}^{k-1} [(Z_i + \delta_i) \wedge 0]^2 + \sum_{i=1}^{k-1} [(Z_i + \delta_i) \vee 0]^2} \quad (3.7)$$

where $Z_1, Z_2, \dots, Z_{k-1}$ are independent standard normal random variables and

$$\delta_i = (\theta_{i+1} - \bar{\theta}_i) w_{i+1}^{1/2} w_i^{1/2} w_{i+1}^{-1/2}; \quad i = 1, 2, \dots, k-1. \quad (3.8)$$

Note that if the parameter vector θ satisfies H_1 then $\delta_i \geq 0$; $i = 1, 2, \dots, k-1$. Moreover if $\delta_i \geq 0$ then $[(Z_i + \delta_i) \wedge 0]^2 \geq [Z_i \wedge 0]^2$ and $[(Z_i + \delta_i) \vee 0]^2 \leq [Z_i \vee 0]^2$. This implies that (3.7) $\geq$ (3.6) and shows that our likelihood ratio test is unbiased.

Of course the same distribution theory holds for testing $H_0: \theta_1 = \theta_2 = \dots = \theta_k$ against the alternative $H_1 - H_0$ if H_1 specifies that θ is IAL, IAR or DAR. It is perhaps somewhat surprising to note that we may intermix the inequality signs. For example, the same distribution theory would hold for testing H_0 against $H_1 - H_0$ if H_1 specified that $\theta_1 \leq \theta_2 \leq \dots \leq \theta_h \geq \theta_{h+1} \geq \dots \geq \theta_k$ for some value of h .

If σ^2 is known, a likelihood ratio test of H_0 against $H_1 - H_0$ would reject for large values of

$$R = -2 \ln \Lambda = \sum_{i=1}^k (\hat{\theta}_i - \bar{Y})^2 w_i.$$

The distribution of R is the same as that of

$$\sum_{i=1}^{k-1} [(Z_i + \delta_i) \wedge 0]^2 \quad (3.9)$$

where Z_i and δ_i ; $i = 1, 2, \dots, k-1$ are defined as before. If the null hypothesis is satisfied then $\delta_i = 0$; $i = 1, 2, \dots, k-1$ and if H_1 is satisfied then $\delta_i \geq 0$; $i = 1, 2, \dots, k-1$. Thus, the test is unbiased and the null hypothesis distribution of R is given by

$$P[R \geq t] = \sum_{m=0}^{k-1} \binom{k-1}{m} (1/2)^{k-1} P[\chi_m^2 \geq t] \quad (3.10)$$

where χ_m^2 denotes a standard chi-square random variable with m degrees freedom ($\chi_0^2 \equiv 0$).

Consider the problem of testing H_1 as a null hypothesis. If σ^2 is known then a likelihood ratio test rejects H_1 for large values of

$$R' = \sum_{i=1}^k (Y_i - \hat{\theta}_i)^2 w_i.$$

The random variable R' is distributed as

$$\sum_{i=1}^{k-1} [(Z_i + \delta_i) \vee 0]^2 \quad (3.11)$$

and $\delta_i \leq 0$ if H_1 is true. Thus $[(Z_i + \delta_i) \vee 0]^2 \leq [Z_i \vee 0]^2$ and

$$\sup_{\theta \in H_1} P_{\theta}[R' \geq t] = P_{H_0}[R' \geq t] = \sum_{m=0}^{k-1} \binom{k-1}{m} (1/2)^{k-1} P[\chi_m^2 \geq t]. \quad (3.12)$$

If σ^2 is unknown, we cannot estimate σ^2 in the denominator of the likelihood ratio since we have only one item from each population. Thus, in this case we cannot construct a likelihood ratio test.

Assume we have a random sample of size n from each of our k normal populations and let $\bar{X}_i$ denote the mean of the items of the sample corresponding to the population with mean θ_i ; $i=1, 2, \dots, k$. The maximum likelihood estimators subject to the constraints, (2.1), are given by (2.4) and (2.5) with y_j replaced by $\bar{X}_j$.

