79-NI-AX-0068

DEVELOPING IMPROVED TECHNIQUES FOR EVALUATING CORRECTIONAL PROGRAMS

Michael D. Maltz
Center for Research in Criminal Justice
University of Illinois at Chicago Circle

89250

Center for Research in Criminal Justice

University of Illinois at Chicago Circle
Box 4348 Chicago 60680

DEVELOPING IMPROVED TECHNIQUES FOR EVALUATING CORRECTIONAL PROGRAMS

Michael D. Maltz
Center for Research in Criminal Justice
University of Illinois at Chicago Circle

May 1981

Final Report submitted to the National Institute of Justice on Grant No. 79-NI-AX-0068. Points of view or opinions expressed herein are those of the authors and do not necessarily reflect the official position or policies of the National Institute of Justice or the U.S. Department of Justice.

This report describes the nature of the problems we have studied and our progress to date. A number of papers and memoranda give the technical details of the research; they are included as appendices.

Since the complete development of a new statistical technique is a lengthy process, not all of the research is complete. However, our efforts will continue beyond the funding provided by this contract. A number of papers, based on the appendices, have been or will be submitted for publication.

Our proposal described four areas of research involving mathematical models of criminal recidivism:

1. estimation of the model's parameters, and associated statements, using maximum likelihood and Bayesian procedures for:
   a. the split population model; and
   b. the mixed exponential and Weibull models;
2. investigation of ways to select appropriate models of recidivism from among candidate models;
3. development of covariate models of recidivism; i.e., descriptions by which the recidivism probabilities of each member of the group under study is determined by parameters, based on his own unique characteristics;
4. critical analysis of certain pretest - posttest designs in evaluating delinquency programs.

These four areas of research, and our progress in each, are described below. Future research is suggested in the conclusion of this report.

I. ESTIMATION

Estimation procedures for the split population model were developed in an earlier contract, using both the continuous version

$$F_c(t) = \gamma(1 - \exp(-\phi t)) \qquad (1)$$

and the equivalent discrete version

$$F_d(i) = \gamma(1 - q^i) \qquad (1)$$

where $F_c(t)$ $[F_d(i)]$ is the probability of recidivism at or before time t [interval i]. We have used the discrete version in estimating confidence intervals for the split population model, and the continuous version for the population mixture model. The continuous version is more appropriate for a mixed Weibull distribution because the model parameters are more easily estimated than with the discrete model. Procedures for estimation of the parameters and their associated confidence intervals for a mixed Weibull distribution are developed in Appendices A-E.

For the split population model the discrete version was used for a number of reasons:

1. The data often are presented in discrete form -- e.g., number failing in month 1, month 2, etc.
2. Little if any information is lost, since the time interval used (months) is small enough to capture the essence of the data.
3. If large data sets are analyzed, computation is appreciably easier when using data grouped by months

than using the failure or exposure time of every individual in the group.

4. Since γ and q are both probabilities, the only allowable values of γ and q lie in the unit square $0 \leq \gamma \leq 1$, $0 < q < 1$; this makes their interpretation, and their visualization and presentation of their confidence intervals, much easier.
5. Correctional officials are likely to be more comfortable dealing with the discrete model, with a constant conditional failure probability q than the continuous model with an exponential failure process.

The programs used in estimating the parameters and in computing the confidence intervals are given in Appendix F.

Standard statistical practice for producing confidence intervals is to assume that the maximum likelihood estimator is asymptotically normally distributed, and to use the information matrix to estimate the variance-covariance matrix. However, the parameters γ and q are defined only on the unit square. This restriction produces extreme non-normality for many cases of interest -- even those with large sample sizes -- thus precluding us from using this standard technique for estimating variance.

To understand this problem more fully, we made simulations of cohorts of different sizes, in which all of the members of the cohort have fixed, given values of γ and q, and fail accordingly. We then plotted the distribution of the resulting maximum likelihood estimates, γ and q, as a basis for forming confidence regions. More important, however, as can be seen from Figures 1-8, the (normalized) likelihood

function is often an excellent representation of the joint density function of γ and q. Based on this (admittedly limited) empirical analysis we concluded that the likelihood function could be used to generate confidence regions for the likelihood estimates γ and q, and thus also supported the more general use of the Bayes estimates and confidence intervals.

II. MODEL SELECTION

One of the more difficult problems of statistical analysis is the development of selection criteria for choosing an appropriate model of the process under study. Although non-parametric models avoid this difficulty, this approach often ignores prior information about the nature of the process that can be used to furnish insight about the process. Since the reason for statistical analysis in the first place is to gain insight, we feel that this is a short-sighted approach.

Models of recidivism other than the ones we studied have been suggested and analyzed. We have prepared a paper for publication on a comparison of models; it forms Appendix G. This paper will be revised and submitted for publication.

III. COVARIATE ANALYSIS

The implicit assumption in the foregoing is that the failure times of the cohort can be characterized by a model based on aggregate properties of the cohort; that is, by γ and q. When comparing different correctional programs, or the effect of the same correctional program on different types of individuals, this is a reasonable assumption. However, it is also possible to study the way that the parameters γ and q are affected by characteristics of the individuals. That is, we may assume that each individual i has unique values γ_i and q_i, and that

$$\gamma_i = g\,(x_{i1}, x_{i2},\ldots x_{ik})$$

$$q_i = h\,(x_{i1}, x_{i2},\ldots x_{ik}) \tag{1}$$

where x_{ij} is the value of the jth characteristics for individual i. We have used standard analytic techniques to investigate such relationships, using data sets obtained by Georgia, Iowa, North Carolina, and the U.S. Bureau of Prisons.

This aspect of our research is described in Appendix H. The results have not been encouraging. First, the maxima are relatively flat; that is, the solution is not very sensitive to relatively large changes in the x's. Although multivariate analysis employing many variables may appear intellectually attractive, it does not lead to insights into post-release behavior. We found that sufficient insight was furnished by looking only at the marginals and two-way crosstabulations, and going beyond that increased complexity and computation cost with no commensurate increase in explanatory power.

Second, the four different data sets we used have resulted in four different relationships between the model parameters and the x's. This finding supports the contention we made in our report from an earlier grant: comparing recidivism rates across jurisdictions is meaningless because they have different release criteria, different laws and policies, and different definitions of recidivism. In addition, because of the degree of diversity we found in the relationships, the notion that there is one single underlying structure that will be valid for all jurisdictions is open to question.

IV. PRETEST-POSTTEST DESIGNS

We had initially intended to continue the work described in two previous papers (Maltz & Pollock, "Artificial Inflation of a Delinquency Rate by a Selection Artifact," <u>Operations Research</u> <u>28</u>, 3, May-June, 1980, 547-559; Maltz, Gordon, McDowall & McCleary, "An Artifact in Pretest-Posttest Designs: How It Can Mistakenly Make Delinquency Programs Look Effective," <u>Evaluation Review</u> <u>4</u>, 2, April, 1980, 225-240). However, the release of a book by the authors of the study we criticized (Murray & Cox, <u>Beyond Probation</u>, Sage Publications, 1980) effectively prevented us from doing so. In their book they tried to explain away the regression artifact, thus promoting their conclusion that it was the correctional program and not possible selection artifacts that caused a 70 percent decline in post-release arrests of juveniles. This finding has led many correctional policy-makers to push for a hard-line approach to juvenile corrections, citing scientific justification.

The finding is wrong. Rather than extend our previous work we decided to explain the limitations of the Murray-Cox finding in greater detail, and to a different audience. Appendix I is the result of this effort.

V. FURTHER RESEARCH

Although we have accomplished most of what we proposed to do in our grant application, we have still not exhausted the basic and applied research problems in this area. Additional areas of research include:

1. The investigation of properties of the joint density function. Although it is not always normally distributed, when it is one can use standard tables for estimating confidence intervals and for testing for significance. The region where it can adequately be approximated by a bivariate normal distribution should be determined. In addition, the form of the function should be investigated to determine if other standardized functions can be used to approximate it.
2. Investigation of the cases when $\gamma = 1$. As can be seen from Figures 1-8, under some circumstances a number of the MLE solutions are on the line $\gamma = 1$. It may be possible to estimate the fraction of solutions on this border using some simple relationship between, say, the height of the likelihood function at the border and the height at the maximum.
3. Investigation of the effect of distributions on γ and q. It may be that the model's parameters for each individual i obey the relationships (1) above, but we have not included the appropriate characteristics x_i (or are unable to measure them accurately). Suppose that each individual in the population under study has parameters γ and q drawn from a known distribution. We can calculate

the mean values ($\bar{\gamma}$ and $\bar{q}$) of these γs and qs; we can also use these γs and qs to generate failure and exposure times, and then use these times to estimate $\hat{\gamma}$ and $\hat{q}$, and MLE values. What is the relationship between $\bar{\gamma}$, $\bar{q}$ and $\hat{\gamma}$, $\hat{q}$, and their variances? For some distributions it may be possible to determine this relationship analytically.

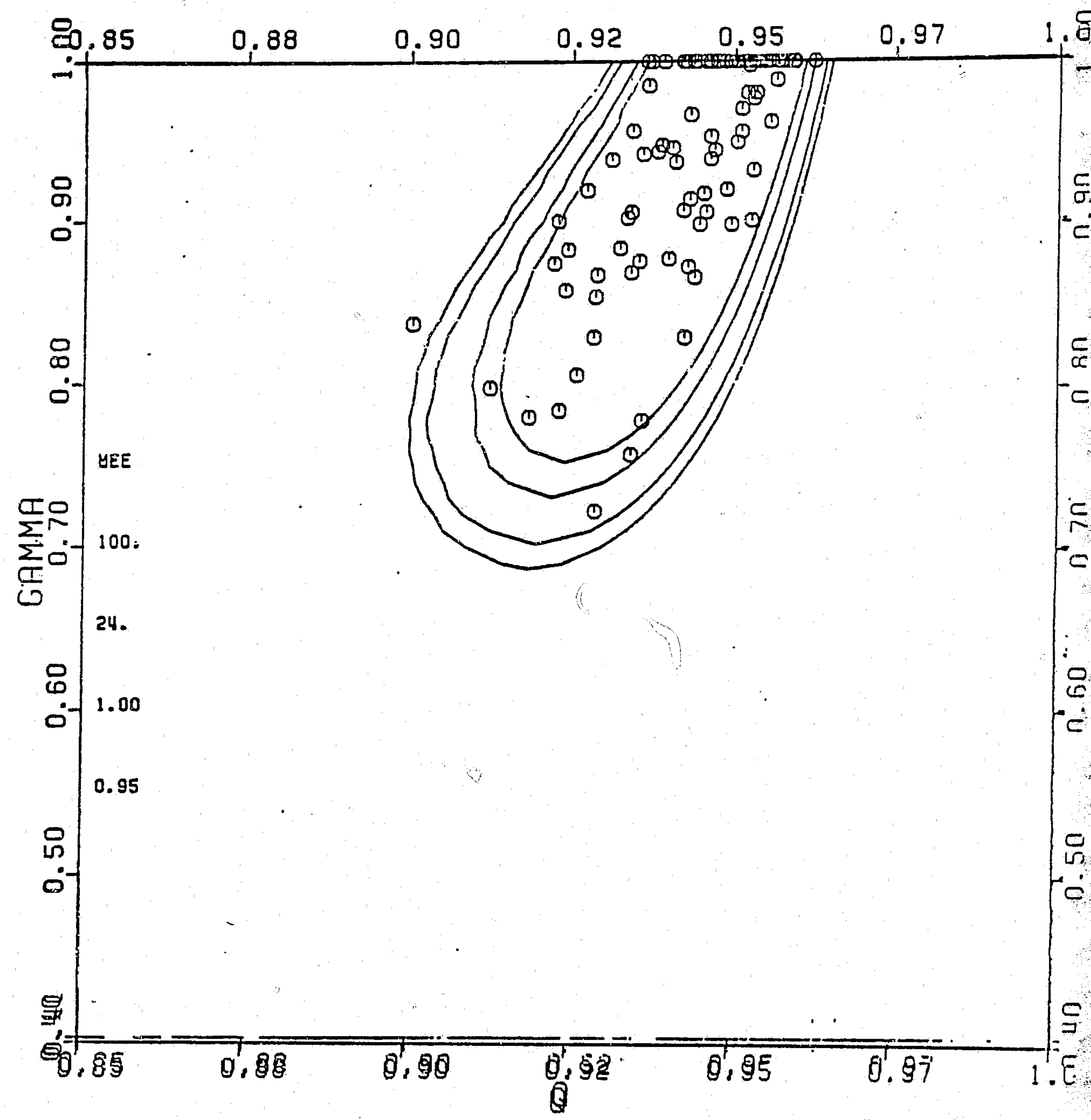


FIGURE 1. Maximum likelihood solutions (circles) of 100 simulations of cohorts with the following characteristics:

N: 100 subjects
τ: 24 months censoring time
γ: 1.0 , probability of eventually failing
q: 0.95 , probability of surviving to the next month

Likelihood function contours superimposed on these solutions, using confidence levels of 0.99, 0.98, 0.95, and 0.90.

Blurring of coordinates due to shrinkage of transparent overlay of contours.

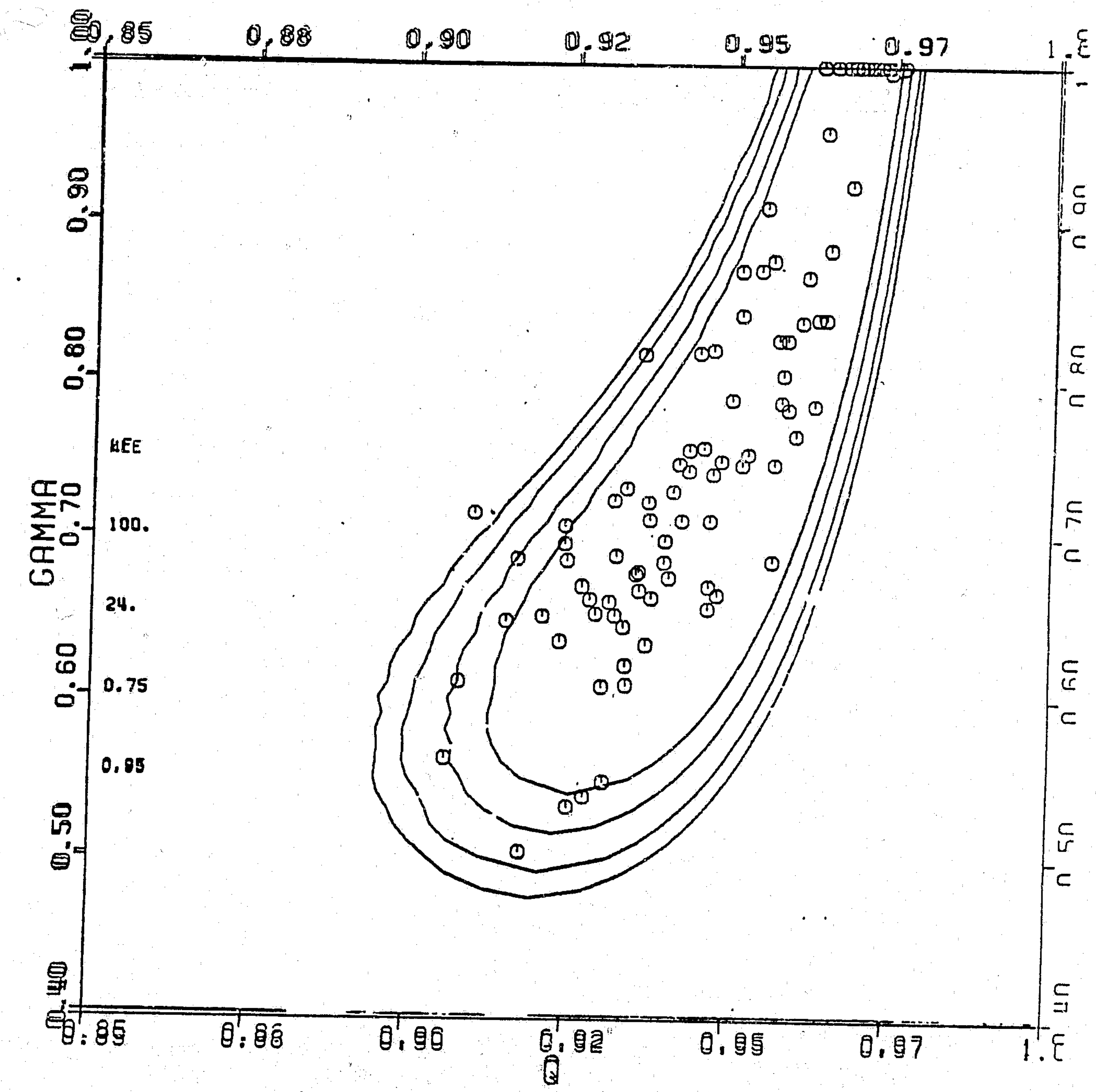


FIGURE 2. Maximum likelihood solutions (circles) of 100 simulations of cohorts with the following characteristics:

N: 100 subjects
τ: 24 months censoring time
γ: 0.75 , probability of eventually failing
q: 0.95 , probability of surviving to the next month

Likelihood function contours superimposed on these solutions, using confidence levels of 0.99, 0.98, 0.95, and 0.90.

Blurring of coordinates due to shrinkage of transparent overlay of contours.

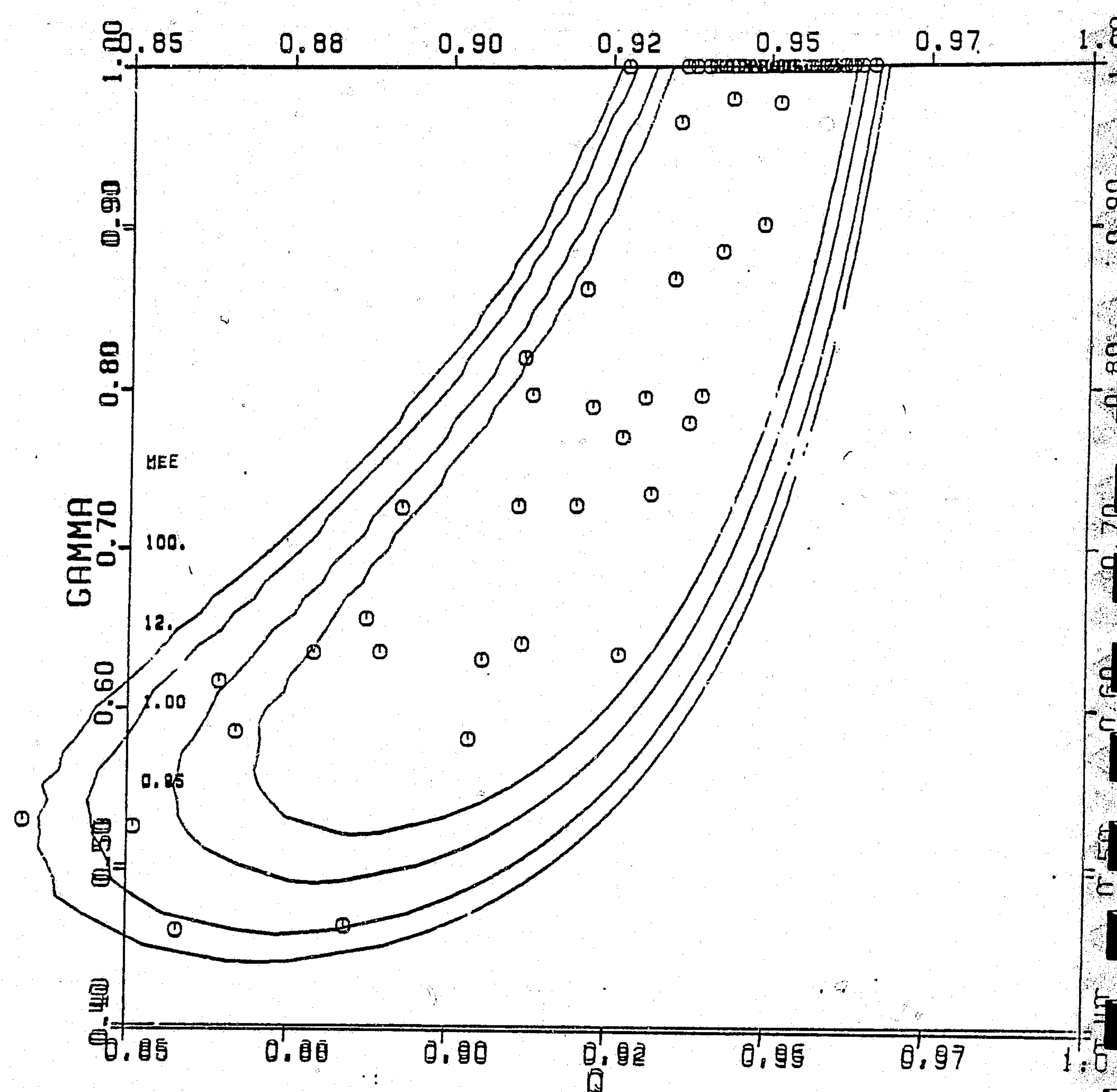


FIGURE 3. Maximum likelihood solutions (circles) of 100 simulations of cohorts with the following characteristics:

N: 100 subjects
τ: 12 months censoring time
γ: 1.0 , probability of eventually failing
q: 0.95 , probability of surviving to the next month

Likelihood function contours superimposed on these solutions, using confidence levels of 0.99, 0.98, 0.95, and 0.90.

Blurring of coordinates due to shrinkage of transparent overlay of contours.

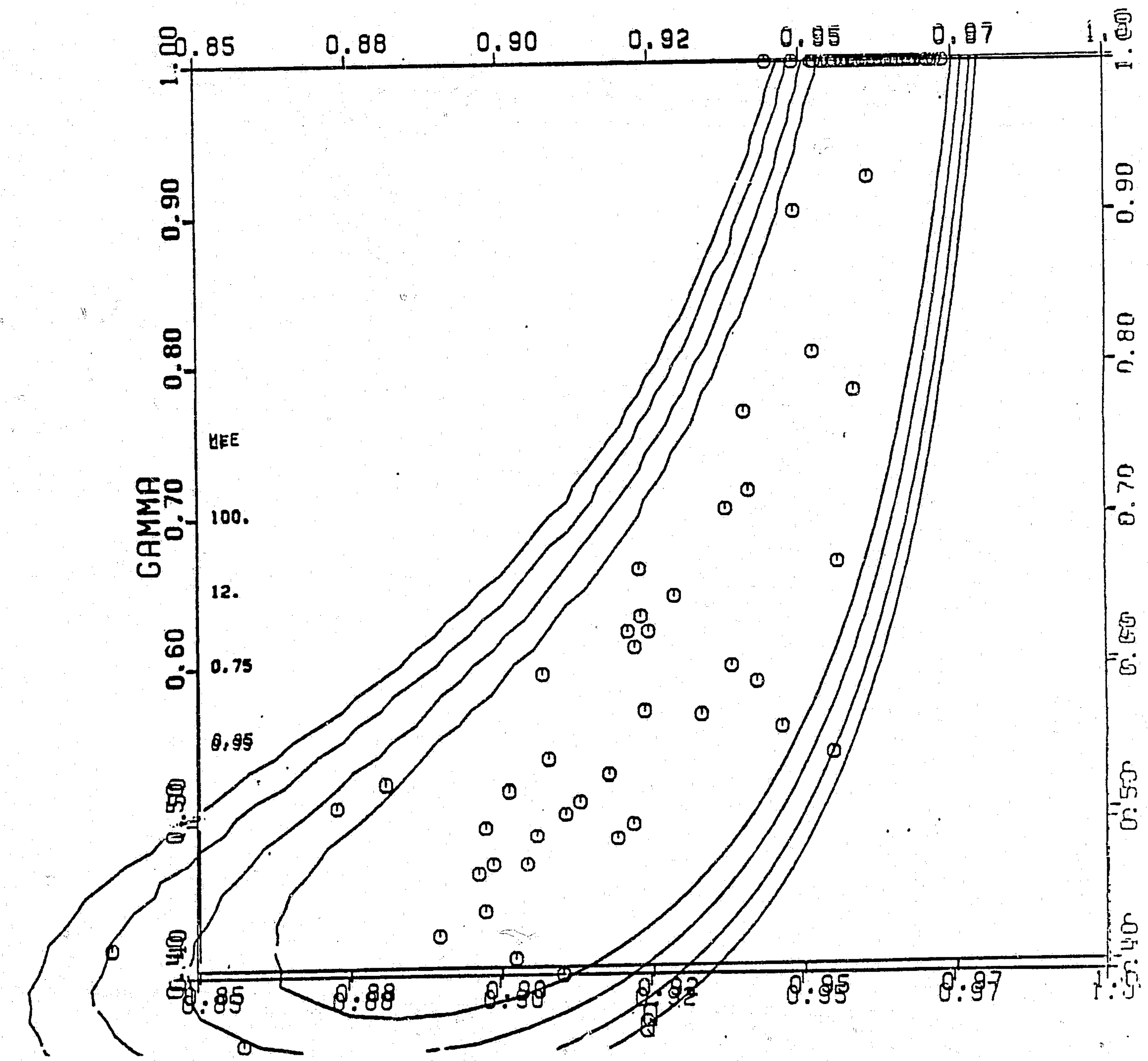


FIGURE 4. Maximum likelihood solutions (circles) of 100 simulations of cohorts with the following characteristics:

N: 100 subjects
τ: 12 months censoring time
γ: 0.75 , probability of eventually failing
q: 0.95 , probability of surviving to the next month

Likelihood function contours superimposed on these solutions, using confidence levels of 0.99, 0.98, 0.95, and 0.90.

Blurring of coordinates due to shrinkage of transparent overlay of contours.

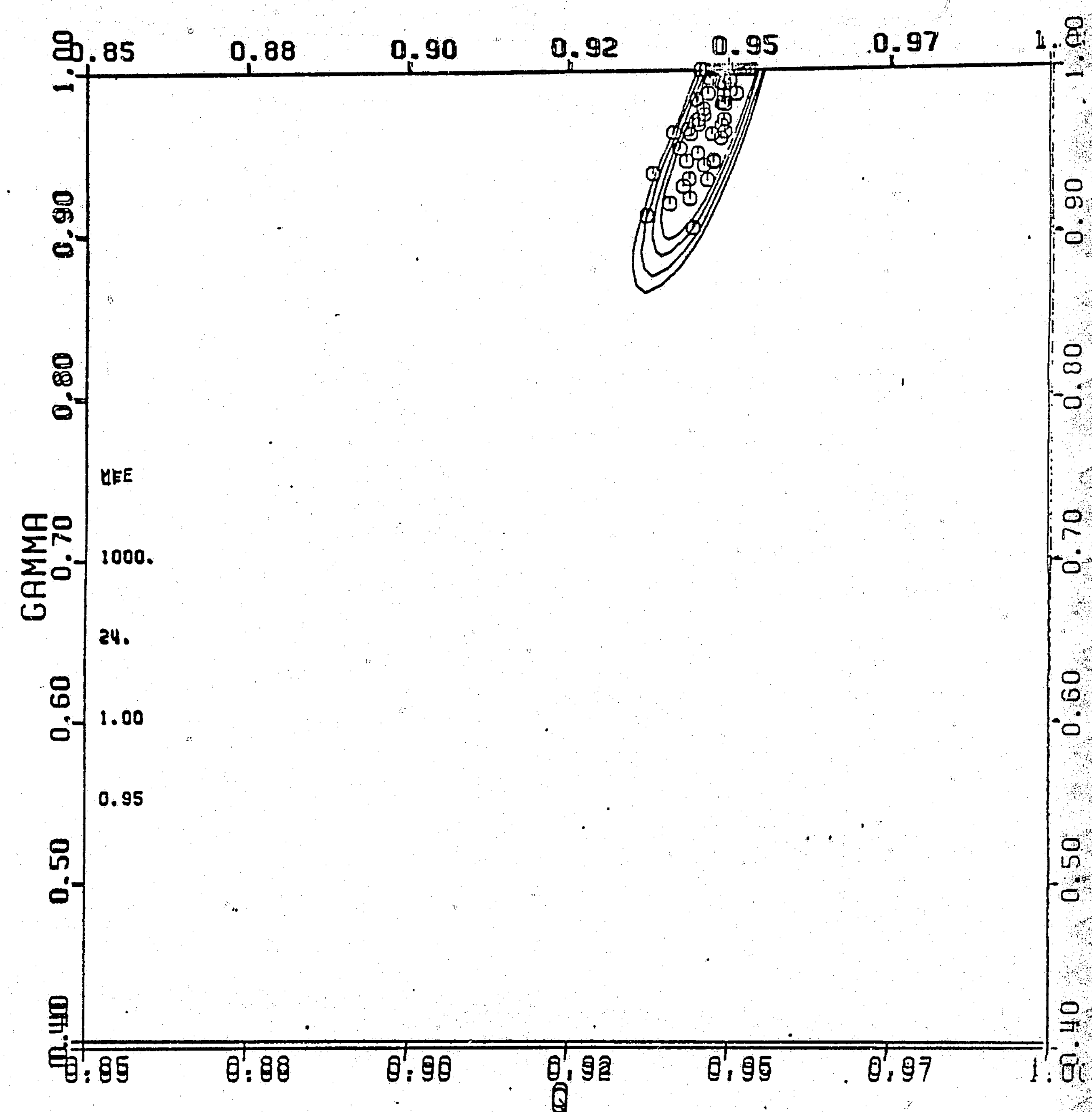


FIGURE 5. Maximum likelihood solutions (circles) of 100 simulations of cohorts with the following characteristics:

N: 1000 subjects
τ: 24 months censoring time
γ: 1.0 , probability of eventually failing
q: 0.95 , probability of surviving to the next month

Likelihood function contours superimposed on these solutions, using confidence levels of 0.99, 0.98, 0.95, and 0.90.

Blurring of coordinates due to shrinkage of transparent overlay of contours.

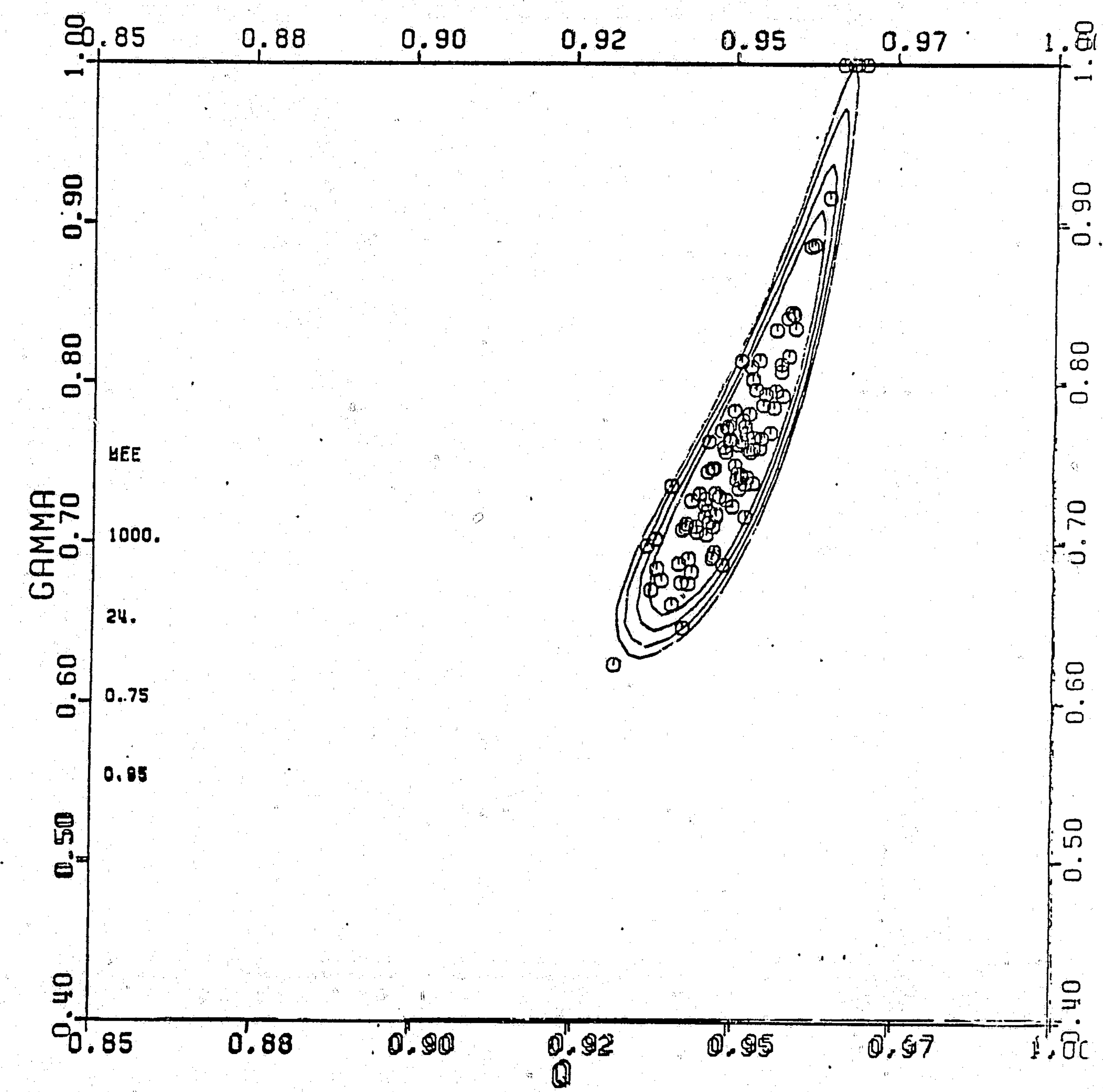


FIGURE 6. Maximum likelihood solutions (circles) of 100 simulations of cohorts with the following characteristics:

N: 1000 subjects
τ: 24 months censoring time
γ: 0.75 , probability of eventually failing
q: 0.95 , probability of surviving to the next month

Likelihood function contours superimposed on these solutions, using confidence levels of 0.99, 0.98, 0.95, and 0.90.

Blurring of coordinates due to shrinkage of transparent overlay of contours.

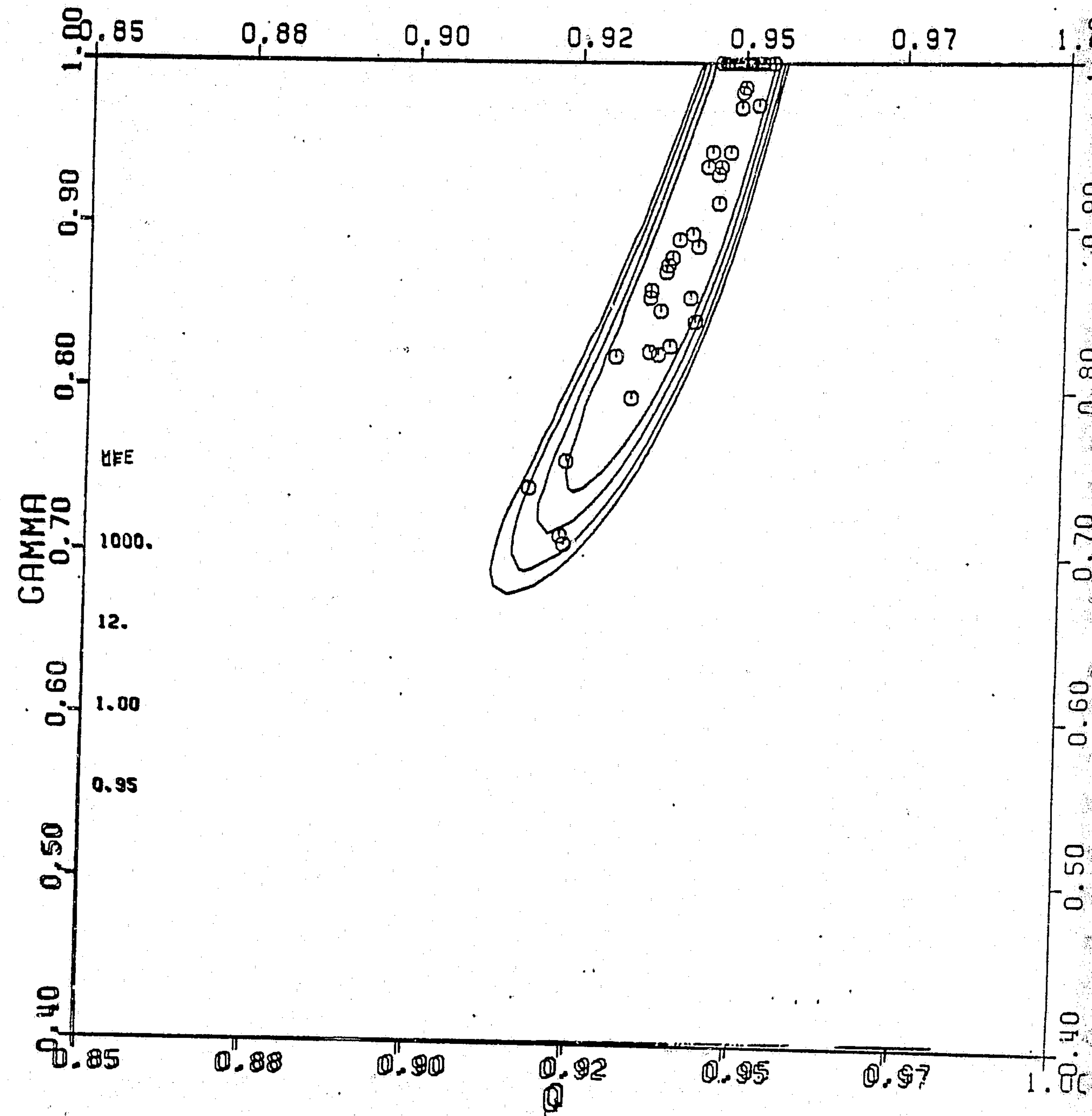


FIGURE 7. Maximum likelihood solutions (circles) of 100 simulations of cohorts with the following characteristics:

N: 1000 subjects
τ: 12 months censoring time
γ: 1.0 , probability of eventually failing
q: 0.95 , probability of surviving to the next month

Likelihood function contours superimposed on these solutions, using confidence levels of 0.99, 0.98, 0.95, and 0.90.

Blurring of coordinates due to shrinkage of transparent overlay of contours.

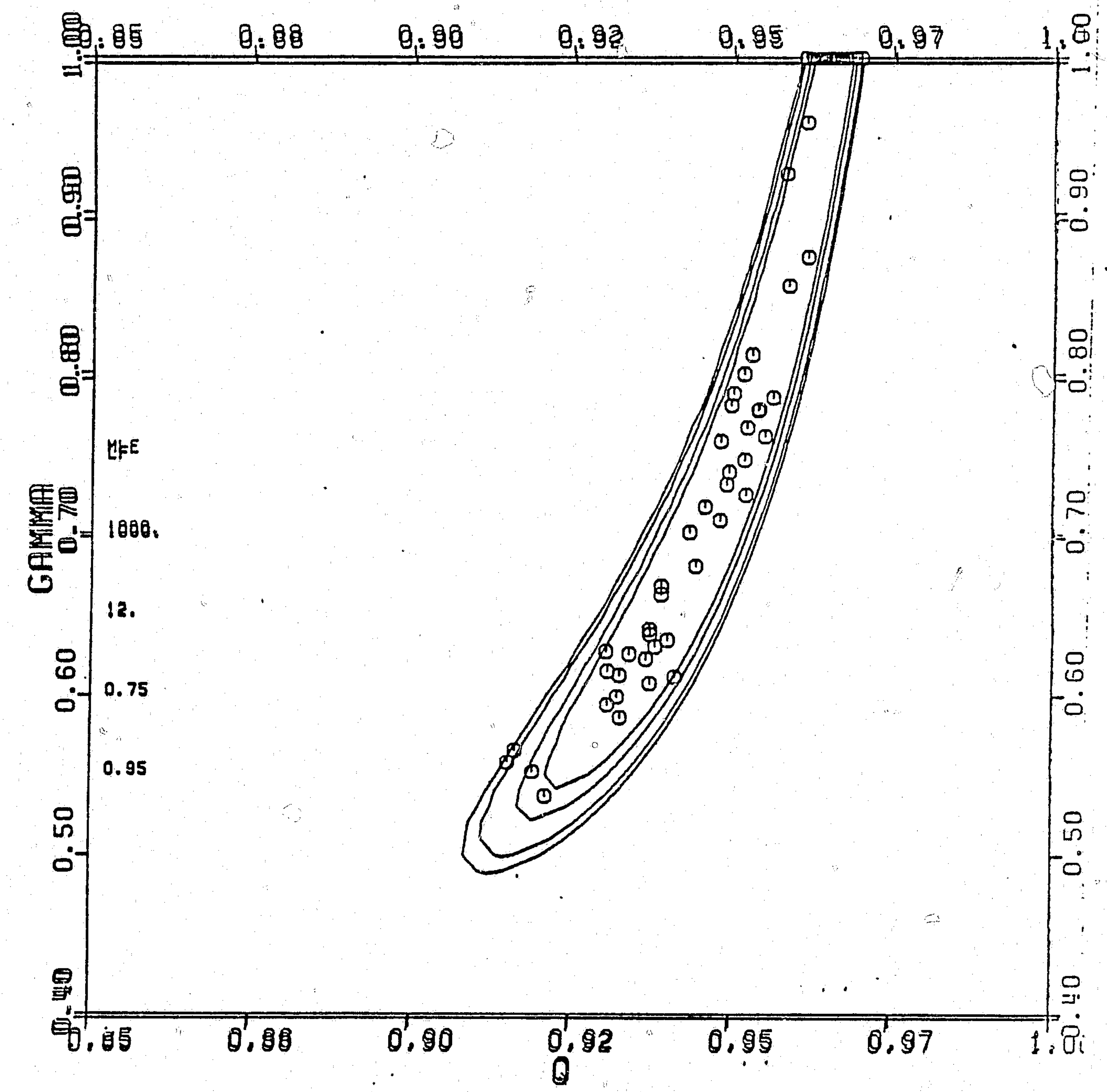


FIGURE 8. Maximum likelihood solutions (circles) of 100 simulations of cohorts with the following characteristics:

N: 1000 subjects
τ: 12 months censoring time
γ: 0.75 , probability of eventually failing
q: 0.95 , probability of surviving to the next month

Likelihood function contours superimposed on these solutions, using confidence levels of 0.99, 0.98, 0.95, and 0.90.

Blurring of coordinates due to shrinkage of transparent overlay of contours.

APPENDIX A

PROJECT MEMO CMH1-UICC-DOJ

ASYMPTOTIC BEHAVIOR OF MAXIMUM-LIKELIHOOD ESTIMATORS OF MIXED WEIBULL PARAMETERS

by

Carl M. Harris
Consultant
Washington, D.C.

and

Jay Mandelbaum
U.S. Department of Transportation
Washington, D.C.

ABSTRACT

When dealing with finite mixtures of distributions, there is an inherent symmetry, whereby several essentially identical but different maxima of the likelihood equation can be obtained. On the other hand, theory states that under a wide variety of circumstances, there is a unique consistent root of the likelihood equation. To avoid this difficulty, this paper proposes an ordering convention for the unknown parameter which eliminates all but one of the equivalent solutions. Asymptotic normality is unaffected.

1. PROBLEM DEFINITION

The issue of possible multiple maxima for the likelihood equation is very important in the case of finite mixtures of distributions. There is an inherent symmetry whereby two essentially identical but different solutions can be obtained. Consider the situation in which there is a mixture of two distributions of a single parameter. The probability density function (pdf) will be[1/]

$$g(p_1,\theta_1,p_2,\theta_2) = p_1 f(\theta_1) + p_2 f(\theta_2)$$

where $p_2 = 1-p_1$. The problem is to estimate the parameters p_1, θ_1, p_2, and θ_2.

Assume the solution to the likelihood equation is

$$\begin{cases} p_1 = a \\ \theta_1 = b \\ p_2 = 1-a \\ \theta_2 = c \end{cases} .$$

An identical $g(p_1,\theta_1,p_2,\theta_2)$ would be obtained if the solution were

$$\begin{cases} p_1 = 1-a \\ \theta_1 = c \\ p_2 = a \\ \theta_2 = b \end{cases} .$$

In both cases, the pdf of the mixture is

$$g(p_1,\theta_1,p_2,\theta_2) = af(b) + (1-a)f(c).$$

Since both solutions yield the same pdf and both solutions maximize the likelihood function, the fundamental question of consistency is raised. Are both solutions consistent? As a result, what can we say with regard to the usual property of asymptotic normality? Also, suppose only one is consistent, then which one do we choose.

[1/] We have omitted the functional argument here, but of course it should be understood that the p_i and θ_i are only parameters and are not representative of the underlying random variable.

2. CONSISTENCY AND MAXIMUM - LIKELIHOOD ESTIMATORS

Under mild regularity conditions, Cramer (1946) proved the existence of a solution of the likelihood equation and that that solution is consistent as $n \to \infty$. Using the same regularity conditions, Huzurbazar (1948) showed that there is a unique consistent solution to the likelihood in the case of a single unknown parameter. If we assume for the moment that Huzurbazar's results can be extended to the multi-dimensional case, then a dilemma arises. For mixtures, only one of the two solutions which yield the same pdf is consistent. How then do we determine the correct one?

Perlman (1969) points out that there are ambiquities in the Huzurbazar result that "a consistent root of the likelihood equation is unique." Perlman claims that consistency is a limiting property of a sequence of estimators. He goes on to say that if T_n is a strongly consistent sequence of estimators of θ, and if $\{T_n{}^*\}$ is another sequence such that the probability is one that $T_n{}^* = T_n$ for all sufficiently large n, then $\{T_n{}^*\}$ is also strongly consistent for θ. Let $S_n(x_1, \ldots\ldots, x_n)$ denote the set of all solutions (roots) to the likelihood equation for the given sample. If $\{S_n\}$ contains more than one element, then there are infinitely many roots as defined here. Also if $\{S_n\}$ is a strongly consistent root, then by the above, the initial terms in the sequence can be changed to produce another consistent root. Therefore Perlman concludes that there may be many consistent roots.

Perlman avoids these ambiquities by showing under slightly weaker regularity conditions, that as n goes to infinity, all but one element is bounded away from the true parameter value and the remaining elements

approach the true parameter value. Unfortunately, Perlman's results are not directly applicable to the mixture situation. He only deals with the case of a single parameter and he assumes that if $\theta_1 \neq \theta_2$, The $f(x_1, \theta_1)$ and $f(x_1, \theta_2)$ determine distinct distributions. This assumption does not hold in our situation.

3. SOLUTION

We are still left with our original dilemma. Although previous work is not directly applicable to the case of mixtures, there is strong evidence that only one of the two solutions identified in Section 1 is consistent. The solution which would be obtained via our methods would be a function of the choice of initial parameter values only.

The cause of over dilemma lies with a basic lack of specificity of $g(p_1,\theta_1,p_2,\theta_2)$. In our current numerical procedures (see Kaylan and Harris, 1979), we are not free to define an "ordering" of the parameter sets for the mixed distributions. Thus this suggests that we adopt an ordering convention of $\theta_1<\theta_2<\theta_3$..... These would be strict inequalities since we do not allow the number of mixed distributions to be a random variable. In the case of a vector of parameters, the above would be extended to

$$\theta_{11} \leq \theta_{21} \leq \theta_{31}\ldots\ldots, \quad \theta_{12} \leq \theta_{22} \leq \theta_{32} \ldots\ldots,$$

$$\ldots\ldots \theta_{1n} < \theta_{2n} < \theta_{3n}\ldots\ldots$$

This ordering completely eliminates the dual solution problem. If $b < c$, then

$$p_1 = a$$
$$\theta_1 = b$$
$$p_2 = 1-a$$
$$\theta_2 = c$$

is feasible. The symmetric point

$$p_1 = 1-a$$
$$\theta_1 = c$$
$$p_2 = a$$
$$\theta_2 = b$$

is not allowable since $\theta_2 > \theta_1$. Operationally, nothing changes.

We employ our usual methodology to calculate the solution. If that solution has $\theta_2 > \theta_1$, we would simply use the symmetric counterpart.

4. IMPLICATIONS FOR ASYMPTOTIC NORMALITY

Given that the usual regularity conditions hold, asymptotic normality should not be affected by the ordering which we have adopted. If α is a vector of unknown parameters and $\hat{\alpha}$ is the likelihood estimator for α based on n observations, then $(\hat{\alpha} - \alpha)/\sqrt{n}$ has a limiting (as $n \to \infty$) multivariate normal distribution with matrix mean 0 and variance-covariance matrix V given as follows. The (i,j) entry of the inverse of V (call it R) is given by

$$r_{ij} = -E\left[\frac{\partial^2}{\partial\alpha_i \partial\alpha_j} \log f(X,\alpha)\right] .$$

Given an initial starting point the iterative algorithm will converge to one and only one of the symmetric maxima that convergence point will be a function of the starting point alone. There will be no shifting between equivalent maxima. After K iterations, some α will have been obtained, and $(\hat{\alpha} - \alpha)/\sqrt{n}$ will be approximately multivariate normal with mean 0 and variance-covariance matrix V as above. But if $\hat{\alpha}$ does not satisfy the ordering conditions, then we determine its symmetric counterpart. This, however, implies an implicit re-ordering of α as well and consequently asymptotic normality will be maintained.

REFERENCES

Cramer, H. (1946), Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey.

Huzurbazar, V.S. (1948), "The Likelihood Equation, Consistency and the Maxima of the Likelihood Equation", Annals of Eugenics, Vol. 14, p. 185.

Kaylan, A.R. and C.M. Harris (1979), "Efficients Algorithms to Derive Maximum-Likelihood Estimates for Finite Exponential and Weibull Mixtures", pending review with Computers and Operations Research.

Perlman, M.D. (1969), "The Limiting Behavior of Multiple Roots of the Likelihood Equation", University of Minnesota Technical Report No. 125.

APPENDIX B

PROJECT MEMO CMH2-UICC-DOJ

LARGE-SAMPLE SIGNIFICANCE TESTS FOR MIXED WEIBULL POPULATIONS

by

Carl M. Harris
Consultant
Washington, D.C.

and

Jay Mandelbaum
U.S. Department of Transportation
Washington, D.C.

Abstract

The importance of recidivism models in the analysis of recidivism rates has been established in the literature. Maximum likelihood estimates for unknown parameters within these models are obtained. The topic of tests of significance for these estimates is addressed in this paper.