If σ^2 is unknown, a likelihood ratio test rejects for large values of

$$Q = 1 - A^{2/N} = \frac{\sum_{i=1}^k (\hat{\theta}_i - \bar{X})^2 w_i \cdot n}{\sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 w_i + \sum_{i=1}^k (\bar{X}_i - \hat{\theta}_i)^2 w_i \cdot n + \sum_{i=1}^k (\hat{\theta}_i - \bar{X})^2 w_i \cdot n}$$

where $\bar{X} = k^{-1} \sum_{i=1}^k \bar{X}_i$. Since the $X_{ij} - \bar{X}_i$ are independent of the $\bar{X}_i$ and since $\sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 w_i / n$ has a chi-square distribution with $k(n-1)$ degrees of freedom it follows that Q has the same distribution as

$$\frac{\sum_{i=1}^{k-1} [(Z_i + \delta_i) \wedge 0]^2}{\sum_{i=1}^{k-1} (Z_i + \delta_i)^2 + \sum_{i=k}^{N-1} Z_i^2}$$

where $N = k \cdot n$, $Z_1, Z_2, \dots, Z_{N-1}$ are independent standard normal variables, and $\sqrt{n} \delta_i$ is defined by (3.8). The following theorem is a consequence of this representation.

Theorem 3.2. The likelihood ratio test of H_0 against $H_1 - H_0$ based upon Q is unbiased and the null hypothesis distribution of Q is given by

$$P[Q \geq t] = \sum_{m=0}^{k-1} \binom{k-1}{m} (1/2)^{k-1} P[B_{(k-m-1)/2, (N-k+m)/2} \geq t].$$

If σ^2 is known then the likelihood ratio statistic $R = -2 \ln \Lambda$ gives rise to an unbiased test and its null hypothesis distribution is given by (3.10).

In testing H_1 as a null hypothesis when σ^2 is known the likelihood ratio statistic $R' = -2 \ln \Lambda = \sum_{i=1}^k (\bar{X}_i - \hat{\theta}_i)^2 \cdot n w_i$ has a distribution which may be represented as in (3.11). Thus, the null hypothesis distribution is as in (3.12).

If the common sample size, n , is larger than one, we can test H_1 as a null hypothesis when σ^2 is unknown. In testing H_1 against $\sim H_1$, reject H_1 for large values of

$$Q' = 1 - \Lambda^{2/N} = \frac{\sum_{i=1}^k (\bar{X}_i - \hat{\theta}_i)^2 \cdot n w_i}{\sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 w_i + \sum_{i=1}^k (\bar{X}_i - \hat{\theta}_i)^2 \cdot n w_i + \sum_{i=1}^k (\hat{\theta}_i - \bar{X})^2 \cdot n w_i} \quad (3.13)$$

It can be argued that Q' is distributed as

$$\frac{\sum_{i=1}^{k-1} [(Z_i + \delta_i) \vee 0]^2}{\sum_{i=1}^{k-1} (Z_i + \delta_i)^2 + \sum_{i=k}^{N-1} Z_i^2}$$

where $\sqrt{n}\delta_i$ is defined by (3.8).

The following theorem is a result of this representation.

Theorem 3.3. A likelihood ratio test of H_1 against $\sim H_1$ rejects for large values of Q' (as given in (3.13)), is unbiased and

$$\sup_{\theta \in H_1} P[Q' \geq t] = P_{H_0}[Q' \geq t] = \sum_{m=0}^{k-1} \binom{k-1}{m} (1/2)^{k-1} P[B_{m/2, (N-m-1)/2} \geq t].$$

4. CONCLUDING REMARKS. If we wish to replace " $\leq$ " or " $\geq$ " by "=" in the restriction imposed by our alternative hypothesis, H_1 , the appropriate distribution theory may be found from the results in Section 3 by appropriate adjustments in the degrees of freedom.

We conducted a Monte Carlo study of the power of the test statistic, $R = \sum_{i=1}^k (\hat{\theta}_i - \bar{Y})^2 w_i$, for testing H_0 against $H_1 - H_0$ ($H_1: \theta$ is DAL) when σ^2 is known. Some of the results of that study are given in Table 1. In this study we let $k=5$, $a_i=1$; $i=1,2,\dots,5$ and $\sigma^2=1$. We approximated the power of each of three test statistics at each of 28 parameter vectors, 6. This was done by randomly generating 2000 5-tuples of normal random numbers where the i^{th} entry in the 5-tuple has a normal distribution with

mean θ_i and variance 1. The entries in the table are the fraction of times the test statistic exceeded the .05 critical value computed from its null hypothesis distribution ($\chi^2_{q_4}$ for χ^2 , using (3.10) for R and using Theorem 3.1 in Barlow et al. for $\bar{\chi}^2$). Corresponding to each parameter vector we have given its spacing and a measure, Δ , ($\Delta^2 = \sum_{i=1}^k (\theta_i - \bar{\theta})^2$) of its distance from the null hypothesis H_0 .

The first thirteen θ vectors in the table are decreasing and the power of $\bar{\chi}^2$ is significantly greater than that of either R or χ^2 . However, for these vectors R is significantly more powerful than is χ^2 . The last fifteen θ vectors are DAL but not strictly decreasing. Here R is significantly more powerful than either χ^2 or $\bar{\chi}^2$. Note that χ^2 is more powerful than $\bar{\chi}^2$ for the last six θ entries.

We have been unable to substantially relax the assumption of equal sample sizes in Sections 2 and 3. In other words, this analysis depends very heavily on the assumption that the weights in the restriction imposed by H_1 are proportional to the variances of the sample means. If we relax the assumption of equal sample sizes we must be willing to use weights $w_i = n_i(a_i\sigma^2)^{-1}$ in our restriction, H_1 . This latter approach is the one taken in Shaked (1979).

A different definition of decreasing on the average can be found in the work of Robertson and Wright (1981). They define the parameter vector, θ , to be decreasing on the average (DA) if $i^{-1} \sum_{j=1}^i \theta_j \geq (k-i)^{-1} \sum_{j=i+1}^k \theta_j$; $i=1,2,\dots,k-1$. If θ is DAL or IAR then θ is DA, in the above sense. Robertson and Wright (1981) derive maximum likelihood estimates of a vector of normal means subject to the restriction that it is DA and discuss testing homogeneity of θ when it is assumed to be DA and testing DA as a null

Table 1. Power Functions of χ^2 , $\bar{\chi}^2$ and R.