1. PROBLEM DEFINITION

When studying the differences between or the effectiveness of two different recidivism programs, one often would like to compare them in terms of failure rates or interoccurrence distributions. In our work thus far, we have characterized recidivists as coming from a single PDF which is a mixture of two other PDFs. This note outlines methods for testing hypotheses concerning the equality of the complete mixed PDFs as well as the individual components and mixing proportions for the two programs.

Assume that the first program is characterized by a mixture of two PDFs whose failure rates are a_1 and b_1 respectively. Also let the mixing proportion be p_1. Then the PDF for this program will be

$$g_1(x,a_1,b_1,p_1) = p_1 f(x,a_1) + (1-p_1) f(x,b_1).$$

If we make analogous assumptions about the second, it will have the PDF

$$g_2(x,a_2,b_2,p_2) = p_2 f(x,a_2) + (1-p_2) f(x,b_2).$$

We obtain maximum-likelihood estimators of the unknown parameters in the usual way and denote them as vectors $\hat{\theta}_1$ and $\hat{\theta}_2$ for programs one and two respectively, where

$$\hat{\theta}_i = \begin{pmatrix} \hat{a}_i \\ \hat{b}_i \\ \hat{p}_i \end{pmatrix} \qquad (i = 1,2) \quad .$$

The estimated variance-covariance matrices for these values will be

$$C^i = -\left[H_i(\hat{\theta}_i)\right]^{-1} \qquad (i = 1,2)$$

where

$$H_i(\hat{\theta}_i) = \left.\begin{pmatrix} \frac{\partial^2 L}{\partial a^2} & \frac{\partial^2 L}{\partial a \partial b} & \frac{\partial^2 L}{\partial a \partial p} \\ \frac{\partial^2 L}{\partial b \partial a} & \frac{\partial^2 L}{\partial b^2} & \frac{\partial^2 L}{\partial b \partial p} \\ \frac{\partial^2 L}{\partial p \partial a} & \frac{\partial^2 L}{\partial p \partial b} & \frac{\partial^2 L}{\partial p^2} \end{pmatrix}\right|_{\substack{a = a_i \\ b = b_i \\ c = c_i}}$$

and L is the standard log-likelihood given by

$$L = \ln \prod_j g(x_j, a, b, p) \quad .$$

Consequently, the terms of C^i will be

$$C^i = \begin{pmatrix} Var(a_i) & Cov(a_i,b_i) & Cov(a_i,p_i) \\ Cov(b_i,a_i) & Var(b_i) & Cov(b_i,p_i) \\ Cov(p_i,a_i) & Cov(p_i,b_i) & Var(p_i) \end{pmatrix}$$

We are now able to do hypothesis testing based on this structure. There are seven possible problems of concern, with alternative hypotheses given as the inequality of the H_o in each case.

H_o: $a_1 = a_2$

H_o: $b_1 = b_2$

H_o: $p_1 = p_2$

H_o: $\underline{\theta}_{41} = \underline{\theta}_{42}$ where $\underline{\theta}_{4i} = \begin{pmatrix} a_i \\ b_i \end{pmatrix}$

H_o: $\underline{\theta}_{51} = \underline{\theta}_{52}$ where $\underline{\theta}_{5i} = \begin{pmatrix} a_i \\ p_i \end{pmatrix}$

H_o: $\underline{\theta}_{61} = \underline{\theta}_{62}$ where $\underline{\theta}_{6i} = \begin{pmatrix} b_i \\ p_i \end{pmatrix}$

H_o: $\underline{\theta}_{71} = \underline{\theta}_{72}$ where $\underline{\theta}_{7i} = \begin{pmatrix} a_i \\ b_i \\ p_i \end{pmatrix}$

For sufficiently large sample sizes the maximum-likelihood estimators are normally distributed with variance-covariance matrix as shown above. Since we also have asymptotic unbiasedness, it follows that

$$\left.\begin{aligned} E(\hat{a}_i) &= a_i \\ E(\hat{b}_i) &= b_i \\ E(\hat{p}_i) &= p_i \end{aligned}\right\} \quad i = 1,2 \quad .$$

If we denote the variance-covariance matrix for the i^{th} sample as C^i whose $(k,j)^{th}$ element is $c^i(k,j)$, then the following hold when H_o is true in the first three of the hypotheses:

$$\hat{a}_1 - \hat{a}_2 \sim N(0, c^1_{11} + c^2_{11})$$

$$\hat{b}_1 - \hat{b}_2 \sim N(0, c^1_{22} + c^2_{22})$$

$$\hat{p}_1 - \hat{p}_2 \sim N(0, c^1_{33} + c^2_{33}) \quad .$$

The hypothesis would then be rejected if the test statistic falls in the tails of its corresponding distribution.

We note that for the final four hypotheses under H_o it turns out that

$$\underline{\theta}_{41} - \underline{\theta}_{42} \sim N(\underline{0}, C_4) \quad \text{where} \quad C_4 = \begin{pmatrix} c^1_{11} + c^2_{11} & c^1_{12} + c^2_{12} \\ c^1_{21} + c^2_{21} & c^1_{22} \quad c^2_{22} \end{pmatrix}$$

$$\underline{\theta}_{51} - \underline{\theta}_{52} \sim N(\underline{0}, C_5) \quad \text{where} \quad C_5 = \begin{pmatrix} c^1_{11} + c^2_{11} & c^1_{13} + c^2_{13} \\ c^1_{31} + c^2_{31} & c^1_{33} + c^2_{33} \end{pmatrix}$$

$$\underline{\theta}_{61} - \underline{\theta}_{62} \sim N(\underline{0}, C_6) \quad \text{where} \quad C_6 = \begin{pmatrix} c^1_{22} + c^2_{22} & c^1_{23} + c^2_{23} \\ c^1_{32} + c^2_{32} & c^1_{33} + c^2_{33} \end{pmatrix}$$

$$\underline{\theta}_{71} - \underline{\theta}_{72} \sim N(\underline{0}, C^1 + C^2)$$

The resultant test follows from the following observation.[1/] If $\underline{X}$ is a vector of m random variables, and $\underline{X}$ has a multivariate normal distribution with mean $\underline{\mu}$ and variance-covariance matrix $\underline{B}$ of rank m

$$T^2 = (\underline{X} - \underline{\mu})' \, \underline{B}^{-1} \, (\underline{X} - \underline{\mu})$$

1/ Scheffe, Henry, The Analysis of Variance, John Wiley & Sons, Inc., New York, 1959.

will be distributed as a chi-squared random variable with m degrees of freedom. If we replace $\underline{\mu}$ by $\underline{0}$, $\underline{X}$ by any one of the above $\underline{\theta}$ vectors, and $\underline{B}$ by the corresponding variance-covariance matrix, then a chi-squared statistic can be used to test the final four hypotheses. A hypothesis would then be rejected if the test statistic is in the tails of the approximate distribution.

APPENDIX C

PROJECT MEMO CMH3-UICC-DOJ

MORE ON SIGNIFICANCE TESTS FOR MIXED WEIBULL POPULATIONS

by

Carl M. Harris

Center for Management and
Policy Research, Inc.
Washington, D.C.

and

Jay Mandelbaum

U.S. Department of Transportation
Washington, D.C.

Abstract

The importance of formalized statistical models in the analysis of recidivism has been well established in the literature. Duration distributions of the mixed Weibull form for such problems have been a special focus of this study. Numerical methods for the maximum-likelihood estimation of their parameters under complex sampling scenarios were the subject of some prior efforts. Here, we provide a thorough discussion of the use of such estimators for population inferences.

1. PROBLEM STATEMENT

To study the differences between or the effectiveness of two different recidivism programs, one usually wishes to compare them in terms of their respective failure rates or interoccurrence distribution functions. For our basic model formulation, we have characterized recidivists as samples from a single population with probability density function (PDF) which is a mixture of two other simple PDFs. This paper outlines methods and theory for testing hypotheses concerning the equality of the complete mixed PDFs as well as the individual components and mixing proportions for the two programs. This work provides more complete details of the theory than provided in our earlier memo on the subject (Harris and Mandelbaum, 1979).

To begin, let us assume that the first program is characterized by a mixture of two PDFs whose failure rates are a_1 and b_1 respectively. Also let the mixing proportion be p_1. Then the PDF for this program will be

$$g_1(x;a_1,b_1,p_1) = p_1 f(x;a_1) + (1-p_1)f(x;b_1).$$

Similarly, let the second PDF be

$$g_2(x;a_2,b_2,p_2) = p_2 f(x;a_2) + (1-p_2)f(x;b_2).$$

We numerically obtain maximum-likelihood estimators of the unknown parameters in the usual way and denote them as vectors $\hat{\theta}_1$ and $\hat{\theta}_2$ for programs one and two respectively, where

$$\hat{\theta}_i = \begin{pmatrix} \hat{a}_i \\ \hat{b}_i \\ \hat{p}_i \end{pmatrix} \qquad (i = 1,2) .$$

The estimated variance-covariance matrices for these estimates will be

$$C^i = -\left[H_i(\hat{\theta}_i)\right]^{-1} \qquad (i = 1,2)$$

where

$$H_i(\hat{\theta}_i) = \left.\begin{pmatrix} \frac{\partial^2 L}{\partial a^2} & \frac{\partial^2 L}{\partial a \partial b} & \frac{\partial^2 L}{\partial a \partial p} \\ \frac{\partial^2 L}{\partial b \partial a} & \frac{\partial^2 L}{\partial b^2} & \frac{\partial^2 L}{\partial b \partial p} \\ \frac{\partial^2 L}{\partial p \partial a} & \frac{\partial^2 L}{\partial p \partial b} & \frac{\partial^2 L}{\partial p^2} \end{pmatrix}\right|_{\substack{a = a_i \\ b = b_i \\ c = c_i}}$$

And L is the standard log-likelihood function given by

$$L = \ln \prod_j g(x_j; a,b,p) \quad .$$

Consequently, the terms of c^i will be

$$C^i = \begin{pmatrix} Var(a_i) & Cov(a_i,b_i) & Cov(a_i,p_i) \\ Cov(b_i,a_i) & Var(b_i) & Cov(b_i,p_i) \\ Cov(p_i,a_i) & Cov(p_i,b_i) & Var(p_i) \end{pmatrix}$$

There are seven possible (large-sample) hypothesis tests with alternative hypotheses given as the inequality of the null H_o in each case. These are:

H_o: $a_1 = a_2$

H_o: $b_1 = b_2$

H_o: $p_1 = p_2$

H_o: $\underline{\theta}_{41} = \underline{\theta}_{42}$ where $\underline{\theta}_{4i} = \begin{pmatrix} a_i \\ b_i \end{pmatrix}$

H_o: $\underline{\theta}_{51} = \underline{\theta}_{52}$ where $\underline{\theta}_{5i} = \begin{pmatrix} a_i \\ p_i \end{pmatrix}$

H_o: $\underline{\theta}_{61} = \underline{\theta}_{62}$ where $\underline{\theta}_{6i} = \begin{pmatrix} b_i \\ p_i \end{pmatrix}$

H_o: $\underline{\theta}_{71} = \underline{\theta}_{72}$ where $\underline{\theta}_{7i} = \begin{pmatrix} a_i \\ b_i \\ p_i \end{pmatrix}$

For sufficiently large sample sizes the maximum-likelihood estimators are normally distributed with variance-covariance matrix as shown above. Since we also have asymptotic unbiasedness, it follows that

$$\left.\begin{aligned} E(\hat{a}_i) &= a_i \\ E(\hat{b}_i) &= b_i \\ E(\hat{p}_i) &= p_i \end{aligned}\right\} \quad i = 1,2 \ .$$

If we denote the variance-covariance matrix for the i^{th} sample as C^i whose $(k,j)^{th}$ element is $c^i(k,j)$, then the following hold when H_o is true in the first three of the hypotheses:

$$\begin{cases} \hat{a}_1 - \hat{a}_2 \sim N(0, c^1_{11} + c^2_{11}) \\ \hat{b}_1 - \hat{b}_2 \sim N(0, c^1_{22} + c^2_{22}) \\ \hat{p}_1 - \hat{p}_2 \sim N(0, c^1_{33} + c^2_{33}) \end{cases}$$

An hypothesis would then be rejected if its test statistic falls in the tails of its corresponding distribution.

We note for the final four hypotheses under H_o that it turns out that

$$\underline{\theta}_{41} - \underline{\theta}_{42} \sim N(\underline{0}, C_4) \quad \text{where} \quad C_4 = \begin{pmatrix} c^1_{11} + c^2_{11} & c^1_{12} + c^2_{12} \\ c^1_{21} + c^2_{21} & c^1_{22} \quad c^2_{22} \end{pmatrix}$$

$$\underline{\theta}_{51} - \underline{\theta}_{52} \sim N(\underline{0}, C_5) \quad \text{where} \quad C_5 = \begin{pmatrix} c^1_{11} + c^2_{11} & c^1_{13} + c^2_{13} \\ c^1_{31} + c^2_{31} & c^1_{33} + c^2_{33} \end{pmatrix}$$

$$\underline{\theta}_{61} - \underline{\theta}_{62} \sim N(\underline{0}, C_6) \quad \text{where} \quad C_6 = \begin{pmatrix} c^1_{22} + c^2_{22} & c^1_{23} + c^2_{23} \\ c^1_{32} + c^2_{32} & c^1_{33} + c^2_{33} \end{pmatrix}$$

$$\underline{\theta}_{71} - \underline{\theta}_{72} \sim N(\underline{0}, C^1 + C^2)$$

The resultant significance test follows from the following observation (for example, see Scheffé, 1959). If $\underline{X}$ is a vector of m random variables such that $\underline{X}$ has a multivariate normal distribution with mean $\underline{\mu}$ and variance-covariance matrix $\underline{B}$ of rank m, then the quantity

$$T^2 = (\underline{X} - \underline{\mu})' \ \underline{B}^{-1} \ (\underline{X} - \underline{\mu})$$

will be distributed as a chi-squared random variable with m degrees of freedom. If we replace $\underline{\mu}$ by $\underline{0}$, $\underline{X}$ by any one of the foregoing $\underline{\theta}$ vectors, and $\underline{B}$ by the corresponding variance-covariance matrix, then a chi-squared statistic can be used to test the final four hypotheses. A hypothesis would then be rejected if the test statistic is in the tails of the appropriate distribution.

2. ASYMPTOTIC THEORY

Of course, these (large-sample) tests cannot be performed unless a set of regularity conditions holds and the sample sizes are indeed adequately large. Exactly how big "large" must be is very much a function of the number of parameters involved and the complexity of the sampling situation. To date, our empirical experience has been that convergence to normality occurs quite rapidly. Our data sets all seem to be of adequate size. However, in general, there is no guarantee that normality will indeed obtain. One must verify that the appropriate regularity conditions are in fact satisfied.

The two major references for univariate regularity are Cramér (1946) and Kulldorf (1957). These authors deal with the univariate case where θ is the parameter being estimated for the PDF $f(x;\theta)$, $\theta\varepsilon\Omega$ (henceforth denoted by f). The regularity conditions are then given as follows:

(i) $\partial \log f/\partial\theta, \partial^2\log f/\partial\theta^2, \partial^3\log f/\partial\theta^3$ exist for all $\theta\varepsilon\Omega$ and every x. Also $\int_{-\infty}^{\infty} \frac{\partial f}{\partial\theta}\, dx = E_\theta \frac{\partial \log f}{\partial\theta} = 0$ for all $\theta\varepsilon\Omega$.

(ii) $\int_{\infty}^{\infty} \frac{\partial^2 f}{\partial\theta^2}\, dx = 0$ for all $\theta\varepsilon\Omega$.

(iii) $-\infty<\int_{-\infty}^{\infty} \frac{\partial^2\log f}{\partial\theta^2}\, f\, dx<0$ for all θ.

(iv) There exists a function H(x) such that for all $\theta\varepsilon\Omega$

$$\left|\frac{\partial^3\log f}{\partial\theta^3}\right| < H(x) \quad \text{and} \quad \int_{-\infty}^{\infty} H(x)f\, dx = M(\theta)< \infty.$$

(v) There exists a function g(θ) that is positive and twice differentiable for every $\theta\varepsilon\Omega$ and a function H(x) such that for all θ

$$\left|\frac{\partial^2}{\partial\theta^2}\left[g(\theta)\frac{\partial\log f}{\partial\theta}\right]\right| < H(x) \quad \text{and} \quad \int_{-\infty}^{\infty} H(x)f\, dx < \infty.$$

Note that condition (v) is equivalent to condition (iv) with the added qualification that g(θ) = 1.

In the multidimensional case, Chanda (1954) shows that the following regularity conditions should be satisfied for asymptotic consistency and normality of the MLEs for the underlying density f(x;θ) where now θ is the vector of parameters $\theta = (\theta_1, \ldots , \theta_k)$.

(i) The point represented by the vector θ lies in a k-dimensional interval Ω; for almost all x and for all $\theta\varepsilon\Omega$

$$\frac{\partial\log f}{\partial\theta_r},\ \frac{\partial^2\log f}{\partial\theta_r\,\partial\theta_s},\ \frac{\partial^3\log f}{\partial\theta_r\partial\theta_s\partial\theta_t}$$

exist for all r, s, t = 1, 2,...k.

(ii) For almost all x and for every point $\theta\varepsilon\Omega$

$$\left|\frac{\partial f}{\partial\theta}\right| < G_r(x), \left|\frac{\partial^2 f}{\partial\theta_r\,\partial\theta_s}\right| < G_{rs}(x)$$

and

$$\left|\frac{\partial^3 \log f}{\partial\theta_r\,\partial\theta_s\,\partial\theta_t}\right| < H_{rst}(x)$$

where $G_r(x)$ and $G_{rs}(x)$ are integrable on $(-\infty,\infty)$ and $H_{rst}(x)$ is such that

$$\int_{-\infty}^{\infty} H_{rst}(x)f\, dx < M \quad \text{(M a finite positive constant).}$$

(iii) For all $\theta\varepsilon\Omega$ the matrix $J=(J_{rs}(\theta))$, defined by

$$J_{rs}(\theta) = \int_{-\infty}^{\infty} \frac{\partial\log f}{\partial\theta_r} \cdot \frac{\partial\log f}{\partial\theta_r}\, f\, dx ,$$

is positive definite, and |J| is finite.

In general, it is extremely burdensome to show that these regularity conditions hold. See Parekh (1972) for a sample of the type of computations involved. In the case of a two-parameter Weibull, as is being considered in this paper, the regularity conditions hold (see Harter and Moore, 1972). When a location parameter is also included, the regularity conditions hold for values of the shape parameter greater than two.

Thus far, this discussion has centered around the regularity conditions for classical maximum-likelihood estimation. Halperin (1952) considered the problem of the estimation of a single parameter under type I censoring. In this case, not all observations result in failures. The successes, however, are all constrained to survive for the same amount of time. The conditions are as follows:

Assumption A. For almost all x, the derivatives

$$\frac{\partial \log f}{\partial \theta}, \quad \frac{\partial^2 \log f}{\partial \theta^2}, \quad \frac{\partial^3 \log f}{\partial \theta^3}$$

exist for all $\theta \varepsilon \Omega$.

Assumption B. For every $\theta \varepsilon \Omega$ we have

$$\left|\frac{\partial f}{\partial \theta}\right| < F_1(x), \qquad \left|\frac{\partial^2 f}{\partial \theta^2}\right| < F_2(x),$$

$$\left|\frac{\partial^3 f}{\partial \theta^3}\right| < F_2(x), \qquad \left|\frac{\partial^3 \log f}{\partial \theta^3}\right| < H(x),$$

where $F_1(x)$, $F_2(x)$ are integrable on $(-\infty, \infty)$, while $\int_{-\infty}^{\infty} H(x) f\, dx < M$, where M is independent of θ.

Assumption C. For every $\theta \varepsilon \Omega$

$$K^2 = \int_{-\infty}^{\lambda} \left(\frac{\partial \log f}{\partial \theta} \right)^2 f\, dx + \frac{1}{p} \left(\int_{-\infty}^{\lambda} \frac{\partial f}{\partial \theta}\, dx\right)^2$$

is greater than zero. Here, if θ_0 is the true value of θ, λ is defined by $q = \int_{-\infty}^{\lambda} f(x,\theta_0)dx$. That is, λ is the population 100q percentage point.

Assumption D. f is continuous in the neighborhood of $x = \lambda$ and has a continuous derivative in x,f', while

$$\frac{\partial \log f}{\partial \theta}, \quad \frac{\partial^2 \log f}{\partial \theta^2}, \quad \frac{\partial^3 \log f}{\partial \theta^3},$$

are continuous in the neighborhood of $x = \lambda$.

Halperin also discussed the extension to the multiparameter case. In this instance, the assumptions necessary to obtain the result are the natural analogues of Assumptions A–D. Thus A, B, D are extended by imposing similar conditions upon the various derivatives up to third order, that is those with respect to each θ, and also the mixed derivatives. The condition C becomes a requirement that the matrix with elements

$$A_{ij} = \int_{\theta}^{\lambda} \left(\frac{\partial \log f}{\partial \theta_i} \right) \left(\frac{\partial \log f}{\partial \theta_j} \right) f\, dx$$
$$+ \frac{1}{p} \left(\int_0^{\lambda} \frac{\partial f}{\partial \theta_i}\, dx \right) \left(\int_0^{\lambda} \frac{\partial f}{\partial \theta_j} \right), \qquad i,j = 1,2,\ldots, p,$$

be positive definite.

Finally, Halperin suggested that the results should be generalizable to the case of several points of truncation, each truncation point being a sample percentage point. Due to the asymptotic normality of sample percentage points, it appeared to Halperin that the results should be extendable. However, their direct analytic verification is a brutal task.

Given the complexity of these conditions and the experience of others that such direct checking is indeed very messy, what can be done? There are three basic things which come to mind. First is the realization that these are only sufficient conditions, and likely not necessary. In other words, it may well be ultimately possible to find a simpler set of sufficient conditions which may be computationally feasible. Second, a number of Monte Carlo experiments can be set up to see whether normality seems to obtain for estimates derived from samples simulated from a variety of (known) populations covering a wide range of possible parameter values. The final thought is to search for approximations which might simplify the verification process. Since direct checking may require numerical integration anyway, an analytic approximation may just do the job.

References

Chanda, K.C. (1954), "A Note on the Consistency and Maxima of the Roots of Likelihood Equations," Biometrika, 41, p. 56-61.

Cramér, H. (1946), Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey.

Halperin, M. (1952), "Maximum Likelihood Estimation in Truncated Samples," Annals of Mathematical Statistics 23, p. 226-238.

Harris, C.M. and J. Mandelbaum (1979), "Significance Tests for Mixed Weibull Populations," Project Memo CMH2-UICC-DOJ, Center for Management & Policy Research, Inc., Washington, D.C.

Harter, H.L. and A.H. Moore (1967), "Asymptotic Variances and Covariances of Maximum-Likelihood Estimators, from Censored Samples of the Parameters of Weibull and Gamma Populations," Technometrics 9, p. 557-565.

Kulldorf, G. (1957), "On the Conditions for Consistency and Asymptotic Efficiency of Maximum-Likelihood Estimates," Skand. Aktuarietidskr 40, p. 129-144.

Parekh, S.C. (1972), Parametric Estimation for the Compound Weibull and Related Distributions, New York University, School of Engineering and Science, Ph.D. dissertation.

Scheffé, H. (1959), The Analysis of Variance, John Wiley & Sons, Inc. New York.

APPENDIX D

PROJECT MEMO CMH4-UICC-DOJ

ASYMPTOTIC BEHAVIOR OF MAXIMUM-LIKELIHOOD ESTIMATORS OF MIXED WEIBULL PARAMETERS

by

Carl M. Harris
Center for Management & Policy Research, Inc.
Washington, D.C.

and

Jay Mandelbaum
U.S. Department of Transportation
Washington, D.C.

August 26, 1980

ABSTRACT

When dealing with finite mixtures of distributions, there is an inherent symmetry, whereby several essentially identical but different maxima of the likelihood equation can be obtained. On the other hand, theory states that under a wide variety of circumstances, there is a unique consistent root of the likelihood equation. To avoid this difficulty, this paper proposes an ordering convention for the unknown parameters which eliminates all but one of the equivalent solutions. Asymptotic normality is unaffected.

1. PROBLEM DEFINITION

The issue of possible multiple maxima for the likelihood equation is very important in the case of finite mixtures of distributions. There is an inherent symmetry whereby two essentially identical but different solutions can be obtained. Consider the situation in which there is a mixture of two distributions of a single parameter. The probability density function (pdf) will be[1/]

$$g(p_1,\theta_1,p_2,\theta_2) = p_1 f(\theta_1) + p_2 f(\theta_2)$$

where $p_2 = 1-p_1$. The problem is to estimate the parameters p_1, θ_1, p_2, and θ_2.

Assume the solution to the likelihood equation is

$$\begin{cases} p_1 = a \\ \theta_1 = b \\ p_2 = 1-a \\ \theta_2 = c \end{cases}$$

An identical $g(p_1,\theta_1,p_2,\theta_2)$ would be obtained if the solution were

$$\begin{cases} p_1 = 1-a \\ \theta_1 = c \\ p_2 = a \\ \theta_2 = b \end{cases}$$

In both cases, the pdf of the mixture is

$$g(p_1,\theta_1,p_2,\theta_2) = af(b) + (1-a)f(c).$$

Since both solutions yield the same pdf and both solutions maximize the likelihood function, the fundamental question of consistency is raised. Are both solutions consistent? As a result, what can we say with regard to the usual property of asymptotic normality? Also, suppose only one is consistent, then which one do we choose?

1/ We have omitted the functional argument here, but of course it should be understood that the p_i and θ_i are only parameters and are not representative of the underlying random variable.

2. CONSISTENCY AND MAXIMUM - LIKELIHOOD ESTIMATORS

Under mild regularity conditions, Cramér (1946) proved the existence of a solution of the likelihood equation and that that solution is consistent as $n \to \infty$. Using the same regularity conditions, Huzurbazar (1948) showed that there is a unique consistent solution to the likelihood in the case of a single unknown parameter. If we assume for the moment that Huzurbazar's results can be extended to the multi-dimensional case, then a dilemma arises. For mixtures, only one of the two solutions which yield the same pdf is consistent. How then do we determine the correct one?

Perlman (1969) shows that there are ambiguities in the Huzurbazar result that "a consistent root of the likelihood equation is unique." Perlman claims that consistency is a limiting property of a sequence of estimators. He goes on to say that if T_n is a strongly consistent sequence of estimators of θ, and if $\{T_n{}^*\}$ is another sequence such that the probability is one that $T_n{}^* = T_n$ for all sufficiently large n, then $\{T_n{}^*\}$ is also strongly consistent for θ. Let $S_n(x_1,\ldots\ldots,x_n)$ denote the set of all solutions (roots) to the likelihood equation for the given sample. If $\{S_n\}$ contains more than one element, then there are infinitely many roots as defined here. Also if $\{T_n\}$ is a strongly consistent root, then by the above, the initial terms in the sequence can be changed to produce another consistent root. Therefore Perlman concludes that there may be many consistent roots.

Perlman avoids these ambiguities by showing under slightly weaker regularity conditions, that as n goes to infinity, all but one element is bounded away from the true parameter value and the remaining elements

approach the true parameter value. Unfortunately, Perlman's results are not directly applicable to the mixture situation. He only deals with the case of a single parameter and he assumes that if $\theta_1 \neq \theta_2$, The $f(x_1,\theta_1)$ and $f(x_1,\theta_2)$ determine distinct distributions. This assumption does not hold in our situation.

In the multiparameter case, Chanda (1954) proved under similar conditions that of all possible solutions to the likelihood equations, one and only one is consistent. Chanda's proof, however, relies on an extension of Rolle's theorem that does not exist. Tarone and Bruenhage (1975) prove an alternative result. They show that of all possible solutions to the likelihood equations which are relative maxima, one and only one is consistent. The extension of these results to the case of mixtures implies that only one of the solutions is consistent.

3. SOLUTION

We are still left with our original dilemma. Although previous work is not directly applicable to the case of mixtures, there is strong evidence that only one of the two solutions identified in Section 1 is consistent. The solution which would be obtained via our methods would be a function of the choice of initial parameter values only.

The cause of over dilemma lies with a basic lack of specificity of $g(p_1,\theta_1,p_2,\theta_2)$. In our current numerical procedures (see Kaylan and Harris, 1979), we are not free to define an "ordering" of the parameter sets for the mixed distributions. Thus this suggests that we adopt an ordering convention of $\theta_1 < \theta_2 < \theta_3 \ldots\ldots$ These would be strict inequalities since we do not allow the number of mixed distributions to be a random variable. In the case of a vector of parameters, the above would be extended to

$$\theta_{11} \leq \theta_{21} \leq \theta_{31} \ldots\ldots, \quad \theta_{12} \leq \theta_{22} \leq \theta_{32} \ldots\ldots,$$

$$\ldots\ldots \theta_{1n} < \theta_{2n} < \theta_{3n} \ldots\ldots$$

This ordering completely eliminates the dual solution problem. If $b < c$, then

$$p_1 = a$$
$$\theta_1 = b$$
$$p_2 = 1-a$$
$$\theta_2 = c$$

is feasible. The symmetric point

$$p_1 = 1-a$$
$$\theta_1 = c$$
$$p_2 = a$$
$$\theta_2 = b$$

is not allowable since $\theta_2 > \theta_1$. Operationally, nothing changes.

We employ our usual methodology to calculate the solution. If that solution has $\theta_2 > \theta_1$, we would simply use the symmetric counterpart.

4. IMPLICATIONS FOR ASYMPTOTIC NORMALITY

Given that the usual regularity conditions hold, asymptotic normality should not be affected by the ordering which we have adopted. If α is a vector of unknown parameters and $\hat{\alpha}$ is the likelihood estimator for α based on n observations, then $(\hat{\alpha} - \alpha)/\sqrt{n}$ has a limiting (as $n \to \infty$) multivariate normal distribution with matrix mean 0 and variance-covariance matrix V given as follows. The (i,j) entry of the inverse of V (call it R) is given by

$$r_{ij} = -E\left[\frac{\partial^2}{\partial\alpha_i\partial\alpha_j} \log f(X,\alpha)\right] .$$

Given an initial starting point the iterative algorithm will converge to one and only one of the symmetric maxima. That convergence point will be a function of the starting point alone. There will be no shifting between equivalent maxima. After K iterations, some α will have been obtained, and $(\hat{\alpha} - \alpha)/\sqrt{n}$ will be approximately multivariate normal with mean O and variance-covariance matrix V as above. But if $\hat{\alpha}$ does not satisfy the ordering conditions, then we determine its symmetric counterpart. This, however, implies an implicit re-ordering of α as well and consequently asymptotic normality will be maintained.

5. CONVERGENCE TO THE CONSISTENT SOLUTION

Thus far in this paper, we have suggested a technique whereby our methodology can differentiate between equivalent symmetric solutions to the likelihood equation. Given that this problem has been laid to rest, we have a method which moves along a direction of improvement to the log-likelihood function. Its convergence properties have also been shown. However, there is no assurance that the convergence will be to a global solution.

Only when the likelihood function is concave can one guarantee that a local solution (or an extreme point) is also a global solution. In general, we cannot guarantee the concavity of the likelihood function. This is shown by counterexample where a situation in which the Hessian of the likelihood function is shown to be positive definite. A function is concave if and only if its Hessian is negative semidefinite.

Consider the mixture of two exponentials

$$f_1(x) = \frac{1}{\eta_1} e^{-x/\eta_1}$$

and

$$f_2(x) = \frac{1}{\eta_2} e^{-x/\eta_2}$$

with p as the mixing proportion. In this case we set the shape parameters of the Weibulls to unity, thus making them both exponential. For further simplicity we also assume that η_1 is known and set to unity without loss of generality. For ease of notation we use η instead of η_2. Thus the mixture can be written as

$$g(x,\eta,p) = pe^{-x} + \frac{(1-p)}{\eta} e^{-x/\eta} .$$

In general, the elements of the Hessian (H) of the log-likelihood function are as follows:

$$\begin{cases} H_{11} = \dfrac{\partial^2 L}{\partial \eta^2} = \displaystyle\sum_{L=1}^{R} \left[-\frac{1}{g^2}\left(\frac{\partial g}{\partial \eta}\right)^2 \right] + \frac{1}{g}\frac{\partial^2 g}{2\eta^2} + \sum_{\ell=1}^{N-R} \left[-\frac{1}{\bar{G}^2}\left(\frac{\partial \bar{G}}{\partial \eta}\right)^2 + \frac{1}{\bar{G}}\frac{\partial^2 \bar{G}}{2\eta^2} \right] \\ H_{22} = \dfrac{\partial^2}{\partial p^2} = \displaystyle\sum_{L=1}^{R} \left[-\frac{1}{g^2}\left(\frac{\partial g}{\partial p}\right)^2 \right] + \sum_{\ell=1}^{N-R} \left[-\frac{1}{\bar{G}^2}\left(\frac{\partial \bar{G}}{\partial p}\right)^2 \right] \\ H_{12} = H_{21} = \dfrac{\partial^2 L}{\partial \eta \partial p} = \displaystyle\sum_{L=1}^{R} \left[-\frac{1}{g^2}\left(\frac{\partial}{\partial \eta}\right)\left(\frac{\partial g}{\partial p}\right) + \frac{1}{g}\frac{\partial^2 g}{\partial \eta^2 p} \right] + \sum_{\ell=1}^{N-R} \left[-\frac{1}{\bar{G}^2}\left(\frac{\partial \bar{G}}{\partial \eta}\right)\left(\frac{\partial \bar{G}}{\partial p}\right) + \frac{1}{\bar{G}}\frac{\partial^2 \bar{G}}{\partial \eta \partial p} \right] \end{cases}$$

In our counterexample, we assume that there is only observation and that observation is a failure. These assumptions eliminate the $\bar{G}$ terms and the summation signs. We now write the terms that we need to further consider this Hessian matrix:

$$\begin{cases} \dfrac{\partial g}{\partial \eta} = \dfrac{(1-p)}{\eta^2} e^{-x/\eta} \left(\dfrac{x}{\eta} - 1\right) \\ \dfrac{\partial^2 g}{\partial \eta^2} = \dfrac{(1-p)}{\eta^3} e^{-x/\eta} \left(\left(\dfrac{x}{\eta}\right)^2 - 4\dfrac{x}{\eta} + 2 \right) \\ \dfrac{\partial g}{\partial p} = e^{-x} - \dfrac{e^{-x/\eta}}{\eta} \\ \dfrac{\partial^2 g}{\partial p \partial \eta} = \dfrac{-e^{-x/\eta}}{\eta^2} \left(\dfrac{x}{\eta} - 1 \right) \end{cases}$$

A necessary condition for H to be negative semidefinite is that $H_{11} \leq 0$. If we can find an x for which $H_{11} > 0$, then we have our counterexample. To begin,

$$H_{11} = -\frac{1}{g^2}\left[\frac{(1-p)}{\eta} e^{-x/\eta}\right]^2 \frac{1}{\eta^2}\left(\frac{x}{\eta} - 1\right)^2 + \frac{1}{g}\left[\frac{1-p}{\eta} e^{-x/\eta}\right] \frac{1}{\eta^2}\left[\left(\frac{x}{\eta}\right)^2 - 4\frac{x}{\eta} + 2\right].$$

Denote $\frac{1-p}{\eta} e^{-x/\eta}$ as f'. Also recognize that $f' \leq g$. We then have

$$H_{11} = -\left(\frac{f'}{g}\right)^2 \frac{1}{\eta^2}\left(\frac{x}{\eta} - 1\right)^2 + \frac{f}{g}\frac{1}{\eta^2}\left[\left(\frac{x}{\eta}\right)^2 - \frac{4x}{\eta} + 2\right].$$

Note that $H_{11} > 0$ implies that

$$\frac{f'}{g}\left[\left(\frac{x}{\eta}\right)^2 - \frac{4x}{\eta} + 2\right] > \left(\frac{f'}{g}\right)^2\left[\left(\frac{x}{\eta}\right)^2 - \frac{2x}{\eta} + 1\right].$$

Since $f' < g$, $f'/g > (f'/g)^2$ for any x. Thus if we can find an x such that

$$\left(\frac{x}{\eta}\right)^2 - \frac{4x}{\eta} + 2 > \left(\frac{x}{\eta}\right)^2 - \frac{2x}{\eta} + 1$$

the counterexample is complete. Any $x < \eta/2$ is such a point.

The counterexample is not truly complete however. We next show the existence of local solutions. It is convenient now to look at the likelihood function itself rather than the log-likelihood. The results are not affected. For a single observation, the likelihood function is g itself:

$$g = pe^{-x} + \frac{(1-p)}{\eta} e^{-x/\eta}.$$

We first locate points where the gradient of g is equal to zero. In the following analysis, x is treated somewhat like a variable. We solve for points where the gradient of the likelihood function is zero, allowing x to assume the most suitable value. Once these points are found, we treat x as an observation and assume that its value is that for which we originally solved.

The first partial with respect to p is

$$\frac{\partial g}{\partial p} = e^{-x} - \frac{1}{\eta} e^{-x/\eta} \quad ;$$

thus

$$\frac{\partial g}{\partial p} = 0 \Rightarrow \eta = 1 \quad \text{or} \quad \frac{\eta}{\eta-1} \ln\eta = x \quad .$$

If we let $x = 2\ln 2$, then $\eta = 2$ also implies $\partial g/\partial p = 0$. Next,

$$\frac{\partial g}{\partial \eta} = \frac{(1-p)}{\eta^2} e^{-x/\eta} \left[\frac{x}{\eta} - 1 \right]$$

and

$$\frac{\partial g}{\partial \eta} = 0 \Rightarrow p = 1 \quad \text{or} \quad \eta = x = 2\ln 2. \text{[1]}$$

Thus the points at which $\nabla g = 0$ are as follows for the observation $x = 2\ln 2$:

$(\eta = 1, \ p = 1)$

$(\eta = 2, \ p = 1)$

At both points, which turn out to be saddle points as seen on the graph shown later in the section, the value of the log-likelihood is e^{-x} or .25. Also note that the log-likelihood has the same value along the lines $\eta = 1$, $\eta = 2$, and $p = 1$.

Let us now further examine the shape of the surface. First consider what happens when, for a fixed $p (\neq 1)$, one maximizes with respect to η:

$$\begin{cases} \dfrac{\partial g}{\partial \eta} = \dfrac{(1-p)}{\eta^2} e^{-x/\eta} \left(\dfrac{x}{\eta} - 1 \right) \\ \dfrac{\partial g}{\partial \eta} = 0 \Rightarrow \eta = x = 2\ln 2 \\ \dfrac{\partial^2 g}{\partial \eta^2} = \dfrac{(1-p)}{\eta^3} e^{-x/\eta} \left[\left(\dfrac{x}{\eta} \right)^2 - \dfrac{4x}{\eta} + 2 \right] \end{cases}$$

At the point $\eta = x = 2\ln 2$,

$$\frac{\partial^2 g}{\partial \eta^2} = - \frac{(1-p)e^{-1}}{(2\ln 2)^3} < 0. \ .$$

Therefore we have a maximum.

---

[1] This is the value at the observation as previously solved.

Now consider what the surface is doing for fixed η. We have already seen that $\frac{\partial g}{\partial p} = 0$ when $\eta = 1$ and $\eta = 2$. Therefore the surface is constant for all values of p at those points. When $0<\eta<1$ or $\eta>2$

$$\frac{\partial g}{\partial p} > 1.$$

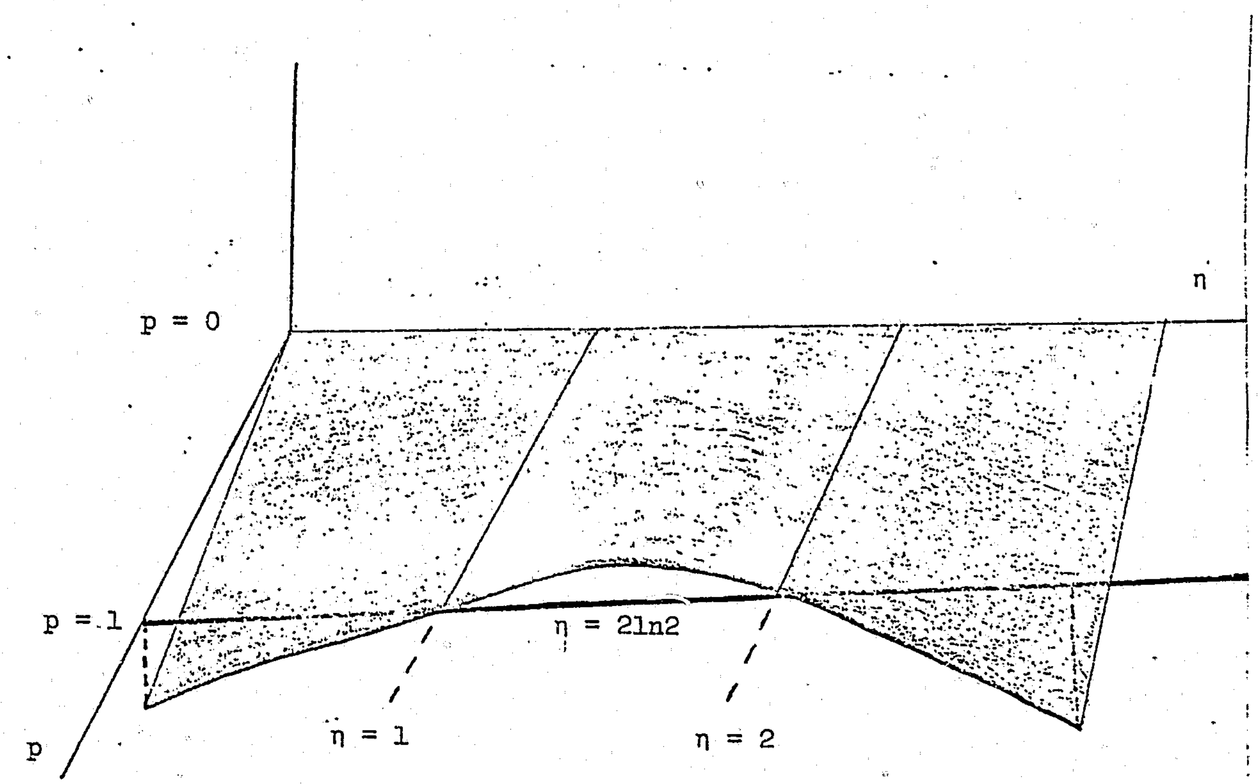
Therefore the likelihood function is increasing when p increases. When $1<\eta<2$,

$$\frac{\partial g}{\partial p} < 1$$

and consequently the function increases as p approaches zero.

The following is a rough graph of the likelihood function.

Figure 1
SAMPLE LIKELIHOOD FUNCTION



From the above discussion, it is clear that the maximum is at the point $p = 0$ and $\eta = 2\ell n2$. The objective function value is

$$\frac{e^{-1}}{2\ell n2} = .2654.$$

From the graph, it is also clear that there are local solutions for points where $p = 1$ and $0<\eta<1$. Consequently, local solutions do exist and these solutions may have objective functions values less than the global maximum.

Interestingly however, as pathological as this example is, the method still works. As long as the starting point is not on the boundary, the method will reach the global solution. This is because the iterative equations are as follows.

$$\begin{cases} \eta^{v+1} = x \\ p^{v+1} = p^{v} \dfrac{e^{-x}}{pe^{-x} + (1-p)\,\dfrac{e^{-x/\eta^{v}}}{\eta^{v}}} \end{cases}$$

After one iteration the point will move to the line $\eta = 2\ell n2$ and stay there. After that, the p value will move toward zero.

REFERENCES

Chanda, K.C. (1954), "A Note on the Consistency and Maxima of the Roots of Likelihood Equations", Biometrika, Vol. 41, pp. 56-61.

Cramer, H. (1946), Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey.

Huzurbazar, V.S. (1948), "The Likelihood Equation, Consistency and the Maxima of the Likelihood Equation", Annals of Eugenics, Vol. 14, p. 185.

Kaylan, A.R. and C.M. Harris (1979), "Efficients Algorithms to Derive Maximum-Likelihood Estimates for Finite Exponential and Weibull Mixtures", pending review with Computers and Operations Research.

Perlman, M.D. (1969), "The Limiting Behavior of Multiple Roots of the Likelihood Equation", University of Minnesota Technical Report No. 125.

Tarone, R.E. and G. Gruenhage (1975), "A Note on the Uniqueness of Roots of the Likelihood Equations for Vector-Valued Parameters", Jl. Amer. Stat. Assn. Vol. 70, pp. 903-904.

APPENDIX E

PARAMETER ESTIMATION UNDER PROGRESSIVE CENSORING CONDITIONS FOR A FINITE MIXTURE OF WEIBULL DISTRIBUTION*

by

Jay Mandelbaum
United States Department of Transportation
Washington, D.C.

and

Carl M. Harris
Center for Management & Policy Research, Inc.
Washington, D.C.

ABSTRACT: Algorithms have been devised to maximize the likelihood function for random samples from mixed Weibull populations under progressive censoring. This is done for a number of different sampling environments which arose in an extensive series of criminal justice program evaluations. Convergence results have been obtained in each case, and the question of local vs. global optimality is explored.

*Prepared under Grant No. 79 NI AX 0068 to the Center for Criminal Justice, University of Illinois at Chicago Circle, from the National Institute of Law Enforcement and Criminal Justice, United States Department of Justice.

1. Background

In recent years, several models featuring mixtures of distributions have been structured to describe phenomena in the criminal justice field. Carr-Hill and Carr-Hill (1972) introduced a model for recidivism in which they assumed that each released prisoner belongs to either a "quick" or "slow" reconviction group. They defined constant reconviction rates for the two subpopulations and further assumed a random process governing the transfer of members from the "quick" to the "slow" group. These assumptions yield a mixture of two exponential distributions as the probability distribution function of reconvictions over time. Carr-Hill and Carr-Hill did not, however, directly face the issue of estimation of parameter and mixing proportions. Estimates were found by a bit of guesswork and then partly verified to be reasonable via a chi-square test.

Greenberg (1978) proposed a modification of the Carr-Hill/Carr-Hill model to take permanently law-abiding people into consideration. He suggested that the population be viewed as three groups; i.e., strictly law-abiding, potential reconviction, and uncommitted. Once again, however, estimation issues were not resolved.

Another model with two groups in the released population was developed by Maltz and McCleary (1977). Maltz and McCleary treated members of the first group as "successes" and assigned them a zero probability of reconviction. The second group is assumed to fail exponentially with a constant failure rate. Iterative equations to obtain maximum-likelihood estimators for the failure rate and the proportion of the population which failed were derived and illustrated in their work.

The major early efforts on the estimation of parameters for mixture-type models were due to Hasselblad (1969) and Mendenhall and Hader (1958). However, as is typical in these kinds of problems, the algorithms developed have very poor convergence properties and are not generally applicable to a wide range of sampling situations which arise in the real world.

To eliminate some of these problems, Kaylan (1979) developed and tested a special iterative scheme to calculate maximum-likelihood estimates of mixing pro-

portions and probability-density-function parameters for a finite mixture of exponential or Weibull distributions where all individuals are assumed to fail (complete sampling). Previous work in the exponential area can be found in the aforementioned work of Hasselblad and Mendenhall and Hader. Kaylan also treated the case commonly called Type I censoring, whereby all observations of non-failures are assumed to begin and observation terminates at the same time whether or not failure has actually occurred.

2. Introduction

This paper deals with the estimation of mixing proportions and parameters of a finite number of mixed Weibull distributions under conditions of progressive censoring. By progressive censoring, we mean the following. Observations of objects or individuals may start at an arbitrary time. If there is a failure during the observation period, then the total operating time is recorded and denoted by x_i. However, an individual does not have to fail during the observation period since observations are allowed to be terminated at any point in time. The total observation time for an individual who does not fail is also recorded and denoted by y_j which is the difference between the points in time at which observation was started and ended for that individual.

Our goal is to obtain maximum-likelihood estimators of the parameters and mixing proportions under the assumptions that the data come from a mixed Weibull population. Under fairly general conditions, the estimators are efficient, invariant, consistent, unbiased, and asymptotically normal. They are also functions of sufficient statistics if such exist and have minimum variance. The problem is that it is not possible to obtain an explicit form for the estimator by taking partial derivatives of the likelihood function and setting them equal to zero. In addition, we encounter the constraint that the sum of the mixing proportions must equal unity. The resultant problem can be described as a mathematical program with a nonlinear objective function and linear constraints. Before one attempts to use a mathematical programming algorithm, however, it is advisable to attempt to maximize the log-likelihood function without taking the constraints into account. If the answer is feasible, the problem is solved with much less computational effort. In any event, we utilize iterative numerical procedures to calculate parameter estimates.

Section 3 establishes notation and also presents some basic results which will be referenced throughout the paper. Parameter estimates can be made under two different assumptions concerning the failed individuals. After a failure we may assume either: (1) we know the true density in the mixture from which the failure came; or, (2) we do not know the true parent. The former case is labelled "non-post-mortem" and the latter is termed "post-mortem." Sections 4 and 5, respectively, present a first and second-order iterative scheme for parameter estimation under non-post-mortem conditions. Convergence proofs are also given. Section 6 derives an unconstrained method for estimation under special post-mortem conditions. Then, a first-order iterative scheme for the post-mortem case is presented in Section 7 along with a convergence proof. Section 8 gives the iterative formulas developed in 4 and 5 for the exponential. Finally, a two-phase second-order method is described in Section 9.

3. Notation

Since we are working with mixtures of K Weibull density functions, the probability density function of the jth Weibull in the mixture will be given as (for $x, \beta_j, \eta_j > 0$)

$$f_j(x;\beta_j,\eta_j) = (\beta_j/\eta_j)(x/\eta_j)^{\beta_j - 1}\exp[-(x/\eta_j)^{\beta_j}] \quad .$$

The mixtures of K Weibulls can thus be expressed as

$$g(x,\alpha) = \sum_{j=1}^{K} p_j f_j(x) \qquad (p_j \geq 0, \sum_{j=1}^{K} p_j = 1)$$

where α is a vector of the 3K-1 unknown parameters.