$(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$	Spacing	Δ	Power of χ^2	Power of $\bar{\chi}^2$	Power of R
(0.0,0.0,0.0,0.0,0.0)	1,1,1,1,1	0	.061	.047	.058
(0.4,0.3,0.2,0.1,0.0)	5,4,3,2,1	.316	.064	.083	.084
(1.0,.75,0.5,.25,0.0)	5,4,3,2,1	.791	.089	.173	.147
(2.0,1.5,1.0,0.5,0.0)	5,4,3,2,1	1.58	.206	.459	.381
(4.0,3.0,2.0,1.0,0.0)	5,4,3,2,1	3.16	.743	.913	.874
(1.0, .9,.8,.7,0.1)	10,9,8,7,1	.707	.082	.146	.118
(2.0,1.8,1.6,1.4,0.2)	10,9,8,7,1	1.41	.172	.359	.286
(5.0,4.5,4.0,3.5,0.5)	10,9,8,7,1	3.54	.838	.950	.926
(.1,.1,.1,.1,0.0)	2,2,2,2,1	.089	.059	.052	.060
(.2,.2,.2,.2,0.0)	2,2,2,2,1	.179	.061	.056	.064
(.5,.5,.5,.5,0.0)	2,2,2,2,1	.447	.073	.081	.079
(1.0,1.0,1.0,1.0,0.0)	2,2,2,2,1	.894	.101	.166	.126
(5.0,5.0,5.0,5.0,0.0)	2,2,2,2,1	4.47	.967	.992	.984
(.3,0.0,.1,0.0,.1)	4,1,2,1,2	.225	.061	.060	.071
(.6,0.0,.2,0.0,.2)	4,1,2,1,2	.490	.071	.083	.101
(1.5,0.0,.5,0.0,.5)	4,1,2,1,2	1.22	.144	.197	.225
(3.0,0.0,1.0,0.0,1.0)	4,1,2,1,2	2.45	.471	.583	.653
(6.0,0.0,2.0,0.0,2.0)	4,1,2,1,2	4.90	.988	.989	.997
(.4,.3,0.0,.1,.2)	(5,4,1,2,3)	.316	.063	.066	.079
(.8,.6,0.0,.2,.4)	(5,4,1,2,3)	.632	.075	.098	.110
(2.4, 1.8,0.0,.6,1.2)	(5,4,1,2,3)	1.90	.296	.397	.446
(4.8,3.6,0.0,1.2,2.4)	(5,4,1,2,3)	3.79	.880	.913	.946
(.2,0.0,.1,.1,.1)	(3,1,2,2,2)	.141	.059	.052	.063
(.4,0.0,.2,.2,.2)	(3,1,2,2,2)	.283	.064	.058	.070
(1.2,0.0,.6,.6,.6)	(3,1,2,2,2)	.849	.099	.095	.128
(2.4,0.0,1.2,1.2,1.2)	(3,1,2,2,2)	1.70	.244	.221	.329
(4.8,0.0,2.4,2.4,2.4)	(3,1,2,2,2)	3.39	.795	.696	.867
(7.2,0.0,3.6,3.6,3.6)	(3,1,2,2,2)	5.09	.994	.969	.997

hypothesis. They assume that σ^2 is known and consider likelihood ratio statistics whose null hypothesis distributions are chi-bar-square distributions.

The maximum likelihood estimates of mortality rates discussed in Section 1 subject to the restrictions that they are IAL and DAR are given in Figures 2 and 3. Surely, actuaries would feel that either one of these estimates requires additional smoothing. However, they both give an indication of the "bump" at age 20 and either one (or their average) might provide a better starting point than the crude mortality rate or its "isotonic regression" (which is oversmoothed) for the graduation.