The cumulative distribution functions for $f_j(x)$ and $g(x,\alpha)$ can also be written respectively as

$$F_j(x) = 1 - \exp[-(x/\eta_j)^{\beta_j}]$$

and

$$G_j(x,\alpha) = \sum_{j=1}^{K} p_j F_j(x) \quad .$$

It turns out that it is more convenient to work with the complements of the above cumulative distribution functions, namely,

$$\overline{F}_j(x) = 1 - F_j(x) = \exp[-(x/\eta_j)^{\beta_j}]$$

and

$$\overline{G}_j(x,\alpha) = 1 - G(x,\alpha) = \sum_{j=1}^{K} p_j \overline{F}_j(x) \quad .$$

Since we shall constantly be taking partial derivatives of $g(x,\alpha)$ and $\overline{G}(x,\alpha)$ throughout the remainder of this paper, we establish these functions now:

$$\frac{\partial g(x,\alpha)}{\partial \beta_j} = p_j f_j(x)\left[\frac{1}{\beta_j} + \ln\frac{x}{\eta_j} - \left(\ln\frac{x}{\eta_j}\right)\left(\frac{x}{\eta_j}\right)^{\beta_j}\right] \tag{1}$$

$$\frac{\partial \overline{G}(x,\alpha)}{\partial \beta_j} = -p_j \overline{F}_j(x)\left(\ln\frac{x}{\eta_j}\right)\left(\frac{x}{\eta_j}\right)^{\beta_j} \tag{2}$$

$$\frac{\partial g(x,\alpha)}{\partial \eta_j} = p_j f_j(x)\left[\left(\frac{x}{\eta_j}\right)^{\beta_j} - 1\right]\frac{\beta_j}{\eta_j} \tag{3}$$

$$\frac{\partial \overline{G}(x,\alpha)}{\partial \eta_j} = p_j \overline{F}_j(x)\left(\frac{x}{\eta_j}\right)^{\beta_j}\frac{\beta_j}{\eta_j} \tag{4}$$

Equations (1) through (4) hold for j = 1, 2, ..., K. When we differentiate with respect to p_j, recalling that the p_j sum to 1, we find that

$$\left\{\begin{array}{lll} \frac{\partial g(x,\alpha)}{\partial p_j} = f_j(x) - f_k(x) & (j = 1,2,\ldots,K-1) & (5) \\ \frac{\partial \bar{G}(x,\alpha)}{\partial p_j} = \bar{F}_j(x) - \bar{F}_k(x) & (j = 1,2,\ldots,K-1) & (6) \end{array}\right.$$

4. A First-Order Method -- Non-Post-Mortem

Our first method for estimating the mixing proportions and the parameters of the K Weibulls is a first-order iterative method. We assume that there are N observations, R of which are failures during the observation period. We also assume that when there is a failure, we do not know from which of the K Weibull density functions it came. Consistent with the literature, this is called non-post-mortem sampling. Later on, we do treat a special case of post-mortem analysis with K = 2.

The likelihood equation for this problem will be

$$\mathcal{L}(\alpha) = \frac{N!}{(N-R)!} \prod_{i=1}^{R} g(x_i,\alpha) \prod_{\ell=1}^{N-R} \bar{G}(y_j,\alpha) \qquad (7)$$

From this point on, we shall adopt a shorthand notation, by setting $f_j = f_j(x_i)$, $g = g(x_i,\alpha)$, $\bar{G} = G(y_\ell,\alpha)$, $\bar{F}_j = \bar{F}_j(y_\ell)$. So we have

$$\mathcal{L} = \frac{N!}{(N-R)!} \prod_{i=1}^{R} g \prod_{\ell=1}^{N-R} \bar{G} .$$

Consistent with standard methods of finding maximum-likelihood estimators, we take logarithms to obtain

$$L = \ln \mathcal{L} = \ln\left[\frac{N!}{(N-R!)}\right] + \sum_{i=1}^{R} \ln g + \sum_{\ell=1}^{N-R} \ln \bar{G}$$

We now take partial derivatives with respect to β_j, η_j, and p_j and then set them equal to zero. The first term of the log-likelihood is constant and hence does not affect the differentiation. As is shown below, we find that it is not possible to solve explicitly for the parameter; thus, we invoke the following numerical procedure. If there are m equations in m unknowns, we separate each of the m unknowns to the left-hand side of the m equations. Each right-hand side however, will not be independent of the variable on its corresponding left-hand side. The standard procedure is to solve iteratively for the unknowns with the right-hand sides containing values of the νth iteration and the left-hand sides being the values at the (ν+1)st iteration. For example assume there are two parameters a_1 and a_2, such that

$$\left\{\begin{array}{l} a_1 = h_1(a_1,a_2) \\ a_2 = h_2(a_1,a_2) \end{array}\right.$$

Then we may solve iteratively for a_1 and a_2. If we use a superscript to index the iterations, then

$$a_1^{\nu+1} = h_1(a_1^\nu, a_2^\nu) , \qquad a_2^{\nu+1} = h_2(a_1^\nu, a_2^\nu) .$$

The likelihood analysis thus proceeds as follows:

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{R} \frac{1}{g}\frac{\partial g}{\partial \beta_j} + \sum_{\ell=1}^{N-R} \frac{1}{G}\frac{\partial \bar{G}}{\partial \beta_j} \qquad (j = 1,2,\ldots,K) .$$

From Equations (1) and (2) we get

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{R} \frac{p_j f_j}{g} \left[\frac{1}{\beta_j} + \ln \frac{x_i}{\eta_j} - \left(\ln \frac{x_i}{\eta_j}\right) \left(\frac{x_i}{\eta_j}\right)^{\beta_j} \right]$$

$$- \sum_{\ell=1}^{N-R} \frac{p_j \overline{F}_j}{\overline{G}} \left(\ln \frac{y_\ell}{\eta_j}\right) \left(\frac{y_\ell}{\eta_j}\right)^{\beta_j} \qquad (j = 1,2,\ldots,K). \quad (8)$$

If we now set $\partial L/\partial \beta_j = 0$ for all j, we obtain

$$\beta_j = \frac{\sum_{i=1}^{R} \left(\frac{f_j}{g}\right)}{\sum_{i=1}^{R} \left[\frac{f_j}{g}\right] \left(\ln \frac{x_i}{\eta_j}\right) \left[\left(\frac{x_i}{\eta_j}\right)^{\beta_j} - 1\right] + \sum_{\ell=1}^{N-R} \frac{\overline{F}_j}{\overline{G}} \left(\ln \frac{y_\ell}{\eta_j}\right) \left(\frac{y_\ell}{\eta_j}\right)^{\beta_j}} \quad . \quad (9)$$

Similarly,

$$\frac{\partial L}{\partial \eta_j} = \sum_{i=1}^{R} \frac{1}{g} \frac{\partial g}{\partial \eta_j} + \sum_{\ell=1}^{N-R} \frac{1}{\overline{G}} \frac{\partial \overline{G}}{\partial \eta_j} \qquad (j = 1,2,\ldots,K) .$$

From Equations (3) and (4) we then obtain (for $j = 1, 2, \ldots, K$)

$$\frac{\partial L}{\partial \eta_j} = \sum_{i=1}^{R} \frac{p_j f_j}{g} \left[\left(\frac{x_i}{\eta_j}\right)^{\beta_j} - 1\right] \frac{\beta_j}{\eta_j} + \sum_{\ell=1}^{N-R} \frac{p_j \overline{F}_j}{\overline{G}} \left(\frac{y_\ell}{\eta_j}\right)^{\beta_j} \frac{\beta_j}{\eta_j} \quad . \quad (10)$$

When these are set to zero, we find that

$$\eta_j = \left[\frac{\left[\sum_{i=1}^{R} \left(\frac{f_j}{g}\right) x_i^{\beta_j} + \sum_{\ell=1}^{N-R} \left(\frac{\overline{F}_j}{\overline{G}}\right) y_\ell^{\beta_j}\right]}{\sum_{i=1}^{R} \frac{f_j}{g}} \right]^{\frac{1}{\beta_j}} \quad . \quad (11)$$

For the mixing proportions we have

$$\frac{\partial L}{\partial p_j} = \sum_{i=1}^{R} \frac{1}{g} \frac{g}{p_j} + \sum_{\ell=1}^{N-R} \frac{1}{\overline{G}} \frac{\partial \overline{G}}{\partial p_j} \qquad (j = 1,2,\ldots,K-1) ,$$

which becomes via Equation (4) and (5)

$$\frac{\partial L}{\partial p_j} = \sum_{i=1}^{R} \frac{1}{g} (f_j - f_K) + \sum_{\ell=1}^{N-R} \frac{1}{\overline{G}} (\overline{F}_j - \overline{F}_K) \qquad (j = 1,2,\ldots,K-1).$$

When we set $\partial L/\partial p_j$ to zero we arrive at

$$\sum_{i=1}^{R} \frac{f_j}{g} + \sum_{\ell=1}^{N-R} \frac{\overline{F}_j}{\overline{G}} = C \qquad (j = 1,2,\ldots,K), \qquad (13)$$

where C is an appropriate constant. If we multiply both sides of Equation (13) by p_j, sum over j, and simplify, then Equation (13) becomes

$$p_j = \frac{p_j}{N} \left[\sum_{i=1}^{R} \frac{f_j}{g} + \sum_{\ell=1}^{N-R} \frac{\overline{F}_j}{\overline{G}} \right] \qquad (j = 1,2,\ldots,K) . \qquad (14)$$

Equations (9), (11), and (14) are the basis of the iterative procedure for finding mixing proportions and distribution parameters. The left-hand sides represent their values at the (v+1)<u>st</u> iteration. The functions on the right-hand sides contain values at the v<u>th</u> iteration.

The iterative scheme, as it has been developed, can be improved via techniques commonly used in the mathematical programming environment. A math programming algorithm is composed of two main features -- the generation of a direction which will lead to improvement in the objective function and the step size or line search problem which indicates how far to move in the prescribed direction. For a non-concave problem, such a direction is only guaranteed to instantaneously lead to improvement in the objective function. Thus a step of arbitrary length may or may not be beneficial. Consequently, math programming algorithms commonly generate a step size, s*, as the solution of

$$\max_{s} \; L\left[\alpha^{v+1} + s\,(\alpha^{v+1} - \alpha^{v})\right] .$$

As our iterative scheme has been posed, the step size is automatically calculated. We have developed equations which lead to α^{v+1}. In the following section we show that the vector $(\alpha^{v+1} - \alpha^{v})$ points in a direction of increasing log-likelihood. We still encounter the possibility that the selection of α^{v+1} does not lead to improvement. In order to insure such improvement, and at the same time avoid the additional computation required to solve the line search problem, we heuristically will bisect the step until an increase is realized. For example, we will first try

$$\alpha = \alpha^{v} + (\alpha^{v+1} - \alpha^{v}); \qquad \text{then,} \quad \alpha = \alpha^{v} + \tfrac{1}{2}\,(\alpha^{v+1} - \alpha^{v}) \text{, etc.}$$

4.1 <u>Convergence Properties</u>

In this section, we prove the convergence of the foregoing algorithm. The conditions which constitute global convergence (see Luenberger, 1973) form the basis of the proof. We need to show that:

1) α^{v} belongs to a compact set;
2) the algorithm generates a sequence of points such that each new point causes the log-likelihood to increase in value;
3) α^{v} is feasible.

As in Kaylan (1979), Equations (9), (11), and (14) constitute a mapping from α^{v} to α^{v+1}. Since all of the functions in the equations are continuous, the mapping is closed. Hence α^{v} belongs to a compact set.

To show that the algorithm generates a sequence of points so that the log-likelihood increases in value, it is sufficient to show that the following inner product is nonnegative:

$$\left[\alpha^{v+1} - \alpha^{v}\right] \cdot \nabla L_{\alpha}^{v} \geq 0 . \qquad (15)$$

We show this in three stages. First,

$$\left[\beta_j^{v+1} - \beta_j^{v}\right] \cdot \nabla L_{\beta}^{v} \geq 0 \qquad (j = 1,2,\ldots,K) . \qquad (16)$$

From Equation (9), after some algebra, we obtain

$$\beta_j^{v+1} - \beta_j^{v} = \left[\frac{(\beta_j^{v+1})(\beta_j^{v})}{p_j^{v} \sum_{i=1}^{R} f_i^{v}/g^{v}}\right] \left[\frac{\partial L}{\partial \beta_j}\right]^{v} \qquad (j = 1,2,\ldots,K) .$$

Since the coefficient of $[\partial L/\partial \beta_j]^{v}$ here is nonnegative, it is clear that the condition in Equation (16) is satisfied. The second stage is to show that

$$\left[\eta_j^{v+1} - \eta_j^{v}\right] \cdot \nabla L^{v} \geq 0 \qquad (j = 1,2,\ldots,K) . \qquad (17)$$

After some algebra building from Equation (11), we find that

$$\text{Sign}\,(\eta_j^{v+1} - \eta_j^{v}) = \frac{\beta_j^{v+1}(\eta_j^{v})}{p_j^{v}\,\beta_j^{v}\,(\sum_{i=1}^{R} f_j^{v}/g^{v})} \left[\frac{\partial L}{\partial \eta_j}\right]^{v} \qquad (j = 1,2,\ldots,K) .$$

Since the coefficient in the above is nonnegative, Equation (17) is satisfied.

The final stage is to show that

$$\sum_{j=1}^{K} \left[p_j^{v+1} - p_j^{v}\right] \cdot \nabla L_p^{v} \geq 0 . \qquad (18)$$

The proof is patterned after Hasselblad (1967). After substituting for ∇L_p^v from Equation (12), the left-hand side becomes

$$\sum_{j=1}^{K}\left[p_j^{v+1}-p_j^{v}\right]\cdot\left\{\sum_{i=1}^{R}\frac{(f_j^v-f_K^v)}{g^v}+\sum_{\ell=1}^{N-R}\frac{(\overline{F}_j^v-F_K^v)}{\overline{G}^v}\right\}.$$

Equation (14) tells us, however, that

$$\frac{N\,p_j^{v+1}}{p_j^v}=\sum_{i=1}^{R}\frac{f_j^v}{g}+\sum_{\ell=1}^{N-R}\frac{\overline{F}_j^v}{\overline{G}^v}\,.$$

Therefore the left-hand side is equal to

$$N\sum_{j=1}^{K}\left[p_j^{v+1}-p_j^{v}\right]\cdot\left\{\frac{p_j^{v+1}}{p_j^v}-\frac{p_K^{v+1}}{p_K^v}\right\} \qquad (19)$$

$$= N\left[\sum_{j=1}^{K}\frac{(p_j^{v+1})^2}{p_j^v}-1\right].$$

If we now let

$$p_j^{v+1}=p_j+\delta_j \qquad (j=1,2,\ldots,K)$$

where the δ_j sum to zero, then after some algebra (19) becomes

$$N\left[1+\sum_{j=1}^{K}\frac{\delta_j^2}{p_j}\right].$$

This is guaranteed to be positive and hence Equation (18) is true. Equations (16), (17), and (18) together show that Equation (15) holds and thus that the algorithm does generate a sequence of points such that each new point causes the log-likelihood to increase in value.

The final step in the convergence proof is to show α^v is feasible. This implies that $p_j \geq 0$, β_j, $\eta_j > 0$ $(j = 1,2,\ldots,K)$ and that $\Sigma p_j = 1$. If α^o is such that $p_j \geq 0$, β_j, $\eta_j > 0$ $(j = 1,2,\ldots,K)$, then this condition will be maintained through all iterations since the right-hand sides of Equations (11) and (14) must be positive.

Under many conditions, the right-hand side of Equation (9) will also be positive, but this is not guaranteed. Thus we resort to a heuristic method to avoid this difficulty -- a bisection of the step size as described in the beginning of this section. We have

$$\alpha^{v+1}=\alpha^v+s\,(\alpha^{v+1}-\alpha^v).$$

We initially set s to unity. If some β_j is negative, then we will set s to 1/2, 1/4, 1/8, ... until feasibility is achieved. To show that $\sum_{j=1}^{K} p_j = 1$, it is sufficient to show that $\sum_{j=1}^{K-1} p_j < 1$. If we take Equation (14) and sum over j, then it is sufficient to show that $\sum_{j=1}^{K-1} p_j < 1$. If we take Equation (14) and sum over j, then

$$\sum_{j=1}^{K-1}p_j=\frac{1}{N}\left[\sum_{i=1}^{R}\frac{\sum_{j=1}^{K-1}p_j f_j}{g}+\sum_{\ell=1}^{N-R}\frac{\sum_{j=1}^{K-1}p_j\overline{F}_j}{\overline{G}}\right].$$

Since

$$g=\sum_{j=1}^{K}p_j f_j \quad \text{and} \quad \overline{G}=\sum_{j=1}^{K}p_j\overline{F}_j\,,$$

then

$$\sum_{j=1}^{K-1}\frac{p_j f_j}{g}<1 \quad \text{and} \quad \sum_{j=1}^{K-1}\frac{p_j\overline{F}_j}{g}<1\,.$$

Thus

$$\sum_{j=1}^{K-1}p_j<\frac{1}{N}\left(\sum_{i=1}^{R}1+\sum_{\ell=1}^{N-R}1\right)=1.$$

5. A Second Order Method -- Non-Post-Mortem

Here the log-likelihood equation may be written as

$$L = \ln \frac{N!}{(N-R)!} + \sum_{i=1}^{R} \ln \sum_{j=1}^{K} p_j f_j + \sum_{\ell=1}^{N-R} \ln \sum_{j=1}^{K} p_j \bar{F}_j \ .$$

We want to make use of the fact that a monotone increasing concave function of a concave function is concave. Any function which is linear in the $\{p_j\}$ is concave with respect to the $\{p_j\}$. Also the logarithm is a monotone increasing concave function. Thus

$$\ln \sum_{j=1}^{K} p_j f_j \quad \text{and} \quad \ln \sum_{j=1}^{K} p_j \bar{F}_j$$

are both concave functions. Since the sum of concave functions is convave, the log-likelihood function is concave with respect to the $\{p_j\}$.

We next look at the sub-Hessian matrix as

$$\nabla^2 L_p = \frac{\partial^2 L}{\partial p_{j_1} \partial p_{j_2}} \qquad (j_1 = 1,2,\ldots,K-1; \ j_2 = 1,2,\ldots,K-1) \ .$$

Equation (12) defined $\nabla L_p = \partial L/\partial p_j$; therefore

$$\nabla L_p = \sum_{i=1}^{R} \frac{1}{g} (f_{j_1} - f_K) + \sum_{\ell=1}^{N-R} \frac{1}{G} (\bar{F}_{j_1} - \bar{F}_K) \qquad (j_1 = 1,2,\ldots,K-1).$$

Thus (for j_1, j_2 = 1,2, ... K-1)

$$\nabla^2 L = \sum_{i=1}^{R} \frac{(f_{j_1} - f_K)(f_{j_2} - f_K)}{g^2} - \sum_{\ell=1}^{N-R} \frac{(\bar{F}_{j_1} - \bar{F}_K)(\bar{F}_{j_2} - \bar{F}_K)}{\bar{G}^2}$$

On the basis of this scheme we may use Newton's Method for generating the mixing proportions vector equation

$$p^{v+1} = p^v - (\nabla^2 L_p^v)^{-1} (\nabla L_p^v) \ . \qquad (20)$$

Therefore the second-order scheme uses Equation (20) instead of Equation (14). The step-size issue, as discussed within the context of the first-order method, is applicable here as well. In order to guarantee improvement in the log-likelihood at the (v+1)st iteration, we will bisect the step size until an increase is realized.

5.1 Convergence Properties

The properties listed in Section 4.1 to show convergence of the first-order method must also be shown to hold in this second-order method. But we need only examine the properties for the differences between the two algorithms. Clearly, α^v belongs to a compact set since the functions in Equation (20) are continuous. We must next prove that

$$\left[\alpha^{v+1} - \alpha^v\right] \cdot \nabla L_\alpha^v \geq 0 \ .$$

Section 4.1 showed that the above holds for the β_j and η_j components of α. Thus we only need to show that

$$\left[p^{v+1} - p^v\right] \cdot \nabla L_p^v \geq 0 \ .$$

From Equation (20) we obtain

$$(\nabla L_p^v) \cdot \left[p^{v+1} - p^v\right] = \left[\nabla L_p^v\right] \left[\nabla^2 L_p^v\right]^{-1} \left[\nabla L_p^v\right] \ .$$

CONTINUED

1 OF 4

Since L is concave with respect to the mixing proportions, $\left[\nabla^2 L_p\right]^{-1}$ is negative semidefinite. Therefore the right-hand side of the above is nonnegative.

The last step in the convergence proof is to show feasibility. Since the β_j and η_j equations are the same as in the first-order method, we need only show that the $\{p_j\}$ are nonnegative and sum to unity. Unfortunately the Newton step does not take the constraints into account. There is no guarantee that the $\{p_j\}$ will sum to unity or remain nonnegative. We may however resort to a heuristic method to avoid this problem. If we consider p^v and p^{v+1} to be (K-1)-dimensional vectors and let $\left[p_K = 1 - \sum_{j=1}^{K-1} p_j\right]$, then Equation (20) represents a move in (K-1) space. But Equation (20) is also a step-size equation in which s has been set to one. It can be rewritten as

$$p^{v+1} = p^v - s \;(\nabla^2 L_p^v)^{-1} \;(\nabla L_p)^v \quad .$$

If a value of s implies either

$$\sum_{j=1}^{K-1} p_j > 1 \qquad \text{or} \qquad p_j < 0 \quad \text{for any} \quad j = 1,2,\ldots,K-1,$$

then we will bisect s and try again for feasibility. Thus we will begin with s = 1; if the resulting p^{v+1} is infeasible, we will try s = 1/2, s = 1/4, etc., until feasibility is achieved. Any of these steps is guaranteed to lead to an improvement in the log-likelihood, since the log-likelihood function is concave in the $\{p_j\}$.

6. An Unconstrained Problem -- Post Mortem

Up to this point, we have been dealing with N observations, R of which fail during the observation period. We also have assumed that the parents of the R failures are unknown. In a "post-mortem" case, we assume instead that the underlying distribution of a failure is known. For this problem, we need to slightly amend our previous notation. For the R failures, assume that R_j of them were found to belong to the jth parent $f_j(x)$ where j = 1,2,...,K, and $\sum_{j=1}^{K} R_j = R$. Previously, f_j was $f_j(x_i)$ - but now f_j will denote $f_j(x_{ij})$ where x_{ij} is the failure time for the ith object with parent $f_j(x)$.

The likelihood function for the post-mortem problem is thus

$$L = \frac{N!}{(N-R)!} \prod_{j=1}^{N-R} \overline{G} \prod_{j=1}^{K} p_j^{R_j} \cdot \prod_{j=1}^{K} \prod_{i=1}^{R_j} f_j \;;$$

and

$$\ln L = \ln \frac{N!}{(N-R)!} + \sum_{\ell=1}^{N-R} \ln \overline{G} \sum_{j=1}^{K} R_j \ln p_j + \sum_{j=1}^{K} \sum_{i=1}^{R_j} \ln f_j \quad . \qquad (21)$$

We first pose an unconstrained post-mortem problem in which L will be maximized. Since the first term in L is constant, we define

$$L_{mod} = \sum_{\ell=1}^{N-R} \ln \overline{G} + \sum_{j=1}^{K} R_j \ln p_j + \sum_{j=1}^{K} \sum_{i=1}^{R_j} \ln f_j \quad .$$

The set of parameter values which maximizes L_{mod} will also maximize L. In the special case where K = 2, the following transformations of variables are made:

$$\begin{cases} \beta_j = u_j^2 & j = 1,2 \\ \eta_j = v_j^2 & j = 1,2 \end{cases}$$

This transformation guarantees that β_j and $\eta_j > 0$. If we also set $p_1 = \sin^2 w$ and $p_2 = \cos^2 w$, then $p_1 + p_2$ must equal unity. When the above is substituted into L_{mod}, an unconstrained maximization problem results since all of the constraints are guaranteed to hold. The solution may be gotten from any standard unconstrained nonlinear optimization procedure.

7. A First-Order Method -- Post Mortem

Let us now draw a parallel to the first-order method of Section 4, but for the post-mortem analysis. The restriction (K = 2) of the previous section is dropped. We use the procedure of differentiating the log-likelihood function with respect to β_j, η_j and p_j, setting the derivatives to zero, and then separating values to form an iterative procedure. The log-likelihood function of Equation (21) is the starting point:

$$\frac{\partial L}{\partial \beta_j} = \sum_{\ell=1}^{N-R} \frac{1}{\bar{G}} \frac{\partial \bar{G}}{\partial \beta_j} + \sum_{i=1}^{R_j} \frac{1}{f_j} \frac{\partial f_j}{\partial \beta_j} \qquad (j = 1,2,\ldots,K) .$$

After substituting from Equations (1) and (2) this becomes

$$\frac{\partial L}{\partial \beta_j} = \sum_{\ell=1}^{N-R} \frac{p_j}{\bar{G}} \left\{ -\bar{F}_j \left(\ln \frac{y_\ell}{\eta_j}\right)\left(\frac{y_\ell}{\eta_j}\right)^{\beta_j} \right\} + \sum_{i=1}^{R_j} \left\{ \frac{1}{\beta_j} + \left(\ln \frac{x_{ij}}{\eta_j}\right)\left[1 - \left(\frac{x_{ij}}{\eta_j}\right)^{\beta_j}\right]\right\} \qquad (j = 1,2,\ldots,K). \qquad (22)$$

If $\frac{\partial L}{\partial \beta_j}$ is set to zero, the following holds for all j:

$$\beta_j = \frac{R_j}{\sum_{i=1}^{R_j} \left(\ln \frac{x_{ij}}{\eta_j}\right)\left[\left(\frac{x_{ij}}{\eta_j}\right)^{\beta_j} - 1\right] + \sum_{\ell=1}^{N-R} \frac{p_j \bar{F}_j}{\bar{G}} \left(\ln \frac{y_\ell}{\eta_j}\right)\left(\frac{y_\ell}{\eta_j}\right)^{\beta_j}} \qquad (23)$$

Similarly,

$$\frac{\partial L}{\partial \eta_j} = \sum_{\ell=1}^{N-R} \frac{1}{\bar{G}} \frac{\partial \bar{G}}{\partial \eta_j} + \sum_{i=1}^{R_j} \frac{1}{f_j} \frac{\partial f_j}{\partial \beta_j} \qquad (j = 1,2,\ldots,K) ,$$

which via Equations (3) and (4) becomes for all j

$$\frac{\partial L}{\partial \eta_j} = \sum_{\ell=1}^{N-R} \frac{p_j \bar{F}_j}{\bar{G}} \left(\frac{y_\ell}{\eta_j}\right)^{\beta_j} \frac{\beta_j}{\eta_j} + \sum_{i=1}^{R_j} \left[\left(\frac{x_{ij}}{\eta_j}\right)^{\beta_j} - 1\right] \frac{\beta_j}{\eta_j} . \qquad (24)$$

When Equation (24) is set to zero we find that

$$\eta_j = \left[\frac{\sum_{\ell=1}^{N-R} \frac{p_j \bar{F}_j}{\bar{G}} y_\ell^{\beta_j} + \sum_{i=1}^{R_j} x_{ij}^{\beta_j}}{R_j}\right]^{1/\beta_j} \qquad (j = 1,2,\ldots,K) . \qquad (25)$$

Finally, the mixing proportion equation is

$$\frac{\partial L}{\partial p_j} = \sum_{\ell=1}^{N-R} \frac{1}{\bar{G}} \frac{\partial \bar{G}}{\partial p_j} + \frac{R_j}{p_j} - \frac{R_K}{p_K} \qquad (j = 1,2,\ldots,K-1) .$$

Upon substitution of Equation (6) and setting the derivative to zero, we get

$$\sum_{\ell=1}^{N-R} \frac{\bar{F}_j}{\bar{G}} + \frac{R_j}{p_j} = \text{constant} = C \qquad (j = 1,2,\ldots,K) .$$

If we multiply both sides by p_j and sum, this becomes

$$\sum_{j=1}^{K} p_j C = C = \sum_{\ell=1}^{N-R} \sum_{p=1}^{K} \frac{p_j F_j}{\bar{G}} + \sum_{j=1}^{K} R_j = N . \qquad (26)$$

Thus, finally,

$$p_j = \frac{R_j}{N} + \frac{1}{N} \sum_{\ell=1}^{N-R} \frac{p_j \bar{F}_j}{\bar{G}} \qquad (j = 1,2,\ldots,K) . \qquad (27)$$

Equations (23), (25) and (27) are the basis for finding the distribution parameters and mixing proportions in the post-mortem case. As before, the left-hand sides represent the values at the (v+1)st iteration, and the right-hand side functions are evaluated with values at the vth iteration. In addition, we will bisect the step size if no improvement in the log-likelihood function is realized.

7.1 Convergence Properties

The convergence proof is patterned after Section 4.1 by showing that

(1) α^v belongs to a compact set;

(2) $\left[\alpha^{v+1} - \alpha^v\right] \cdot \nabla L^v \geq 0$; and

(3) α^v is feasible.

Since all functions in Equations (23), (25) and (27) are continuous, the mapping from α^v to α^{v+1} is closed and α^v thus belongs to a compact set.

For the second property we first show

$$\left[\beta_j^{v+1} - \beta_j^v\right] \cdot \nabla L_\beta^v \geq 0 \qquad (j = 1,2,\ldots,K) \quad . \tag{28}$$

Equation (23) implies for all j that

$$\frac{1}{\beta_j^{v+1}} - \frac{1}{\beta_j^v} = \sum_{\ell=1}^{N-R} \frac{p_j^v \bar{F}_j^v}{\bar{G}^v R_j} \left(\ln \frac{y_\ell}{\eta_j^v}\right) \left(\frac{y_\ell}{\eta_j^v}\right)^{\beta_j^v} - \frac{R_j}{R_j \beta_j}$$

$$- \sum_{i=1}^{R_j} \frac{\left(\ln \frac{x_{ij}}{\eta_j^v}\right)\left[1 - \left(\frac{x_{ij}}{\eta_j^v}\right)^{\beta_j^v}\right]}{R_j} = - \frac{1}{R_j} \frac{\partial L^v}{\partial \beta_j} \qquad [\text{via } (20)] \quad .$$

After some algebra, this becomes

$$\beta_j^{v+1} - \beta_j^v = \frac{\beta_j^v \beta_j^{v+1}}{R_J} \nabla L_\beta^v \qquad (j = 1,2,\ldots,K) \quad .$$

Since the coefficient of ∇L_β^v is positive, Equation (28) holds.

We next need to show that

$$\left[\eta_j^{v+1} - \eta_j^v\right] \cdot \nabla L_\eta^v \geq 0 \qquad (j = 1,2,\ldots,K) \quad . \tag{29}$$

Equation (25) implies for all j that

$$\eta_j^{v+1} - \eta_j^v = \left[\frac{\sum_{\ell=1}^{N-R} \frac{p_j^v \bar{F}_j^v}{\bar{G}^v} y_\ell^{\beta_j^v} + \sum_{i=1}^{R_j} x_{ij}^{\beta_j^v}}{R_j}\right]^{1/\beta_j^v} - \eta_j^v \quad .$$

Since A-B has the same sign as $A^x - B^x$, the right-hand side has the same sign as

$$\frac{\sum_{\ell=1}^{N-R} \frac{p_j^v \bar{F}_j^v}{\bar{G}} y_\ell^{\beta_j^v} + \sum_{i=1}^{R_j} x_{ij}^{\beta_j^v} - \sum_{i=1}^{R_j} \eta_j^{v\,\beta_j^v}}{R_j} \quad ,$$

which [via Equation (24)] equals

$$\frac{\eta_j^v}{R_j} \frac{\eta_j^v}{\beta_j^v} \frac{\partial L^v}{\partial \eta_j^v} \qquad (j = 1,2,\ldots,K) \quad .$$

Since the coefficient of $\left[\partial L/\partial \eta_j\right]^v$ is positive, Equation (29) holds.

In order to demonstrate the second property, it only remains to show that

$$\sum_{j=1}^{K} \left[p_j^{v+1} - p_j^{v}\right] \cdot \nabla L_p^v \geq 0 \quad . \qquad (30)$$

After substituting for ∇L_p^v from Equation (26), the left-hand side of Equation (30) becomes

$$\sum_{j=1}^{K} \left[p_j^{v+1} - p_j^{v}\right] \cdot \left\{ \sum_{\ell=1}^{N-R} \frac{\bar{F}_j^v - F_K^v}{G^v} + \frac{R_j}{p_j^v} - \frac{R_K}{p_K^v} \right\} .$$

Equation (27) may be rewritten as

$$\frac{N\, p_j^{v+1}}{p_j^v} = \frac{R_j}{p_j^v} + \sum_{\ell=1}^{N-R} \frac{\bar{F}_j^v}{\bar{G}^v} \quad .$$

Thus the left-hand side is equal to

$$N \sum_{j=1}^{K} \left[p_j^{v+1} - p_j^{v}\right] \cdot \left\{ \frac{p_j^{v+1}}{p_j^v} - \frac{p_K^{v+1}}{p_K^v} \right\} .$$

This expression is identical to Equation (19), thus the arguments of Section 4.1 apply and consequently the expression is positive and the proof of Equation (30) is complete.

The final step is to show that α^v is feasible. Since the right-hand sides of Equations (23) and (25) are positive and Equation (27) is nonnegative, then $\beta_j\, \eta_j > 0$ and $p_j \geq 0$ for all j. We have only to show that $\sum_{j=1}^{K} p_j = 1$. If both sides of Equation (27) are summed over j, then

$$\sum_{j=1}^{K} p_j^{v+1} = \frac{1}{N}\left[\sum_{j=1}^{K} R_j + \sum_{\ell=1}^{N-R} \sum_{j=1}^{K} \frac{p_j \bar{F}_j}{G}\right] = \frac{1}{N}\left[R + \sum_{\ell=1}^{N-R} 1\right] = 1.$$

Thus convergence is assured.

8. The Exponential Case

If we deal with a mixture of exponentials rather than a mixture of Weibulls, the results are quite similar. The probability density function and complementary CDF for the exponential are $f_j(x) = \frac{1}{\eta_j} \exp(-x/\eta_j)$ and $\bar{F}_j(x) = \exp(-x/\eta_j)$. Since these forms are equivalent to the Weibull when β_j is one, the iterative equations for the exponential case are found by setting β_j to unity in each of the various algorithms. The results are as follows:

First-Order Method -- Non-Post Mortem

$$\left\{ \begin{aligned} \eta_j^{v+1} &= \frac{\sum_{i=1}^{R} (f_j^v / g^v)\, x_i + \sum_{\ell=1}^{N-R} (\bar{F}_j^v / \bar{G}^v)\, y_\ell}{\sum_{i=1}^{R} f_j^v / g^v} & (j = 1,2,\ldots,K) \\ p_j^{v+1} &= \frac{p_j^v}{N}\left[\sum_{i=1}^{R} \frac{f_j^v}{g^v} + \sum_{\ell=1}^{N-R} \frac{\bar{F}_j^v}{\bar{G}^v}\right] & (j = 1,2,\ldots,K) \end{aligned} \right.$$

Second-Order Method -- Non-Post Mortem

$$\left\{ \begin{aligned} \eta_j^{v+1} &= \frac{\sum_{i=1}^{R} (f_j^v / g^v)\, x_i + \sum_{\ell=1}^{N-R} (\bar{F}_j^v / \bar{G})\, y_\ell}{\sum_{i=1}^{R} f_j^v / g^v} & (j = 1,2,\ldots,K) \\ p_j^{v+1} &= p_j^v - (\nabla^2 L_p^v)^{-1} (\nabla L_p^v) & (j = 1,2,\ldots,K) \end{aligned} \right.$$

First-Order Method -- Post Mortem

$$\left\{ \begin{aligned} \eta_j^{v+1} &= \frac{\sum_{\ell=1}^{N-R} \frac{p_j^v \bar{F}_j^v}{\bar{G}^v}\, y_\ell + \sum_{i=1}^{R_j} x_{ij}}{R_j} & (j = 1,2,\ldots,K) \\ p_j^{v+1} &= \frac{R_j}{N} + \frac{1}{N} \sum_{\ell=1}^{N-R} \frac{p_j^v \bar{F}_j^v}{\bar{G}^v} & (j = 1,2,\ldots,K) \end{aligned} \right.$$

9. A Two-Phase Method

In Section 4, we presented a first-order method in the non-post mortem case. Equations (9), (10) and (14) form the basis of this approach. In Section 7, analogous equations were developed in circumstances where a post-mortem was performed. We were able to take advantage of the fact that the log-likelihood function is concave with respect to the mixing proportions in the non-post mortem case in Section 5, and consequently were able to take advantage of second-order convergence by replacing Equation (14) with Equation (20).

If in either of the first-order methods, we are in a neighborhood of a local maximum where concavity is guaranteed, then a Newton step can be made in the vector (α) of all parameters as $\alpha^{v+1} = \alpha^v - s(\nabla^2 L_\alpha^v)^{-1} \nabla L_\alpha^v$, where s is the step size, initially set to unity. Convergence will occur since $\nabla^2 L_\alpha^v$ is assumed to be negative semi-definite. If α^{v+1} is not feasible, the step size will be bisected as was previously the case. Computationally we test the closeness to a solution by the absolute value of the gradient of the log-likelihood being arbitrarily small.

More specifically, when K = 2, we will define the vector as $\alpha = (\beta_1, \beta_2, \eta_1, \eta_2, p_1)$. Thus the (ij)th element of $\nabla^2 L_\alpha^v$ will be the second partial of L with respect to α_i and α_j of the vector α. Lengthy formulas can now be derived to enable us to write all terms which may be encountered in a $\nabla^2 L_\alpha$ matrix. The differentiation is straightforward but messy, so the detailed results are not offered here.

10. Directions for Future Research

Several methods for finding the parameters of the mixture model have been presented. The next step is to test these methods under a wide variety of conditions to determine the most appropriate choice corresponding to the particular circumstances. Once the parameters have been estimated, one would want to make statistical statements about them. Thus another research direction will involve the testing of hypotheses concerning the parameter. Two other issues for consideration are the handling of local solutions and goodness-of-fit testing.

REFERENCES

Carr-Hill, G.A. and R.A. Carr-Hill (1972). "Reconviction as a Process," British Journal of Criminology, Vol. 12, pp. 35-43.

Feller, W. (1966). Theory of Probability, Vol. II, John Wiley and Sons, New York.

Greenberg, David F. (1978). "Recidivism as Radioactive Decay," Journal of Research in Crime and Delinquency, Vol. 15, pp. 124-125.

Hasselblad, Victor (1969). "Estimation of Finite Mixtures of Distributors from the Exponential Family," Journal of the American Statistical Association, Vol. 84, pp. 1459-1471.

Kaylan, Ali (1978). Statistical Analysis of Finite Mixtures of Exponential and Weibull Distributions, Ph.D. Dissertation, Department of Industrial Engineering and Operations Research, Syracuse University.

Luenberger, D.G. (1973). Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Massachusetts.

Maltz, Michael D. and Richard McCleary (1977). "The Mathematics of Behavioral Change, Recidivism and Construct Validity," Evaluation Quarterly, Vol. 1, pp. 421-438.

Mendenhall, W. and R.J. Hader (1958). "Estimation of Parameters of Mixed Exponentially Distributed Failure Time Distributions From Censored Life Test Data," Biometrika, Vol. 45, pp. 504-520.

APPENDIX F

```
      COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
      DIMENSION CL(4),CH(4)
      DATA CL/.99,.98,.95,.90/
      NN=100
      KK=53
      MM=NN-KK
      TT=535
      BB=TT-KK
      NU=24
      ITYPE =1
      CALL MLECAL(GML,QML,SG,SQ,R)
      WRITE(6,*)GML,QML,SG,SQ,R
      CALL BYSCAL(GML,QML,GBA,QBA,SG,SQ,R)
      WRITE(6,*)GBA,QBA,SG,SQ,R
      CALL NTGRTN(GML,QML,CL,CH)
      CALL CCNTOR(GML,QML,GBA,QBA,CL,CH)
      STOP
      END
```

```
PTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
PTIONS IN EFFECT*  NAME = MAIN    , LINECNT =      50
TATISTICS*    SOURCE STATEMENTS =        18,PROGRAM SIZE =      722
STATISTICS*  NO DIAGNOSTICS GENERATED
```

```
      SUBROUTINE MLECAL(GAM,QQ,SG,SQ,RHO)
C***********************************************************************
C*
C*   THIS SUBROUTINE CALCULATESTHE MAXIMUM LIKELIHOOD ESTIMATE OF
C*   GAMMA AND Q AND THEIR (ASYMPTOTIC) COVARIANCE PARAMETERS.  IF
C*   ITYPE=1 THE DATA ARE SINGLY CENSORED; IF ITYPE=2 THE DATA ARE
C*   MULTIPLY CENSORED.
C*
C***********************************************************************
      COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
      ICT=0
      IF(ITYPE.GT.1)GO TO 100
      Z=TT/BB
      RT=BB/KK
   10 T1=1/(Z-1)
      T2=NU/(Z**NU-1)
      TOP=T1-T2-RT
      BOT=T1**2-Z**(NU-1)*T2**2
      DZ=TOP/BOT
      Z=Z+DZ
      ICT=ICT+1
      IF(ICT.GT.30)GO TO 200
      IF(ABS(DZ/Z).GT.1E-6)GO TO 10
      QQ=1/Z
      GAM=KK/(NN*(1-QQ**NU))
      T1=NN*KK*GAM**(-2)/MM
      S1=KK/(1-QQ)-BB/QQ
      T2=-S1*KK/(GAM*MM)
      T3=(S1**2)/MM-S1*(NU-1)/QQ+KK/(1-QQ)**2+BB/QQ**2
      DEL=T1*T3-T2**2
      SG=T3/DEL
      SQ=T1/DEL
      CV=-T2/DEL
      IF(SG.GE.0)SG=SQRT(SG)
      IF(SQ.GE.0)SQ=SQRT(SQ)
      RHO=CV/(SG*SQ)
      RETURN
  100 GAM=(KK+0.0)/NN
      QQ=BB/TT
  110 ICT=ICT+1
      HG=KK/GAM
      HQ=BB/QQ-KK/(1-QQ)
      HGG=-HG/GAM
      HGQ=0
      HQQ=-BB/QQ**2-KK/(1-QQ)**2
      QI=1
      DO 130 I=1,NU
      T0=GAM*I*QI
      QI=QI*QQ
      T1=1-QI
      T2=1-GAM*T1
      T3=T1/T2
      T4=T0/T2
      HG=HG-MI(I)*T3
      HQ=HQ+MI(I)*T4
      HGG=HGG-MI(I)*T3**2
      HGQ=HGQ+MI(I)*T4/(GAM*T2)
  130 HQQ=HQQ-MI(I)*T4*(T4-(I-1)/QQ)
      DET=HGG*HQQ-HGQ**2
      DG=-(HQQ*HG-HGQ*HQ)/DET
      DQ=-(-HGQ*HG+HGG*HQ)/DET
      GAM=GAM+DG
      IF(GAM.GT.1.5)GO TO 200
      QQ=QQ+DQ
      TEST=(DG/GAM)**2+(DQ/QQ)**2
      IF(TEST.LE.1E-10)GO TO 150
      IF(ICT.GT.30)GO TO 200
      GO TO 110
  150 IF(GAM.GT.1.0)GO TO 200
      SG=SQRT(-HQQ/DET)
      SQ=SQRT(-HGG/DET)
      RHO=HGQ/SQRT(HGG*HQQ)
      RETURN
  200 GAM=1
      SG=0
      SQ=0
      RHO=0
      CC=NU*MM
      IF(ITYPE.EQ.1)GO TO 220
      CC=0
      DO 210 I=1,NU
  210 CC=CC+I*MI(I)
  220 QQ=(BB+CC)/(TT+CC)
      RETURN
      END
```

```
*OPTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
 PTIONS IN EFFECT*  NAME = MLECAL   , LINECNT =        50
 TATISTICS*    SOURCE STATEMENTS =          77,PROGRAM SIZE =      2750
 TATISTICS*  NO DIAGNOSTICS GENERATED
```

```
      SUBROUTINE BYSCAL(GML,QML,GAM,QQ,SG,SQ,RHO)
C*****************************************************************
C*
C*    THIS SUBROUTINE CALCULATES THE BAYESIAN ESTIMATES OF GAMMA AN
C*    Q AND THEIR COVARIANCE PARAMETERS, ASSUMING A UNIFORM PRIOR
C*    DISTRIBUTION.  IF THE DATA ARE SINGLY CENSORED ITYPE=1; IF
C*    ARE MULTIPLY CENSORED ITYPE=2.
C*
C*****************************************************************
      COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
      DIMENSION F(3,3),SUM(3,3)
      IF(ITYPE.GT.1)GO TO 100
      DO 10 IR=1,3
      JS=4-IR
      R=IR-1
      DO 10 IS=1,JS
      S=IS-1
      F(IR,IS)=FIRST(R,S)
   10 SUM(IR,IS)=F(IR,IS)
      DO 50 I=1,MM
      DO 20 IR=1,3
      R=IR-1
      JS=4-IR
      DO 20 IS=1,JS
      S=IS-1
      F(IR,IS)=F(IR,IS)*XNEXT(I,R,S)
   20 SUM(IR,IS)=SUM(IR,IS)+F(IR,IS)
      IF(ABS(F(1,1)/SUM(1,1)).LT.1E-6)GO TO 200
   50 CONTINUE
      GO TO 200
  100 TOP=VALLF(GML,QML,1)
      DO 110 I=1,3
      JX=4-I
      DO 110 J=1,JX
  110 SUM(I,J)=0
      GX=1.0
  120 IF(GX.LE.0)GO TO 200
      CALL QLINE(GX,QTOP,SQ)
      HT=VALLF(GX,QTOP,2)
      IF(HT.GT.0)GO TO 130
      IF(GX.LT.GAM)GO TO 200
      GX=GX-0.05
      GO TO 120
  130 DO 140 I=1,3
      JX=4-I
      DO 140 J=1,JX
  140 F(I,J)=HT*GX**(I-1)*QTOP**(J-1)
      DQ=0.05*SQ
      QX=QTOP
      ISW=1
  150 QX=QX+DQ
      IF(QX*(1-QX).LE.0)GO TO(170,180),ISW
      HT=VALLF(GX,QX,2)
      IF(HT.LE.0)GO TO(170,180),ISW
      DO 160 I=1,3
      JX=4-I
      DO 160 J=1,JX
  160 F(I,J)=F(I,J)+HT*GX**(I-1)*QX**(J-1)
      GO TO 150
  170 DQ=-DQ
      QX=QTOP
      ISW=2
      GO TO 150
  180 DO 190 I=1,3
      JX=4-I
      DO 190 J=1,JX
  190 SUM(I,J)=SUM(I,J)+F(I,J)*ABS(DQ)
      GX=GX-0.05
      GO TO 120
  200 GAM=SUM(2,1)/SUM(1,1)
      GSQ=SUM(3,1)/SUM(1,1)
      QQ=SUM(1,2)/SUM(1,1)
      QSQ=SUM(1,3)/SUM(1,1)
      GQB=SUM(2,2)/SUM(1,1)
      SG=GSQ-GAM**2
      IF(SG.GE.0)SG=SQRT(SG)
      SQ=QSQ-QQ**2
      IF(SQ.GE.0)SQ=SQRT(SQ)
      RHO=(GQB-GAM*QQ)/(SG*SQ)
      RETURN
      END
```

```
*OPTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
 PTIONS IN EFFECT*  NAME = BYSCAL   , LINECNT =       50
 TATISTICS*     SOURCE STATEMENTS =        73,PROGRAM SIZE =      2382
 TATISTICS*  NO DIAGNOSTICS GENERATED
```

```
      FUNCTION FIRST(R,S)
C*********************************************************************
C
C  THIS FUNCTION CALCULATES THE FIRST VALUE OF THE BETA FUNCTION
C  WITH INDEXES R AND S, FOR THE SUBROUTINE BYSCAL.
C
C*********************************************************************
      COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
      FIRST=1
      IF(R.EQ.0)GO TO 20
      KR=R
      DO 10 I=1,KR
   10 FIRST=FIRST*(XK+I)/(NN+I+1.0)
   20 IF(S.EQ.0)RETURN
      KS=S
      DO 30 J=1,KS
   30 FIRST=FIRST*(BB+J)/(BB+XK+J+1)
      RETURN
      END

*OPTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
*OPTIONS IN EFFECT*  NAME = FIRST   , LINECNT =      50
*STATISTICS*    SOURCE STATEMENTS =      13,PROGRAM SIZE =     696
*STATISTICS*  NO DIAGNOSTICS GENERATED
```



```
      FUNCTION XNEXT(I,R,S)
C*********************************************************************
C
C  THIS FUNCTION CALULATES THE NEXT VALUE IN THE SERIES, FOR
C  THE SUBROUTINE BYSCAL.
C
C*********************************************************************
      COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
      ZI=I
      XNEXT=(XK+R+ZI)/ZI
      TRM=BB+XK+S+(I-1)*NU
      DO 10 J=1,NU
   10 XNEXT=XNEXT*(TRM-XK+J)/(TRM+1.0+J)
      RETURN
      END

PTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
PTIONS IN EFFECT*  NAME = XNEXT   , LINECNT =      50
TATISTICS*    SOURCE STATEMENTS =       9,PROGRAM SIZE =     590
*STATISTICS*  NO DIAGNOSTICS GENERATED
```

FORTRAN IV G1 RELEASE 2.0 VALLF DATE = 81030 09/4

```
      FUNCTION VALLF(G,Q,IX)
C*****************************************************************
C
C*   THIS FUNCTION CALCULATES THE VALUE OF THE LIKELIHOOD FUNCTIC
C*   AT THE POINT G,Q.  IF IX=1 THEN G AND Q ARE THE VALUES OF TH
C*   L.F. AT THE MAXIMUM AND THE LOG L.F. IS RETURNED.  IF IX=2 T
C*   FUNCTION RETURNS THE VALUE OF THE LF RELATIVE TO THE MAXIMUM
C
C*****************************************************************
      COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
      X=KK*ALOG(G*(1-Q))+BB*ALOG(Q)
      IF(ITYPE.GT.1)GO TO 10
      X=X+MM*ALOG(1-G+G*Q**NU)
      GO TO (30,40),IX
10    QI=1
      DO 20 I=1,NU
      QI=Q*QI
20    X=X+MI(I)*ALOG(1-G+G*QI)
      IF(IX.GT.1)GO TO 40
30    VALLF=X
      HMX=X
      RETURN
40    VALLF=0
      E=X-HMX
      IF(E.LT.-40)RETURN
      VALLF=EXP(E)
      RETURN
      END
```

```
*OPTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
*OPTIONS IN EFFECT*  NAME = VALLF    , LINECNT =       50
*STATISTICS*     SOURCE STATEMENTS =        20,PROGRAM SIZE =       944
*STATISTICS*  NO DIAGNOSTICS GENERATED
```

FORTRAN IV G1 RELEASE 2.0 QLINE DATE = 81030 09/46

```
      SUBROUTINE QLINE(GAM,Q,SQ)
C*****************************************************************
C
C*   THIS SUBROUTINE FINDS THE VALUE OF Q THAT MAXIMIZES THE L.F.
C*   ALONG THE LINE GAM=CONSTANT.  IT ALSO CALCULATES THE (ASYMP-
C*   TOTIC) STANDARD DEVIATION OF Q AT THAT POINT.
C
C*****************************************************************
      COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
      IF(GAM.LT.0.999)GO TO 40
      IF(ITYPE.GT.1)GO TO 10
      C=BB+MM*NU
      GO TO 30
10    C=BB
      DO 20 I=1,NU
20    C=C+MI(I)*I
30    Q=C/(C+KK)
      T=C*KK/(C+KK)**3
      GO TO 100
40    ICT=0
50    ICT=ICT+1
      IF(Q*(1-Q).LE.0)Q=0.99
      TOP=BB/Q-KK/(1-Q)
      BOT=BB/Q**2+KK/(1-Q)**2
      IF(ITYPE.GT.1)GO TO 60
      T1=1-GAM+GAM*Q**NU
      T2=GAM*NU*Q**(NU-1)/T1
      TOP=TOP+MM*T2
      BOT=BOT+MM*T2*(T2-(NU-1)/Q)
      GO TO 80
60    QI=1
      DO 70 I=1,NU
      QX=QI
      QI=QI*Q
      T1=1-GAM+GAM*QI
      T2=GAM*I*QX/T1
      TOP=TOP+MI(I)*T2
70    BOT=BOT+MI(I)*T2*(T2-(I-1)/Q)
80    DQ=TOP/BOT
      Q=Q+DQ
      IF(ABS(DQ/Q).LT.1E-5)GO TO 90
      IF(ICT.LT.30)GO TO 50
      Q=0.0
      SQ=-1
      RETURN
90    T=1/BOT
100   SQ=SQRT(T)
      RETURN
      END
```

```
*OPTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
*OPTIONS IN EFFECT*  NAME = QLINE    , LINECNT =       50
*STATISTICS*     SOURCE STATEMENTS =        42,PROGRAM SIZE =      1638
*STATISTICS*  NO DIAGNOSTICS GENERATED
```

```
      SUBROUTINE NTGRTN(GMX,QMX,CL,CH)
C*****************************************************************
C
C   THIS SUBROUTINE CALCULATES THE VOLUME OF THE LIKELIHOOD
C   FUNCTION WITHIN A SET OF HEIGHTS.  WHEN HT=0 THE VOLUME
C   CALCULATED IS THE ENTIRE VOLUME.  THE FRACTION OF VOLUME
C   WITHIN EACH CONTOUR CAN THEN BE CALCULATED FOR THE NINE
C   REMAINING HEIGHTS.  THESE HEIGHTS ARE THEN INTERPOLATED TO
C   APPROXIMATE THE HEIGHT OF THE CONTOUR (CH) THAT WILL PRODUCE
C   APPROPRIATE CONFIDENCE LEVELS (CL).  AT PRESENT FOUR CONFIDENC
C   LEVELS ARE CALCULATED -- .99, .98, .95, AND .90.
C
C*****************************************************************
      COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
      DIMENSION HT(10),SLICE(10),VOL(10),CL(4),CH(4)
      DATA HT/0.,.005,.01,.015,.02,.03,.05,.1,.15,.2/
C
C   CALCULATE THE HEIGHT OF THE LIKELIHOOD FUNCTION AT ITS MAXIMUM
C
      HX=VALLF(GMX,QMX,1)
      DO 10 I=1,10
   10 VOL(I)=0
      GAM=1.05
C
C   THE L.F. IS DIVIDED INTO 100 "SLICES" FOR NUMERICAL INTEGRATIO
C
   20 GAM=GAM-0.05
      IF(GAM.LE.0)GO TO 100
C
C   FIND QTOP THAT MAXIMIZES L.F. FOR THAT GAMMA.
C
      CALL QLINE(GAM,QTOP,SQ)
C
C   CALCULATE RELATIVE HEIGHT OF L.F. AT THAT POINT.
C
      HREL=VALLF(GAM,QTOP,2)
      IF(HREL.GT.0)GO TO 30
      IF(GAM.LT.GMX)GO TO 100
      GO TO 20
   30 DO 40 I=1,10
      SLICE(I)=0
      IF(HREL.GE.HT(I))SLICE(I)=HREL
   40 CONTINUE
C
C   FIRST INTEGRATE SLICE TO RIGHT OF SLICE MAXIMUM, FOR EACH HEIG
C
      DQ=.02*SQ
      Q=QTOP
      ISW=1
   50 Q=Q+DQ
      IF(Q*(1-Q).LE.0)GO TO (70,80),ISW
      HREL=VALLF(GAM,Q,2)
      IF(HREL.LE.0)GO TO(70,80),ISW
      DO 60 I=1,10
      IF(HREL.LT.HT(I))GO TO 50
   60 SLICE(I)=SLICE(I)+HREL
      GO TO 50
C
C   NEXT INTEGRATE SLICE TO LEFT OF SLICE MAXIMUM.
C
   70 DQ=-DQ
      Q=QTOP
      ISW=2
      GO TO 50
   80 DO 90 I=1,10
   90 VOL(I)=VOL(I)+SLICE(I)*ABS(DQ)
      GO TO 20
  100 DO 110 I=2,10
      VOL(I)=VOL(I)/VOL(1)
  110 WRITE(6,120)I,HT(I),VOL(I)
  120 FORMAT(I5,F6.3,F10.6)
      VOL(1)=1
      DO 200 IC=1,4
      DO 150 I=1,10
      IF(VOL(I).LE.CL(IC))GO TO 160
  150 CONTINUE
      STOP 2
C
C   THIS NEXT SECTION INTERPOLATES TO OBTAIN THE APPROPRIATE
C   CONTOUR HEIGHTS (CH) FOR THE SPECIFIED CONFIDENCE LEVELS (CL).
C
  160 J=I-1
      SL=(HT(I)-HT(J))/(VOL(I)-VOL(J))
  200 CH(IC)=HT(J)+SL*(CL(IC)-VOL(J))
      WRITE(6,210)CL,CH
  210 FORMAT(/4F10.3/3X,4F10.6)
      RETURN
      END
```

```
 PTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
*OPTIONS IN EFFECT*  NAME = NTGRTN   , LINECNT =       50
*STATISTICS*    SOURCE STATEMENTS =        54,PROGRAM SIZE =      1672
 TATISTICS*  NO DIAGNOSTICS GENERATED
```

```
      SUBROUTINE CONTOR(GAM,QQ,GBA,QBA,CL,CH)
      COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
      DIMENSION CL(4),CH(4),IC(4),JC(4),GF(4),GL(4)
      DIMENSION GC(203,4),QC(203,4),GP(203),QP(203)
C
C    INITIALIZE
C
      CALL BGNPLT(7,'CONTOUR')
      HX=VALLF(GAM,QQ,1)
      GX=1.0
      DG=.01
      DO 10 KL=1,4
      IC(KL)=2
      JC(KL)=204
10    GF(KL)=0
C
C    FIND CONTOUR POINTS ON THE RIGHT SIDE.
C
20    CALL QLINE(GX,QT,SQ)
      IF(GX.EQ.1.0)QSTRT=AMIN1(0.9999,QT+5.0*SQ)
      RX=VALLF(GX,QT,2)
C
C    IS THE HIGHEST POINT ON THE LINE GX=CONST ABOVE THE LOWEST
C    CONTOUR LINE?
C
      IF(RX.GT.CH(1))GO TO 30
C
C    IF NOT, AND IF GX IS BELOW THE PEAK, FIND CONTOUR POINTS
C    ON THE LEFT SIDE.
C
      IF(GX.LT.GAM)GO TO 100
      GO TO 70
30    DQ=0.05*SQ
      KL=1
      QL=QSTRT
      QX=QL
C
C    FIND A POINT LOWER THAN CONTOUR.
C
40    RL=VALLF(GX,QL,2)
      IF(RL.LT.CH(KL))GO TO 50
      QL=QL+DQ
      QX=QL
      IF(QL.LT.1.0)GO TO 40
      GO TO 70
C
C    BRACKET CONTOUR HEIGHT BETWEEN QL AND QX.
C
50    QX=QX-DQ
      IF(QX.LT.QT)GO TO 70
      RX=VALLF(GX,QX,2)
      IF(RX.GT.CH(KL))GO TO 60
      RL=RX
      QL=QX
      GO TO 50
C
C    INTERPOLATE TO FIND CORRECT Q VALUE.
C
60    II=IC(KL)
      QC(II,KL)=QX+(QL-QX)*(RX-CH(KL))/(RX-RL)
      GC(II,KL)=GX
      IF(KL.EQ.1)QSTRT=QL
      IF(GF(KL).EQ.0)GF(KL)=GX
      GL(KL)=GX
      IC(KL)=IC(KL)+1
      KL=KL+1
      QX=QL
      IF(KL.LE.4)GO TO 50
C
C    GO TO NEXT GX LINE
C
70    GX=GX-DG
      IF(GX.GT.0)GO TO 20
C
C    SIMILAR PROCEDURE TO FIND CONTOUR POINTS ON LEFT SIDE.
C
100   GX=GF(1)
110   CALL QLINE(GX,QT,SQ)
      IF(GX.EQ.GF(1))QSTRT=AMAX1(0.01,QT-5.0*SQ)
      RX=VALLF(GX,QT,2)
      IF(RX.GT.CH(1))GO TO 120
      IF(GX.LT.GAM)GO TO 200
      GO TO 160
120   DQ=0.05*SQ
      KL=1
      QL=QSTRT
130   RL=VALLF(GX,QL,2)
      IF(RL.LT.CH(KL))GO TO 140
      QL=QL-DQ
      QX=QL
      IF(QL.GT.0)GO TO 130
      STOP 1
140   QX=QX+DQ
      IF(QX.GT.QT)GO TO 160
      RX=VALLF(GX,QX,2)
      IF(RX.GT.CH(KL))GO TO 150
      RL=RX
      QL=QX
      GO TO 140
150   JC(KL)=JC(KL)-1
      JJ=JC(KL)
      QC(JJ,KL)=QX+DQ*(RX-CH(KL))/(RX-RL)
      GC(JJ,KL)=GX
      IF(KL.EQ.1)QSTRT=QL
      KL=KL+1
      QX=QL
      IF(KL.LE.4)GO TO 140
160   GX=GX-DG
      IF(GX.GE.GL(1))GO TO 110
200   DO 300 KL=1,4
C
C    CALCULATE THE CONTOUR'S FIRST POINT
C
      QC(1,KL)=QC(2,KL)
      GC(1,KL)=GC(2,KL)
      GF1=GF(KL)+DG
      IF(GF1.GT.1)GO TO 220
      CALL QLINE(GF(KL),Q1,SQ)
      IF(Q1.GE.1)GO TO 220
      R1=VALLF(GF(KL),Q1,2)
```

```
      CALL QLINE(GF1,C2,SQ)
      IF(Q2.LT.1)GO TO 210
      Q2=Q1
210   R2=VALLF(GF1,Q2,2)
      FAC=(R1-CH(KL))/(R1-R2)
      QC(1,KL)=Q1+(Q2-Q1)*FAC
      GC(1,KL)=GF(KL)+DG*FAC
C
C   CALCULATE THE CCNTOUR'S MIDPOINT
C
220   II=IC(KL)
      WRITE(6,*)KL,QC(1,KL),GC(1,KL)
      KC=II-1
      QC(II,KL)=QC(KC,KL)
      GC(II,KL)=GC(KC,KL)
      GL1=GL(KL)-DG
      IF(GL1.LT.0)GO TO 230
      CALL QLINE(GL(KL),Q1,SQ)
      IF(Q1.GE.1)GO TO 230
      R1=VALLF(GL(KL),Q1,2)
      CALL QLINE(GL1,C2,SQ)
      IF(Q2.GE.1)Q2=Q1
      R2=VALLF(GL1,Q2,2)
      FAC=(R1-CH(KL))/(R1-R2)
      QC(II,KL)=Q1+(Q2-Q1)*FAC
      GC(II,KL)=GL(KL)-DG*FAC
C
C  PACK THE TWO CONTOUR HALVES TOGETHER
C
230   IL=2*II-2
      WRITE(6,*)II,KL,QC(II,KL),GC(II,KL)
      II=II+1
      JEX=207-2*II
      DO 240 I=II,IL
      JJ=I+JEX
      QC(I,KL)=QC(JJ,KL)
240   GC(I,KL)=GC(JJ,KL)
C
C  COMPLETE THE CONTOUR
C
      IL=IL+1
      QC(IL,KL)=QC(1,KL)
      GC(IL,KL)=GC(1,KL)
      I1=IL+1
      QC(I1,KL)=0
      GC(I1,KL)=0
      I2=I1+1
      QC(I2,KL)=0.125
      GC(I2,KL)=0.125
      DO 250 I=1,I2
      QP(I)=QC(I,KL)
250   GP(I)=GC(I,KL)
      CALL LINE(QP,GP,IL,1,0,1)
      WRITE(6,260)(I,GP(I),QP(I),I=1,I2)
260   FORMAT((I5,2F10.4)/)
300   CONTINUE
      CALL SYMBCL((QQ-QP(I1))/QP(I2),(GAM-GP(I1))/GP(I2),.07,1,0..
      CALL SYMBOL((QBA-QP(I1))/QP(I2),(GBA-GP(I1))/GP(I2),.07,4,C.
      CALL PLOT(8.,8.,3)
      DO 320 I=1,5
      Y=(6-I)*1.6
      CALL PLOT(0.,Y,2)
      CALL PLOT(0.,Y-.8,3)
      CALL PLOT(8.,Y-.8,2)
320   CALL PLOT(8.,Y-1.6,3)
      CALL PLOT(0.,0.,2)
      DO 330 I=1,5
      X=1.6*I
      CALL PLOT(X-1.6,8.,2)
      CALL PLOT(X-.8,8.,3)
      CALL PLOT(X-.8,0.,2)
330   CALL PLOT(X,0.,3)
      CALL PLOT(X,8.,2)
      CALL SYMBOL(1.0,6.5,.14,'N =',0.,3)
      XN=NN
      CALL NUMBER(1.5,6.5,.14,XN,0.,-1)
      CALL SYMBOL(1.0,5.7,.14,'NU =',0.,4)
      UN=NU
      CALL NUMBER(1.64,5.7,.14,UN,0.,-1)
      CALL SYMBOL(1.0,4.9,.14,'GAMMA =',0.,7)
      CALL NUMBER(2.06,4.9,.14,GAM,0.,2)
      CALL SYMBOL(1.0,4.1,.14,'Q =',0.,3)
      CALL NUMBER(1.5,4.1,.14,QQ,0.,2)
      DO 340 I=1,6
      TIC=0.2*(I-1)
340   CALL NUMBER(8.*TIC-.14,-.2,.14,TIC,0.,1)
      DO 350 I=1,6
      TIC=0.2*(I-1)
350   CALL NUMBER(-.1,8.*TIC-.14,.14,TIC,90.,1)
      CALL SYMBOL(-.1,3.7,.14,'GAMMA',90.,5)
      CALL SYMBOL(3.93,-.2,.14,'Q',0.,1)
      CALL ENDPLT
      RETURN
      END

 OPTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,N LIST,NODECK,LOAD,NOMAP,NOTEST
*OPTIONS IN EFFECT*  NAME = CONTOR   , LINECNT          50
 TATISTICS*    SOURCE STATEMENTS =          173,  OGRAM SIZE =     13508
 TATISTICS*  NO DIAGNOSTICS GENERATED
```

```
IRTRAN IV G1   RELEASE 2.0            COHSIM               DATE = 81030         09/4

              SUBROUTINE COHSIM(GAM,QQ,CENS)
         C*****************************************************************
         C*
         C*    THIS SUBROUTINE GENERATES THE FAILURE AND EXPOSURE TIMES OF
         C*    COHORT OF SIZE NN.  THE PROBABILITY OF AN INDIVIDUAL IN THE
         C*    COHORT EVENTUALLY FAILING IS GAM, AND THE PROBABILITY THAT AN
         C*    EVENTUAL FAILURE FAILS IN ANY MONTH IS QQ.  NOT ALL MEMBERS O
         C*    THE COHORT HAVE THE MAXIMUM EXPOSURE TIME NU; A FRACTION CEN
         C*    OF THEM HAVE CENSORED EXPOSURE TIMES, UNIFORMLY DISTRIBUTED
         C*    FROM 1 TO NU MONTHS.
         C*
         C*****************************************************************
              COMMON /BLK/NN,KK,MM,TT,BB,NU,KI(100),MI(100),ITYPE
              DATA I1/43215/,I2/89753/,I3/58742/,I4/57463/
         C
         C   INITIALIZE
         C
              MM=0
              DO 10 I=1,NU
              KI(I)=0
         10   MI(I)=0
              QLOG=ALOG(QQ)
              ITYPE=1
              IF(CENS.GT.0)ITYPE=2
              DO 40 IND=1,NN
         C
         C   MX IS THE MAXIMUM EXPOSURE TIME
         C
              MX=NU
              IF(ITYPE.EQ.1)GO TO 20
         C
         C   IS THIS PERSON SUBJECT TO CENSORING?  IF NOT,GO TO 20.
         C
              IF(URAN1(I1).GT.CENS)GO TO 20
         C
         C   IF CENSORED, CALCULATE HIS MAXIMUM EXPOSURE TIME.
         C
              MX=URAN1(I2)*NU+1.0
         C
         C   WILL THIS PERSON EVENTUALLY FAIL?  IF NOT, GO TO 30.
         C
         20   IF(URAN1(I3).GT.GAM)GO TO 30
         C
         C   IN WHICH MONTH WILL THIS EVENTUAL FAILURE FAIL?  IF BEYOND MX,
         C   WE DO NOT SEE HIM FAIL.
         C
              MF=ALOG(URAN1(I4))/QLOG+1.0
              IF(MF.GT.MX)GO TO 30
              KI(MF)=KI(MF)+1
              GO TO 40
         30   MI(MX)=MI(MX)+1
         40   CONTINUE
              IF(ITYPE.EQ.1)MM=MI(NU)
         C
         C   CALCULATE COHORT STATISTICS.
         C
              TT=0
              KK=0
              DO 50 I=1,NU
              KK=KK+KI(I)
         50   TT=TT+I*KI(I)
              BB=TT-KK
              WRITE(6,*)NN,KK,MM,NU,TT
              WRITE(6,*)(KI(I),I=1,NU)
              IF(ITYPE.GT.1)WRITE(6,*)(MI(I),I=1,NU)
              RETURN
              END

PTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
PTIONS IN EFFECT*  NAME = COHSIM   , LINECNT =       50
*STATISTICS*    SOURCE STATEMENTS =        34,PROGRAM SIZE =      1356
*STATISTICS*  NO DIAGNOSTICS GENERATED

TATISTICS*  NO DIAGNOSTICS THIS STEP
```

APPENDIX G

COMPARING MODELS OF SOCIAL PROCESSES

One of the more neglected aspects of the study of social processes is the method used to select appropriate models of the processes (Dhrymes, 1972; GAO, 1979; Gass & Thompson, 1980). Although rules of thumb are often employed, model selection is still more of an art than a codified set of procedures.

For example, model selection is one of the principal issues in the controversy surrounding the putative deterrent effect of the death penalty on homicides. Different authors have developed different models and, not surprisingly, have come to different conclusions about the ability of executions to deter homicide (Baldus & Cole, 1975; Bowers & Pierce, 1975; Ehrlich, 1975; Forst, 1975; Passell, 1975; see also Brier & Fienberg, 1980, and Barnett, forthcoming).

Chi-square goodness-of-fit tests (e.g., Mann et al, 1974: 350) are often used to determine which one, of some set of alternative models, best fits the empirical data. However, Harris et al (1981), in comparing different models of recidivism using the chi-square statistic, show how this test can be misleading. In particular, they point out that finding the model which best fits the known data does not necessarily provide a model that will be the best one for extrapolating or forecasting beyond the data.

Selection of a model, then, should not necessarily be based on a statistical best-fit criterion. In this paper we argue that additional considerations, based on theoretical constructs and ease of interpretation of the results, as well as forecasting ability, should be used in determining which model to employ.

Our argument will be demonstrated by the comparison of two almost identical models of criminal recidivism,[1] one described by Maltz and

McCleary (1977), the other proposed by Bloom (1979), referred to hereinafter as M1 and M2, respectively.

The first section discusses the nature of the process under study, presents the assumptions inherent in the two models, and describes how the statistical properties of the process suggest a specific form of the model. The second section discusses the interpretation of the two models' parameters and associated confidence statements. Section 3 describes variants of M1 which are useful for different purposes. In Section 4 the models are applied to different data sets to compare their forecasting ability. In Section 5 another approach to modeling recidivism is described.

1. Comparing the Two Models' Assumptions

Both models represent recidivism as a binary event -- a person either fails or does not fail. Both models also assume that an individual's failure event can occur at a random time, and so both are couched in the language of probability and reliability theory (e.g., Barlow and Proschan, 1975; Mann et al, 1974). In particular, the cumulative distribution of failures for M1 is:

$$P_1(t) = \gamma[1 - \exp(-\phi t)] \qquad (1)$$

where $P_1(t)$ is the probability that the failure time for a randomly selected releasee is less than t. The subscript 1 refers to model 1 and γ and ϕ are the model's parameters, with $0 \leq \gamma \leq 1$ and $\phi > 0$. Note that $P\ (t)$ is an incomplete or defective distribution in that $P\ (\infty) = \gamma < 1$.

The cumulative distribution of failures for model M2 is given by

$$P_2(t) = 1 - \exp\{-(b/c)[1 - \exp(-ct)]\} \qquad (2)$$

which is also an incomplete distribution.

That these two models can produce quite similar results can be seen in Figure 1. The cumulative number of failures estimated by M1 and M2 (using maximum likelihood estimates of their respective parameters) is plotted along with the actual number of failures, for data analyzed in Maltz and McCleary (1977) and Bloom (1979).

Equations (1) and (2) each completely specify a stochastic failure process for an individual. Furthermore, both are decreasing failure rate (DFR) models[2] (Barlow & Proschan, 1975: 55), with respective failure rates:

$$h_1(t) = \gamma\phi \exp(-\phi t)/\{1 - \gamma[1 - \exp(-\phi t)]\} \qquad (3)$$

and

$$h_2(t) = b \exp(-ct) \qquad (4)$$

where $h_1(t)dt$ is the conditional probability of failure in the time interval (t,t+dt), for an individual who has not failed up to time t.[3]

A. A Rationale for M2

Bloom's critique of M1 is based on his assumption that "the longer releasees avoid criminal behavior the less likely they are to commit future crimes" (Bloom & Singer, 1979: 615). This assumption, in turn, led to his imposing a number of conditions on the failure rate (Bloom, 1979: 184):

1. $h(0) \geq 0$
2. $h'(t) \leq 0$
3. $h''(t) \geq 0$
4. $\lim_{t \to \infty} h(t) = 0$

Conditions 1 and 4 are necessary for physical realizability (actually, condition 1 should be $h(t) \geq 0$ for $t \geq 0$). Condition 2 (decreasing failure rate) is based on Bloom's assumption about the behavior of released offenders. Condition 3 (convex failure rate) is appealing on empirical grounds, but is not otherwise justified. Taken together, however, the four conditions are not sufficient to specify a <u>particular</u> functional form from the infinity of functions which exhibit these characteristics. Bloom's choice of a function that has the exponentially decreasing failure rate given by Equation 4 (Partanen, 1969, also suggested this as a possibility) is thus reasonable but is nevertheless arbitrary.

B. <u>A Rationale for M1</u>

In contrast, M1 (which also satisfies assumptions 1-4) is supported by structural considerations of the recidivism process under study.[4] Two features of this process stand out in particular: the fact that not all people fail; and the random nature of the failure event. These features dictate characteristics of the model in the following way:

(1) Not all people should be expected to fail: whether rehabilitation is due to a rehabilitative program or not, at least <u>some</u> of the program's participants should be expected not to recidivate. There is strong empirical evidence to support this contention; see Kitchener et al (1977), Hoffman and Stone-Meierhoefer (1979), and Philpotts and Lancucki (1979) for recent long-term (six to eighteen years) studies which point in this direction.

Belief in rehabilitation aside, there are other reasons for expecting that not everyone will fail. A person may recidivate but do so in another state, in which case the event may not be reported to the state analyzing the recidivism statistics. Or plea bargaining may convert a felony to a misdemeanor which may not be considered a failure event. Or the offender may be placed in a diversion program, ensuring that the failure event is unrecorded. Or he may be granted immunity in exchange for testimony, and again the event may not be recorded. In short, there are many reasons to expect that not all releasees will ultimately register a failure event.

Both models take this into consideration and produce $P(\infty)$, the probability that an individual will eventually fail: for M1, this is the parameter γ; for M2, it is $1-\exp(-b/c)$.

(2) When considering those who do fail, it is important to examine more closely what constitutes a failure event. It is clearly not the commission of a crime (or the violation of conditions of parole). Indeed, we do not have information on all crimes or parole violations that were committed. However, we would expect to have information (e.g., arrests or recorded parole violations) on some small fraction of the events: those that come to the attention of the authorities. This recorded information thus becomes the failure event data used for evaluative purposes.

We now make the following assertion: The time to the first failure event of an individual, <u>given</u> that he will eventually fail, is distributed exponentially.

Consider first a single individual who commits crimes according to some (unknown) point process.[5] One can realistically assume that most of these crimes do not lead to detection and subsequent arrest. Thus, the <u>crime process</u> is said to have been "thinned" to produce the <u>arrest</u> process. It can be shown (Haight, 1967: 22) that (a) if the thinning process is independent of the original process, and (b) if most of the points are thinned out (i.e., not detected), then whatever the statistics of the crime process are, the thinned points tend to occur as a Poisson process; i.e., the interoccurrence times tend to be distributed exponentially.

These two conditions are quite reasonable for the thinning of the crime process by arrests. The probability of an individual being arrested for a post-release crime is quite low, so most offenses <u>are</u>

thinned out.[6] Furthermore, the arrest process (a function of police activity) <u>is</u> relatively independent of the crime process (a function of offender behavior). Therefore, since the thinned process tends toward a Poisson process, the time to <u>first</u> arrest is distributed exponentially.

The two models' rationales differ most markedly in this aspect, the characterization of the recidivism process. M2 considers only the rehabilitative effect of a correctional program, an effect which appears to be more elusive (Sechrest et al, 1979) or illusory (Maltz & Pollock, 1980) than real. On the other hand, model M1 derives its form from consideration of the characteristics of the process under study, which includes the effect of the criminal justice process on the data. This gives rise to an incomplete distribution (not all people fail) and an exponential distribution of times to failure for those who do fail.

2. <u>Interpreting the Models</u>

The two parameters of M1 are γ, the probability that an individual will fail, and ϕ, the failure rate of those who do fail. Both parameters are amenable to straightforward interpretation -- γ is a central tendency estimate for the expected fraction of people in a group that eventually will fail, and ϕ describes how fast they are likely to fail. The same is not true of parameters b and c of M2. According to Bloom (1979: 186), "the substantive interpretations of b and c are not useful to policy makers."

Model M1 provides an additional advantage. Statistical confidence intervals for γ and ϕ are of course useful in their own right; i.e., knowing that γ, the probability of failure, has a 95 percent confidence interval of (.4, .6) needs no further interpretation. Furthermore, it is relatively easy to use these parameters' confidence intervals to produce confidence statements about P (t) for any given time t. In contrast, confidence limits on b and c of M2 cannot easily be translated into confidence limits on measures of policy significance.

A further advantage of M1 is that when the data are singly censored (that is, all releasees are observed for the same maximum length of time) they may be summarized by the four sufficient statistics: N, the number in the group under study; τ, the maximum observation time; K, the total number of failures by time τ; and T, the total time between release and failure for the K failures. No similar small set of sufficient statistics exists for M2.

3. <u>Variants of M1</u>

One of the more persuasive reasons for using a structurally derived model is that extensions, if necessary, follow in a straightforward way. In this section we describe three such variants of M1: a geometric model (M1a), a mixed exponential model (M1b), and a "critical time" model (M1c).

a. <u>Geometric model</u>. In most cases recidivism data obtained for evaluative purposes are grouped by months. That is, one rarely is given the number of days to failure for each individual, but rather the number of individuals who failed within each month. To represent this discrete-time behavior, we can define p to be the probability that an individual fails <u>within</u> any month, given that he will eventually fail.[7] Then q = 1 - p is the probability that an individual who will eventually fail does not do so in any month. Again, γ is the probability that an individual will fail. In this case, the cumulative distribution of failures becomes

$$P_{1a}(i) = \gamma(1-q\ ) \qquad i = 1, 2, ... \qquad (5)$$

where $P_{1a}(i)$ is the probability of failing at or before month i.

The parameters of this model are conveniently obtained by maximum likelihood techniques (Maltz, 1981), since the likelihood function is defined only in the unit square $0 \leq \gamma \leq 1$, $0 < q < 1$. An analogous discrete version of M2, on the other hand, does not lend itself to a convenient estimation of parameters. This is due to the fact that the a priori specification of the discrete hazard function for M2 is of a geometrically decreasing form, which does not result in mathematical simplifications (as in Equation 5).

b. A Mixed Exponential Model. A different extension of the assumptions leading to model M1 leads to a mixed exponential model M1b (Harris et al, 1980). M1 is predicated on the assumption that each individual will undergo failure (at rate ϕ) with some probability γ. Consequently, each individual has probability $1-\gamma$ of not failing at all (or, equivalently, of failing at a zero failure rate).

However, it is also possible to consider that with probability $1-\gamma$ an individual is still subject to failure, not with a zero rate but with a lower rate than ϕ. In other words, for some individuals the failure rate may be nonzero but small.[8] This leads to the following expression for the probability of failure by time t:

$$P_{1b}(t) = \gamma[1-\exp(-\phi_1 t)] + (1-\gamma)[1-\exp(-\phi_2 t)] \qquad (6)$$

with associated failure rate

$$h_{1b}(t) = [\gamma\phi_1 \exp(-\phi_1 t) - \gamma\phi_2 \exp(-\phi_2 t)]/[1-P_{1b}(t)] \qquad (7)$$

The parameters ϕ_1 and ϕ_2 again have a "natural" interpretation: ϕ_1 is a primary failure rate (for "failures") and ϕ_2 is a residual "ambient" or "background" non-zero failure rate for the "non-failures". Note that setting $\phi_1 = \phi$ and $\phi_2 = 0$ will reduce (6) and (7) to (1) and (3), respectively.

c. A "Critical Time" Model. It is also possible to develop an alternative model of the observed behavior of released offenders by use of a presumed underlying structure of the recidivism process quite different than that used above. As is shown below, however, this new model structure readily reduces to Equation 6. Thus, it leads to observations operationally indistinguishable from model M1, although it affords additional insights.

This model, M1c, can be called a "critical time" model. It is assumed that each individual in a cohort is subject to random failure (i.e., exponential time to first failure) with failure rate λ_1, until some critical time θ. After this time the individual's failure rate drops to λ_2. The rate λ_1 again may be interpreted to be the failure rate of the general population: an "ambient" failure rate. In other words, if the individual can survive to time θ (during which he has a failure rate λ_1) without failing he is then in some sense "rehabilitated", and is thereafter subject to failure only with a rate λ_2 experienced by the general population.

Given a specific value of θ, the probability distribution of failure time t for this model can be shown to be

$$P_{1c}(t/\theta) = \begin{cases} 1 - \exp(-\lambda_1 t), & 0 \leq t < \theta \\ 1 - \exp[-(\lambda_1 - \lambda_2)\theta - \lambda_2 t], & \theta \leq t < \infty \end{cases} \qquad (8)$$

However, θ may not be known with certainty for all individuals. Rather, we can consider θ to be a random variable with some probability density function $g(\theta)$. The unconditional probability distribution for the time to failure t becomes, then,

$$P_{1c}(t) = \int g(\theta)\, P_{1c}(t/\theta)\, d\theta \qquad (9)$$

Moreover, if θ has an exponential distribution, so that $g(\theta) = \mu \exp(-\mu\theta)$, Equation 9 becomes

$$P_{1c}(t) = 1 - \frac{\lambda_1 - \lambda_2}{\mu + \lambda_1 - \lambda_2} \exp[-(\mu + \lambda_1)t] - \frac{\mu}{\mu + \lambda_1 - \lambda_2} \exp(-\lambda_2 t) \qquad (10)$$

This equation is identical to the mixed exponential distribution of Equation 6, with the relationship between the two sets of parameters given by:

$$\begin{aligned} \phi_1 &= \mu + \lambda_1 & \qquad \lambda_1 &= \gamma\phi_1 + (1-\gamma)\phi_2 \\ \phi_2 &= \lambda_2 & \lambda_2 &= \phi_2 \\ \gamma &= \frac{\lambda_1 - \lambda_2}{\mu + \lambda_1 - \lambda_2} & \mu &= (1-\gamma)(\phi_1 - \phi_2) \end{aligned} \qquad (11)$$

Note that setting $\lambda_2 = 0$ produces Equation 1, the split population distribution of M1.

4. Forecasting

Thus far we have discussed the structural and practical underpinnings of the models. In this section we apply the two models to a number of data sets to test their ability to extrapolate beyond the given data.

The method used to compare the models is relatively straightforward. If failure data are available for each of 22 months, as they are for the cohort in Figure 1, we can use the 22-month data point as the target point. Then, if we wish to test a model's forecasting ability using six months of data, we use only the first six months of data to estimate the model's parameters. Using these parameters, we can forecast the number of failures at month 22. This can be done using varying cutoff points, from 21 months down to the point where the model is no longer valid.

Figure 2 shows such forecasts of the number of failures at 22 months for models M1 and M2, using the data described in Maltz & McCleary (1977). Since the forecasts from both models are within one standard deviation of each other from two through 21 months, there is no relative advantage for using either model for forecasting.

A similar comparison, using data from a North Carolina study (Witte & Schmidt, 1977; see also Harris et al, 1980), is shown in Figure 3. In this case model M1a clearly provides a better forecast, over a greater range, than does M2.

Four additional comparisons are shown in Figures 4-7, using data from four cohorts released on parole from federal prisons (Hoffman & Stone-Meierhoefer, 1979). The four cohorts are distinguished from each other by risk level, obtained from a Salient Factor Score (Hoffman & Beck, 1974). Figure 4 shows the forecasts of the two models for the "very good risk" cohort; Figure 5, the "good risk" cohort; Figure 6, the "fair risk" cohort; and Figure 7, the "poor risk" cohort. As can be seen, Model M1 generally provides a better forecast than does M2.

5. Continuous Failure Rate Distribution

Morrison (1980) has recently proposed a recidivism model (M3) that at first glance appears to be quite different from those discussed above. Rather than assuming that all members of the population have the same characteristics (γ and ϕ for M1, b and c for M2), he assumes that each individual has his own constant (but unknown to us) failure rate. He posits that all that is known about an individual's failure rate ϕ is the distribution $G(\phi)$ of failure rates for the population. This results in a cumulative distribution of failure times of

$$P_3(t) = \int \left[1 - \exp(-\phi t)\right] dG(\phi) \qquad (12)$$

If $G(\phi)$ can be represented in some parametric form, then estimates of the parameters can be obtained. Although this will permit one to characterize the failure process for the <u>collection</u> of individuals, it will say little about any individual's failure process.[9]

Morrison demonstrates this technique using a gamma function for $g(\phi)=dG(\phi)/d\phi$. This permits Equation 12 to be easily integrated, resulting in a Pareto distribution for $P_3(t)$. However, there are a number of difficulties with this approach.

- no explicit allowance is made for the possibility that some individuals do not fail;
- a specific form of $P_3(t)$ (e.g., a Pareto) is not sufficient to specify $G(\phi)$ uniquely; and
- there is little <u>a priori</u> justification for using any particular $G(\phi)$, such as a gamma distribution.

This approach can be related to the ones discussed above. Consider the following distribution $G(\phi)$ of failure rates:

$$G(\phi) = \begin{cases} 0, & 0 \le \phi < \phi_1 \\ \gamma, & \phi_1 \le \phi < \phi_2 \\ 1, & \phi_2 \le \phi < \infty \end{cases}$$

Using this distribution, Equation 12 produces

$$P_3(t) = \gamma[1 - \exp(-\phi_1 t)] + (1-\gamma)\,[1 - \exp(-\phi_2 t)]$$

which is M1b, Equation 6. Thus, M1b can be viewed as a special case of a mixture of rates for which only three values (0, ϕ_1 and ϕ_2) are allowed. More complicated mixtures of failure rates are, of course, possible, but as in all modeling activity parsimony is to be sought.

6. Conclusion

In this paper we have described the characteristics of similar models of a social process. Rather than comparing them using a purely statistical "goodness of fit" criterion, we have examined the models according to different criteria: how well each model typifies the process under study; interpretation of each model's parameters; the ease with which confidence statements can be obtained; the adaptability of each model to changes in assumptions; and the ability of each model to forecast beyond the available data. Such comparisons, we maintain, are appropriate for choosing among competitive models in a wide variety of social process modeling situations.

NOTES

1. A person is considered to have recidivated if, after having been released from custody, he commits another crime. However, there are many ways to operationalize this definition: based on violation of the conditions of release (parole, probation, halfway house, etc.), on arrest, on prosecution, on conviction, or on return to prison (Maltz, 1980).

2. Bloom (1979: 184) erroneously implies that M1 is a constant failure rate model.

3. This is called the "failure rate" or "hazard rate". It is the probability density function of failures ($P'(t)$) divided by the complementary cumulative distribution of failures ($1 - P(t)$).

4. We neglected to point this out in our earlier papers. We welcome the opportunity to do so now.

5. A point process is a random event-generating process whose primary characteristic is the time of occurrence of each event.

6. Based on Boland and Wilson's (1978) review of the literature on criminality, one can estimate the fraction of crimes resulting in arrest as somewhere between 0.2 and 0.05.

7. The exponential distribution is a limiting form of the geometric distribution as the interval between time periods becomes small.

8. When a distribution is a mixture of two exponential distributions whose failure rates (ϕ_1, ϕ_2) are of the same order of magnitude, it cannot be easily distinguished from a single exponential distribution whose failure rate is $w\phi_1 + (1-w)\phi_2$, where $0 \leq w \leq 1$ (see Harris et al, 1980). If the data are noisy, distinguishing the two situations becomes more difficult. Thus the smaller failure rate should be much smaller than the initial failure rate for this model to provide additional information over one with a single failure rate.

9. Yet another approach is to posit that an individual's failure parameters are functions of certain of his characteristics (e.g., Witte & Schmidt, 1977).

APPENDIX: CONDITIONAL FORECASTS

This appendix describes the computation of forecasts of the total number of recidivists by a target date, conditional upon the data observed up to some earlier time.

Geometric Model (M1a)

Since the concepts are easiest to present for a discrete-time model, we first develop the forecasting procedure for model 1a, in which the relevant variables are:

τ: number of months for which data is available

N: number in cohort

K: total number of failures at or before the τth month

t: target date (in months) for which forecast is desired

M = N - K: number of people who have not failed by the τth month

In order to compute a forecast of the total number of recidivists at the target month t, given that a total of K failures have been observed up to month τ, it is sufficient to consider what can happen to the M non-failures in the $(t - \tau)$ months of the forecast interval.

Based on this model, for each of the M non-failures there will be a conditional probability u of his failing in the interval between the τth month and the tth month. This probability can be expressed:

$$u = \text{Prob}\{\text{ failure month} \le t \mid \text{failure month} > \tau\}$$

and can be computed by using the definition of conditional probability

and equation (5):

$$u = \frac{\text{Prob}\{\text{ failure month} \le t \cap \text{failure month} > \tau\}}{\text{Prob}\{\text{ failure month} > \tau\}}$$

$$= \frac{P_{1a}(t) - P_{1a}(\tau)}{1 - P_{1a}(\tau)}$$

$$= \frac{\gamma q^{\tau}\,(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}} \qquad \text{(A2)}$$

Thus, if γ and q are known, u is known.

To complete the forecast, we now note that the number of people who will fail in the interval between the month τ and month t, of those M people who had not failed by month τ, is a binomially distributed random variable R, with

$$\text{Prob}\{R = r\} = \binom{M}{r} u^{r} (1 - u)^{M-r} \qquad r = 0,1,2,\ldots, M$$

Thus, given a value of u, the expected value of R is:

$$E[R] = Mu \qquad \text{(A3)}$$

and $E[R^2] = Mu + M(M-1)\,u^2$ (A4)

from which the variance Var [R] may be found:

$$\text{Var}[R] = E[R^2] - E^2[R] = Mu(1-u) \qquad \text{(A5)}$$

This allows an estimate $\hat{k}(t)$ of the total number of failures by time t, given K failures by time τ, to be obtained from (A3) and the definition of M:

$$\hat{k}(t) = K + Mu \qquad \text{(A6)}$$

with associated variance

$$\sigma_k^2(t) = M\,u(1-u) \qquad \text{(A7)}$$

Equations (A6) and (A7) were used to plot Figures 2-5.

Continuous Model (M1)

Forecasting the total number of recidivists by time t, given that a total of K have been observed by time τ — where t and τ can take on continuous time values as in model M1 — follows a parallel computation. Again we define u to be the probability that an individual will fail at or before time t, given that he has not yet failed at or before time τ. Now, however, equation (1) is used and (assuming γ and $\emptyset$ are known) this probability becomes:

$$u = \frac{P_1(t) - P_1(\tau)}{1 - P_1(\tau)}$$

$$u = \frac{\gamma\, e^{-\emptyset\tau}\,[1 - e^{-\emptyset(t-\tau)}]}{1 - \gamma + \gamma\, e^{-\emptyset\tau}} \qquad \text{(A8)}$$

The estimate $\hat{k}(t)$ and associated variance $\sigma_k^2(t)$ is again computed from equations (A6) and (A7), with u obtained from (A8).

Adjustment for Uncertain Parameter Values

The forecasts above were derived under the assumption that values of γ and q (or $\emptyset$) are known. For example, they may be hypothesized, or they can be the results of a statistical analysis yielding precise estimates. However, in the case where these parameters are not

previously known the variance given in equation (A7) is an underestimate.

One approach to determining a legitimate probabilistic forecast is available, however. If the uncertainty about the values of γ and q can be represented by means of a probability distribution $F(\gamma,q)$ on these variables — as is done when using Bayesian inference methods — then R is still a random variable, but with a more general distribution than the binomial given above.

In this case, it is possible to compute the expectation and variance of R. In particular, taking the expectation* of equations (A3) and (A4) results in

$$E[R] = M\,E_{\Gamma Q}\left[\frac{\gamma\, q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]$$

$$E[R^2] = E[R] + M(M-1)\,E_{\Gamma Q}\left[\frac{\gamma\, q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]^2$$

The resulting forecast is

$$\hat{k}(t) = K + E[R]$$

with associated variance

$$\sigma_k^2(t) = E\,[R^2] - E^2\,[R].$$

Similiar results are, of course, obtained for the continuous model M1 with parameters γ and $\emptyset$.

---

* $E_{\Gamma Q}\, g(\gamma,q) \equiv \iint g(\gamma,q)\, dF(\gamma,q)$

BIBLIOGRAPHY

Baldus, D. and J.W. Cole. "A Comparison of the Work of Thorsten Sellin and Isaac Ehrlich on the Deterrent Effect of Capital Punishment. Yale Law Journal 85 (1975): 170-186.

Barlow, Richard E., and Frank Proschan. Statistical Theory of Reliability and Life Testing: Probability Models. New York: Holt, Rinehart and Winston, Inc., 1975.

Barnett, Arnold. "The Deterrent Effect of Capital Punishment: A Test of Some Recent Studies." Operations Research (forthcoming).

Bloom, H.S. "Evaluating Human Service and Criminal Justice Programs by Modeling the Probability and Timing of Recidivism." Sociological Methods & Research 8 (1979): 179-208

Bloom, Howard S. and Neil M. Singer. "Determining the Cost-Effectiveness of Correctional Programs: The Case of Patuxent Institution." Evaluation Quarterly 3, 4 (1979): 609-627.

Boland, Barbara and J.W. Wilson. "Age, Crime, and Punishment." The Public Interest 54 (Spring 1978): 22-34.

Bowers, W.J. and G.L. Pierce. "The Illusion of Deterrence in Issac Ehrlich's Research on Capital Punishment." Yale Law Journal 85 (1975): 187-208.

Brier, Stephen S. and Stephen E. Fienberg. "Recent Econometric Modeling of Crime and Punishment: Support for the Deterrence Hypothesis? Evaluation Review 4 (1980): 147-192.

Dhrymes, Phoebus J. et al, "Criteria for Evaluation of Econometric Models," Annals of Economic and Social Measurement, 1, 3 (1972), 291-324.

Ehrlich, I. "Deterrence: Evidence and Inference." Yale Law Journal 85 (1975): 209-227.

Forst, B. "Participation in Illegitimate Actvities: Further Empirical Findings." Policy Analysis 2 (1976): 447-492.

previously known the variance given in equation (A7) is an underestimate.

One approach to determining a legitimate probabilistic forecast is available, however. If the uncertainty about the values of γ and q can be represented by means of a probability distribution $F(\gamma,q)$ on these variables — as is done when using Bayesian inference methods — then R is still a random variable, but with a more general distribution than the binomial given above.

In this case, it is possible to compute the expectation and variance of R. In particular, taking the expectation* of equations (A3) and (A4) results in

$$E[R] = M\, E_{\Gamma Q}\left[\frac{\gamma q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]$$

$$E[R^2] = E[R] + M(M-1)\, E_{\Gamma Q}\left[\frac{\gamma q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]^2$$

The resulting forecast is

$$\hat{k}(t) = K + E[R]$$

with associated variance

$$\sigma_k^2(t) = E\,[R^2] - E^2\,[R].$$

Similiar results are, of course, obtained for the continuous model M1 with parameters γ and ϕ.

---

* $E_{\Gamma Q}\; g(\gamma,q) \;\equiv \iint g(\gamma,q)\; dF(\gamma,q)$

previously known the variance given in equation (A7) is an underestimate.

One approach to determining a legitimate probabilistic forecast is available, however. If the uncertainty about the values of γ and q can be represented by means of a probability distribution $F(\gamma,q)$ on these variables — as is done when using Bayesian inference methods — then R is still a random variable, but with a more general distribution than the binomial given above.

In this case, it is possible to compute the expectation and variance of R. In particular, taking the expectation* of equations (A3) and (A4) results in

$$E[R] = M\, E_{TQ}\left[\frac{\gamma\, q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]$$

$$E[R^2] = E[R] + M(M-1)\, E_{TQ}\left[\frac{\gamma\, q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]^2$$

The resulting forecast is

$$\hat{k}(t) = K + E\{R\}$$

with associated variance

$$\sigma_k^2(t) = E\,[R^2] - E^2\,[R].$$

Similiar results are, of course, obtained for the continuous model M1 with parameters γ and $\emptyset$.

---

* $E_{TQ}\, g(\gamma,q) \equiv \iint g(\gamma,q)\, dF(\gamma,q)$

BIBLIOGRAPHY

Baldus, D. and J.W. Cole. "A Comparison of the Work of Thorsten Sellin and Isaac Ehrlich on the Deterrent Effect of Capital Punishment. Yale Law Journal 85 (1975): 170-186.

Barlow, Richard E., and Frank Proschan. Statistical Theory of Reliability and Life Testing: Probability Models. New York: Holt, Rinehart and Winston, Inc., 1975.

Barnett, Arnold. "The Deterrent Effect of Capital Punishment: A Test of Some Recent Studies." Operations Research (forthcoming).

Bloom, H.S. "Evaluating Human Service and Criminal Justice Programs by Modeling the Probability and Timing of Recidivism." Sociological Methods & Research 8 (1979): 179-208

Bloom, Howard S. and Neil M. Singer. "Determining the Cost-Effectiveness of Correctional Programs: The Case of Patuxent Institution." Evaluation Quarterly 3, 4 (1979): 609-627.

Boland, Barbara and J.W. Wilson. "Age, Crime, and Punishment." The Public Interest 54 (Spring 1978): 22-34.

Bowers, W.J. and G.L. Pierce. "The Illusion of Deterrence in Issac Ehrlich's Research on Capital Punishment." Yale Law Journal 85 (1975): 187-208.

Brier, Stephen S. and Stephen E. Fienberg. "Recent Econometric Modeling of Crime and Punishment: Support for the Deterrence Hypothesis? Evaluation Review 4 (1980): 147-192.

Dhrymes, Phoebus J. et al, "Criteria for Evaluation of Econometric Models," Annals of Economic and Social Measurement, 1, 3 (1972), 291-324.

Ehrlich, I. "Deterrence: Evidence and Inference." Yale Law Journal 85 (1975): 209-227.

Forst, B. "Participation in Illegitimate Actvities: Further Empirical Findings." Policy Analysis 2 (1976): 447-492.

BIBLIOGRAPHY

Baldus, D. and J.W. Cole. "A Comparison of the Work of Thorsten Sellin and Isaac Ehrlich on the Deterrent Effect of Capital Punishment. Yale Law Journal 85 (1975): 170-186.

Barlow, Richard E., and Frank Proschan. Statistical Theory of Reliability and Life Testing: Probability Models. New York: Holt, Rinehart and Winston, Inc., 1975.

Barnett, Arnold. "The Deterrent Effect of Capital Punishment: A Test of Some Recent Studies." Operations Research (forthcoming).

Bloom, H.S. "Evaluating Human Service and Criminal Justice Programs by Modeling the Probability and Timing of Recidivism." Sociological Methods & Research 8 (1979): 179-208

Bloom, Howard S. and Neil M. Singer. "Determining the Cost-Effectiveness of Correctional Programs: The Case of Patuxent Institution." Evaluation Quarterly 3, 4 (1979): 609-627.