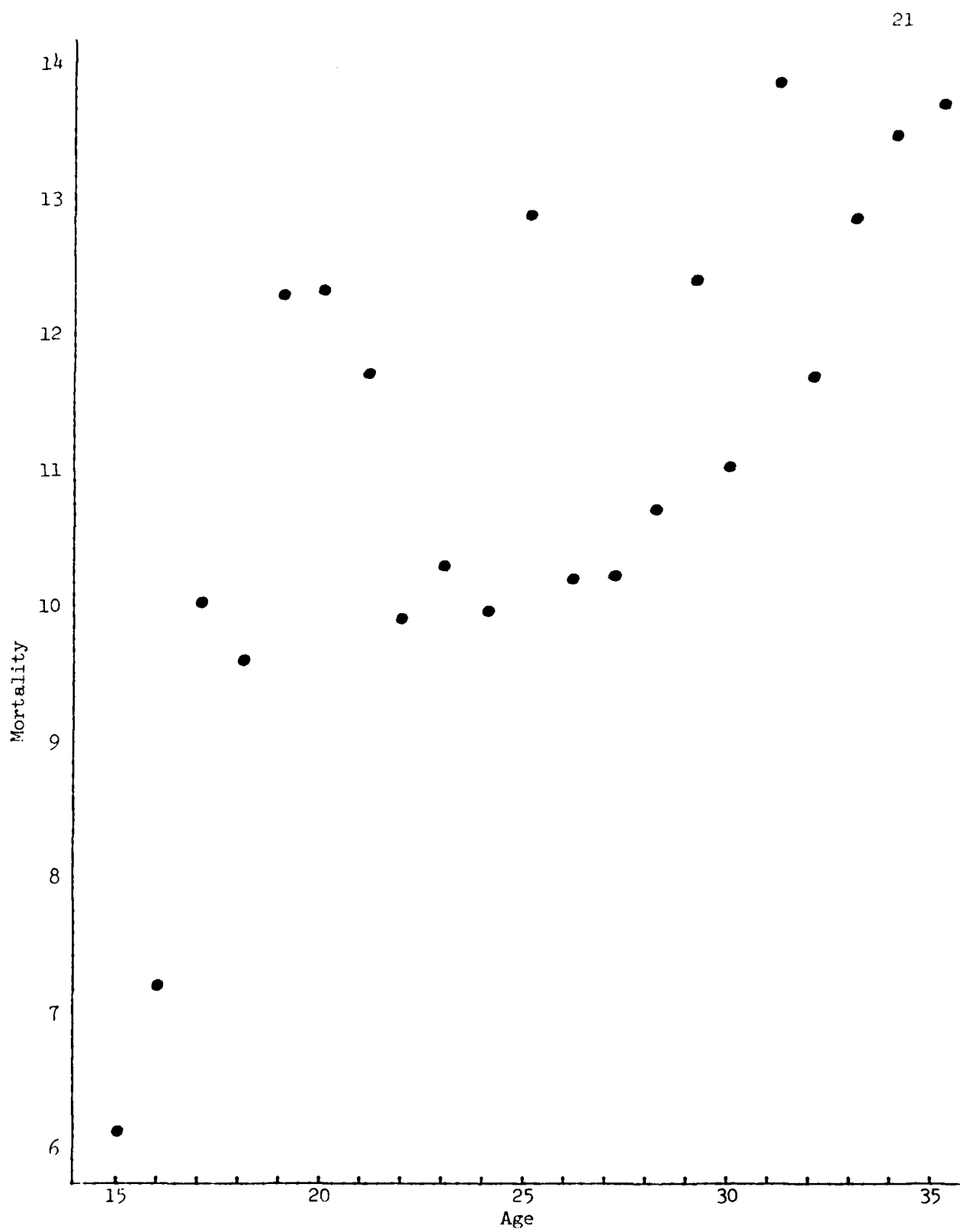


Figure 2. Mortalities Smoothed to be IAL.