Boland, Barbara and J.W. Wilson. "Age, Crime, and Punishment." The Public Interest 54 (Spring 1978): 22-34.

Bowers, W.J. and G.L. Pierce. "The Illusion of Deterrence in Issac Ehrlich's Research on Capital Punishment." Yale Law Journal 85 (1975): 187-208.

Brier, Stephen S. and Stephen E. Fienberg. "Recent Econometric Modeling of Crime and Punishment: Support for the Deterrence Hypothesis? Evaluation Review 4 (1980): 147-192.

Dhrymes, Phoebus J. et al, "Criteria for Evaluation of Econometric Models," Annals of Economic and Social Measurement, 1, 3 (1972), 291-324.

Ehrlich, I. "Deterrence: Evidence and Inference." Yale Law Journal 85 (1975): 209-227.

Forst, B. "Participation in Illegitimate Actvities: Further Empirical Findings." Policy Analysis 2 (1976): 447-492.

previously known the variance given in equation (A7) is an underestimate.

One approach to determining a legitimate probabilistic forecast is available, however. If the uncertainty about the values of γ and q can be represented by means of a probability distribution $F(\gamma,q)$ on these variables — as is done when using Bayesian inference methods — then R is still a random variable, but with a more general distribution than the binomial given above.

In this case, it is possible to compute the expectation and variance of R. In particular, taking the expectation* of equations (A3) and (A4) results in

$$E[R] = M\, E_{TQ}\left[\frac{\gamma q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]$$

$$E[R^2] = E[R] + M(M-1)\, E_{TQ}\left[\frac{\gamma q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]^2$$

The resulting forecast is

$$\hat{k}(t) = K + E[R]$$

with associated variance

$$\sigma_k^2(t) = E[R^2] - E^2[R].$$

Similiar results are, of course, obtained for the continuous model M1 with parameters γ and $\emptyset$.

---

* $E_{TQ}\, g(\gamma,q) \equiv \iint g(\gamma,q)\, dF(\gamma,q)$

BIBLIOGRAPHY

Baldus, D. and J.W. Cole. "A Comparison of the Work of Thorsten Sellin and Isaac Ehrlich on the Deterrent Effect of Capital Punishment. Yale Law Journal 85 (1975): 170-186.

Barlow, Richard E., and Frank Proschan. Statistical Theory of Reliability and Life Testing: Probability Models. New York: Holt, Rinehart and Winston, Inc., 1975.

Barnett, Arnold. "The Deterrent Effect of Capital Punishment: A Test of Some Recent Studies." Operations Research (forthcoming).

Bloom, H.S. "Evaluating Human Service and Criminal Justice Programs by Modeling the Probability and Timing of Recidivism." Sociological Methods & Research 8 (1979): 179-208

Bloom, Howard S. and Neil M. Singer. "Determining the Cost-Effectiveness of Correctional Programs: The Case of Patuxent Institution." Evaluation Quarterly 3, 4 (1979): 609-627.

Boland, Barbara and J.W. Wilson. "Age, Crime, and Punishment." The Public Interest 54 (Spring 1978): 22-34.

Bowers, W.J. and G.L. Pierce. "The Illusion of Deterrence in Issac Ehrlich's Research on Capital Punishment." Yale Law Journal 85 (1975): 187-208.

Brier, Stephen S. and Stephen E. Fienberg. "Recent Econometric Modeling of Crime and Punishment: Support for the Deterrence Hypothesis? Evaluation Review 4 (1980): 147-192.

Dhrymes, Phoebus J. et al, "Criteria for Evaluation of Econometric Models," Annals of Economic and Social Measurement, 1, 3 (1972), 291-324.

Ehrlich, I. "Deterrence: Evidence and Inference." Yale Law Journal 85 (1975): 209-227.

Forst, B. "Participation in Illegitimate Actvities: Further Empirical Findings." Policy Analysis 2 (1976): 447-492.

previously known the variance given in equation (A7) is an underestimate.

One approach to determining a legitimate probabilistic forecast is available, however. If the uncertainty about the values of γ and q can be represented by means of a probability distribution $F(\gamma,q)$ on these variables — as is done when using Bayesian inference methods — then R is still a random variable, but with a more general distribution than the binomial given above.

In this case, it is possible to compute the expectation and variance of R. In particular, taking the expectation* of equations (A3) and (A4) results in

$$E[R] = M\ E_{\Gamma Q}\left[\frac{\gamma\ q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]$$

$$E[R^2] = E[R] + M(M-1)\ E_{\Gamma Q}\left[\frac{\gamma\ q^{\tau}(1 - q^{t-\tau})}{1 - \gamma + \gamma q^{\tau}}\right]^2$$

The resulting forecast is

$$\hat{k}(t) = K + E[R]$$

with associated variance

$$\sigma_k^2(t) = E\,[R^2] - E^2\,[R].$$

Similiar results are, of course, obtained for the continuous model M1 with parameters Υ and $\emptyset$.

* $E_{\Gamma Q}\ g(\gamma,q) \equiv \iint g(\gamma,q)\ dF(\gamma,q)$

BIBLIOGRAPHY

Baldus, D. and J.W. Cole. "A Comparison of the Work of Thorsten Sellin and Isaac Ehrlich on the Deterrent Effect of Capital Punishment. Yale Law Journal 85 (1975): 170-186.

Barlow, Richard E., and Frank Proschan. Statistical Theory of Reliability and Life Testing: Probability Models. New York: Holt, Rinehart and Winston, Inc., 1975.

Barnett, Arnold. "The Deterrent Effect of Capital Punishment: A Test of Some Recent Studies." Operations Research (forthcoming).

Bloom, H.S. "Evaluating Human Service and Criminal Justice Programs by Modeling the Probability and Timing of Recidivism." Sociological Methods & Research 8 (1979): 179-208

Bloom, Howard S. and Neil M. Singer. "Determining the Cost-Effectiveness of Correctional Programs: The Case of Patuxent Institution." Evaluation Quarterly 3, 4 (1979): 609-627.

Boland, Barbara and J.W. Wilson. "Age, Crime, and Punishment." The Public Interest 54 (Spring 1978): 22-34.

Bowers, W.J. and G.L. Pierce. "The Illusion of Deterrence in Issac Ehrlich's Research on Capital Punishment." Yale Law Journal 85 (1975): 187-208.

Brier, Stephen S. and Stephen E. Fienberg. "Recent Econometric Modeling of Crime and Punishment: Support for the Deterrence Hypothesis? Evaluation Review 4 (1980): 147-192.

Dhrymes, Phoebus J. et al, "Criteria for Evaluation of Econometric Models," Annals of Economic and Social Measurement, 1, 3 (1972), 291-324.

Ehrlich, I. "Deterrence: Evidence and Inference." Yale Law Journal 85 (1975): 209-227.

Forst, B. "Participation in Illegitimate Actvities: Further Empirical Findings." Policy Analysis 2 (1976): 447-492.

Gass, Saul I. and Bruce W. Thompson, "Guidelines for Model Evaluation: An Abridged Version of the U.S. General Accounting Office Exposure Draft," Operations Research 28 (1980): 431-439.

Haight, Frank A. Handbook of the Poisson Distribution. New York: John Wiley and Sons, Inc., 1967.

Harris, Carl M., Ali R. Kaylan, and Michael D. Maltz. "Advances in Statistics of Recidivism Measurement." To appear in J.A. Fox, Ed., Frontiers in Quantitative Criminology. New York: Academic Press, 1980.

Hoffman, Peter B. and James L. Beck. "Parole Decision-Making: A Salient Factor Score." Journal of Criminal Justice 2 (1974): 195-206.

Hoffman, Peter B. and Barbara Stone-Meierhoefer. "Post Release Arrest Experiences of Federal Prisoners - A Six Year Followup," Journal of Criminal Justice 7 (1979): 193-216.

Kitchener, Howard, Annesley K. Schmidt, and Daniel Glaser. "How Persistent Is Post-Prison Success?" Federal Probation 41 (1977): 9-15.

Maltz, Michael D. "On Recidivism: Exploring Its Properties as a Measure of Correctional Effectiveness." Final report on Grant Number 77-NI99-0073, the National Institute of Law Enforcement and Criminal Justice: University of Illinois at Chicago Circle: March, 1980.

Maltz, Michael D., and Richard McCleary. "The Mathematics of Behavioral Change: Recidivism and Construct Validity." Evaluation Quarterly 1 (1977): 421-438.

Maltz, Michael D., and Stephen M. Pollock. "Artificial Inflation of a Poisson Rate by a 'Selection Effect.'" Operations Research (May-June, 1980).

Mann, Nancy R., Ray E. Schafer, and Nozer D. Singpurwalla. Methods for Statistical Analysis of Reliability & Life Data. New York: John Wiley and Sons, Inc., 1974.

Morrison, Donald G., "Some New Models for the Analysis of Recidivism Data", Graduate School of Business, Columbia University, New York (1980). Presented at the Joint National Meeting of the Institute of Management Sciences and the Operations Research Society of America, Washington, DC, May 6, 1980.

Partanen, Juha. "On Waiting Time Distributions." Acta Sociologica 12 (1969): 132-143.

Passell, P. "The Deterrent Effect of the Death Penalty: A Statistical Test," Stanford Law Review 28 (1): 61-80.

Philpotts, G.J.O., and L.B. Lancucki. Previous Convictions, Sentence and Reconviction. Home Office Research Study No. 53. London: Her Majesty's Stationery Office, 1979.

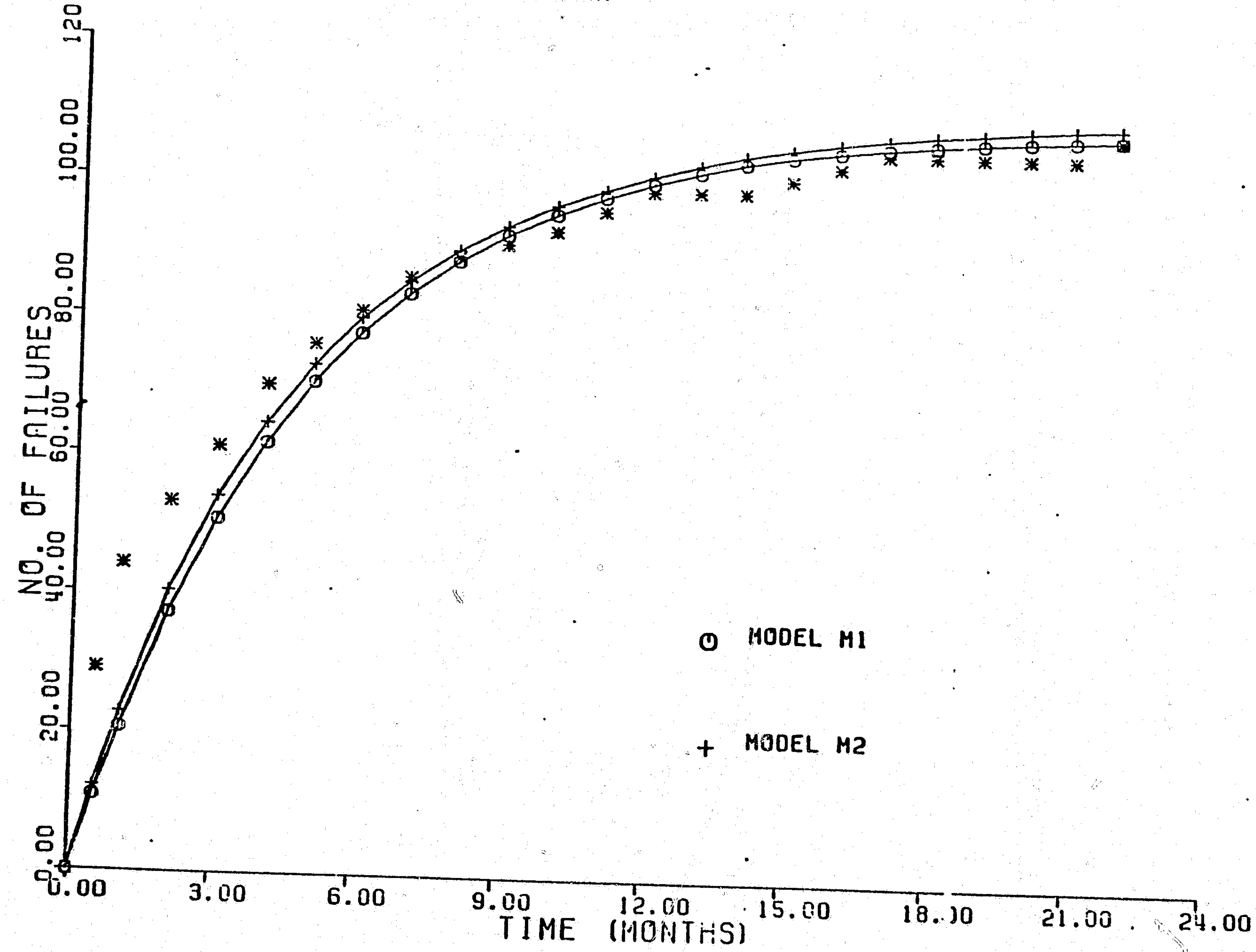
Sechrest, Lee, Susan O. White, and Elizabeth D. Brown, Editors. The Rehabilitation of Criminal Offenders: Problems and Prospects. Washington, DC: National Academy of Sciences, 1979.

Tuma, Nancy B. and Michael T. Hannan, "Dynamic Analysis of Event Histories," American Journal of Sociology 84, 4 (1979): 820-854.

U.S. General Accounting Office. Guidelines for Model Evaluation. Washington DC: U.S. GAO, January, 1979.

Witte, Ann D. and Peter Schmidt. "An Analysis of Recidivism, Using the Truncated Lognormal Distribution." Applied Statistics 26 (1977): 302-311.

FIGURE 1. COMPARING TWO DIFFERENT MODELS OF RECIDIVISM WITH EMPIRICAL DATA



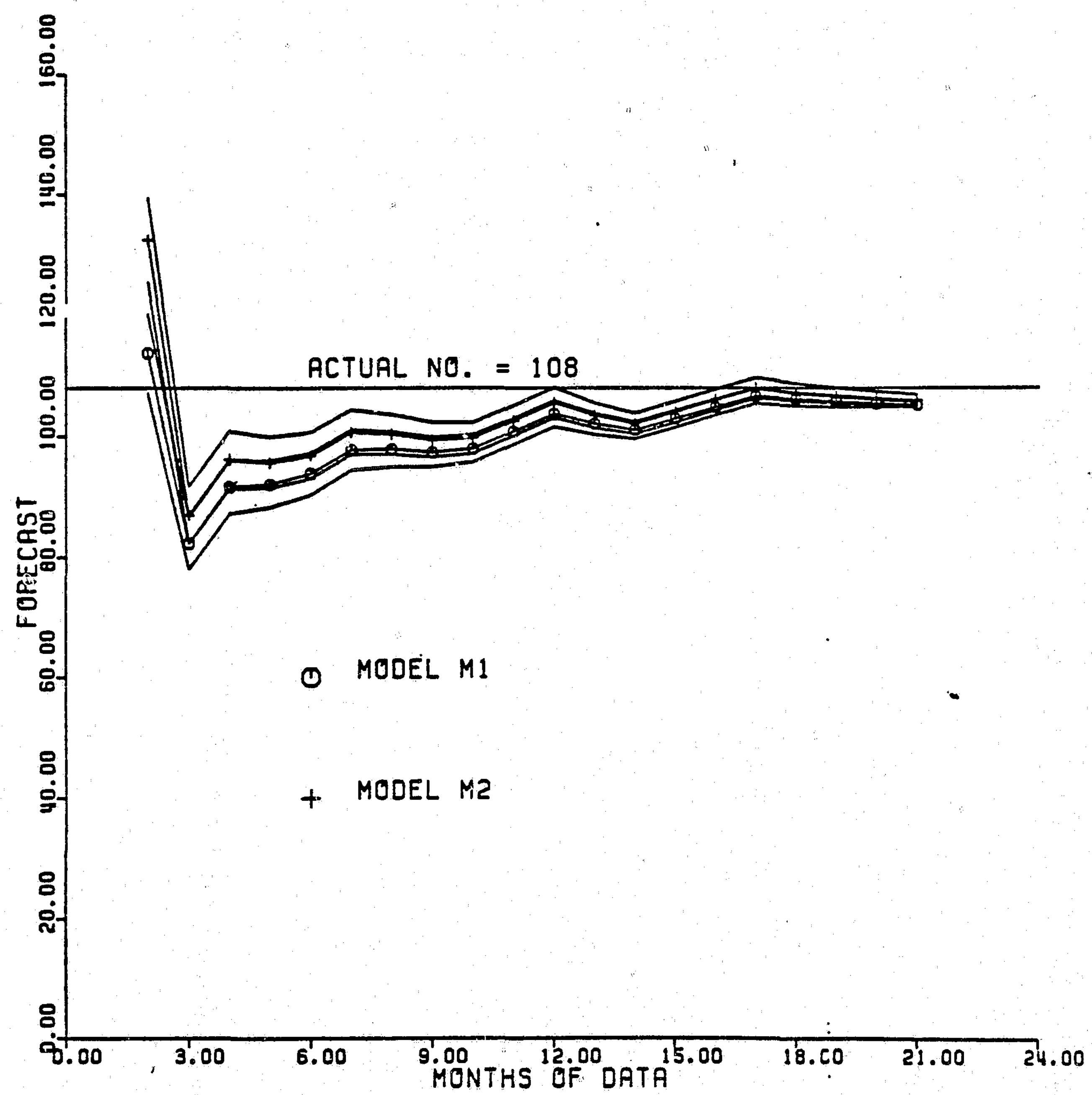


FIGURE 2. COMPARING FORECASTS OF THE NUMBER OF FAILURES AT 22 MONTHS, USING ILLINOIS DATA

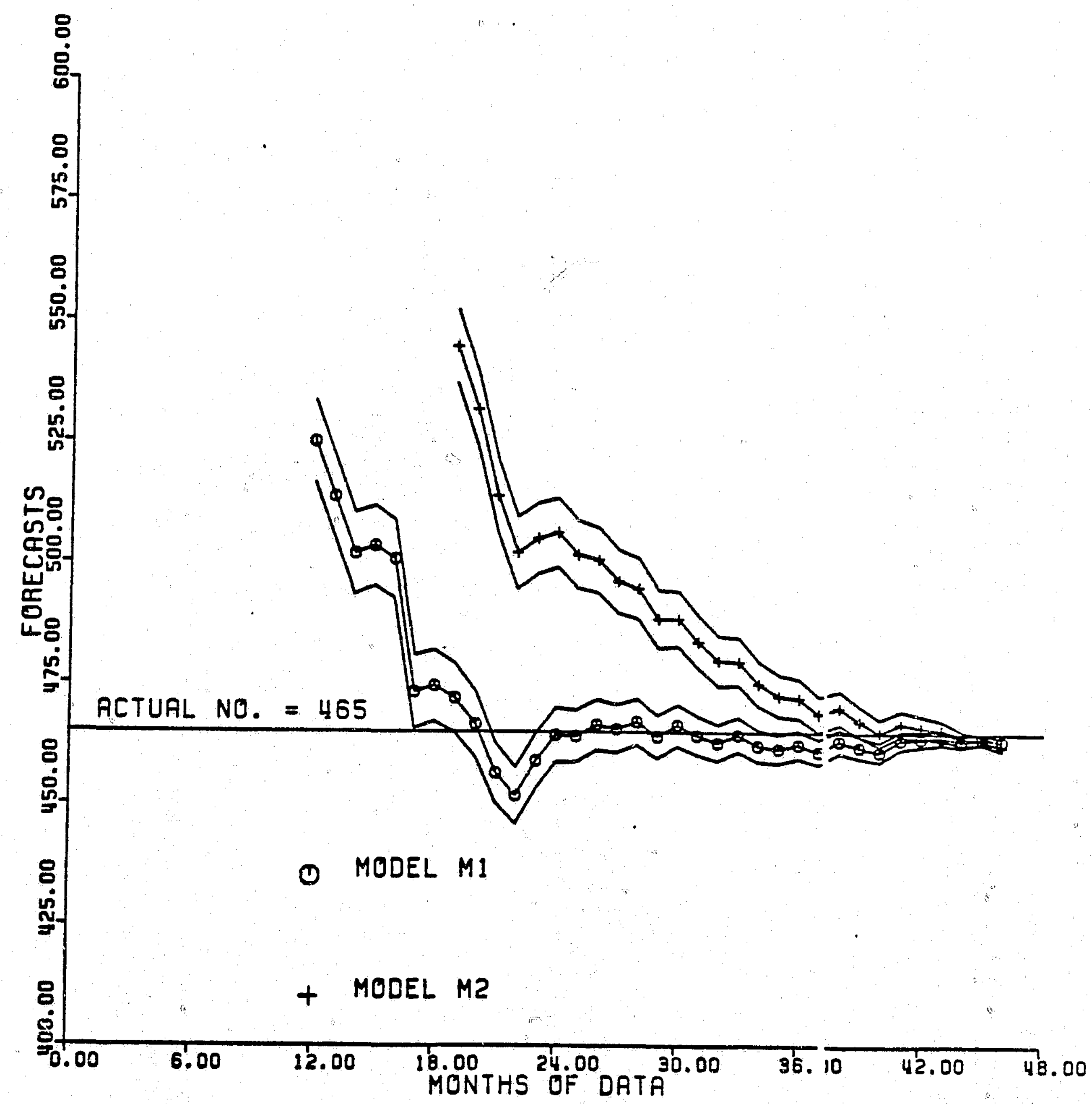


FIGURE 3. COMPARING FORECASTS OF THE NUMBER OF FAILURES AT 47 MONTHS, USING NORTH CAROLINA DATA

FIGURE 4. COMPARING FORECASTS OF THE NUMBER OF FAILURES AT 72 MONTHS, USING US PAROLE COMMISSION DATA ("VERY GOOD RISK" COHORT)

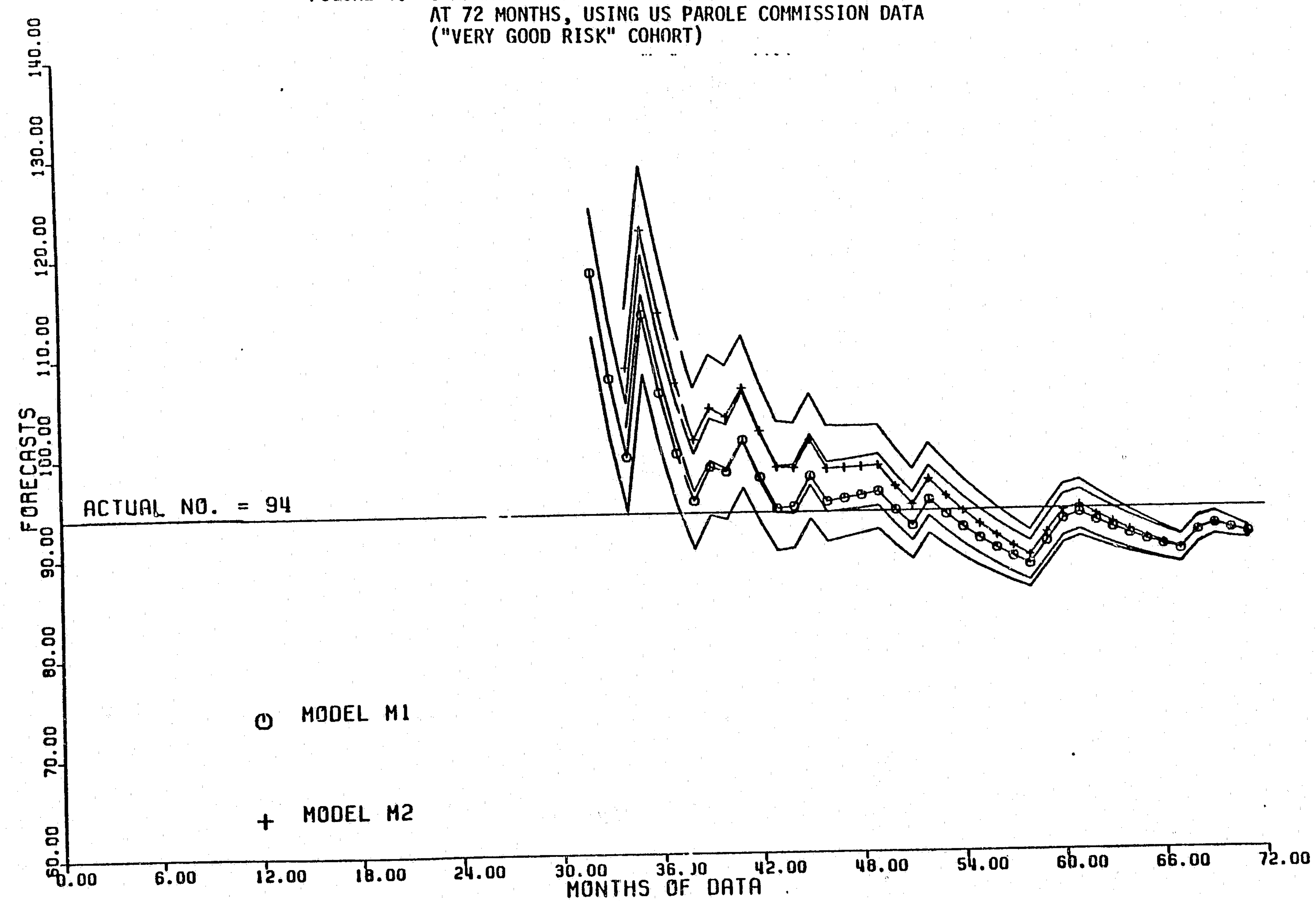
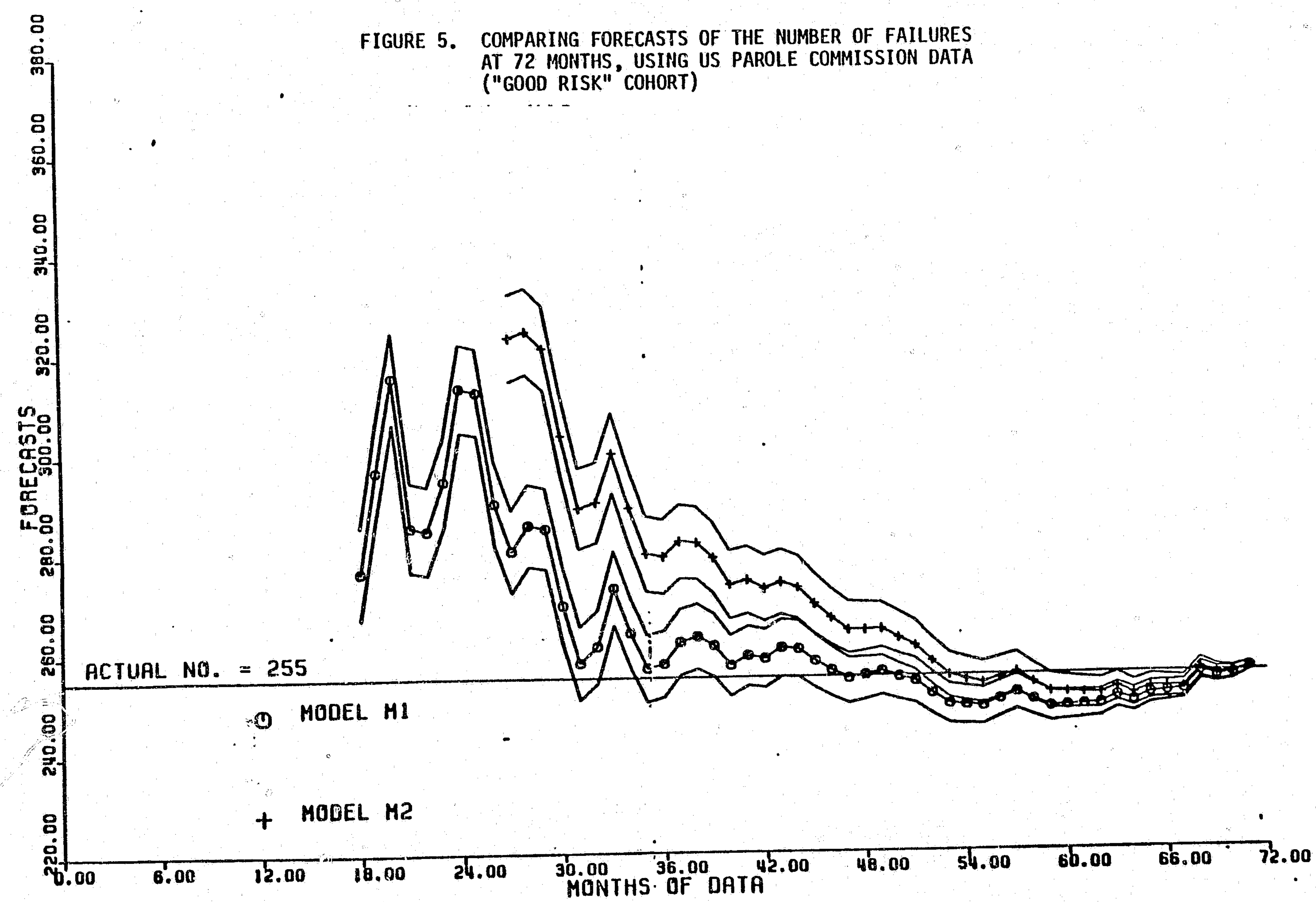


FIGURE 5. COMPARING FORECASTS OF THE NUMBER OF FAILURES AT 72 MONTHS, USING US PAROLE COMMISSION DATA ("GOOD RISK" COHORT)



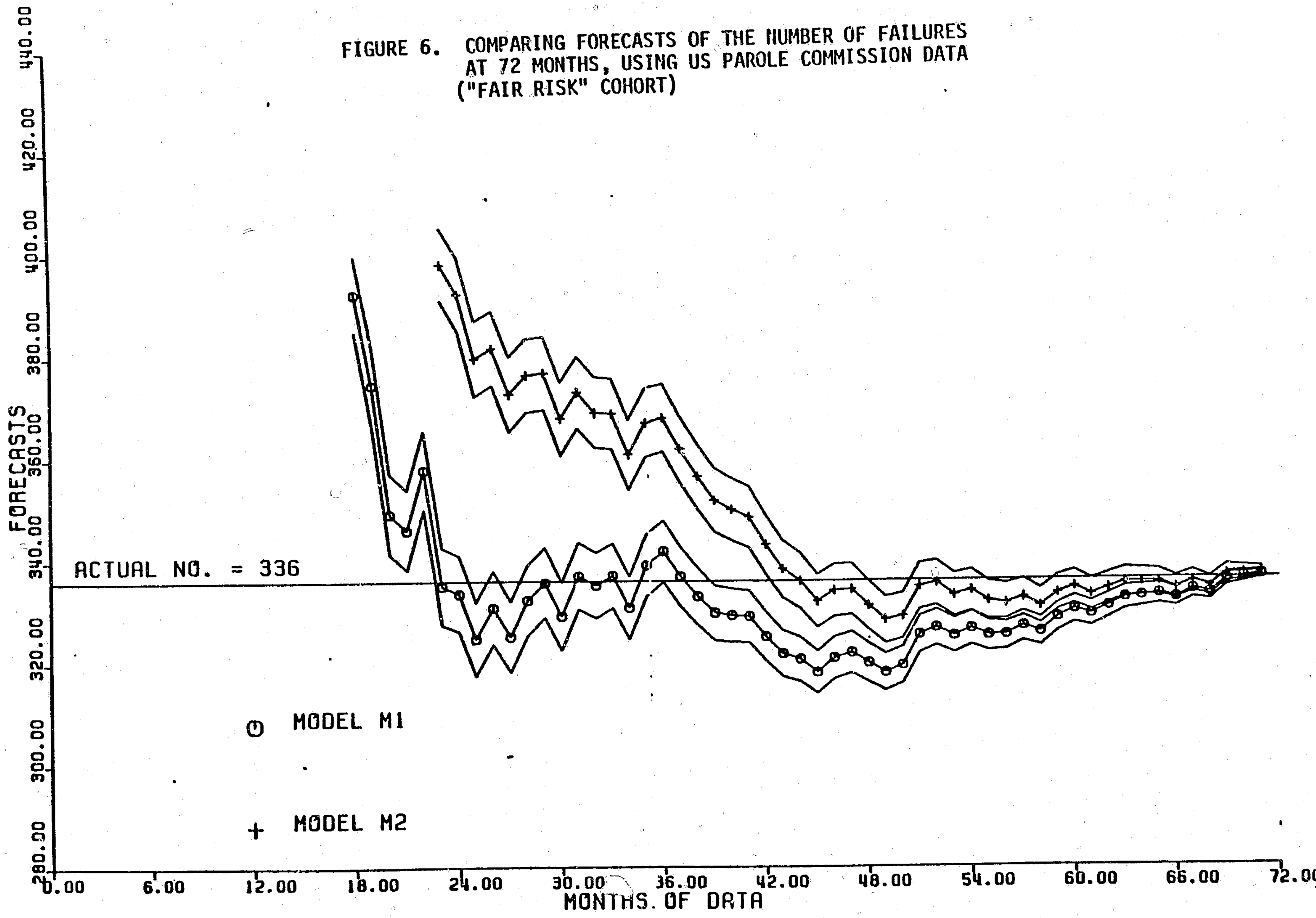
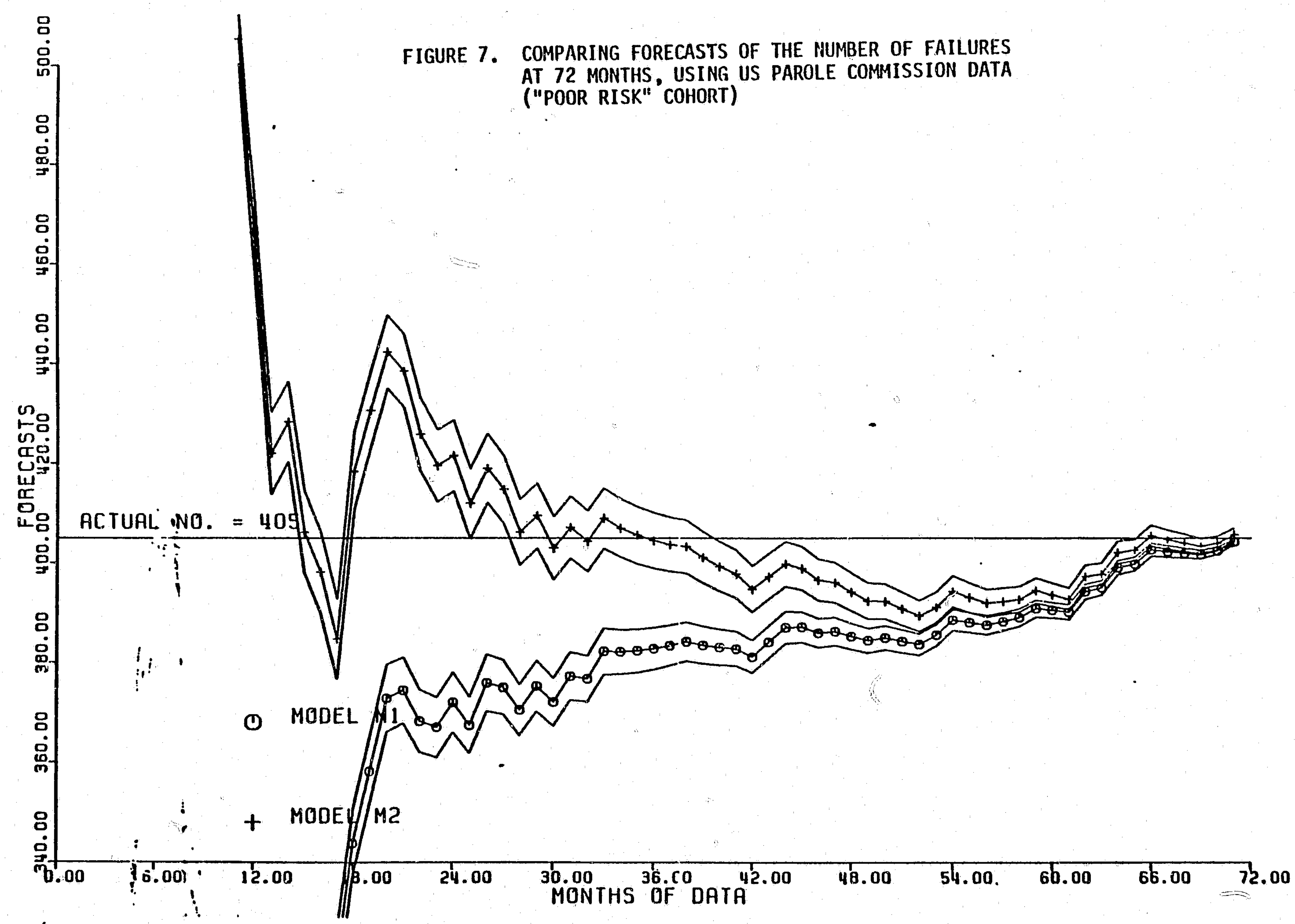


FIGURE 6. COMPARING FORECASTS OF THE NUMBER OF FAILURES AT 72 MONTHS, USING US PAROLE COMMISSION DATA ("FAIR RISK" COHORT)
FORECASTS
440.00
420.00
400.00
380.00
360.00
340.00
320.00
300.00
280.00
ACTUAL NO. = 336
MODEL M1
MODEL M2
0.00
6.00
12.00
18.00
24.00
30.00
36.00
42.00
48.00
54.00
60.00
66.00
72.00
MONTHS OF DATA

FIGURE 7. COMPARING FORECASTS OF THE NUMBER OF FAILURES AT 72 MONTHS, USING US PAROLE COMMISSION DATA ("POOR RISK" COHORT)



APPENDIX H

COVARIATE ANALYSIS

This Appendix describes our efforts to perform covariate analyses of recidivism. Our approach assumes that each individual i has a unique probability γ_i of ultimately recidivating, and (if a recidivist) has a unique probability q_i of surviving another month without recidivating. We further assume that a small number of individual characteristics are primarily responsible for individual variations in γ_i and q_i, so that

$$\gamma_i = \gamma_i(x_{i1}, x_{i2}, \cdots x_{im}) \qquad (1)$$

and

$$q_i = q_i(x_{i1}, x_{i2} \cdots x_{im}) \qquad (2)$$

where x_{ij} is the jth characteristic of individual i.

An initial attempt to use multivariate regression techniques to estimate the above relationships was not successful. Our approach and the reasons for its failure are described in Section I. An alternative approach was used: partitioning the data according to specific covariate values. This is described in Section II. This approach met with more success, and some interesting patterns emerged. Sections 3-7 describes the four data sets we analyzed using this approach and the results of the analyses. Section 8 compares the results with each other and discusses their implications.

I. Multivariate Parametric Modelling

A common approach to multivariate regression is to assume that the relation between independent and dependent variables is linear, e.g., that

$$\gamma_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots b_m x_{im}$$

However, this ignores the fact γ_i is restricted to be between 0 and 1 (as is q).

To insure that the bounds on γ_i and q_i are not exceeded, we used the following forms for equations (1) and (2):

$$1 - \gamma_i = (1 - \alpha_0)\alpha_1^{x_{i1}}\alpha_2^{x_{i2}} \cdots \alpha_m^{x_{im}} \quad (3)$$

$$q_i = \beta_0\beta_1^{x_{i1}}\beta_2^{x_{i2}} \cdots \beta_m^{x_{im}}$$

For γ_i and q_i to be between 0 and 1, we must have $0 < \alpha_j, \beta_j < 1$. Furthermore, the x_{ij} are scaled so that large values imply a "worse" recidivism behavior -- i.e., γ_i increases and q_i decreases as any x_{ij} increases. Equations (3) and (4) are made linear when logarithms are taken of both sides, so standard techniques can be used.

Estimation of the coefficient vectors $\underline{\alpha}$ and $\underline{\beta}$ were attempted by maximizing the (log) likelihood function resulting when the forms of equations (3) and (4) were assumed for each individual in the cohort. That is, solutions were sought for the problem:

$$\max_{\underline{\alpha}, \underline{\beta}} \sum_{i=1}^{k} \ln[\gamma_i q_i^{t_i - 1}(1-q_i)] + \sum_{i=k+1}^{n} \ln(1-\gamma_i + \gamma_i q_i^{t_i}) \quad (5)$$

Subject to $0 \leq \alpha_j \leq 1 \qquad j = 1, 2, \ldots m$

$0 \leq \beta_j \leq 1$

where:

n = Number of individuals in the cohort.

k = Number of failures.

t_i = Failure time for ith individual $(i = 1,2,\ldots k)$
Censoring time for ith individual $(i = k+1, k+2,\ldots n)$

$\gamma_i = \gamma(\underline{\alpha}, \underline{x}_i)$ from equation (3).

$q_i = q(\underline{\beta}, \underline{x}_i)$ from equation (4).

$\underline{x}_i = (x_{1i}, x_{2i}, \ldots, x_{mi})$ = covariate values for the ith individual.

The solution and interpretation of such an optimization problem was made difficult by a number of factors.

a) When m = 4 (a not unreasonable number of influencing covariates) the solution space contains 10 variables.

b) The objective function is extremely flat in the neighborhood of the optimum, thus leading to convergence problems.

c) For the data sets of interest, n is in the order of 1,000, thus requiring appreciable computation <u>each</u> time the log likelihood function needs to be evaluated by <u>whatever</u> optimization algorithm is used.

d) The stability of any solution algorithm requiring essentially numerical computation of gradients will be influenced greatly by the choice of scaling for the covariates $\underline{x}$. (This is not necessarily a problem for 0-1 covariates).

e) Once a solution is found, statistical statements about the <u>quality</u> of estimates for $\underline{\alpha}$ and $\underline{\beta}$ are almost impossible to make.

In spite of these difficulties, attempts were made to solve equation (3) for the four data sets. Two non-linear unconstrained procedures were used: PRAXIS, a NASA originated "Powell-type" conjugate direction method without derivative information; and a Fletcher-Reeves Conjugate Gradient Method. Both were resident on the University of Michigan"s MTS system, and both required modifications to force feasibility. The results were unsatisfactory: When only a single dichotomous covariate was used, results were eventually obtained that were numerically consistent with the estimates obtained by the partition method of Section 2 of this appendix, but without the benefit of producing confidence intervals.

The failure of these methods led us to a different method of studying the relationship between the recidivism parameters and relevant variables. We partitioned the data according to the categories provided in the data, for a number of variable types:*

- race
- age (at first arrest, at release)
- prior record (probation violations, arrest, felonies, commitments)
- drug or alcohol usage
- social stability (employment, home at release, early home environment)
- objective measures (psychological diagnosis, type of release, parole risk scale)

Unfortunately, the four data sets we analyzed (from Georgia, the U.S.

*The data sets had upwards of a hundred or so variables. We selected for analysis those that our prior research had shown to be correlated with recidivism, and those that the literature show to be salient.

Bureau of Prisons, Iowa, and North Carolina) did not use the same categories. However, we were able to obtain information on these variables in most cases.

II. Georgia

A computer file was made available by George H. Cox, Jr., of the Georgia Department of Offender Rehabilitation, containing information on a sample of 1902 individuals released during the early 1970's. A detailed description of the data sources, appropriate caveats on its use (no formal "randomization" was attempted -- the set instead represents all the cases for which certain data were available), and an analysis of one, two and three-year rearrest rates are given in Cox (1977). The definition of recidivism is rearrest (presumably in Georgia), and although time from release to either rearrest or censoring is available in days, data was converted into time units of months. [For a discussion of the magnitude of error introduced by this time truncation into the computation of parameter estimates, see Maltz (1981).]

Of the 232 data items available for each individual, seven were selected as potentially relevant determinants of the parameters γ and q. These are shown in Table 1A, along with the values used to partition the data set. Table 1B shows for each partition the number of individuals, recidivists, total days-to-recidivism and estimates for γ and q (both maximum likelihood and Bayes'). In order to visualize whether or not these partitions produce estimates whose differences are statistically significant, Figures 1a through 1h show the approximate (using normal approximation) 90% confidence intervals for the Bayes estimates.

A detailed discussion of the interpretation of these confidence intervals is in Appendix F. It is important to note here, however, that the greater the disparity between the maximum likelihood and Bayes estimators, the less the likelihood function (and thus the Bayes posterior distribution of γ and q) has a Gaussian shape, and thus the less meaningful these confidence intervals are. These cases for which the likelihood function is distinctly non-Normal (recognized by having $\gamma_{MLE} = 1$) are shown in Table 1b by "*" entries and are indicated in Figures 1a-1h by dashed confidence intervals.

As the Figures show, some differentiation in parameters -- and thus implied different recidivism behavior -- can be seen for some covariates. A convenient (though informal) way of distinguishing two populations (1 and 2) yielding the different estimate pairs (γ_1, q_1) and (γ_2, q_2) is to say that population 1 is "better" than 2 (or 2 is "worse" than 1) if $\gamma_1 < \gamma_2$ and $q_1 > q_2$, since individuals in population 1 are less sure to recidivate, and are slower to do so given they will. On the other hand, if $\gamma_1 < \gamma_2$ but $q_1 < q_2$, population 2 is "surer but slower" to recidivate as compared to population 1 who are "less sure but faster."

Thus, we can observe that:

- Growing up in a SMSA city is "worse" than a farm or town background (Figure 1d).
- No prior arrests are "better" than one or more (Figure 1g).
- The earlier the age at release, the more likely an individual is to recidivate (Figure 1h).

On the other hand, there is little statistically significant within the other covariate partitions.

III. U.S. Bureau of Prisons

A sample of 927 individuals released from Federal Prisons during 1956 and 1957 (Kitchener et al, 1977) was made available to us by the U.S. Bureau of Prisons. Although many definitions of recidivism have been used in the analysis of this data set, we have chosen the criterion of parole revocation or reconviction (regardless of whether reincarcerated). Thus, simple rearrest without conviction is not counted. Because of the length of the followup period (18 years), the time interval used is one quarter (three months). Thus the reader should be careful in comparing q for these data (q is the conditional probability of no failure in a quarter) to those estimates in the other data sets which pertain to months. The estimates of γ, of course, are comparable across the data sets. Table 2a shows selected covariates and their partitions; Table 2b gives their associated statistics and estimates, and Figures 2a-2m show the corresponding confidence intervals.

The data show:

- Whites have a slightly lower recidivism rate (Figure 2b)
- The lower the age at first arrest, the more likely an individual is to recidivate (Figure 2d)
- Prior felony sentences have a major effect (15 percent increase) on recidivism probability (Figure 2e)
- The better the employment record, the better the post-release record (Figure 2h)
- The lower the age at release, the more likely an individual is to recidivate (Figure 2n)

IV. Iowa

Data from a sample of 3372 releases from the Iowa state prison system was obtained from Daryl Fischer of the Iowa Office for Planning and Programming. Data description and analysis can be found in Iowa (1979). The covariate information used is shown in Table 3a, and associated data and parameter estimates in Table 3b. Figures 3a-3r show the 90% confidence regions for γ and q.

Among the findings for this cohort are:

- Nonwhites are more likely than whites to recidivate (Figure 3c)
- Those with prior juvenile or prison records are more likely to recidivate (Figures 3; 3k, 3o, 3p)
- In contrast to other cohorts, the lower the age at release the less likely an individual is to recidivate (Figure 3m)
- The scale used by Iowa for calculating parole risk at admission is better at predicting speed of recidivism (for those who will fail) than it is at predicting probability of recidivism (Figure 3r)

V. North Carolina

A sample of 641 releases into the North Carolina Work Release Program was made available by Ann Witte, and is described in detail in Schmidt & Witte (undated). Of the over 200 covariates contained in this set, the eleven shown in Table 4a were analyzed.

It was possible to use two different definitions of recidivism: rearrest or reconviction (it is clear that the latter definition is more restrictive than the former). Table 4b and Figures 4a-4ℓ give the statistics, estimates and parameter confidence regions for the rearrest definitions; Table 5 and Figures 5a-5ℓ give the statistics, estimates and parameter confidence regions for the reconviction definition.

The patterns of recidivism do not change significantly when using rearrest or reconviction as the definition of recidivism; the primary difference between the two definitions is an approximately five percent lower recidivism rate when reconviction is used.

Other findings include:

- Whites fail faster than (but with about the same likelihood as) nonwhites (Figures 4b, 5b)
- The lower the age at release, the more likely an individual is to recidivate (Figure 4l, 5l)

VI. Discussion of Results

A. Maximum Likelihood vs. Bayes Estimates

Because of progressive censoring and the incomplete nature of the p.d.f. for the time of recidivism, the joint likelihood function for some of the observed data points may not be close to normal. For those data for which the likelihood function is normal, then the MLE and Bayes estimates are close to each other. Moreover, the Bayes confidence regions are nearly ellipses, and the figures show valid 90% posterior probability regions.

When the MLE differs substantially from the Bayes estimates, however, it is an indication that the likelihood function is non-normal. Indeed, the MLE for γ will equal 1.0 when the likelihood function is increasing in γ for $\gamma < 1.0$, (but of course is zero when $\gamma > 0$). In this case the Bayes confidence regions are no longer ellipses. Nevertheless, the use of the ellipse-approximations serve to qualitatively distinguish between data sets, and in the very least can be used to screen out incidences of data which deserve further investigation. Such cases, occurring mainly in the Georgia data set, are indicated by dashed confidence regions.

B. Important Covariates

Although it was not a primary objective of our research, the information contained in Tables 1-5 (or in Figures 1-5) allow us to identify "important" covariates: those that appear to affect the values of the paramaters γ and q. By noting those covariates that provide distinct (non-overlapping) 90 percent confidence regions for at least two partition values (or sets of values), 7 "types" appear:

1. Race
2. Age (at release or first arrest)
3. Prior record (probation violations, arrests, felonies, committments)
4. Current offense (type, etiology, admission category)
5. Drug or alcohol usage
6. Social stability (employment, home at release, early home environment)
7. Objective measures (psychological diagnosis, type of release parole risk scale).

Prior record is clearly a determinant of the recidivism parameters in all four cohorts. In all cases the direction is as expected: the existence or severity of a prior exposure to (or involvement with) the criminal justice system increases both γ (the probability that an individual will recidivate) and 1-q (the rate at which he does so). When the Georgia sample is put aside (due to the fact that γ=1 for almost all partitions, as discussed above) then we see that race, drug and social stability are important covariates. The direction of "badness" is again what would be expected. It is of interest to note that such a priori suggestive covariates as IQ, SES and educational levels have <u>no</u> statistical effect in the recidivism parameters.

C. Differences Between Cohorts

One of the most obvious covariates is, of course, the one distinguishing among the cohorts (or "treatments"): which state sample the individual is from. The parameter estimates are given in Table 6.

As can be seen, the Iowa cohort has a low probability of recidivism (γ), compared to the other cohorts. The fairly wide range in γ is probably more reflective of the variation in laws and regulations governing sentencing and correctional alternatives than of variation in criminal behavior. For example, it may be that Iowa makes less use of probation or other alternatives to incarceration than do Georgia or North Carolina. If this is indeed true, then low-risk offenders who would not be in the Georgia or North Carolina cohorts <u>would</u> be in the Iowa cohort, thus "improving" the Iowa statistics.*

Another finding, concerning the variable q, is noted from Table 6. The North Carolina cohort has a significantly lower value of q (i.e., its eventual failures do so more rapidly than those in the other cohorts). Again, it is not immediately clear why this should be the case. Discrepancies of this sort make state-to-state comparisons less than useful. With different researchers having different research goals, evaluating different programs in different states, one should not expect any degree of comparability.

One comparison that can be made is the effect of two different definitions of recidivism rearrest and reconviction, since the same cohort (North Carolina) was employed in both cases. As previously mentioned, the dominant effect is a difference of about 6 percent in

*This point is sheer conjecture on our part and is contradicted to some extent by Figure 3r. Were our conjecture true we would normally expect more variation in recidivism probability (γ) as a function of parole risk than is seen in that figure.

converting arrests (γ=.87) to convictions (γ=.81). In some cases there was also a shift in q because of the different definition, but this was not as consistent as the shift in γ.

REFERENCES

(1) Cox, George H. "The Relative Impact of Georgia's Institutional Training Programs on the Post Release Behavior of Adult Male Offenders," Program Evaluation Section, Evaluation and Monitoring Services, Georgia Department of Offender Rehabilitation (July 25, 1977).

(2) "Crime and Criminal Justice in Iowa", Volume VII Recidivism, State of Iowa Statistical Aanlysis Center, Office for Planning and Programming, May, 1979.

(3) Kalbfleisch, J. and Prentice, R.L., "The Statistical Analysis of Failure Time Data," Wiley Interscience (1980).

(4) Kitchener, H., Schmidt, A. and Glaser, D., "How Persistent is Post Prison Success?" Federal Probation, Volume 41, No. 1, pp. 9-15 (1977).

(5) Maltz, Michael D. "On Recidivism," Center for Research in Criminal Justice, University of Illinois Chicago Circle, January, 1981.

(6) Schmidt, P. and Witte, A. "Models of Criminal Recidivism and an Illustration of their Use in Evaluating Correctional Programs," Economics Department, University of North Carolina, Chapel Hill (undated).

Table 1a
Georgia Covariate Partitions

RACE
- A. White
- B. Black

SOCIO-ECONOMIC STATUS
- A. Welfare
- B. Occasionally employed
- C. Minimum Standard
- D. Middle Class
- E. Other or Unknown

HOME ENVIRONMENT TO AGE L¢
- A. Rural Farm
- B. Rural Non-farm
- C. SMSA City
- D. Urban town
- E. Small town

MARITAL STATUS
- A. Never married
- B. Married
- C. Separated
- D. Divorced
- E. Widowed or common-law wife

EMPLOYMENT STATUS PRIOR TO ARREST
- A. Full time
- B. Unknown
- C. Other

PRIOR ARRESTS
- A. None
- B. One or more

AGE AT RELEASE
- A. 17.8-22.3 years
- B. 22.3-23.6 years
- C. 23.6-24.6 years
- D. 24.6-25.6 years
- E. 25.6-26.9 years
- F. 26.9-28.7 years
- G. 28.7-31.3 years
- H. 31.3-35.1 years
- I. 35.1-41.2 years
- J. 41.2 and above

TABLE 1B: GEORGIA COVARIATE PARTITIONS

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
				γ	σ_γ	δ	σ_δ	ρ	γ	σ_γ	δ	σ_δ	ρ
ALL INDIVIDUALS													
	1902	479	5733.	1.000	*	0.985	*	*	0.867	0.100	0.981	0.003	*
RACE													
A	825	195	2494.	1.000	*	0.986	*	*	0.851	0.117	0.983	0.003	0.919
B	1077	284	3239.	0.833	0.215	0.979	0.007	0.982	0.796	0.121	0.977	0.005	0.935
SOCIO-ECONOMIC STATUS													
A	71	23	262.	1.000	*	0.977	*	*	0.739	0.159	0.962	0.014	0.750
B	11	3	54.	1.000	*	0.982	*	*	0.637	0.225	0.960	0.025	0.451
C	1195	302	3454.	0.743	0.175	0.978	0.007	0.979	0.749	0.128	0.977	0.005	0.944
D	463	103	1371.	1.000	*	0.987	*	*	0.835	0.131	0.983	0.004	0.878
E	162	48	592.	1.000	*	0.980	*	*	0.775	0.142	0.970	0.009	0.816
HOME ENVIRONMENT TO AGE 16													
A	245	40	527.	1.000	*	0.991	*	*	0.674	0.196	0.983	0.008	0.830
B	60	11	120.	0.418	0.346	0.968	0.037	0.947	0.465	0.213	0.959	0.026	0.719
C	676	215	2459.	1.000	*	0.979	*	*	0.861	0.101	0.974	0.005	0.894
D	342	82	1035.	1.000	*	0.986	*	*	0.769	0.153	0.980	0.006	0.899
E	579	131	1592.	0.964	0.586	0.986	0.010	0.993	0.753	0.152	0.980	0.006	0.903
MARITAL STATUS													
A	656	184	2270.	1.000	*	0.983	*	*	0.853	0.112	0.979	0.004	0.878
B	690	163	1997.	1.000	*	0.986	*	*	0.828	0.127	0.981	0.004	0.905
C	173	33	362.	0.549	0.373	0.976	0.021	0.975	0.558	0.199	0.971	0.015	0.823
D	187	42	443.	0.525	0.215	0.968	0.019	0.947	0.573	0.179	0.966	0.015	0.832
E	139	34	312.	0.433	0.118	0.948	0.022	0.847	0.495	0.156	0.949	0.021	0.788

CONTINUED

2 OF 4

TABLE 1B: GEORGIA COVARIATE PARTITIONS (CONT.)

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
				γ	σ_γ	δ	σ_δ	ς	γ	σ_γ	δ	σ_δ	ς
EMPLOYMENT													
A	1159	263	3242.	1.000	*	0.987	*	*	0.849	0.117	0.983	0.003	0.901
B	283	79	941.	0.999	0.583	0.983	0.013	0.988	0.771	0.144	0.974	0.008	0.867
C	460	137	1550.	0.777	0.195	0.973	0.010	0.964	0.765	0.126	0.970	0.007	0.891
PRIOR ARRESTS													
A	514	116	1544.	1.000	*	0.988	*	*	0.772	0.153	0.983	0.005	0.933
B	1388	363	4189.	1.000	*	0.983	*	*	0.865	0.101	0.979	0.004	0.908
AGE AT RELEASE													
A	231	89	1226.	1.000	*	0.981	*	*	0.845	0.116	0.975	0.006	0.842
B	210	67	796.	1.000	*	0.981	*	*	0.776	0.143	0.972	0.008	0.864
C	189	47	579.	1.000	*	0.984	*	*	0.766	0.157	0.976	0.008	0.823
D	181	47	517.	0.740	0.395	0.975	0.017	0.975	0.678	0.168	0.969	0.012	0.844
E	168	36	391.	0.723	0.575	0.980	0.020	0.984	0.625	0.190	0.971	0.013	0.833
F	179	30	387.	1.000	*	0.989	*	*	0.695	0.191	0.980	0.009	0.825
G	187	40	424.	0.522	0.240	0.970	0.019	0.956	0.567	0.185	0.967	0.015	0.838
H	184	39	418.	0.566	0.305	0.973	0.019	0.967	0.585	0.188	0.969	0.014	0.832
I	178	47	616.	1.000	*	0.984	*	*	0.782	0.150	0.977	0.007	0.823
J	195	37	379.	0.381	0.124	0.960	0.019	0.897	0.458	0.172	0.961	0.018	0.805

Table 2a.

U.S. Bureau of Prisons Covariate Partitions

RACE
- A. White
- B. Other

PROBATION OR SUPERVISION VIOLATIONS
- A. None
- B. One or more

AGE AT FIRST ARREST
- A. 16 or less
- B. 17, 18
- C. 19, 20
- D. 21-23
- E. 24-27
- F. 28-34
- G. 35 and over

NUMBER OF PRIOR FELONY SENTENCES
- A. None
- B. One or more

PRIMARY CURRENT OFFENSE
- A. Vehicle theft for interstate transportation
- B. Fraudulent check, counterfeit, tax fraud, embezzlement
- C. Moonshine
- D. Other

ETIOLOGY OF LAST PATTERN OF CRIMINALITY
- A. Delinquent or criminal orientation (but not narcotic or alcohol related)
- B. Financial straits
- C. Other

EMPLOYMENT DURING LAST 2 YEARS PRECEDING LAST IMPRISONMENT
- A. Less than 25% time
- B. 26-50% time
- C. 51-75% time
- D. 76-100% time
- E. student or unemployed

ALCOHOL PROBLEMS
- A. None
- B. Other

PSYCHOLOGICAL DIAGNOSIS
- A. None or favorable
- B. Other

Table 2a. U.S. Bureau of Prisons Covariate Partitions (continued)

SCHOOLING COMPLETED AT RELEASE
A. 8th grade or below
B. 9th grade or above

IQ
A. Less than or equal to 100
B. Greater than 100

ANTICIPATED HOME AT RELEASE
A. Plans to live alone
B. With wife (or common law wife)
C. With parents
D. Other

AGE AT RELEASE
A. 19 or less
B. 20, 21
C. 22, 23
D. 24-26
E. 27-30
F. 31-35
G. 36-46
H. 47 and over.

TABLE 2B: U.S. BUREAU OF PRISONS COVARIATE PARTITIONS

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
				$\hat{\gamma}$	σ_γ	$\hat{q}$	σ_q	ρ	$\tilde{\gamma}$	σ_γ	$\tilde{q}$	σ_q	ρ
ALL INDIVIDUALS													
	927	661	7544.	0.715	0.015	0.913	0.003	0.022	0.704	0.014	0.913	0.004	0.079
RACE													
A	628	430	5003.	0.687	0.019	0.915	0.004	0.024	0.692	0.018	0.915	0.004	0.014
B	299	231	2540.	0.773	0.024	0.910	0.006	0.017	0.772	0.026	0.909	0.006	0.029
PROBATIONS OR SUPERVISION VIOLATION													
A	173	139	1407.	0.804	0.030	0.902	0.008	0.010	0.800	0.030	0.901	0.008	0.019
B	754	522	6136.	0.695	0.017	0.916	0.004	0.025	0.698	0.010	0.916	0.003	0.212
AGE AT FIRST ARREST													
A	284	234	2478.	0.825	0.023	0.906	0.006	0.016	0.823	0.026	0.906	0.006	0.033
B	159	123	1242.	0.774	0.033	0.901	0.009	0.011	0.771	0.033	0.901	0.009	0.015
C	132	91	1024.	0.695	0.040	0.912	0.009	0.020	0.692	0.040	0.911	0.009	0.037
D	147	105	1215.	0.715	0.037	0.914	0.008	0.021	0.713	0.037	0.914	0.008	0.023
E	94	58	800.	0.623	0.051	0.930	0.010	0.062	0.621	0.050	0.929	0.010	0.073
F	61	35	561.	0.581	0.064	0.941	0.011	0.094	0.581	0.064	0.940	0.011	0.123
G	50	15	225.	0.303	0.066	0.936	0.018	0.059	0.314	0.066	0.934	0.019	0.114
NUMBER OF PRIOR FELONY SENTENCES													
A	385	226	2522.	0.591	0.025	0.924	0.005	0.037	0.592	0.024	0.924	0.006	0.039
B	542	435	4622.	0.803	0.017	0.906	0.004	0.015	0.800	0.009	0.906	0.004	-0.032
PRIMARY CURRENT OFFENSE													
A	323	271	2234.	0.774	0.022	0.879	0.007	0.003	0.772	0.026	0.879	0.007	0.015
B	90	52	679.	0.580	0.052	0.925	0.011	0.034	0.579	0.052	0.924	0.011	0.042
C	105	60	961.	0.579	0.049	0.941	0.009	0.099	0.579	0.049	0.940	0.008	0.110
D	330	238	2971.	0.723	0.025	0.921	0.005	0.034	0.722	0.027	0.921	0.005	0.021
ETIOLOGY OF LAST PATTERN OF CRIMINALITY													
A	536	401	3875.	0.750	0.012	0.897	0.005	0.007	0.749	0.012	0.897	0.005	0.045
B	67	43	894.	0.676	0.056	0.959	0.009	0.352	0.681	0.070	0.959	0.009	0.409
C	324	217	2775.	0.673	0.026	0.924	0.005	0.041	0.672	0.028	0.923	0.005	0.017

TABLE 2B: U.S. BUREAU OF PRISONS COVARIATE PARTITIONS (CONT.)

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES γ	σ_γ	g	σ_g	ρ	BAYES ESTIMATES γ	σ_γ	g	σ_g	ρ
EMPLOYMENT DURING 2 YEARS PRECEDING LAST IMPRISONMENT													
A	244	194	1922.	0.796	0.026	0.900	0.007	0.011	0.793	0.024	0.899	0.007	0.013
B	172	136	1421.	0.791	0.031	0.905	0.008	0.013	0.788	0.031	0.904	0.008	0.042
C	170	125	1372.	0.736	0.034	0.910	0.008	0.025	0.735	0.034	0.910	0.008	0.041
D	242	135	1944.	0.564	0.032	0.933	0.006	0.057	0.563	0.032	0.933	0.006	0.063
E	99	71	865.	0.719	0.045	0.921	0.009	0.031	0.715	0.045	0.920	0.009	0.033
ALCOHOL PROBLEMS													
A	631	430	5105.	0.684	0.019	0.917	0.004	0.025	0.689	0.021	0.917	0.002	-0.099
B	296	231	2438.	0.781	0.024	0.906	0.006	0.016	0.781	0.026	0.906	0.007	0.071
PSYCHOLOGICAL DIAGNOSIS													
A	567	391	4693.	0.693	0.019	0.916	0.004	0.028	0.696	0.016	0.918	0.004	0.130
B	360	270	2851.	0.751	0.023	0.906	0.006	0.014	0.749	0.020	0.906	0.005	0.031
SCHOOLING COMPLETED AT RELEASE													
A	531	373	4266.	0.704	0.020	0.913	0.004	0.019	0.701	0.014	0.913	0.005	0.147
B	396	288	3277.	0.731	0.022	0.913	0.005	0.025	0.731	0.025	0.913	0.004	-0.019
I.Q.													
A	716	515	5890.	0.722	0.017	0.914	0.004	0.022	0.717	0.023	0.913	0.004	0.038
B	211	146	1054.	0.693	0.032	0.913	0.007	0.021	0.692	0.031	0.912	0.007	0.043
ANTICIPATED HOME AT RELEASE													
A	109	88	967.	0.809	0.038	0.910	0.009	0.021	0.803	0.038	0.909	0.009	0.010
B	233	137	1763.	0.592	0.032	0.924	0.007	0.035	0.592	0.032	0.924	0.007	0.040
C	330	239	2666.	0.726	0.025	0.911	0.006	0.020	0.725	0.027	0.911	0.006	0.039
D	255	197	2147.	0.773	0.026	0.909	0.006	0.017	0.772	0.027	0.909	0.006	0.006

Table 3a

Iowa Covariate Partitions

SEX
- A. Female
- B. Male

RACE
- A. Non-white
- B. White

MARITAL STATUS
- A. Widowed or separated
- B. Single
- C. Married or common-law
- D. Divorced

LIVING ARRANGEMENT ON RELEASE
- A. With relatives, foster parents, institution, other
- B. Alone
- C. Spouse and/or children
- D. Parents or step-parents

EDUCATIONAL ATTAINMENT
- A. 13 years or more
- B. 8 years or less
- C. 9 - 12 years

TYPE OF ADMISSION
- A. Direct CRT committment
- B. Probation revocation
- C. Parole revocation - NOA
- D. Parole violation
- E. Safekeeping or Evaluation
- F. Other

ALCOHOL INVOLVEMENT
- A. None
- B. Under intoxication at arrest
- C. History of alcoholism

DRUG INVOLVEMENT
- A. None
- B. Some

JUVENILE COMMITTMENTS
- A. One or more
- B. None

Table 3a

Iowa Covariate Partitions - continued

PRIOR PRISON RECORD
- A. None
- B. One or more

TYPE OF RELEASE
- A. Expiration of sentence
- B. Parole
- C. Safekeeping or evaluation
- D. Other

AGE AT RELEASE
- A. 19 or less
- B. 20-21
- C. 22-23
- D. 24-26
- E. 27-29
- F. 30-35
- G. 36-46
- H. 47 and over

OCCUPATION AT ADMISSION
- A. None or unskilled
- B. Skilled or higher

PRIOR JUVENILE ARRESTS
- A. None
- B. One or more

PRIOR ARRESTS
- A. None
- B. One or more

PRIOR FELONY CONVICTIONS
- A. None
- B. One
- C. Two or more

PAROLE RISK SCALE AT ADMISSION
- A. Ultra High
- B. High
- C. Medium
- D. Low
- E. Nil.

TABLE 3B: IOWA COVARIATE PARTITIONS

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES γ	σ_γ	δ	σ_δ	ρ	BAYES ESTIMATES γ	σ_γ	δ	σ_δ	ρ
ALL INDIVIDUALS													
	3372	838	12477.	0.357	0.018	0.964	0.003	0.797	0.354	0.014	0.964	0.003	0.939
SEX													
A	183	50	799.	0.427	0.099	0.970	0.011	0.857	0.482	0.140	0.971	0.011	0.809
B	3189	788	11678.	0.353	0.018	0.964	0.003	0.792	0.353	0.012	0.964	0.002	0.745
RACE													
A	599	197	3025.	0.492	0.053	0.967	0.006	0.843	0.506	0.062	0.968	0.006	0.866
B	2772	641	9452.	0.328	0.018	0.963	0.003	0.782	0.333	0.024	0.964	[illegible].004	0.849
MARITAL STATUS													
A	230	46	678.	0.298	0.068	0.966	0.013	0.817	0.350	0.126	0.968	0.013	0.765
B	1499	356	5072.	0.319	0.021	0.959	0.005	0.720	0.319	0.025	0.959	0.005	0.801
C	1049	272	4360.	0.441	0.053	0.974	0.005	0.904	0.458	0.066	0.974	0.005	0.892
D	581	161	2314.	0.369	0.035	0.959	0.007	0.717	0.376	0.039	0.959	0.007	0.709
LIVING ARRANGEMENT ON RELEASE													
A	678	150	2200.	0.309	0.034	0.962	0.007	0.763	0.318	0.039	0.963	0.007	0.785
B	958	270	4025.	0.403	0.034	0.964	0.005	0.801	0.410	0.037	0.965	0.005	0.810
C	917	206	3265.	0.380	0.052	0.973	0.006	0.898	0.400	0.070	0.974	0.006	0.869
D	800	207	2908.	0.340	0.028	0.957	0.006	0.688	0.345	0.030	0.957	0.006	0.742
EDUCATIONAL ATTAINMENT													
A	205	32	515.	0.274	0.103	0.975	0.014	0.903	0.370	0.186	0.976	0.013	0.766
B	675	174	2672.	0.357	0.036	0.964	0.006	0.769	0.366	0.041	0.964	0.006	0.775
C	2422	625	9155.	0.368	0.021	0.964	0.003	0.791	0.367	0.024	0.963	0.004	0.734
TYPE OF ADMISSION													
A	1519	464	6860.	0.431	0.027	0.963	0.004	0.788	0.435	0.030	0.963	0.004	0.868
B	526	187	2583.	0.535	0.057	0.964	0.006	0.843	0.549	0.066	0.964	0.006	0.855
C	120	44	589.	0.466	0.074	0.953	0.013	0.664	0.492	0.096	0.953	0.014	0.702
D	321	109	1684.	0.501	0.070	0.967	0.008	0.835	0.527	0.089	0.968	0.008	0.860
E	774	25	663.	1.000	*	0.999	*	*	0.448	0.264	0.996	0.003	0.669
F	111	8	87.	0.078	0.027	0.924	0.036	0.250	0.107	0.092	0.924	0.039	0.441

TABLE 3B: IOWA COVARIATE PARTITIONS (CONT.)

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
				γ	σ_γ	δ	σ_δ	ρ	γ	σ_γ	δ	σ_δ	ρ
ALCOHOL INVOLVEMENT													
A	1174	245	3968.	0.354	0.046	0.974	0.005	0.900	0.370	0.059	0.974	0.005	0.870
B	865	200	2702.	0.295	0.024	0.954	0.006	0.648	0.301	0.023	0.954	0.006	0.627
C	1260	386	5644.	0.438	0.030	0.963	0.004	0.795	0.442	0.033	0.964	0.004	0.810
DRUG INVOLVEMENT													
A	1799	452	7031.	0.373	0.027	0.968	0.004	0.828	0.376	0.030	0.968	0.003	*
B	1501	381	5305.	0.354	0.024	0.960	0.005	0.766	0.358	0.023	0.961	0.004	0.740
JUVENILE COMMITTMENTS													
A	987	310	4304.	0.409	0.027	0.956	0.005	0.699	0.412	0.027	0.956	0.005	0.666
B	2370	525	8129.	0.341	0.024	0.969	0.004	0.847	0.348	0.025	0.969	0.004	0.847
PRIOR PRISON RECORD													
A	2329	512	7594.	0.317	0.020	0.964	0.004	0.794	0.315	0.024	0.964	0.003	*
B	1036	325	4873.	0.444	0.034	0.964	0.005	0.799	0.450	0.036	0.964	0.005	0.737
TYPE OF RELEASE													
A	770	251	3762.	0.471	0.041	0.965	0.005	0.810	0.479	0.045	0.965	0.005	0.804
B	1575	568	8151.	0.506	0.027	0.962	0.004	0.788	0.509	0.027	0.962	0.004	0.652
C	16	16	520.	1.000	*	0.999	*	*	0.423	0.270	0.997	0.003	0.682
D	309	3	44.	0.014	0.011	0.963	0.050	0.736	0.240	0.250	0.995	0.007	0.449

TABLE 3B: IOWA COVARIATE PARTITIONS (CONT.)

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
				γ	σ_γ	δ	σ_δ	ρ	γ	σ_γ	δ	σ_δ	ρ
AGE AT RELEASE													
A	305	36	437.	0.137	0.024	0.940	0.016	0.444	0.148	0.033	0.941	0.016	0.508
B	469	115	1676.	0.353	0.046	0.964	0.008	0.792	0.369	0.060	0.964	0.008	0.800
C	480	127	1852.	0.388	0.050	0.965	0.008	0.815	0.405	0.064	0.965	0.008	0.817
D	665	177	2590.	0.364	0.035	0.961	0.006	0.748	0.372	0.039	0.962	0.006	0.770
E	501	127	1883.	0.365	0.046	0.964	0.007	0.796	0.379	0.057	0.965	0.008	0.801
F	338	101	1522.	0.449	0.069	0.967	0.008	0.844	0.478	0.094	0.968	0.009	0.817
G	404	101	1464.	0.343	0.044	0.961	0.008	0.751	0.357	0.056	0.962	0.009	0.757
H	210	54	1053.	0.706	0.452	0.988	0.010	0.984	0.644	0.180	0.984	0.007	0.849
OCCUPATION AT ADMISSION													
A	1438	462	6491.	0.436	0.025	0.959	0.004	0.743	0.440	0.027	0.959	0.004	0.742
B	962	314	4722.	0.490	0.042	0.967	0.005	0.843	0.498	0.046	0.967	0.005	0.823
PRIOR JUVENILE ARRESTS													
A	1439	399	5925.	0.410	0.030	0.965	0.004	0.818	0.415	0.031	0.966	0.004	0.787
B	987	386	5374.	0.517	0.030	0.957	0.004	0.732	0.519	0.031	0.957	0.005	0.670
PRIOR ARRESTS													
A	986	243	3410.	0.343	0.029	0.960	0.006	0.756	0.349	0.031	0.961	0.006	0.758
B	1435	539	7860.	0.522	0.028	0.962	0.004	0.786	0.524	0.030	0.962	0.004	0.693
PRIOR FELONY CONVICTIONS													
A	1350	384	5729.	0.431	0.034	0.967	0.004	0.840	0.436	0.037	0.967	0.004	0.857
B	531	201	2746.	0.478	0.035	0.953	0.006	0.657	0.482	0.037	0.953	0.006	0.650
C	542	198	2804.	0.495	0.042	0.959	0.006	0.753	0.502	0.046	0.959	0.006	0.741
PAROLE RISK SCALE AT ADMISSION													
A	89	46	480.	0.585	0.066	0.926	0.015	0.430	0.591	0.068	0.925	0.016	0.468
B	554	236	3076.	0.562	0.040	0.954	0.006	0.734	0.568	0.042	0.954	0.006	0.735
C	519	164	2442.	0.461	0.051	0.965	0.007	0.820	0.475	0.061	0.966	0.007	0.812
D	356	76	1298.	0.442	0.145	0.981	0.009	0.952	0.514	0.172	0.981	0.008	0.852
E	208	23	409.	0.394	0.543	0.990	0.017	0.990	0.423	0.234	0.985	0.011	0.745

Table 4a

North Carolina Covariate Partitions

RACE
- A. White
- B. Black or Indian

TYPE OF RELEASE
- A. Other
- B. Unconditional

IQ
- A. 100 or less
- B. Greater than 100

SCHOOL ACHIEVEMENT
- A. 8 years or less
- B. 9 years or more

WORK STABILITY 5 YEARS PRIOR TO SAMPLE TERM
- A. Other
- B. Two or fewer job changes
- C. Student

PRIOR ARRESTS
- A. None
- B. One or two
- C. Three or more

MARITAL STATUS AT INTERVIEW
- A. Single
- B. Married
- C. Divorced
- D. Other

EMPLOYMENT STATUS AT INTERVIEW
- A. No Job
- B. Other

DRINKING PROBLEM
- A. None
- B. Some

DRUG USE
- A. None
- B. Some

AGE AT RELEASE
- A. twenty or less
- B. 21, 22
- C. 22-24.5
- D. 24.5-28
- E. 28-34
- F. 34-40
- G. 40-47
- H. Greater than 47.

TABLE 4B: NORTH CAROLINA COVARIATE PARTITIONS - RE-ARREST

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
				γ	σ_γ	δ	σ_δ	ρ	γ	σ_γ	δ	σ_δ	ρ
ALL INDIVIDUALS	641	513	5649.	0.874	0.018	0.926	0.004	0.444	0.871	0.025	0.925	0.002	*
RACE													
A	315	258	2434.	0.869	0.023	0.908	0.007	0.332	0.866	0.026	0.908	0.007	0.360
B	326	255	3215.	0.884	0.028	0.939	0.005	0.543	0.883	0.029	0.939	0.005	0.557
TYPE OF RELEASE													
A	192	136	1764.	0.832	0.044	0.943	0.007	0.572	0.831	0.044	0.943	0.007	0.561
B	449	377	3885.	0.897	0.019	0.918	0.005	0.411	0.897	0.014	0.918	0.005	0.245
I.Q.													
A	114	95	946.	0.879	0.039	0.913	0.011	0.383	0.874	0.039	0.913	0.011	0.377
B	242	194	2114.	0.924	0.034	0.932	0.007	0.631	0.922	0.034	0.931	0.007	0.649
SCHOOL ACHIEVEMENT													
A	56	47	476.	0.876	0.052	0.912	0.015	0.298	0.866	0.053	0.911	0.015	0.310
B	285	231	2471.	0.919	0.030	0.929	0.007	0.608	0.917	0.030	0.928	0.007	0.604
WORK STABILITY													
A	361	296	3029.	0.870	0.022	0.915	0.006	0.324	0.867	0.024	0.915	0.006	0.353
B	173	132	1412.	0.852	0.040	0.927	0.009	0.510	0.849	0.040	0.926	0.009	0.513
C	36	33	486.	1.000	*	0.941	*	*	0.979	0.036	0.939	0.010	0.109
PRIOR ARRESTS													
A	190	136	1809.	0.855	0.048	0.947	0.007	0.656	0.855	0.048	0.947	0.007	0.646
B	217	183	1899.	0.904	0.027	0.919	0.007	0.410	0.900	0.026	0.918	0.007	0.399
C	234	194	1941.	0.881	0.026	0.913	0.007	0.333	0.879	0.027	0.913	0.008	0.327

TABLE 4B: NORTH CAROLINA COVARIATE PARTITIONS - RE-ARREST (CONT.)

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
				γ	σ_γ	δ	σ_δ	ζ	γ	σ_γ	δ	σ_δ	ζ
MARITAL STATUS													
A	125	101	1043.	0.903	0.044	0.924	0.011	0.573	0.899	0.044	0.923	0.010	0.560
B	195	158	1679.	0.882	0.033	0.923	0.008	0.459	0.879	0.033	0.923	0.008	0.454
C	138	115	1256.	0.917	0.037	0.926	0.009	0.501	0.911	0.036	0.925	0.009	0.475
D	11	7	80.	0.665	0.158	0.924	0.037	0.291	0.662	0.148	0.917	0.036	0.302
EMPLOYMENT STATUS AT INTERVIEW													
A	387	307	3363.	0.875	0.025	0.927	0.006	0.506	0.874	0.027	0.927	0.006	0.563
B	100	86	843.	0.967	0.043	0.920	0.012	0.644	0.954	0.039	0.918	0.011	0.528
DRINKING PROBLEM													
A	333	257	3137.	0.864	0.029	0.936	0.005	0.528	0.862	0.029	0.936	0.006	0.514
B	275	232	2206.	0.893	0.023	0.908	0.007	0.334	0.892	0.021	0.908	0.007	0.300
DRUG USE													
A	576	460	4990.	0.864	0.019	0.923	0.005	0.418	0.857	0.019	0.923	0.005	0.510
B	32	29	353.	1.000	*	0.929	*	*	0.974	0.041	0.926	0.013	0.137
AGE AT RELEASE													
A	75	64	759.	0.950	0.044	0.933	0.011	0.521	0.935	0.043	0.931	0.010	0.466
B	75	58	635.	0.881	0.061	0.930	0.013	0.558	0.872	0.059	0.929	0.013	0.529
C	78	67	730.	0.963	0.040	0.927	0.011	0.497	0.945	0.040	0.925	0.011	0.467
D	81	68	700.	0.886	0.042	0.914	0.012	0.272	0.876	0.042	0.913	0.012	0.265
E	83	71	821.	0.909	0.041	0.925	0.011	0.364	0.900	0.042	0.925	0.011	0.361
F	90	74	745.	0.878	0.044	0.915	0.012	0.368	0.870	0.044	0.914	0.012	0.353
G	81	59	620.	0.784	0.056	0.920	0.013	0.384	0.782	0.057	0.920	0.013	0.406
H	78	52	639.	0.756	0.070	0.939	0.013	0.545	0.760	0.073	0.938	0.013	0.568

TABLE 5B: NORTH CAROLINA COVARIATE PARTITIONS - RE-CONVICTION

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
				γ	σ_γ	g	σ_g	ρ	γ	σ_γ	g	σ_g	ρ
ALL INDIVIDUALS													
	641	465	5383.	0.812	0.022	0.932	0.004	0.492	0.809	0.021	0.932	0.004	0.563
RACE													
A	315	243	2464.	0.833	0.027	0.918	0.007	0.396	0.832	0.028	0.917	0.006	0.431
B	326	222	2919.	0.802	0.036	0.945	0.006	0.598	0.802	0.036	0.945	0.005	0.595
TYPE OF RELEASE													
A	192	120	1536.	0.741	0.048	0.944	0.008	0.563	0.743	0.049	0.943	0.008	0.583
B	449	345	3847.	0.844	0.024	0.928	0.005	0.481	0.845	0.022	0.928	0.005	0.448
I.Q.													
A	114	85	835.	0.797	0.046	0.914	0.012	0.376	0.794	0.047	0.914	0.012	0.387
B	242	175	2040.	0.853	0.040	0.938	0.007	0.623	0.852	0.041	0.938	0.007	0.649
SCHOOL ACHIEVEMENT													
A	56	41	412.	0.765	0.063	0.912	0.016	0.262	0.760	0.064	0.911	0.017	0.284
B	285	210	2377.	0.856	0.036	0.935	0.007	0.615	0.856	0.036	0.935	0.007	0.627
WORK STABILITY													
A	361	264	2790.	0.791	0.026	0.921	0.006	0.358	0.790	0.026	0.921	0.006	0.352
B	173	125	1465.	0.836	0.047	0.937	0.008	0.605	0.835	0.047	0.937	0.009	0.586
C	36	31	494.	1.000	*	0.953	*	*	0.954	0.052	0.948	0.010	0.312
PRIOR ARRESTS													
A	190	126	1788.	0.846	0.062	0.955	0.007	0.750	0.848	0.061	0.955	0.007	0.704
B	217	168	1787.	0.836	0.032	0.922	0.008	0.397	0.834	0.032	0.921	0.008	0.399
C	234	171	1808.	0.794	0.033	0.922	0.008	0.385	0.793	0.033	0.921	0.008	0.388

TABLE 5B: NORTH CAROLINA COVARIATE PARTITIONS - RE-CONVICTION (CONT.)

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
MARITAL STATUS													
A	125	91	1009.	0.880	0.065	0.938	0.011	0.726	0.877	0.061	0.937	0.010	0.686
B	195	149	1829.	0.878	0.042	0.940	0.007	0.625	0.876	0.042	0.940	0.007	0.629
C	138	105	1189.	0.856	0.045	0.931	0.009	0.517	0.853	0.045	0.931	0.009	0.512
D	11	7	87.	0.681	0.169	0.934	0.035	0.403	0.675	0.152	0.926	0.033	0.340
EMPLOYMENT STATUS AT INTERVIEW													
A	387	279	3265.	0.827	0.031	0.937	0.006	0.590	0.827	0.031	0.936	0.005	0.640
B	100	82	964.	1.000	*	0.942	*	*	0.964	0.040	0.937	0.008	0.523
DRINKING PROBLEM													
A	333	224	2966.	0.803	0.039	0.947	0.006	0.653	0.804	0.039	0.947	0.006	0.655
B	275	220	2216.	0.858	0.027	0.916	0.007	0.365	0.855	0.026	0.915	0.007	0.379
DRUG USE													
A	576	417	4795.	0.805	0.023	0.931	0.005	0.482	0.804	0.020	0.931	0.005	0.432
B	32	27	387.	1.000	*	0.947	*	*	0.953	0.056	0.943	0.012	0.275
AGE AT RELEASE													
A	75	62	807.	0.951	0.055	0.943	0.010	0.632	0.933	0.049	0.940	0.010	0.517
B	75	53	599.	0.811	0.069	0.934	0.013	0.559	0.809	0.069	0.933	0.013	0.552
C	78	59	694.	0.899	0.069	0.940	0.012	0.652	0.886	0.063	0.938	0.011	0.573
D	81	53	550.	0.726	0.063	0.923	0.014	0.405	0.725	0.063	0.922	0.014	0.417
E	83	64	806.	0.845	0.054	0.936	0.011	0.440	0.841	0.054	0.935	0.010	0.452
F	90	71	761.	0.846	0.048	0.921	0.011	0.368	0.840	0.047	0.920	0.011	0.361
G	81	55	601.	0.738	0.060	0.925	0.014	0.391	0.738	0.061	0.924	0.014	0.417
H	78	48	565.	0.687	0.068	0.934	0.014	0.467	0.692	0.072	0.933	0.014	0.502

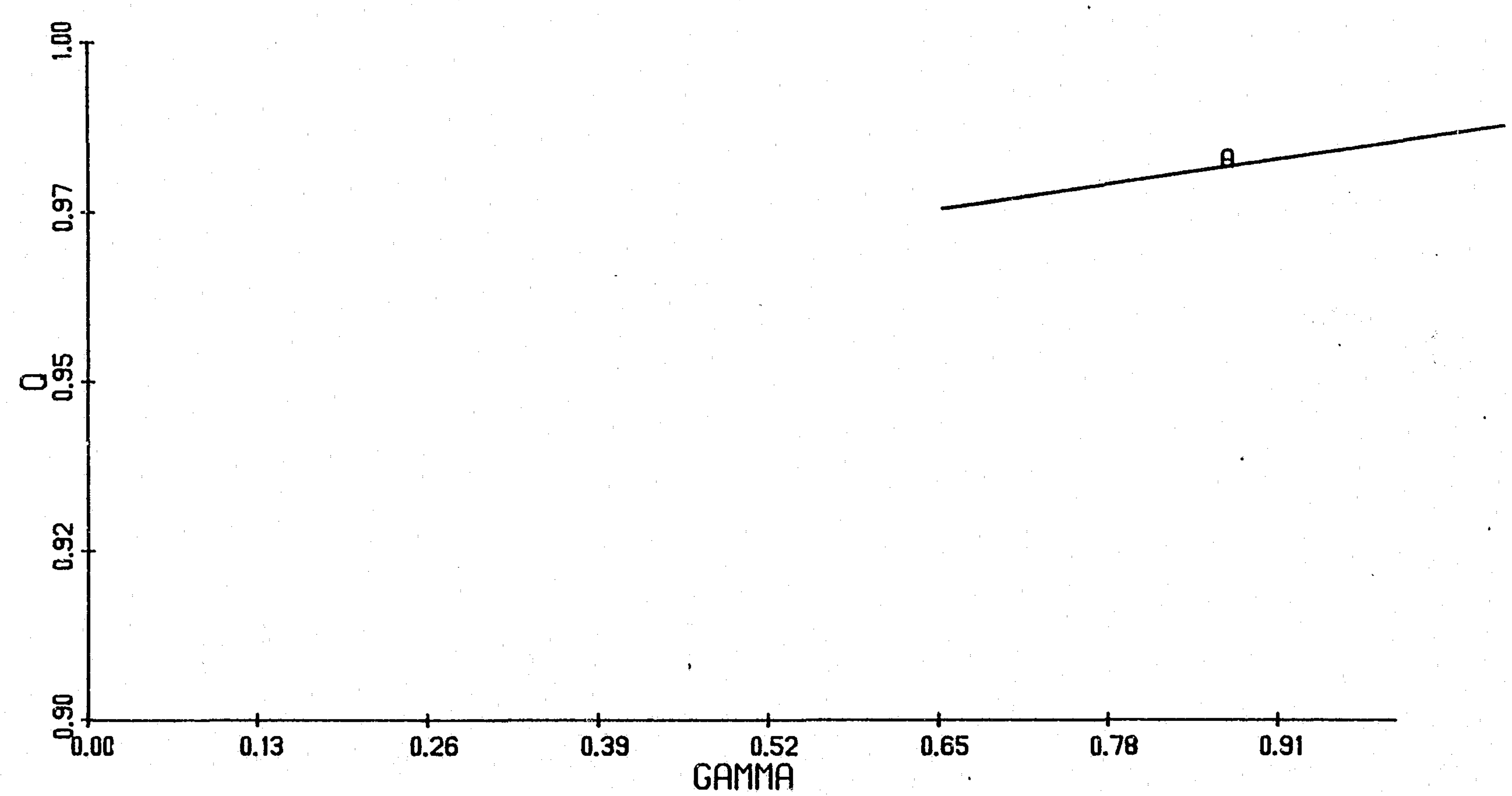


Figure 1a: Georgia Cohort
Bayes 90% Confidence Regions for estimates of γ and q.

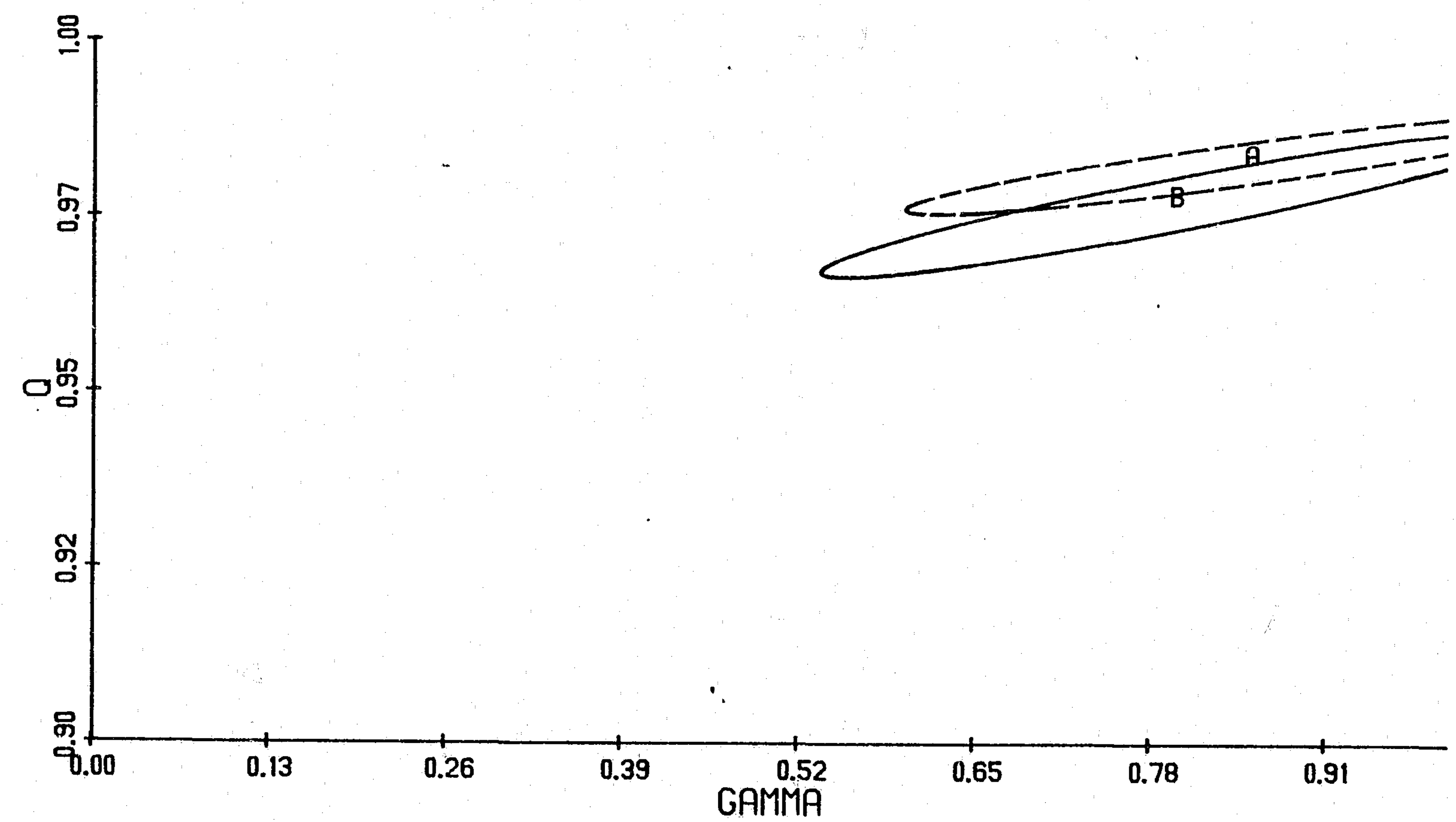


Figure 1b: Georgia Cohort
Bayes 90% Confidence Regions for estimates of γ and q, by RACE
A. White
B. Black

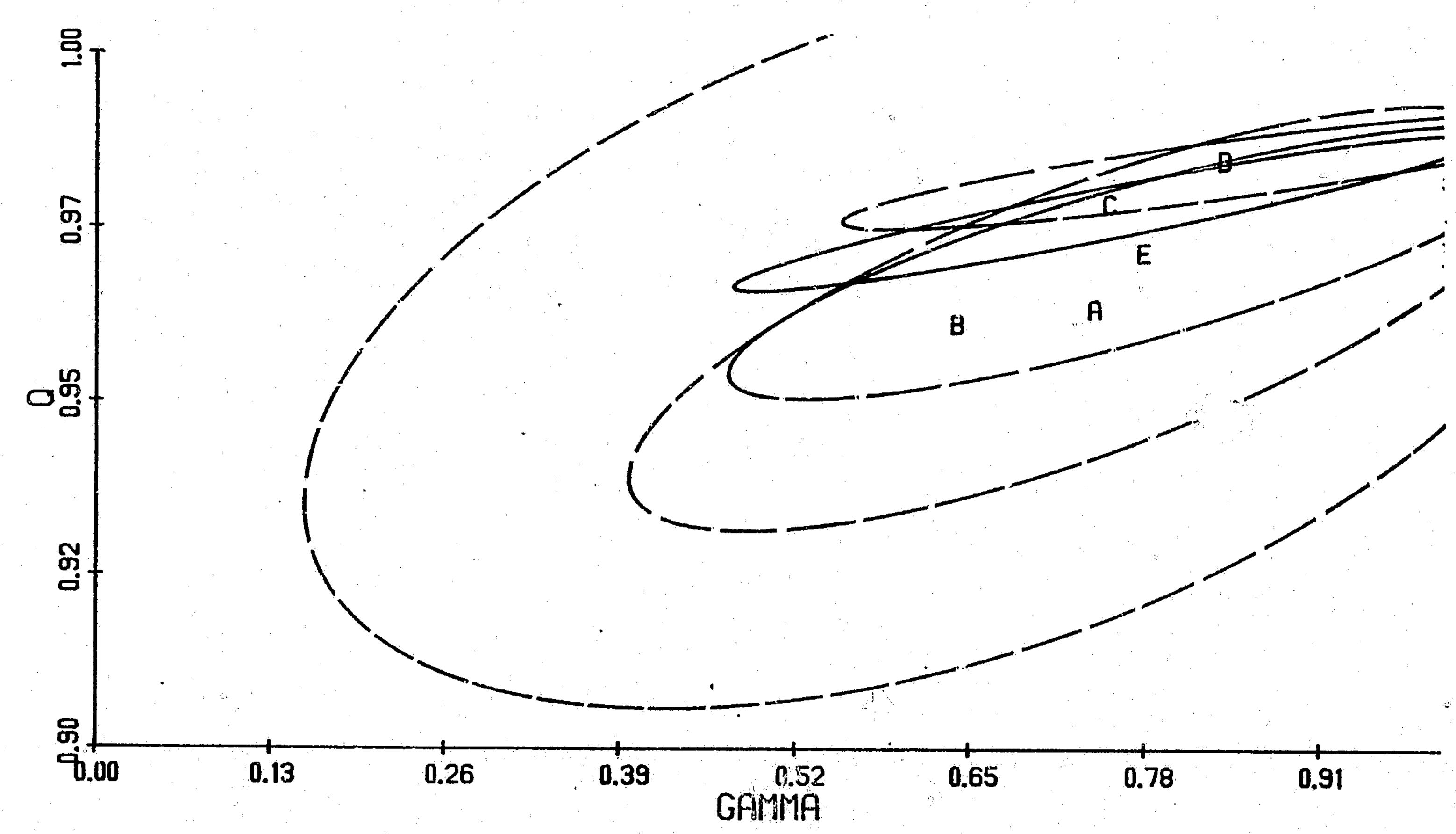


Figure 1c: Georgia Cohort
Bayes 90% Confidence Regions for estimates of γ and q, by SOCIO-ECONOMIC STATUS

A. Welfare
B. Occasionally Employed
C. Minimum Standard
D. Middle Class
E. Other or Unknown

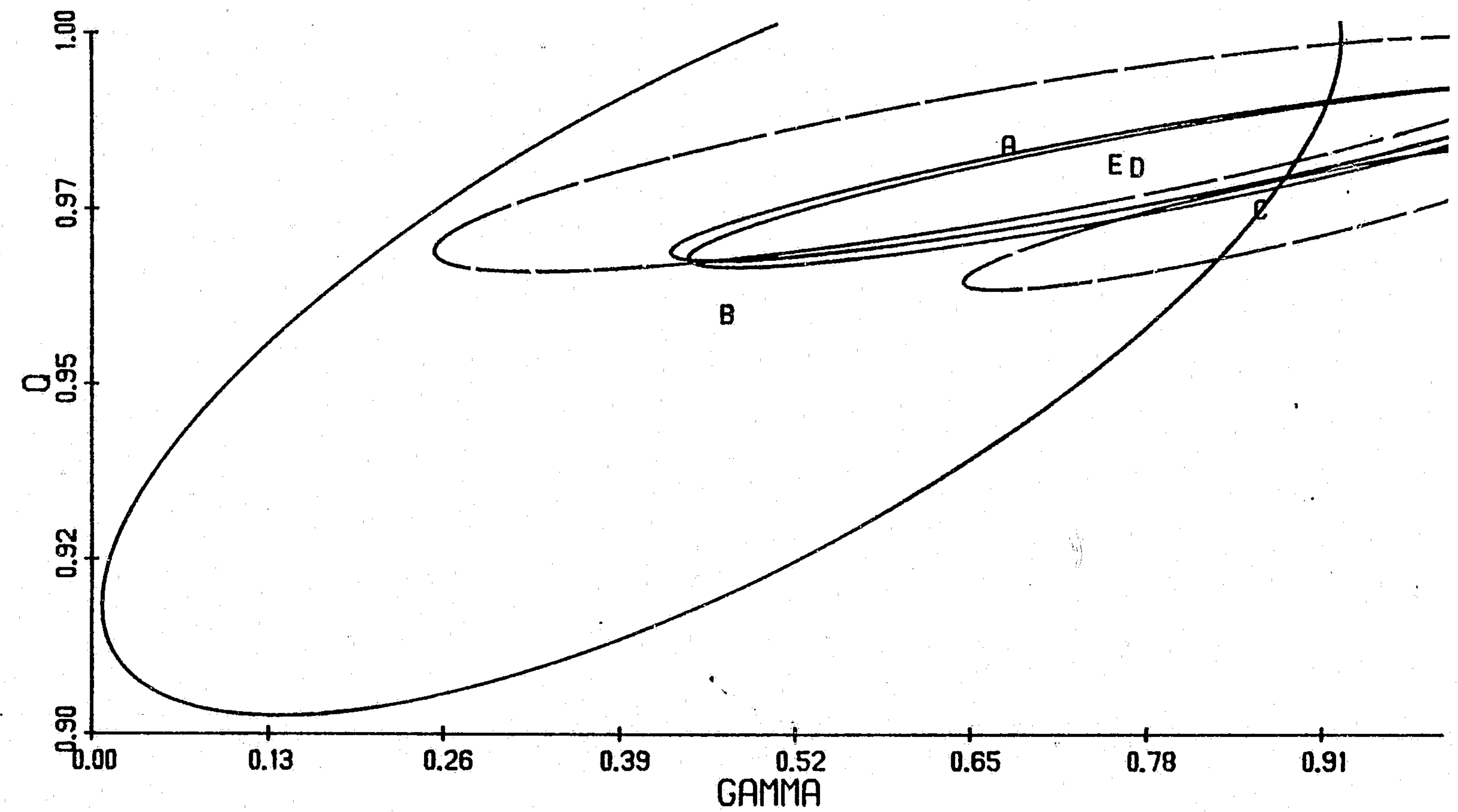


Figure 1d: Georgia Cohort
Bayes 90% Confidence Regions for estimates of γ and q, by HOME ENVIRONMENT TO AGE 16
A. Rural Farm
B. Rural Non-farm
C. SMSA City
D. Urban Town
E. Small Town