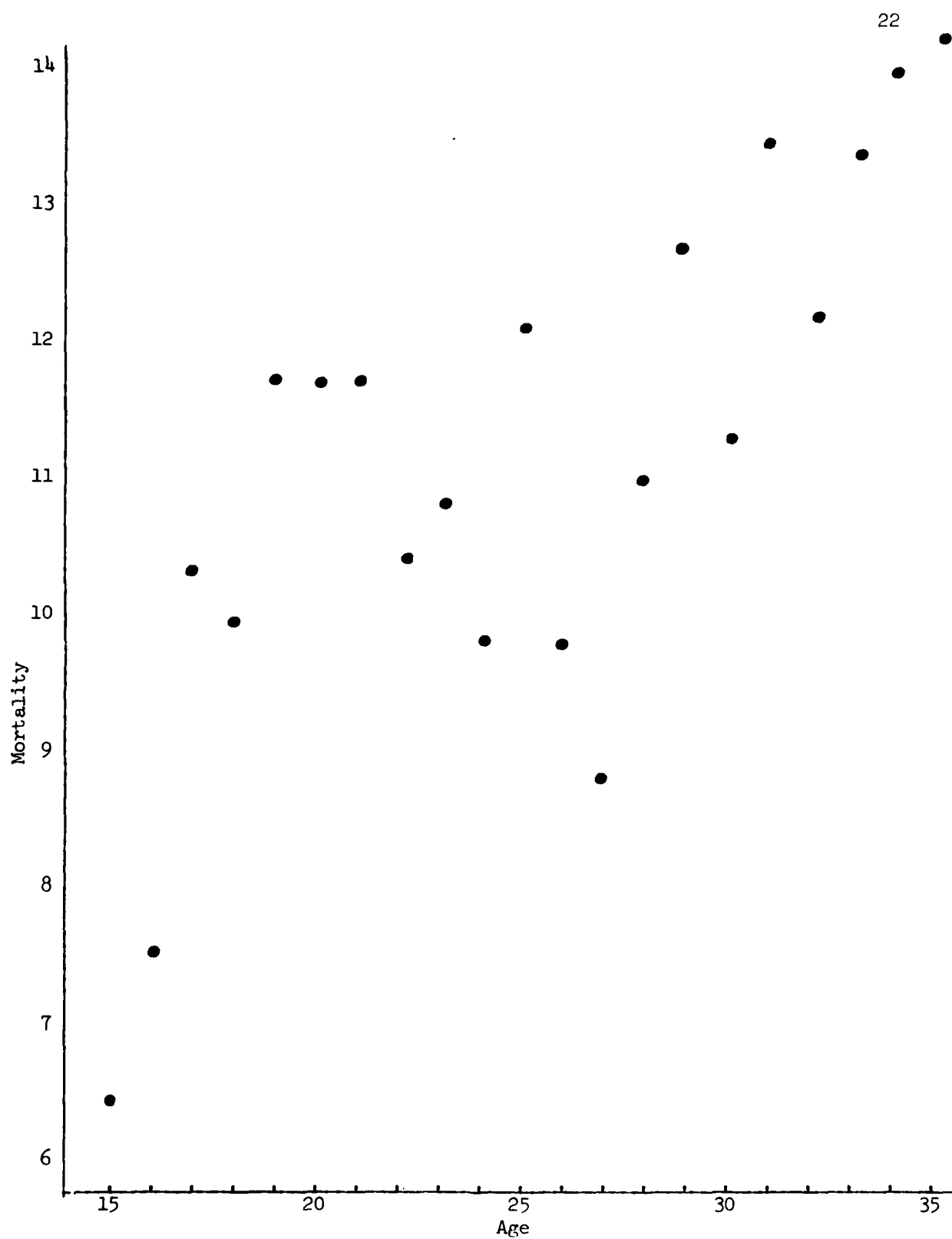


Figure 3. Mortalities Smoothed to be DAR.

REFERENCES

- Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972).
Statistical Inference Under Order Restrictions, John Wiley and Sons,
New York.
- Robertson, Tim and Wright, F.T. (1980). On measuring the conformity of a
parameter set to a trend, with applications, University of Iowa,
Department of Statistics, Technical Report No. 71.
- Shaked, Moshe (1979). Estimation of starshaped sequences of Poisson and
normal means, Ann. Statist. 7, 729-741.
- Translations of the Society of Actuaries, 1973 Reports of Mortality and
Morbidity Experience, published July 1974.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 9	2. GOVT ACCESSION NO. AD-A108 832	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On testing monotone tendencies	5. TYPE OF REPORT & PERIOD COVERED Technical report	
	6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s) Richard L. Dykstra and Tim Robertson	8. CONTRACT OR GRANT NUMBER(s) N00014-78-C-0655 N00014-80-C-0321	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Dept. of Statistics Dept. of Statistics Univ. of Missouri University of Iowa Columbia, MO Iowa City, IA 52242		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Prog. - Code 436 Arlington, VA 22217	12. REPORT DATE October 30, 1981	
	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) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION 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) Order restricted inference, starshaped, decreasing on the average, chi-bar-square distribution, E-bar-square distribution.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) In certain problems, it may be expected that a regression function has a substantial overall tendency to be monotone and yet we may not be certain that all of the restrictions imposed by a simple order are satisfied. Distribution theory for likelihood ratio tests of homogeneity of a collection		

DD FORM 1473

EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

of normal means when the collection is "decreasing on the average" and for testing "decreasing on the average" as a null hypothesis, is presented. The restriction "decreasing on the average" is less restrictive than the usual monotone restriction and allows the data to give rise to "reversals" over short ranges of values of the parameter set. It is closely related to the "starshaped ordering" restriction discussed in Shaked (Ann. Statist. (1979)).