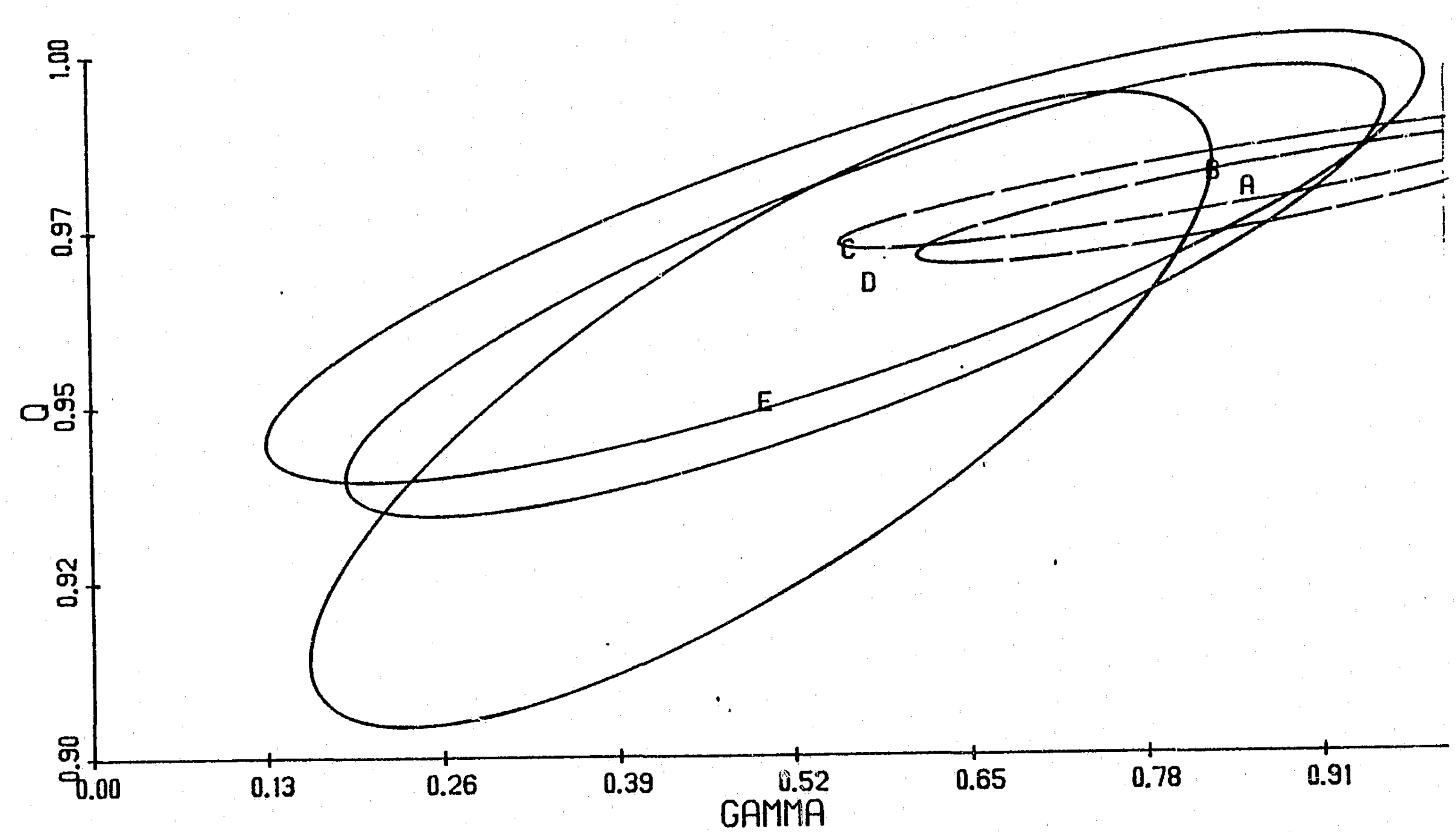


Figure 1e: Georgia Cohort
Bayes 90% Confidence Regions for estimates of γ and q, by MARITAL STATUS

A. Never Married
B. Married
C. Separated
D. Divorced
E. Widowed or Common-law Wife.

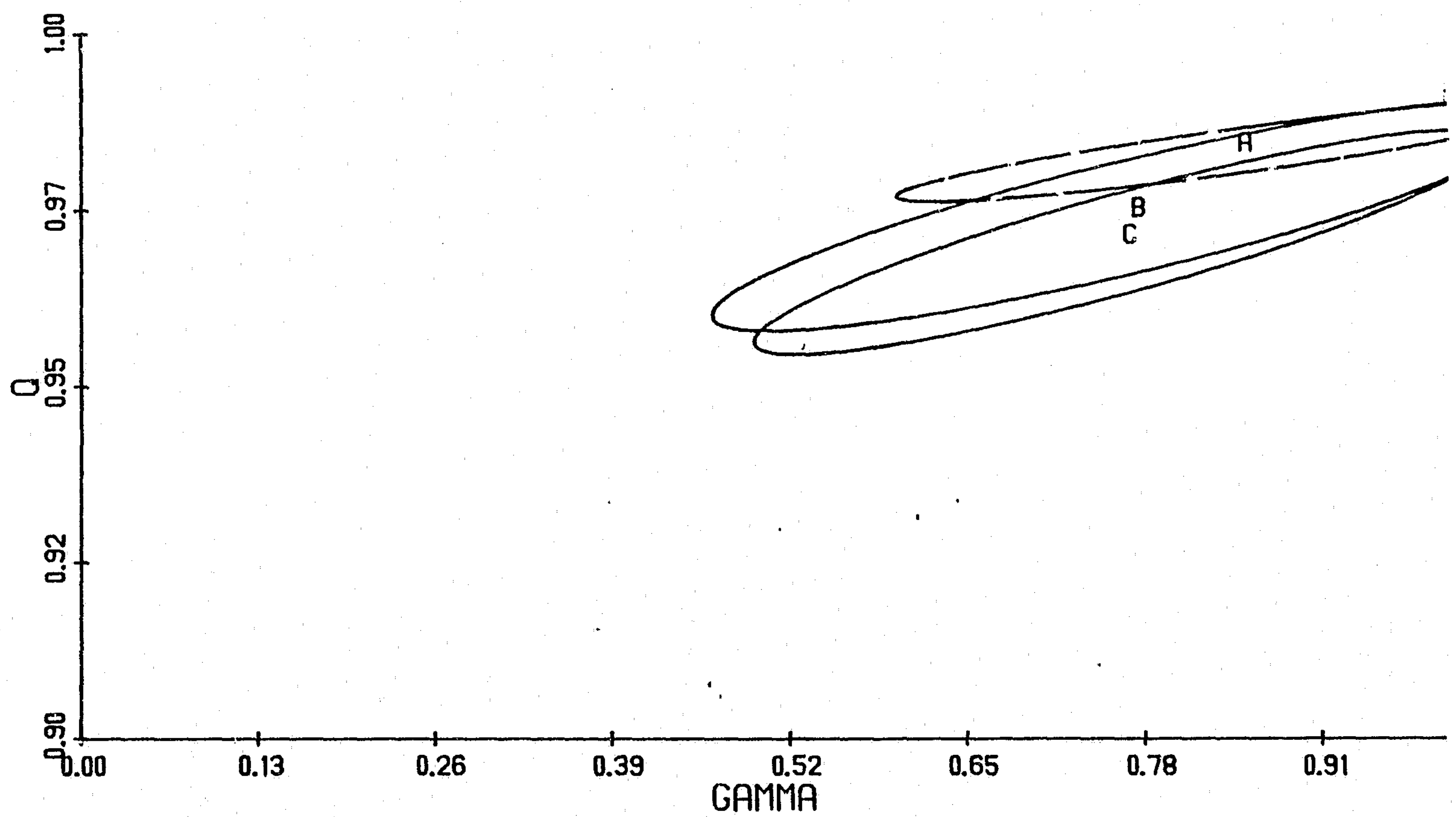


Figure 1f: Georgia Cohort
Bayes 90% Confidence Regions for estimates of γ and q, by
EMPLOYMENT STATUS PRIOR TO ARREST:

A. Full Time
B. Unknown
C. Other

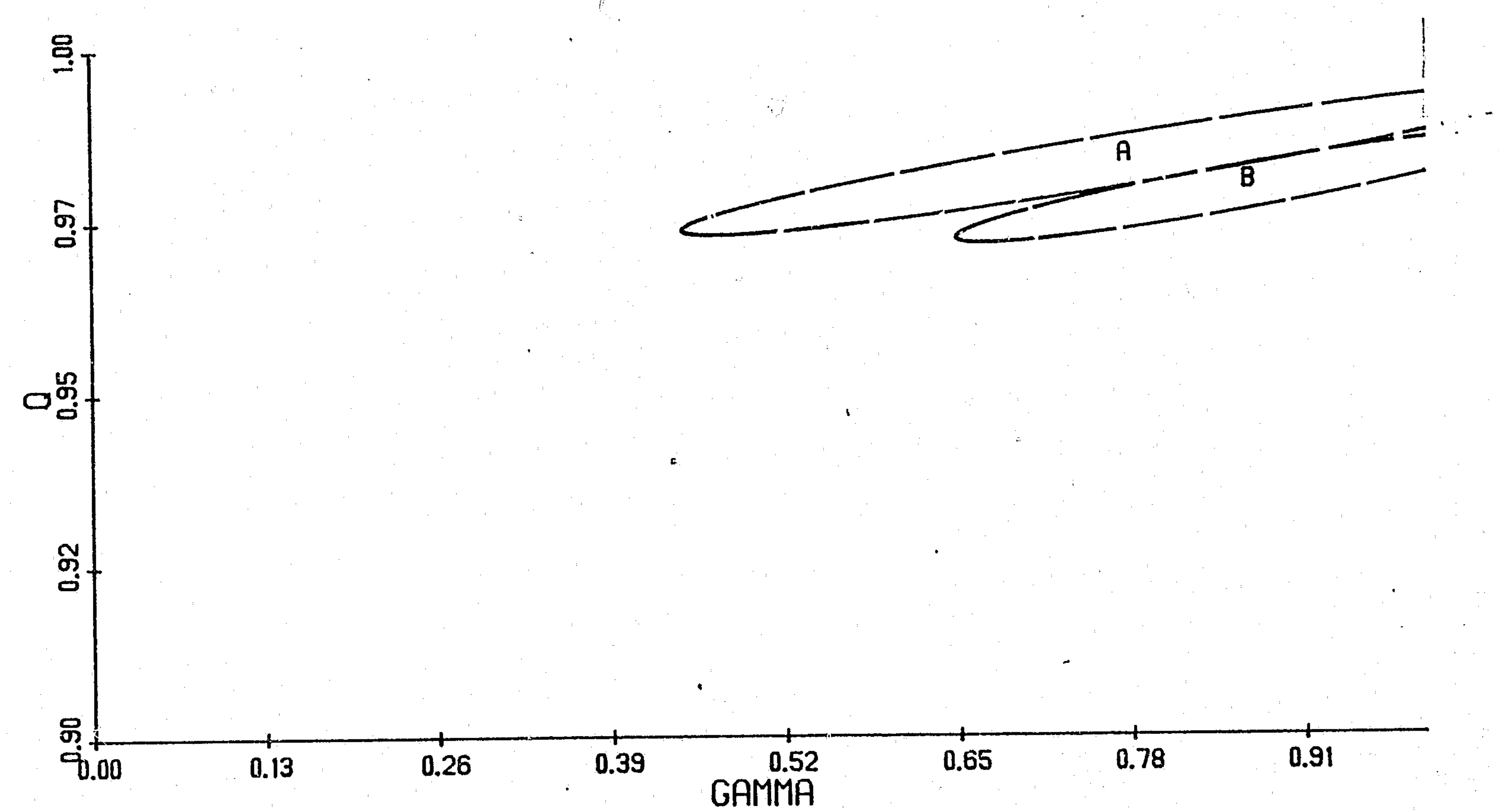


Figure 1g: Georgia Cohort
Bayes 90% Confidence Regions for PRIOR ARRESTS
A. None
B. One or More.

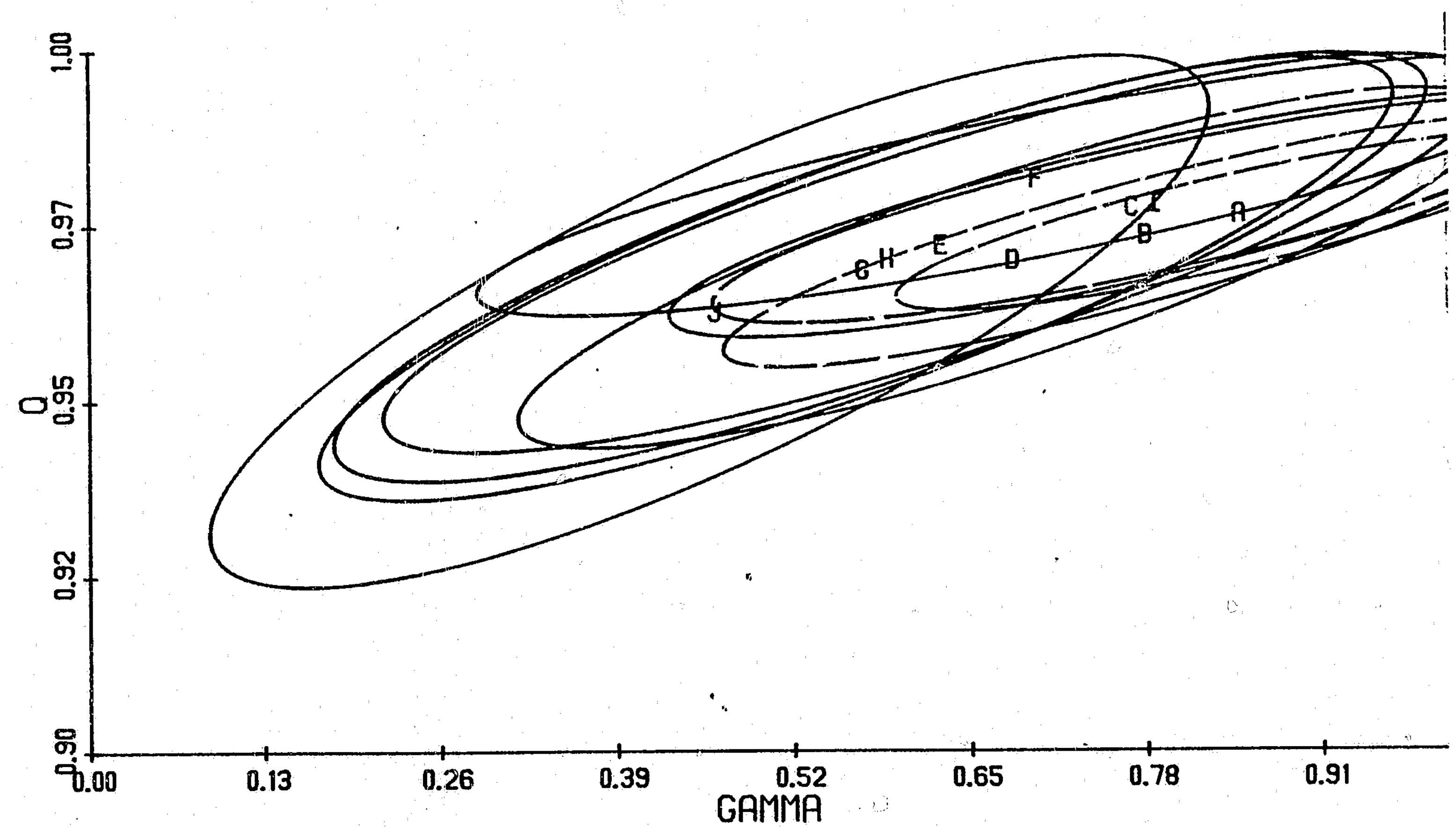


Figure 1h: Georgia Cohort
Bayes 90% Confidence Regions for AGE AT RELEASE

A. 17.8-22.3 years
B. 22.3-23.6 years
C. 23.6-24.6 years
D. 24.6-25.6 years
E. 25.6-26.9 years
F. 26.9-28.7 years
G. 28.7-31.3 years
H. 31.3-35.1 years
I. 35.1-41.2 years
J. 41.2 and above

TABLE 2B: U.S. BUREAU OF PRISONS COVARIATE PARTITIONS (CONT.)

	N	K	TT	MAXIMUM LIKELIHOOD ESTIMATES					BAYES ESTIMATES				
AGE AT RELEASE				γ	σ_γ	q	σ_q	ρ	γ	σ_γ	q	σ_q	ρ
A	61	50	656.	0.823	0.049	0.925	0.011	0.052	0.813	0.049	0.924	0.011	0.063
B	87	70	836.	0.807	0.043	0.918	0.010	0.034	0.800	0.043	0.917	0.010	0.052
C	85	67	565.	0.796	0.044	0.882	0.014	0.008	0.769	0.044	0.881	0.014	0.018
D	118	85	663.	0.721	0.041	0.902	0.010	0.009	0.717	0.041	0.901	0.011	0.013
E	166	128	1417.	0.772	0.033	0.910	0.008	0.020	0.769	0.033	0.910	0.008	0.014
F	142	101	1232.	0.715	0.038	0.920	0.008	0.036	0.712	0.038	0.919	0.008	0.048
G	175	113	1266.	0.647	0.036	0.911	0.008	0.014	0.645	0.036	0.911	0.008	0.013
H	93	47	706.	0.511	0.053	0.937	0.010	0.074	0.512	0.052	0.936	0.010	0.088

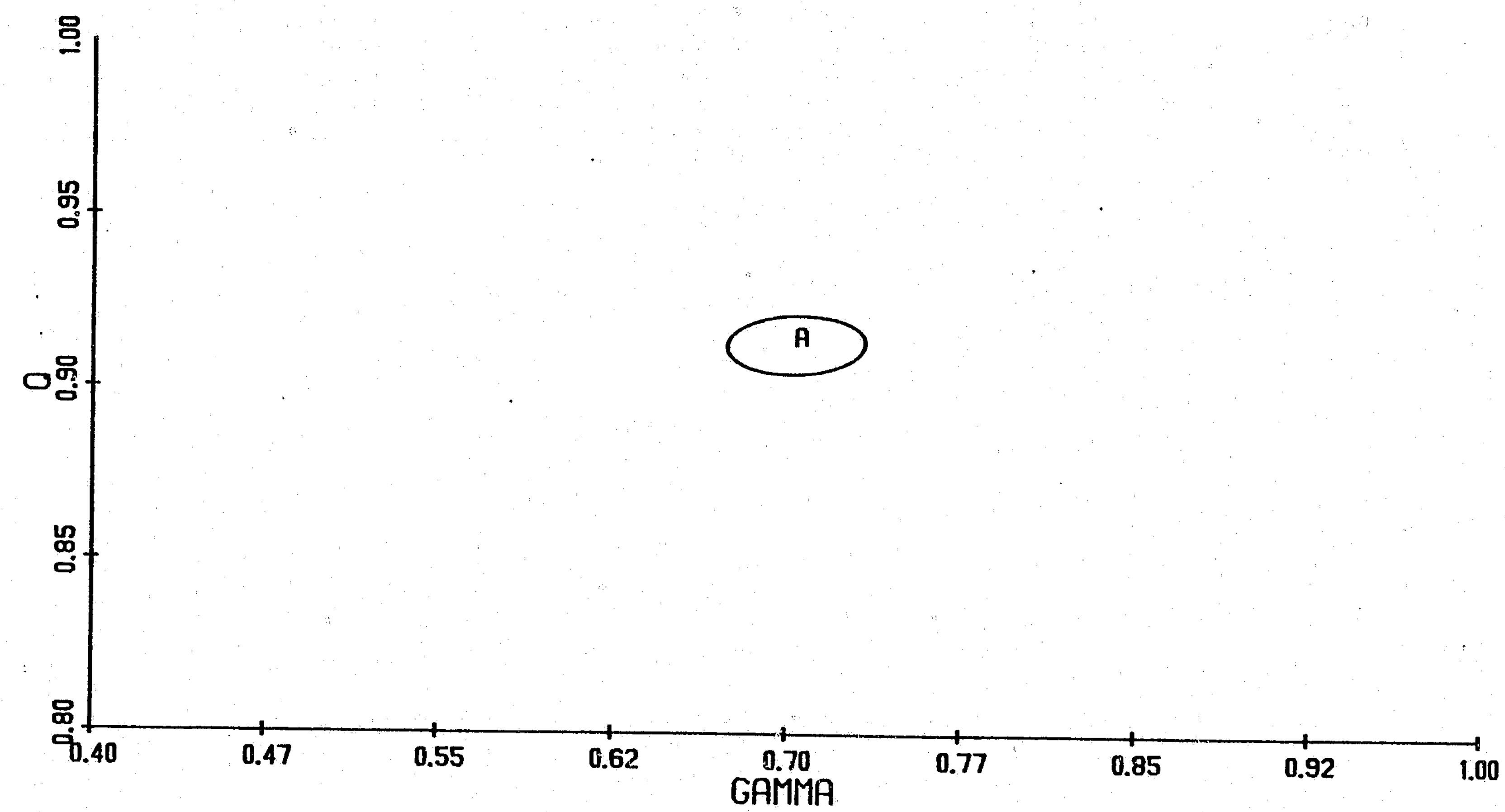


Figure 2a: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q.

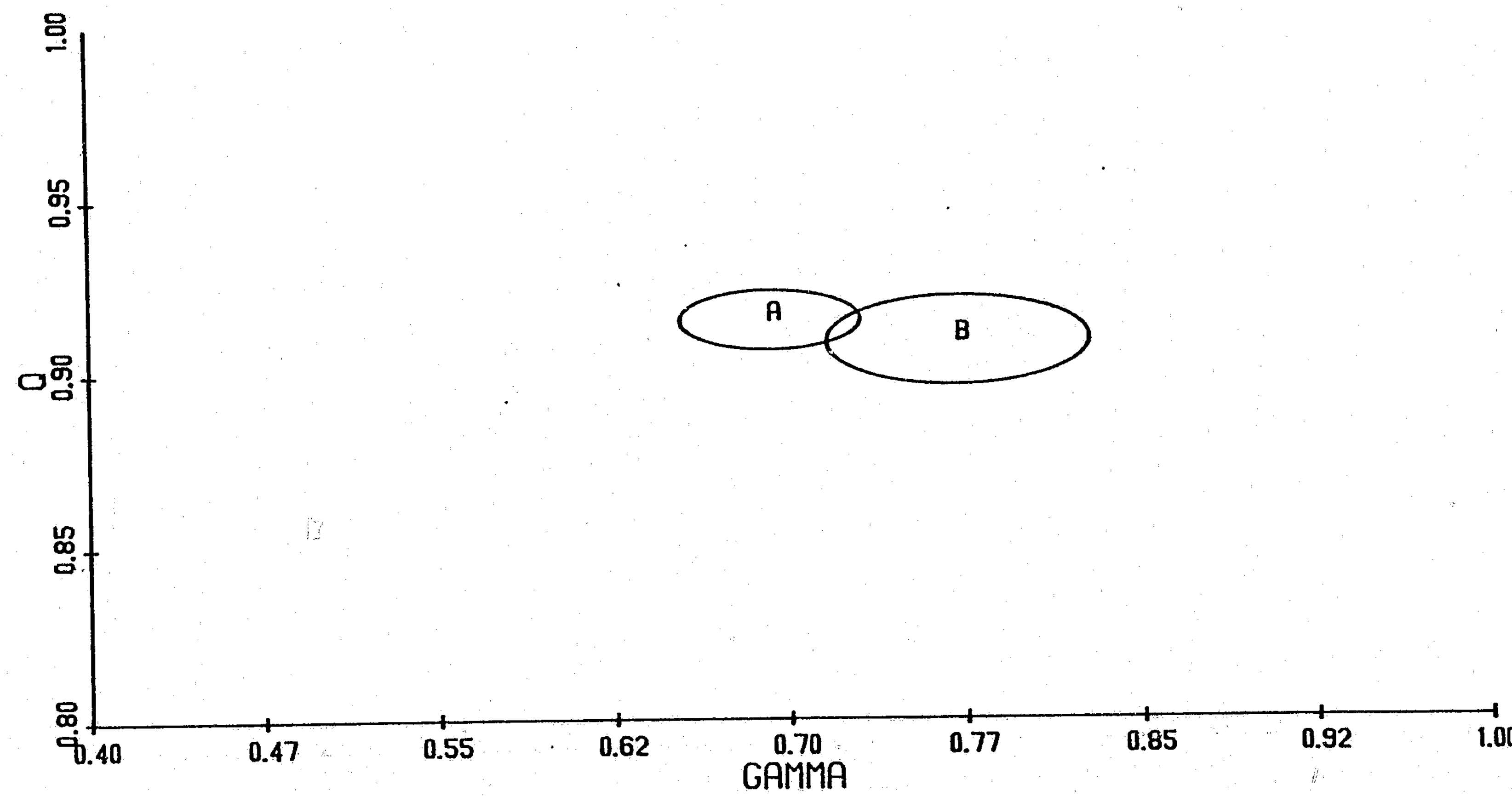


Figure 2b: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q by RACE

A. White
B. Other

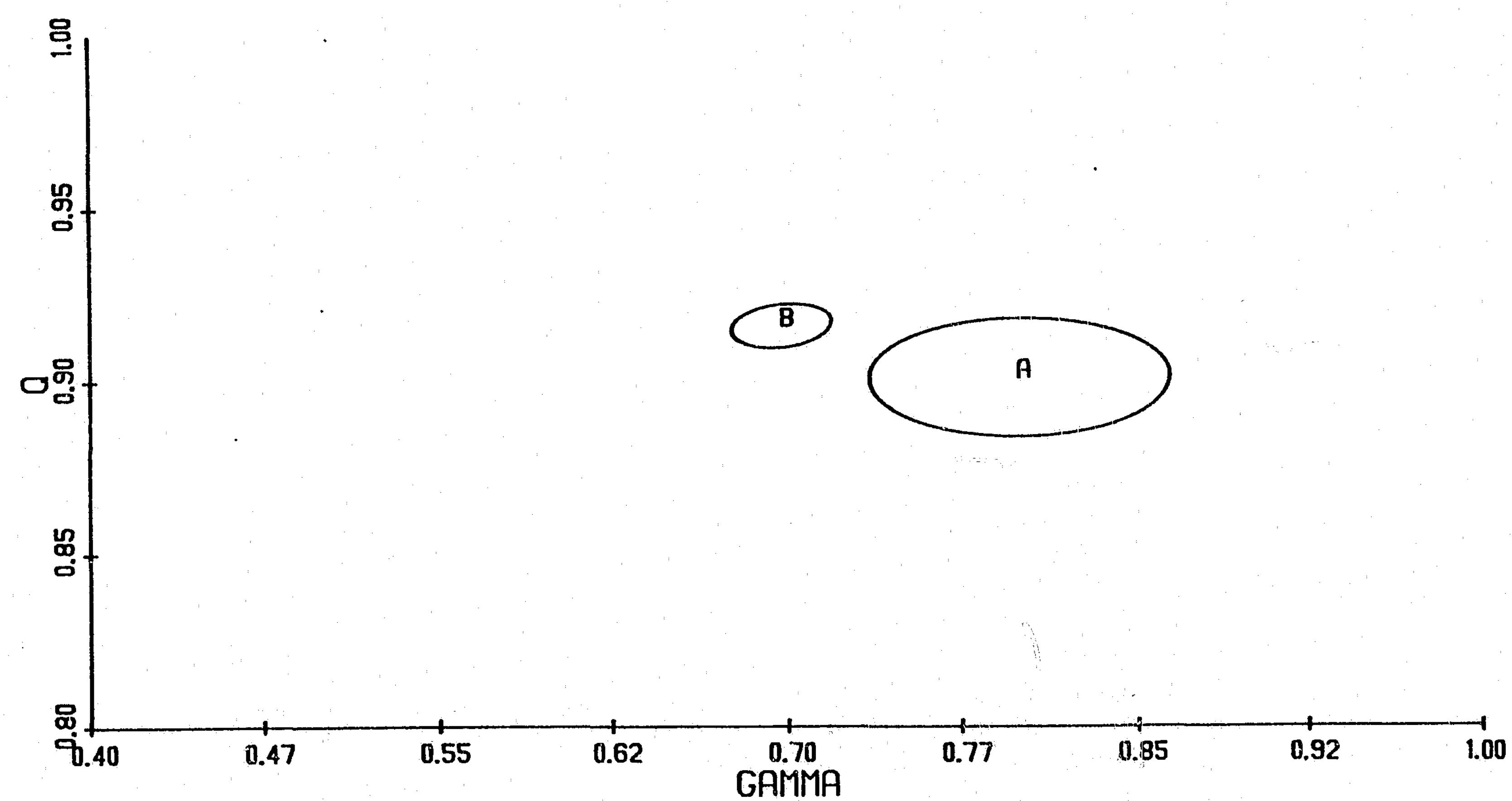


Figure 2c: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q by
PROBATION OR SUPERVISION VIOLATIONS

A. None
B. One or more

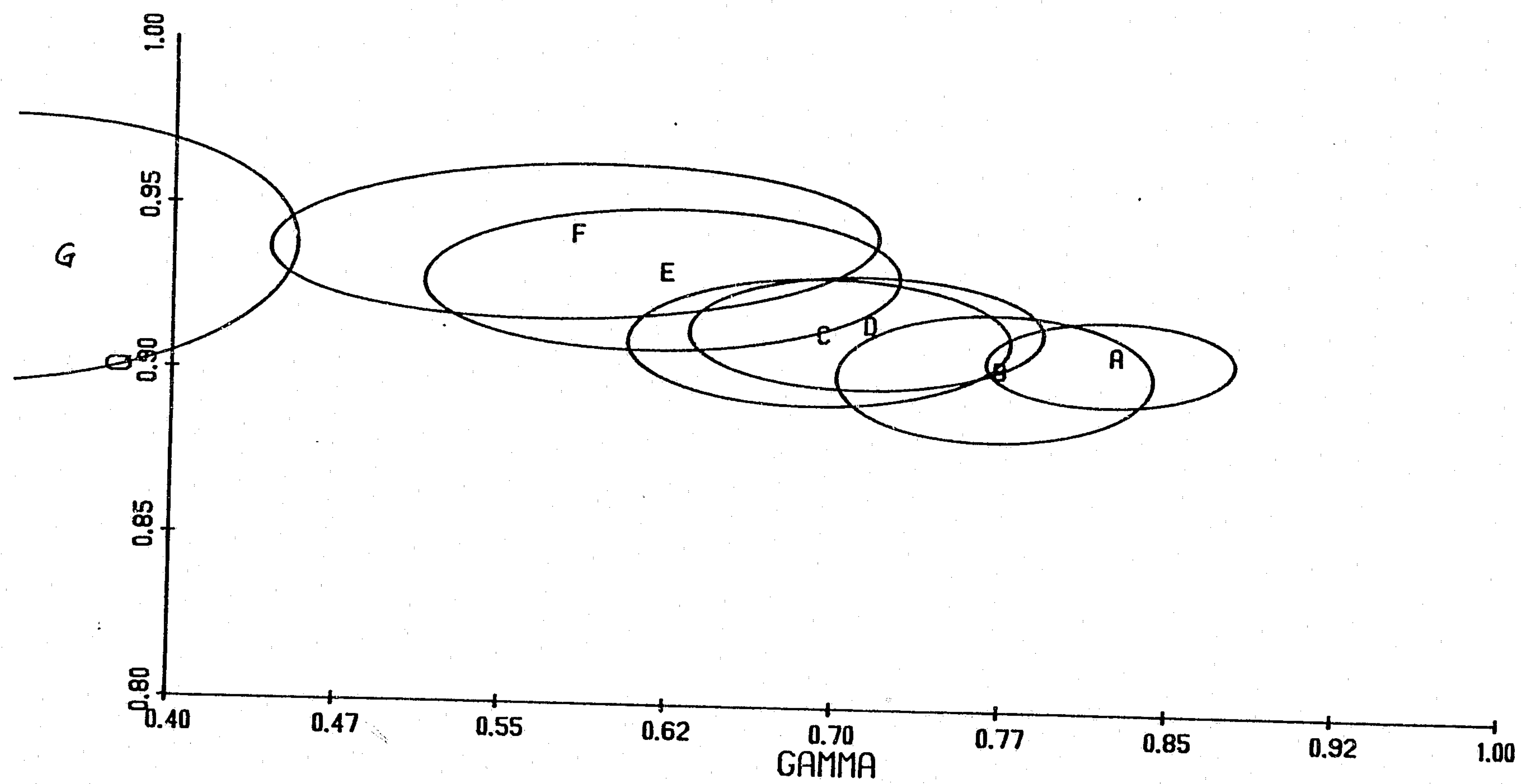


Figure 2d: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q by AGE AT FIRST ARREST

A. 16 or less
B. 17, 18
C. 19, 20
D. 21-23
E. 24-27
F. 28-34
G. 35 and over.

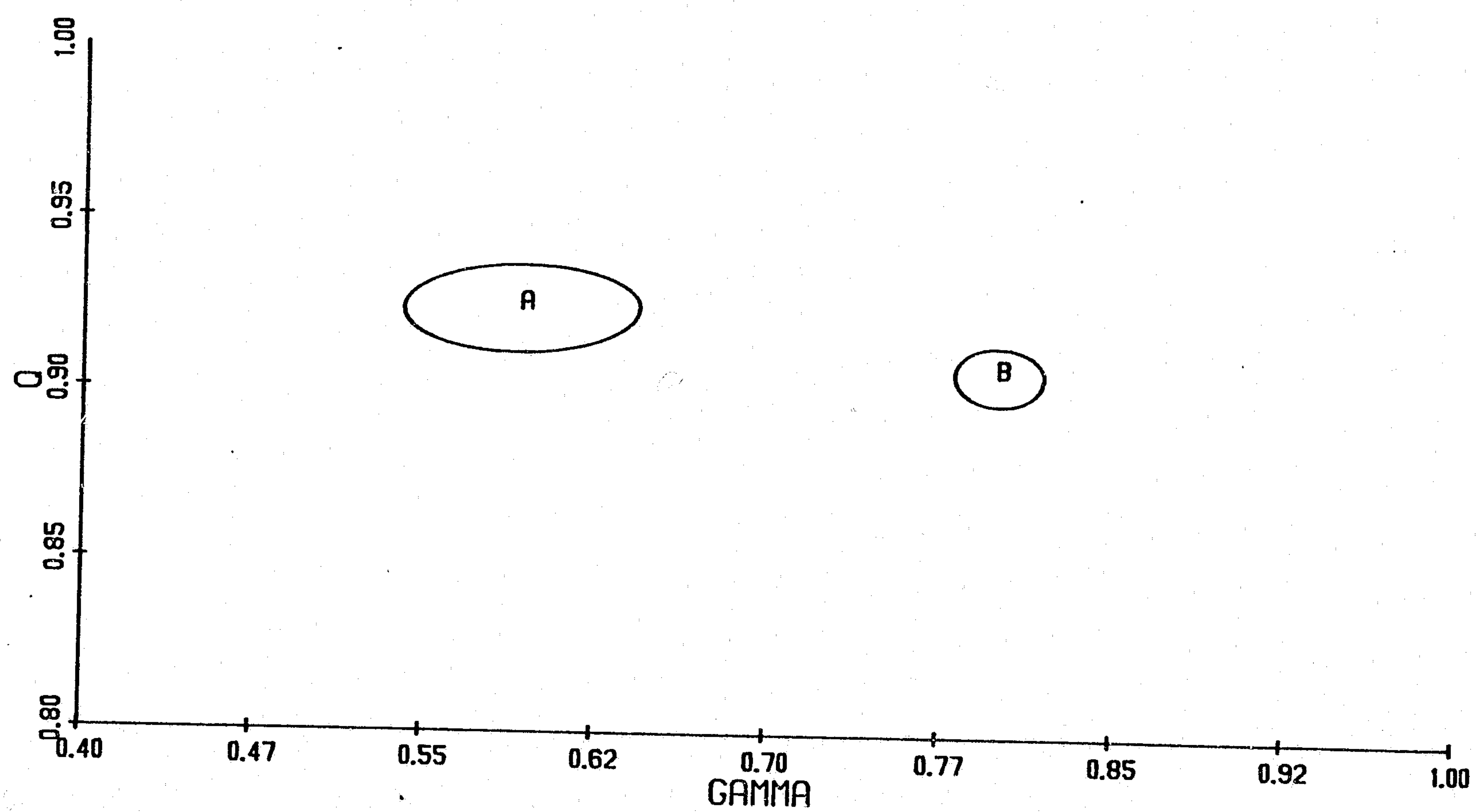


Figure 2e: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q by NUMBER OF PRIOR FELONY SENTENCES
A. None
B. One or more

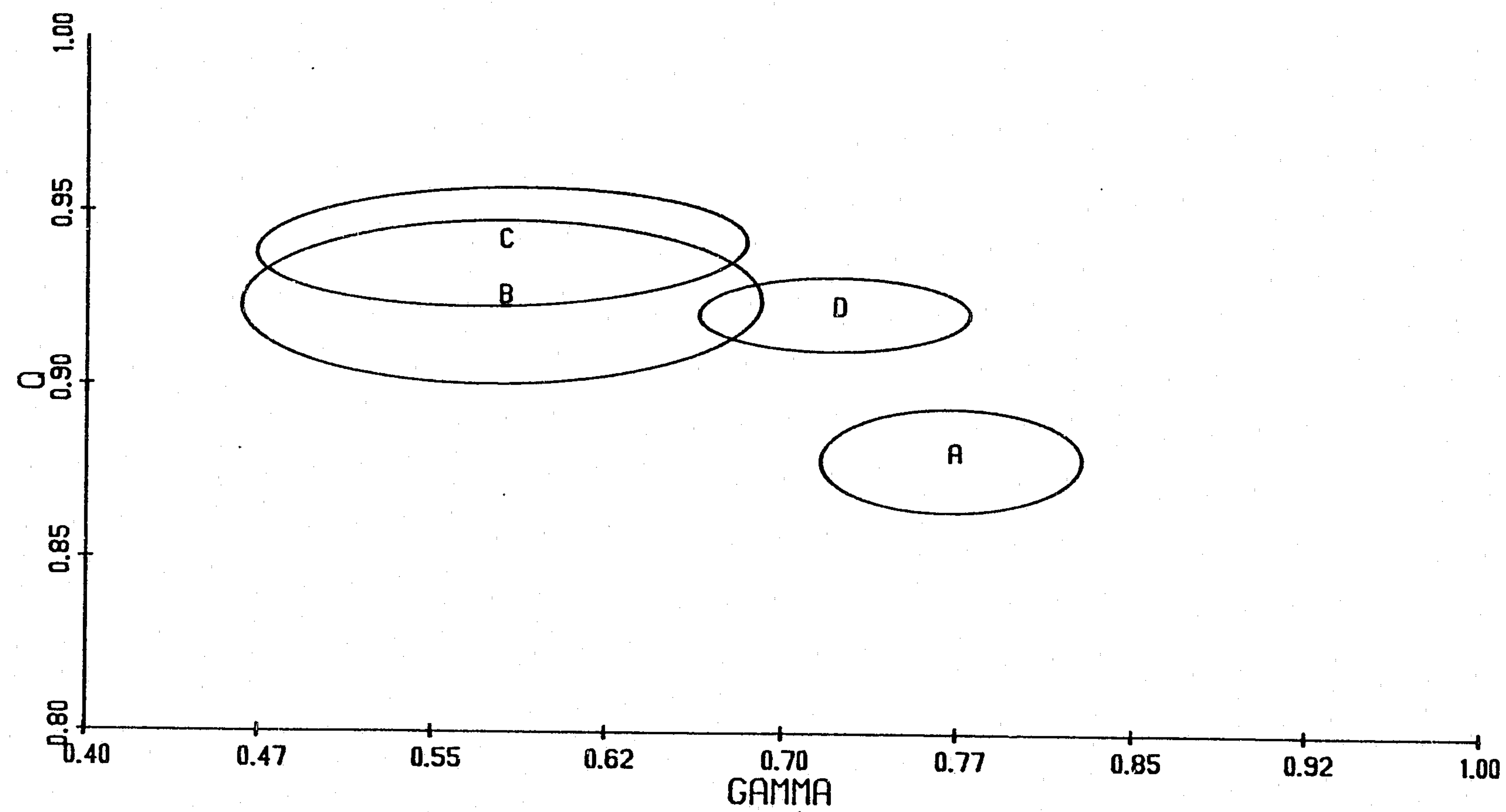


Figure 2f: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q by PRIMARY CURRENT OFFENSE

A. Vehicle theft for interstate transportation
B. Fraudulent check, counterfeit, tax fraud, embezzlement
C. Moonshine
D. Other

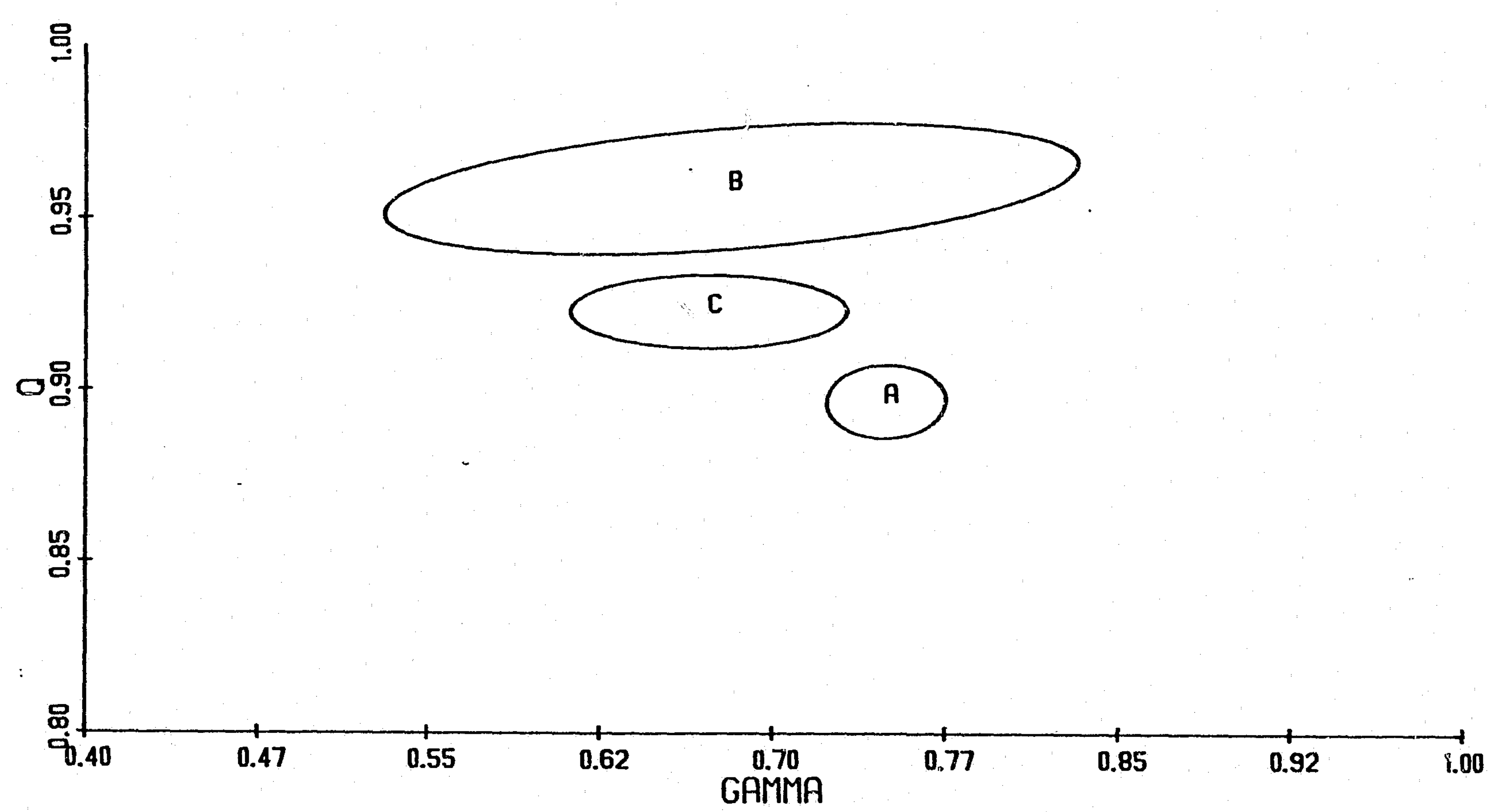


Figure 2g: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q by ETIOLOGY OF LAST PATTERN OF CRIMINALITY

A. Delinquent or criminal orientation (but not narcotic or alcohol related)
B. Financial straits
C. Other

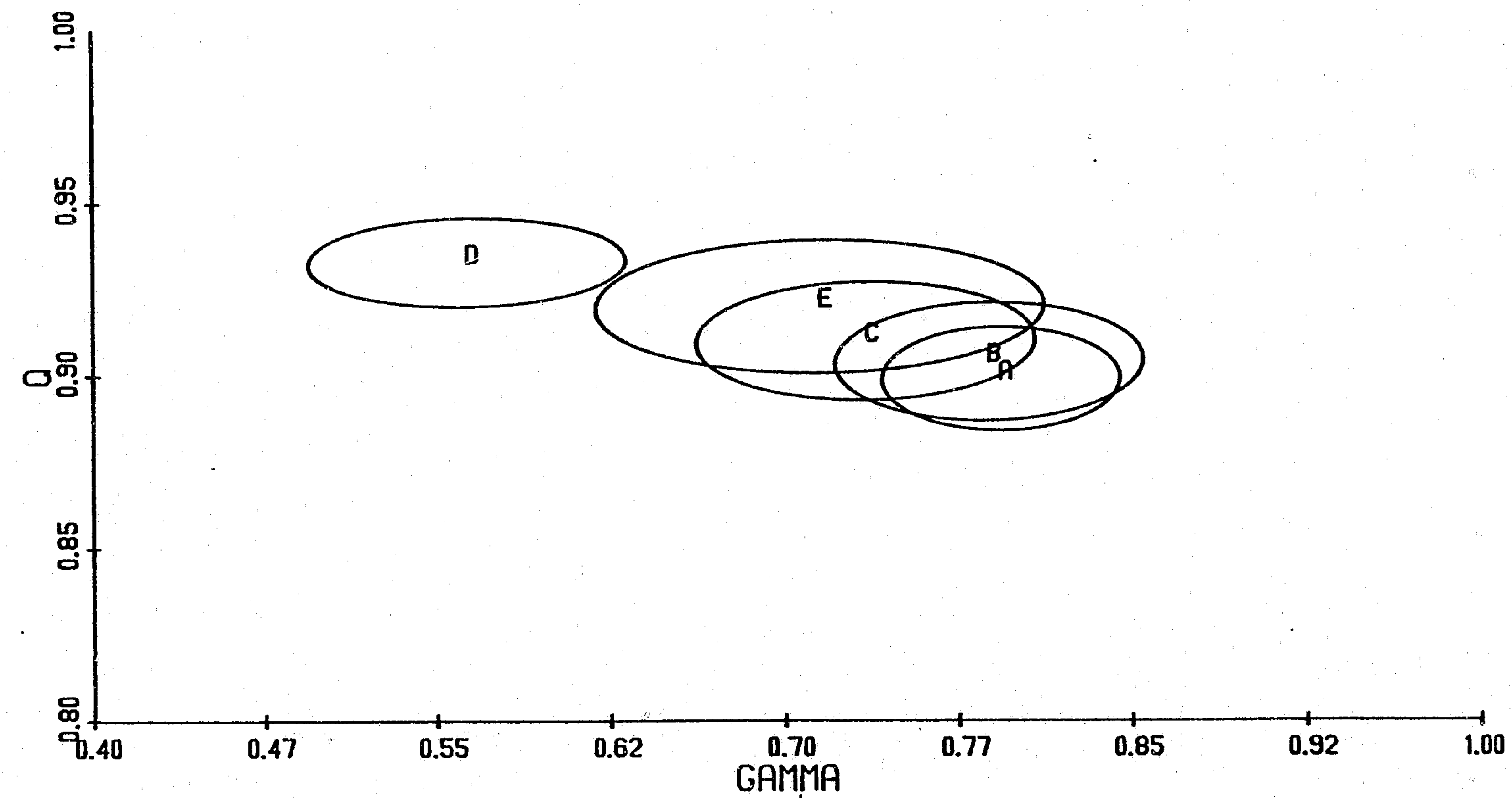


Figure 2h: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q for
EMPLOYMENT DURING LAST 2 YEARS PRECEDING LAST IMPRISONMENT

A. Less than 25% time
B. 26-50% time
C. 51-75% time
D. 76-100% time
E. Student or Unemployed

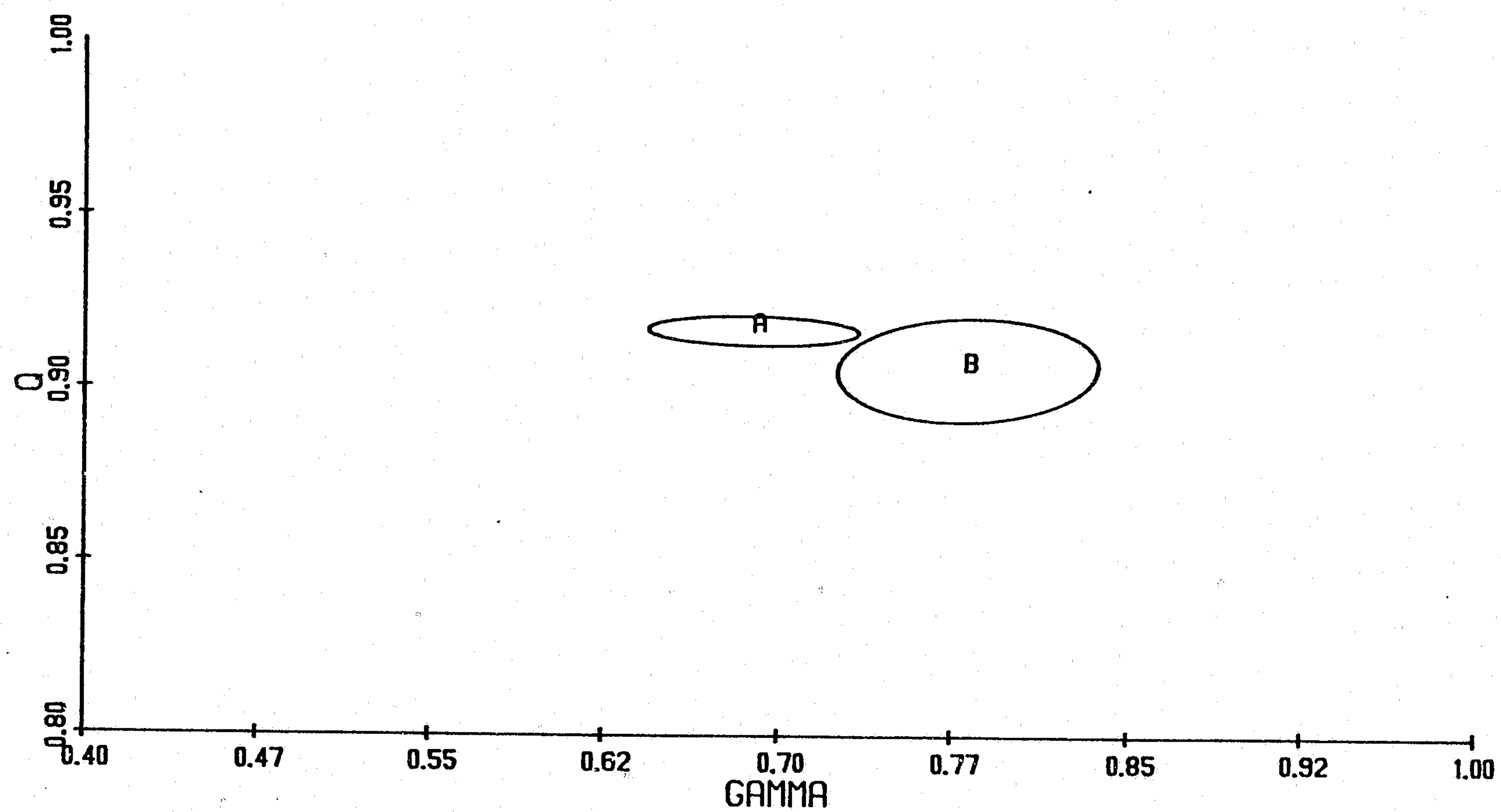


Figure 2i: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q for ALCOHOL PROBLEMS
A. None
B. Other

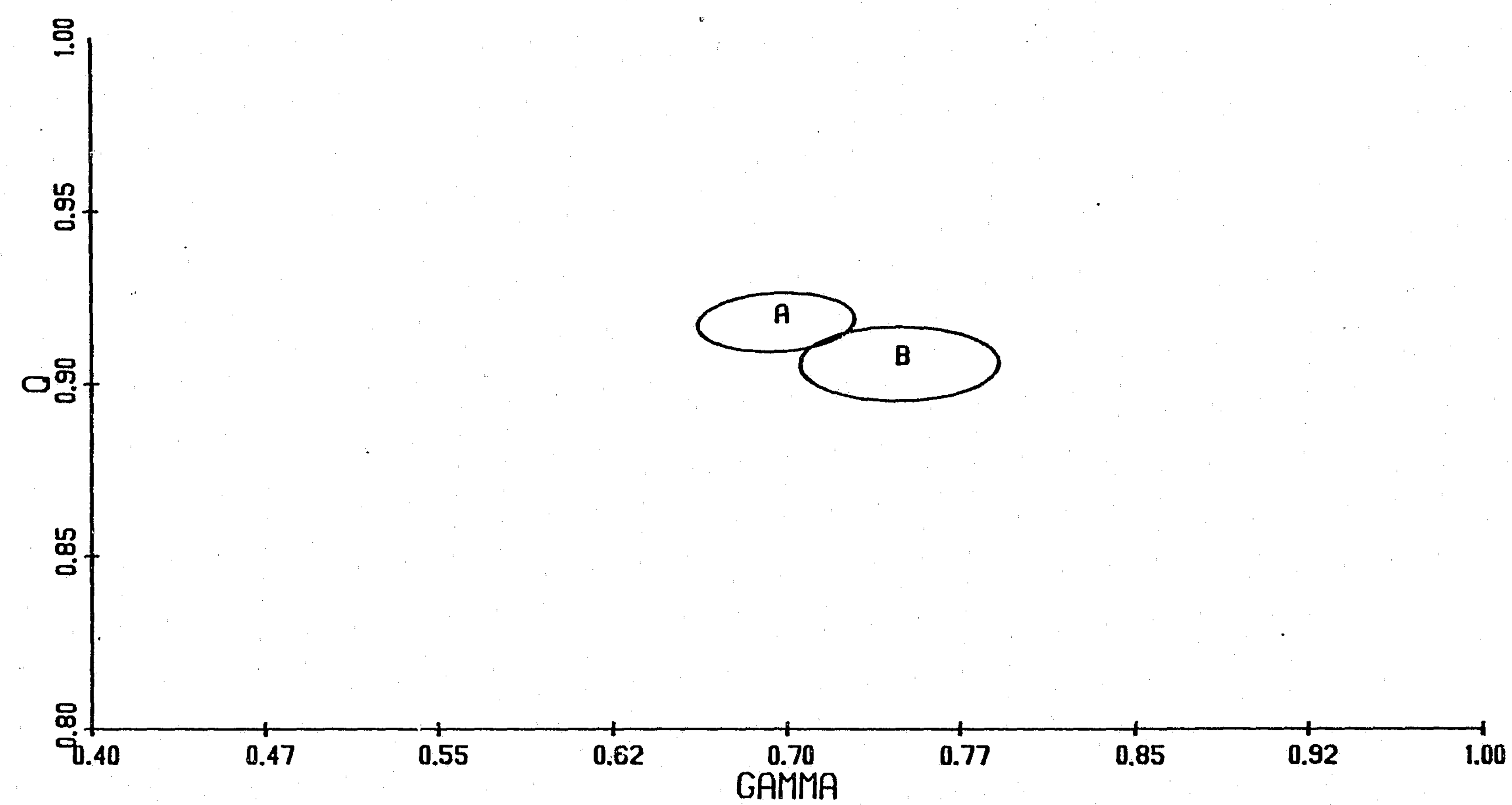


Figure 2j: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q for PSYCHOLOGICAL DIAGNOSIS
A. None or favorable
B. Other

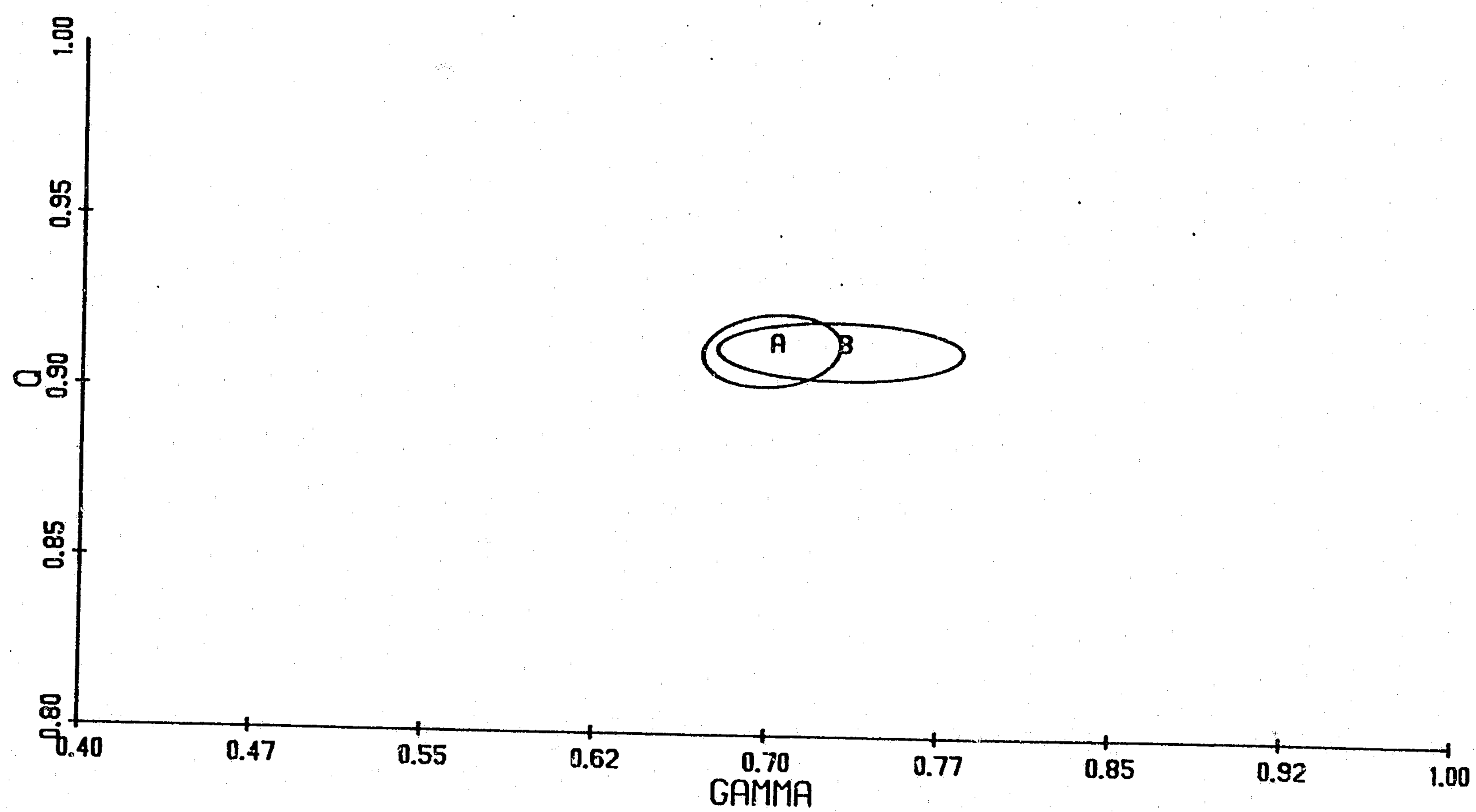


Figure 2k: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q for
SCHOOLING COMPLETED AT RELEASE
A. 8th grade or below
B. 9th grade or above

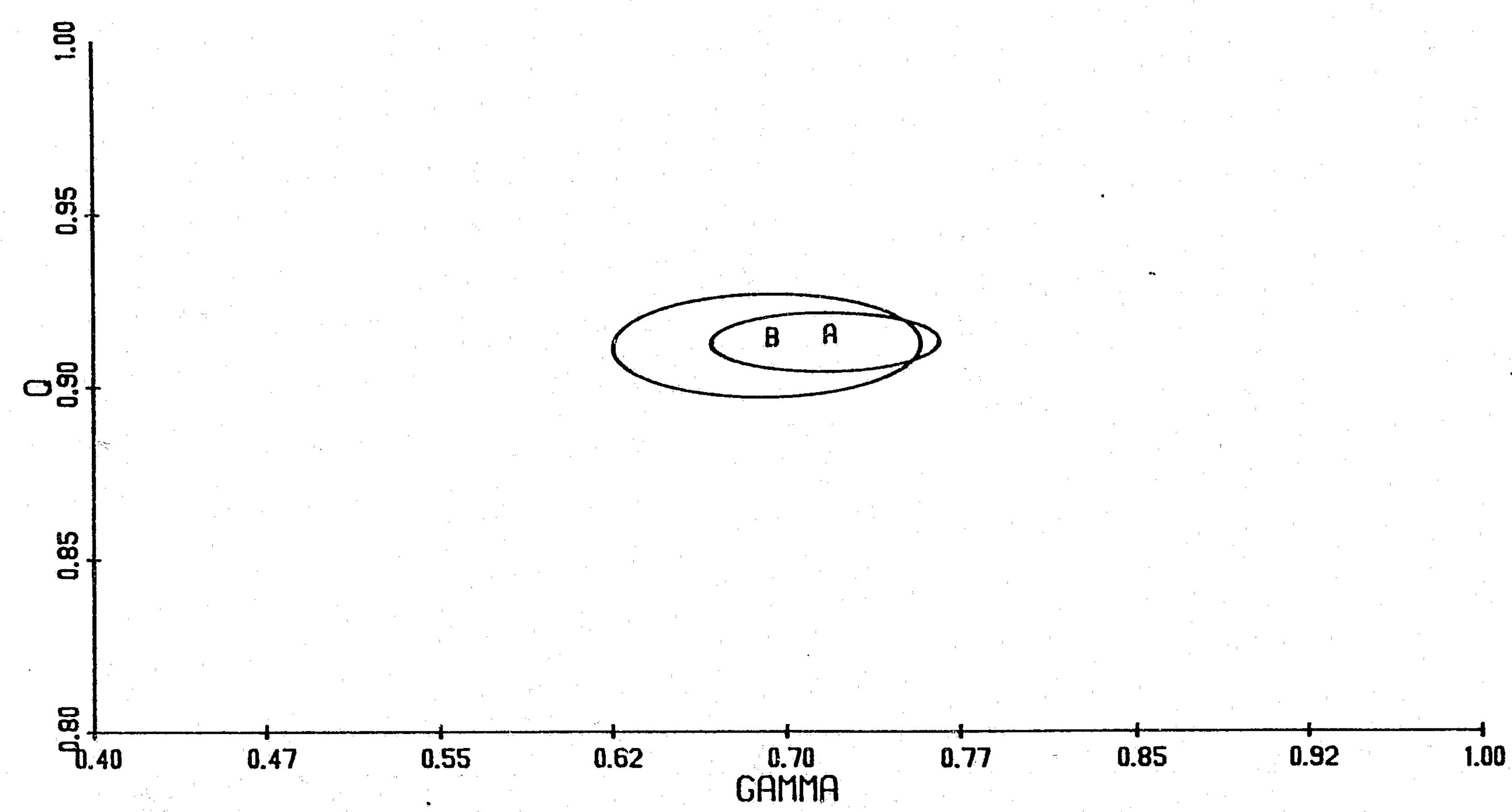


Figure 21: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q for IQ
A. Less than or equal to 100
B. Greater than 100

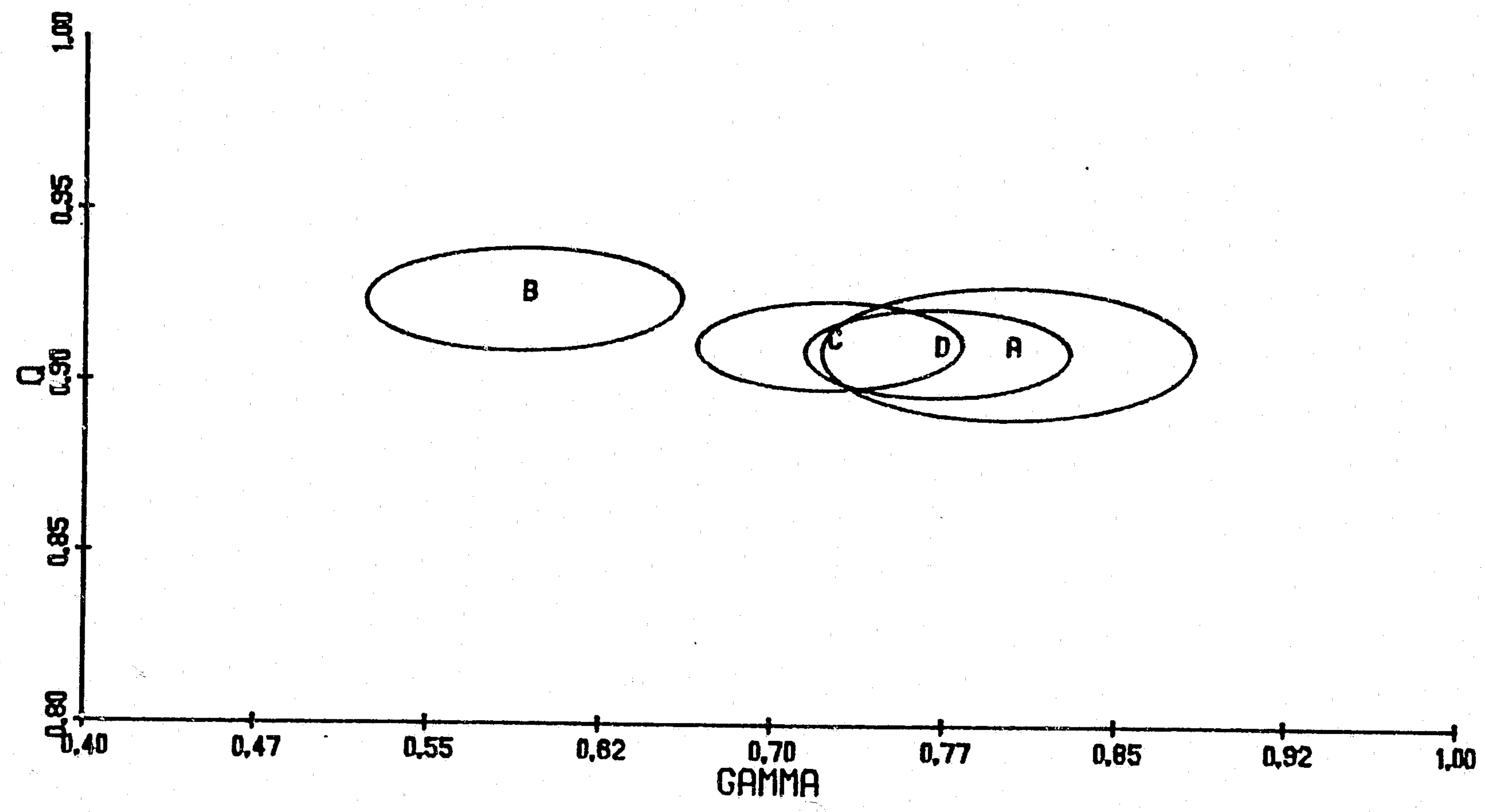


Figure 2m: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q for
ANTICIPATED HOME AT RELEASE
  A. Plans to live alone
  B. With wife (or common law wife)
  C. With parents
  D. Other

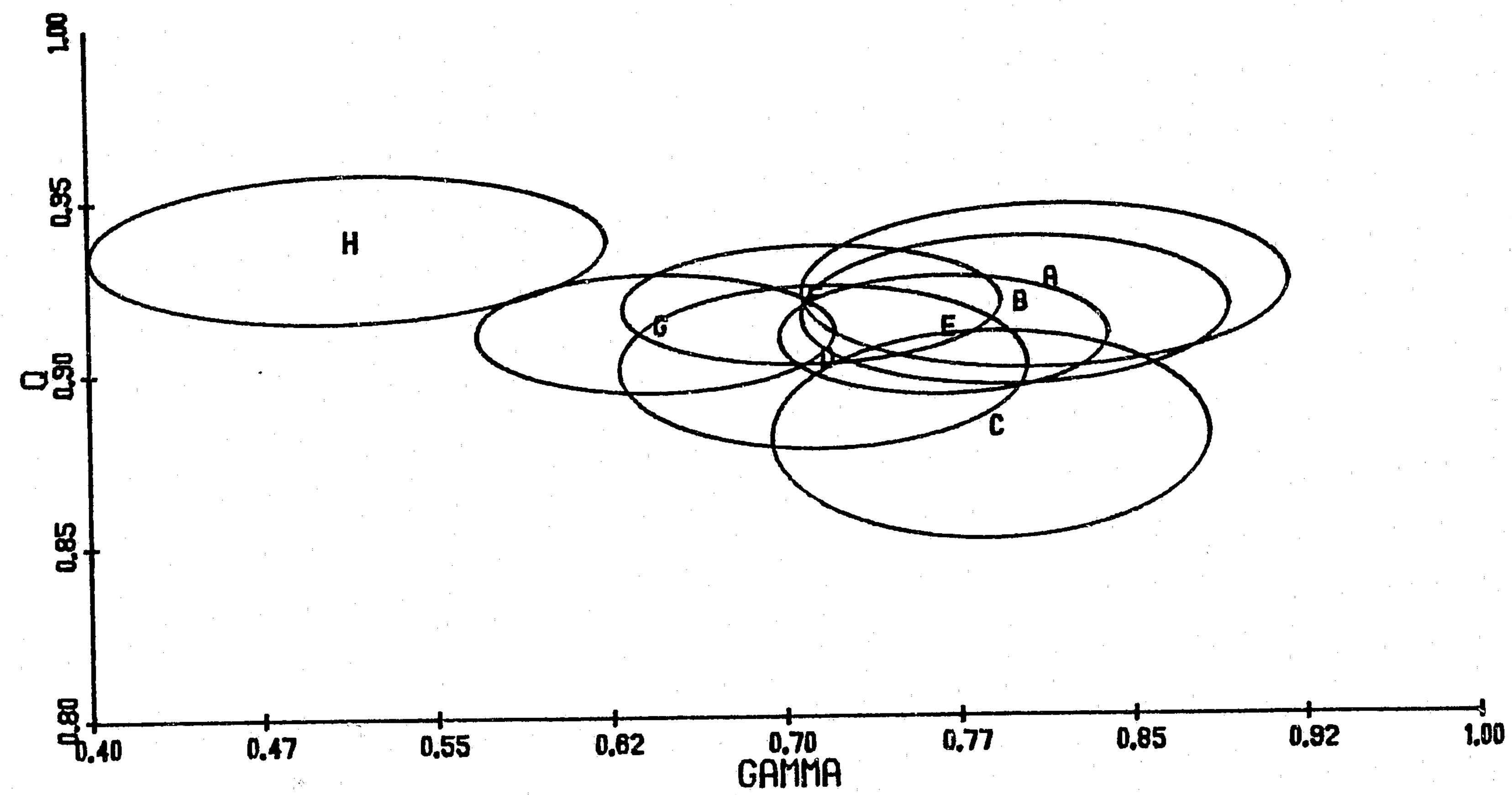


Figure 2n: United States B.O.P. Cohort
Bayes 90% Confidence Regions for estimates of γ and q for
AGE AT RELEASE

A. 19 or less
B. 20, 21
C. 22, 23
D. 24-26
E. 27-30
F. 31-35
G. 36-46
H. 47 and over.

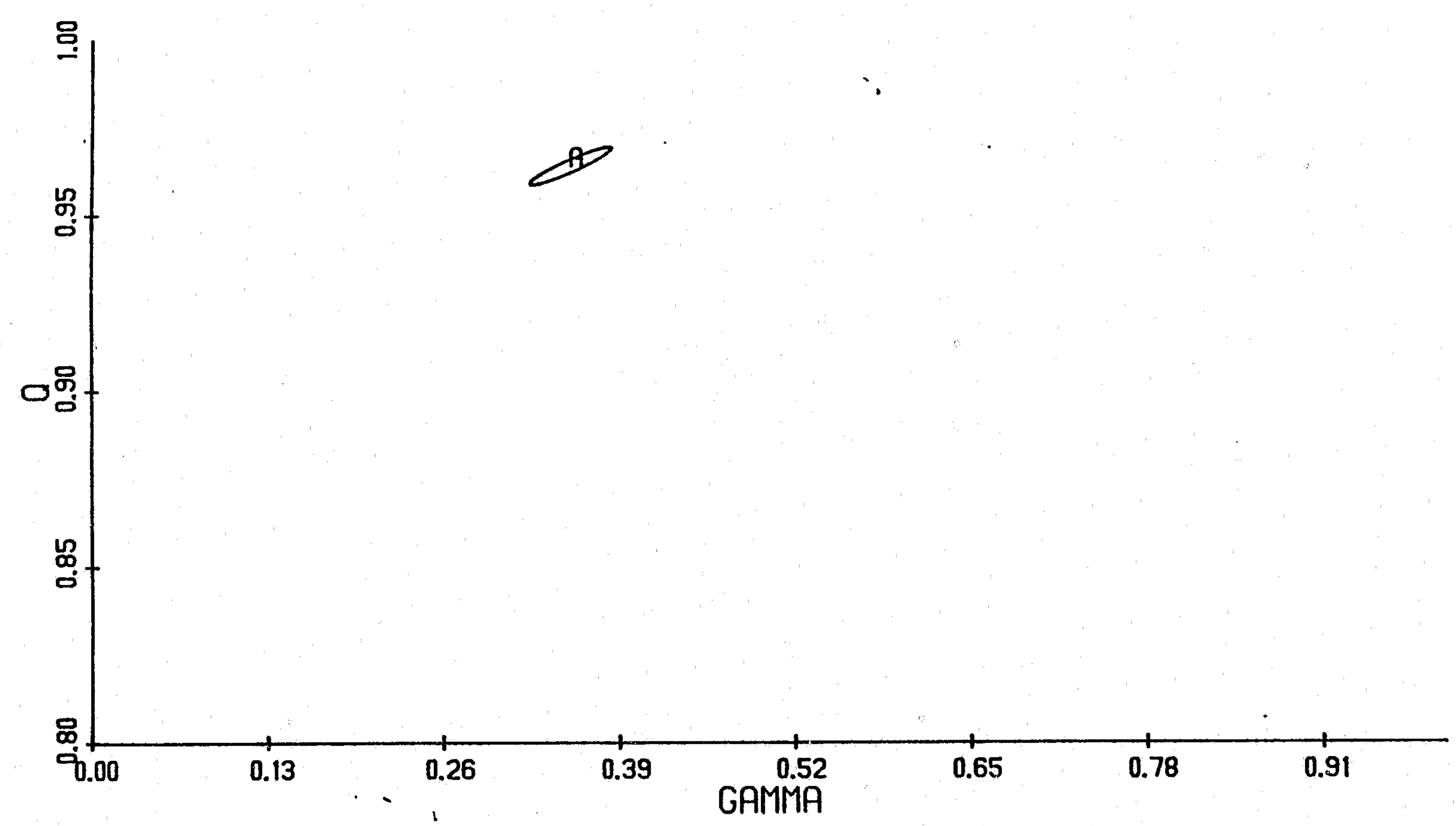


Figure 3a: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q.

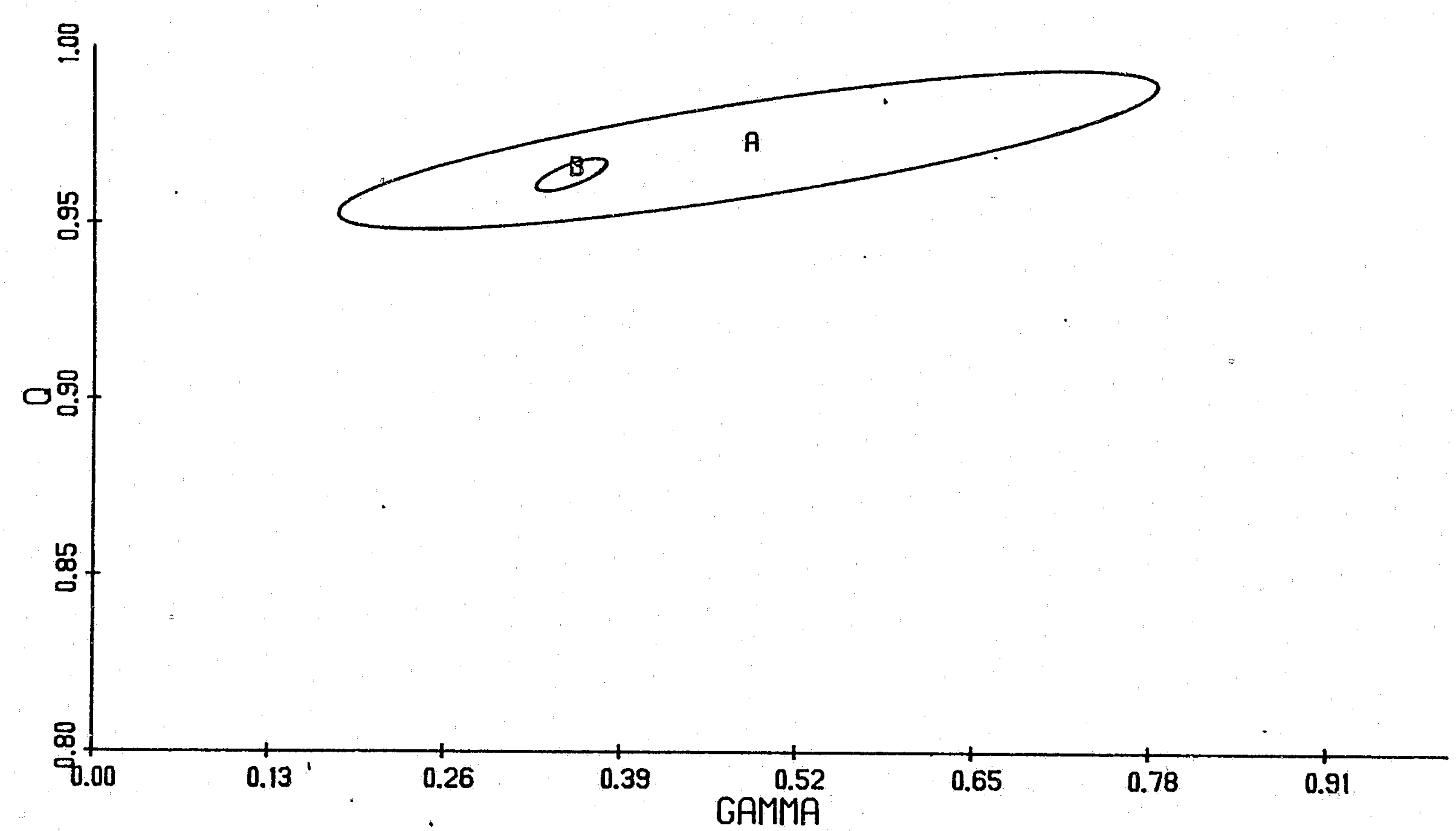


Figure 3b: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by SEX
A. Female
B. Male

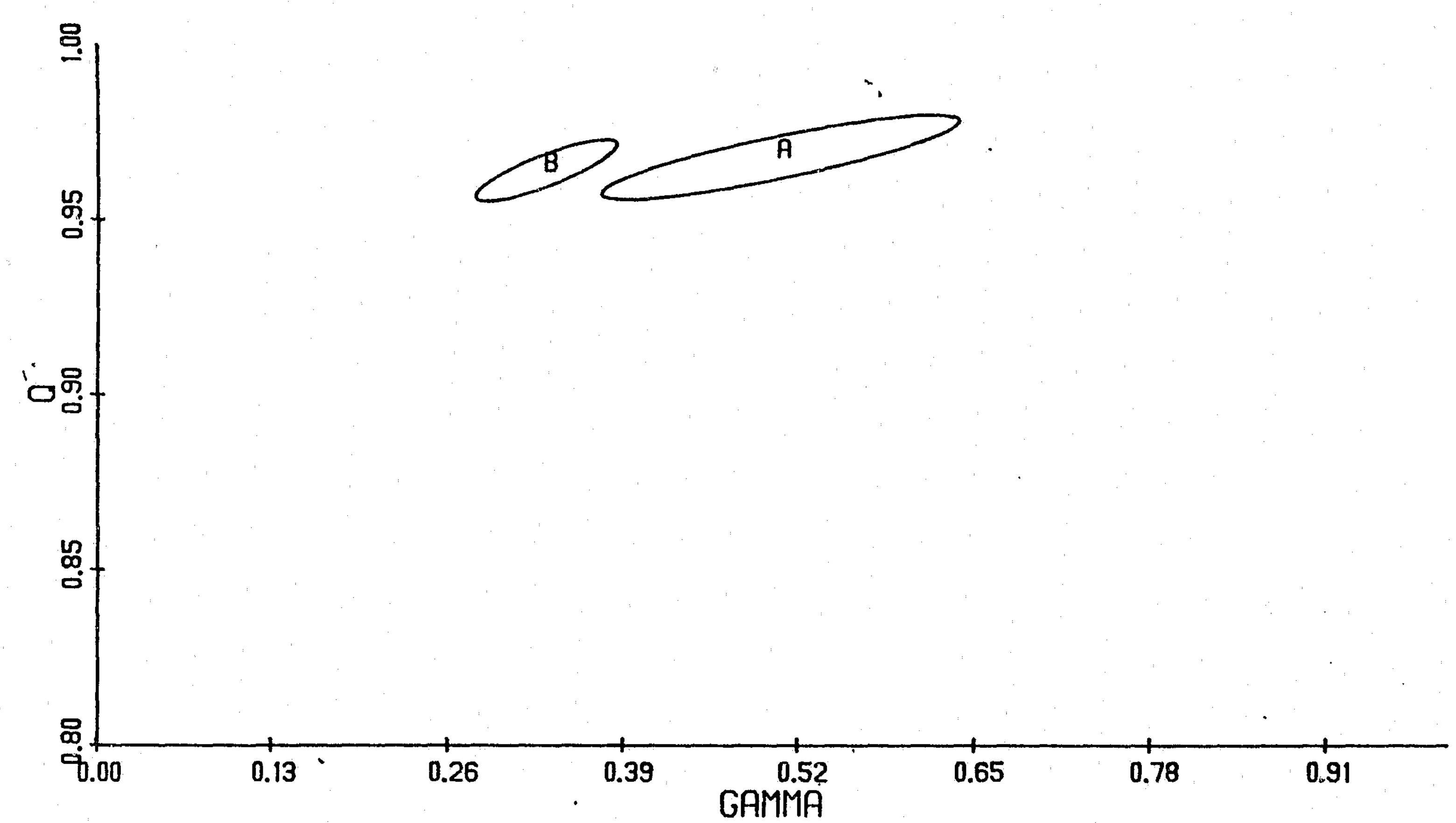


Figure 3c: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by RACE
A. Non-white
B. White

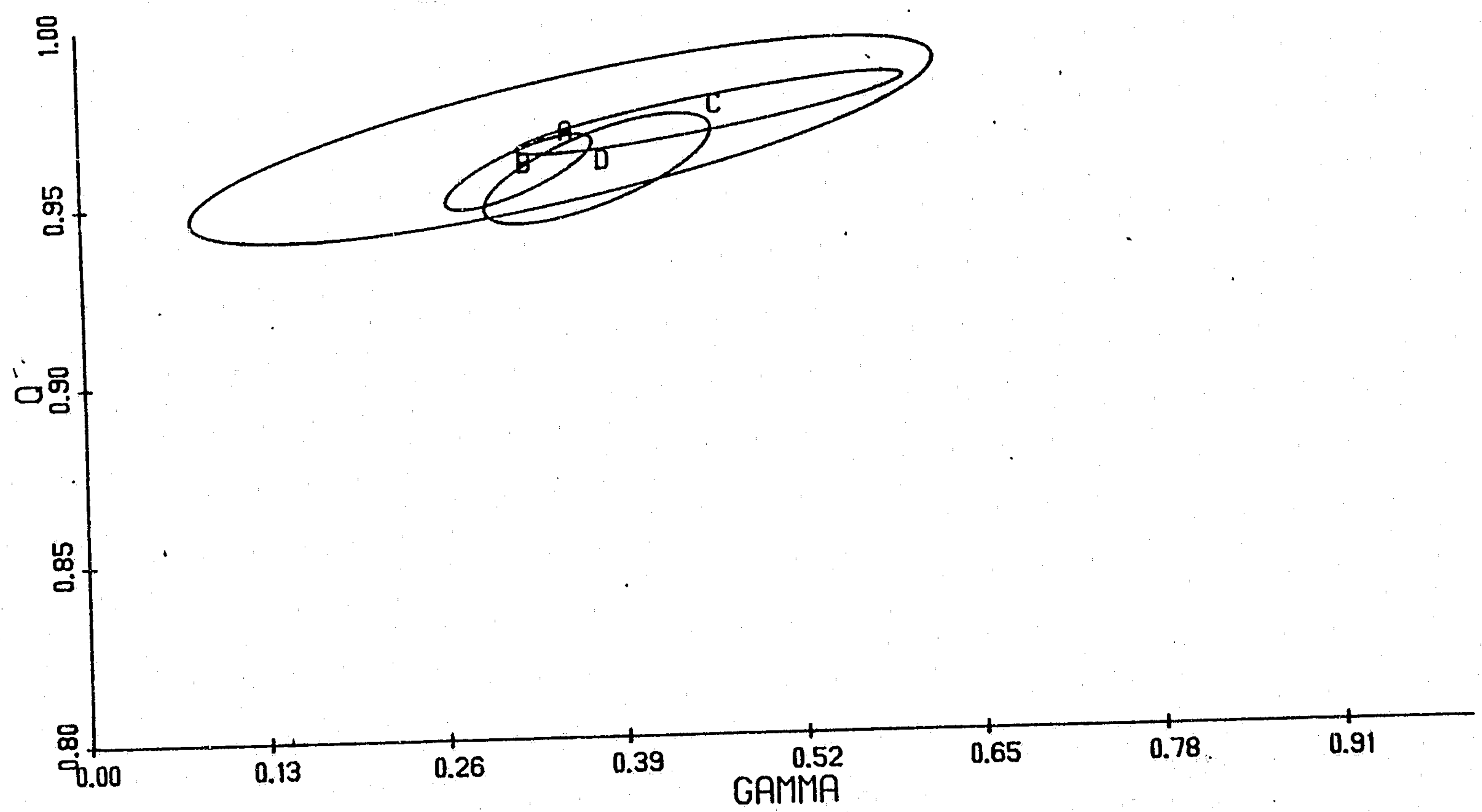


Figure 3d: Iowa Cohort
Bayes 90% Confidence Regions for estimates of Υ and q by MARITAL STATUS

A. Widowed or separated
B. Single
C. Married or common-law
D. Divorced

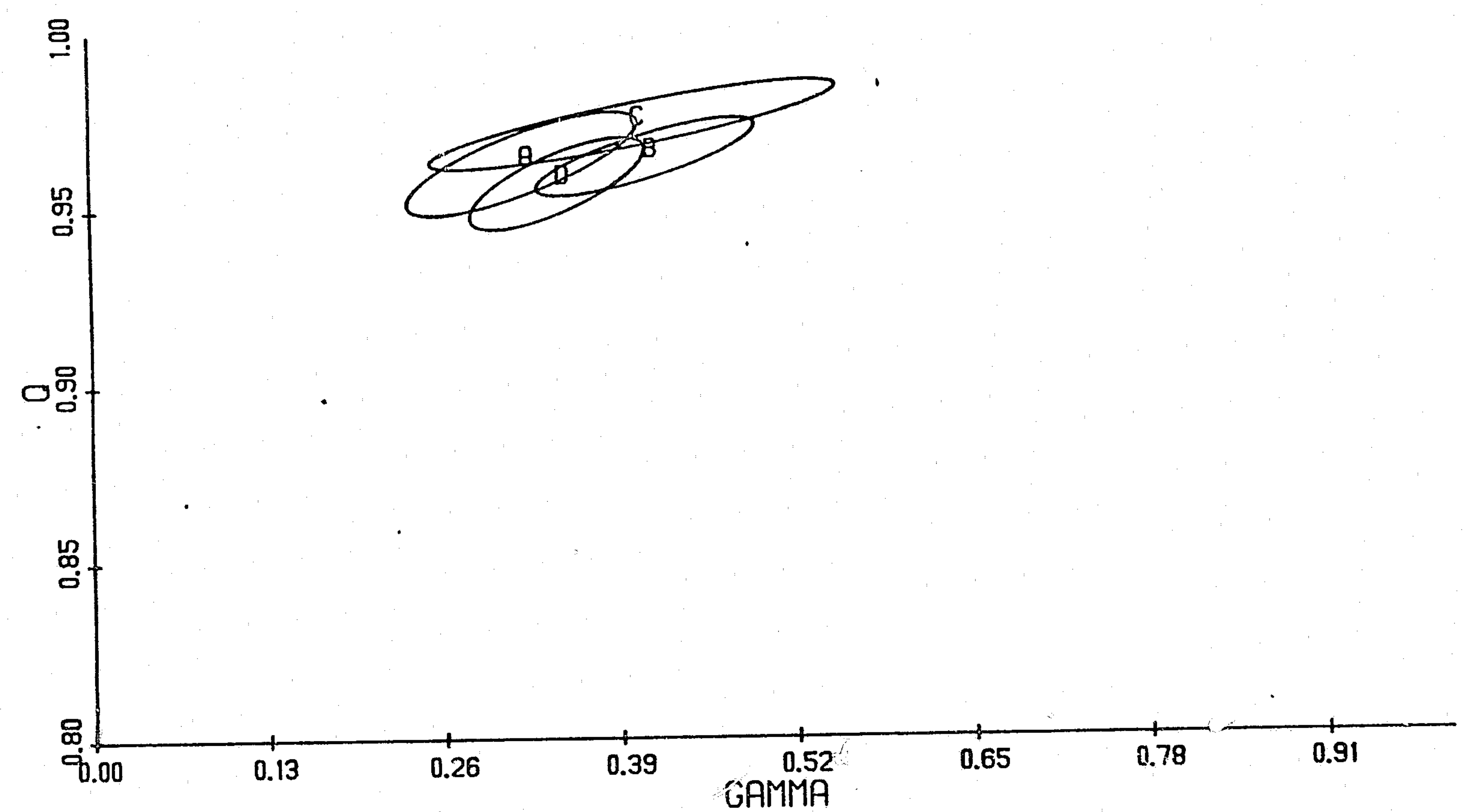


Figure 3e: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by
LIVING ARRANGEMENT ON RELEASE

A. With relatives, foster parents, institution, other
B. Alone
C. Spouse and/or children
D. Parents or step-parents

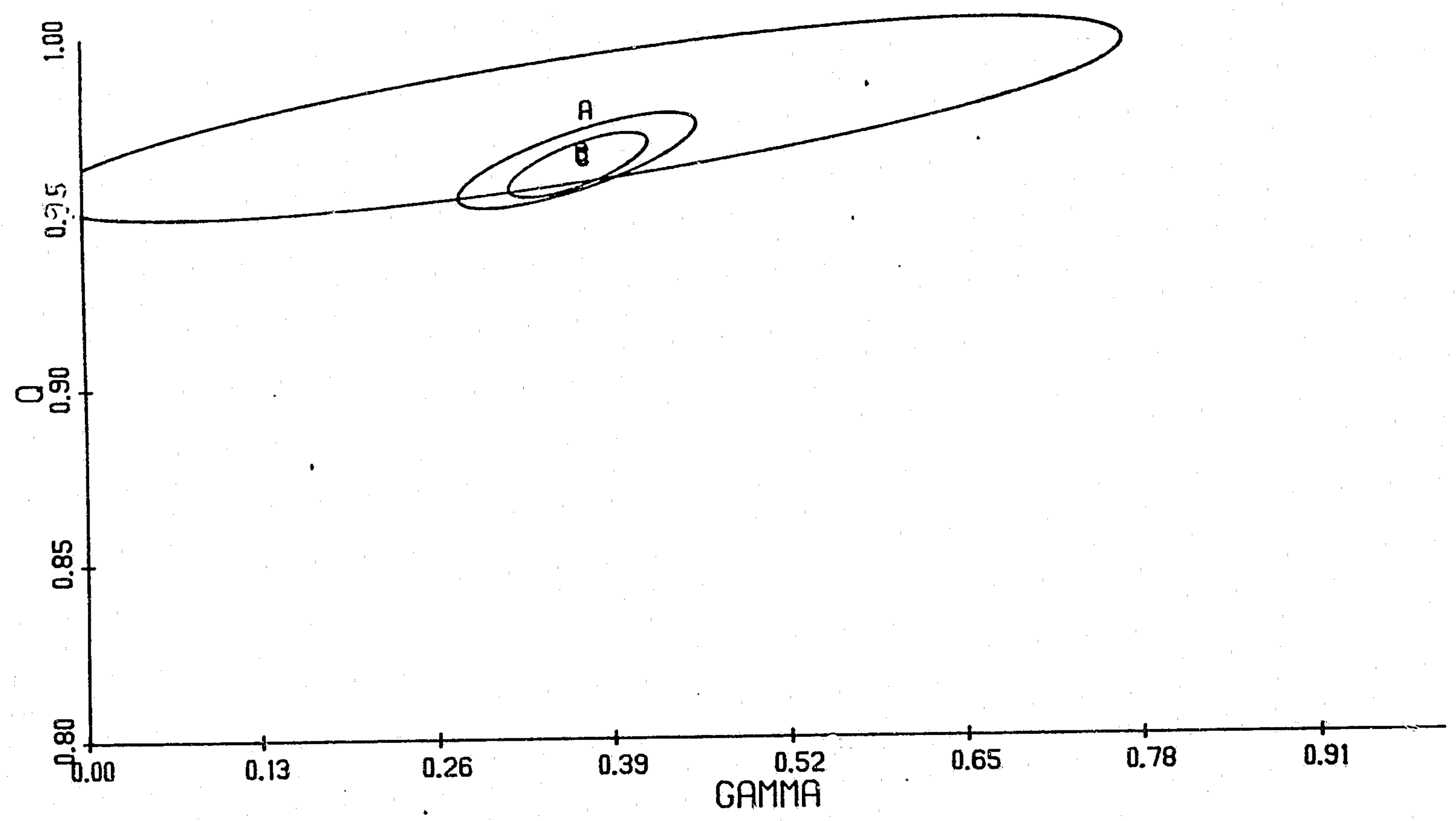


Figure 3f: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by EDUCATIONAL ATTAINMENT

A. 13 years or more
B. 8 years or less
C. 9-12 years

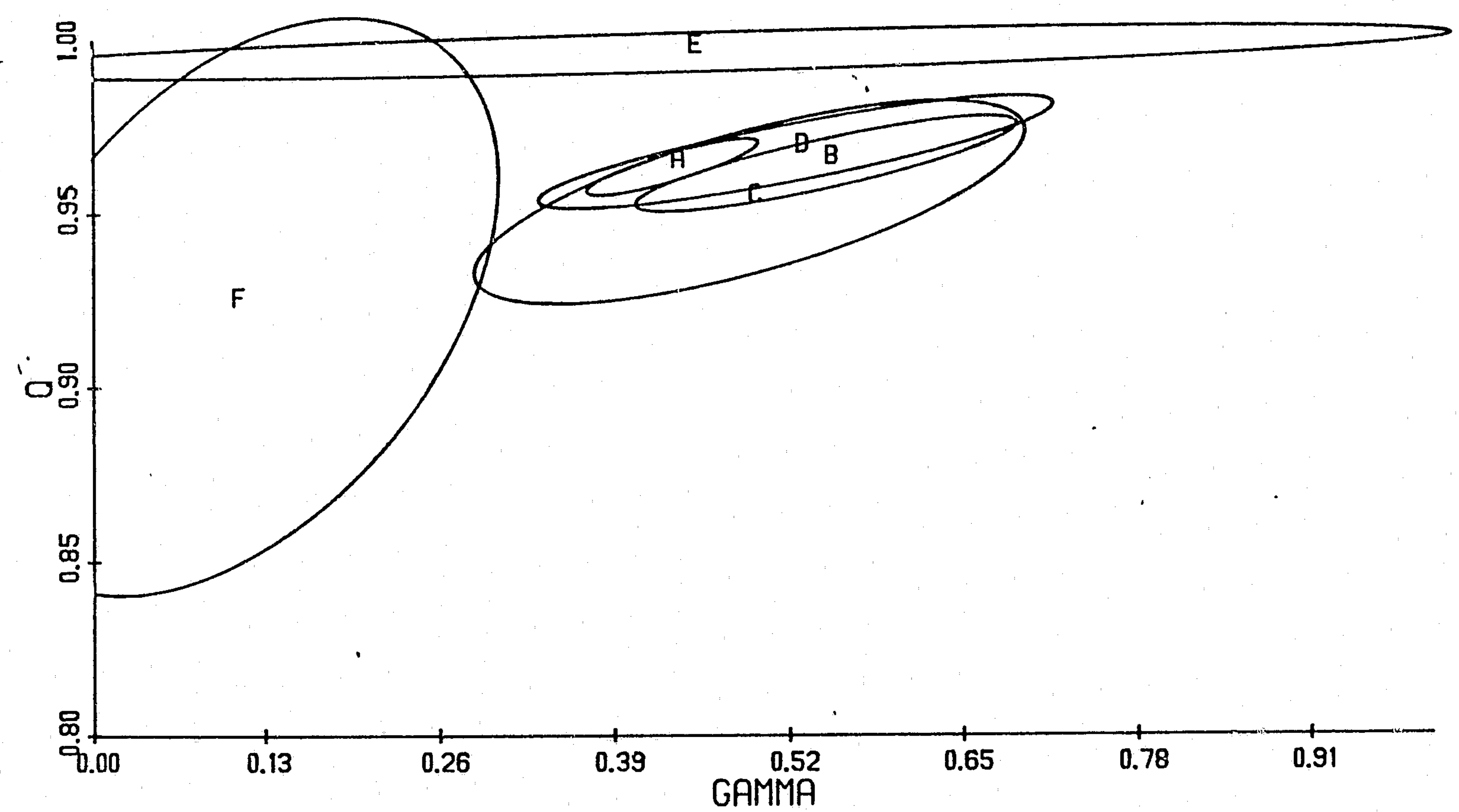


Figure 3g: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by TYPE OF ADMISSION

A. Direct CRT committment
B. Probation revocation
C. Parole revocation - NOA
D. Parole violation
E. Safekeeping or Evaluation
F. Other

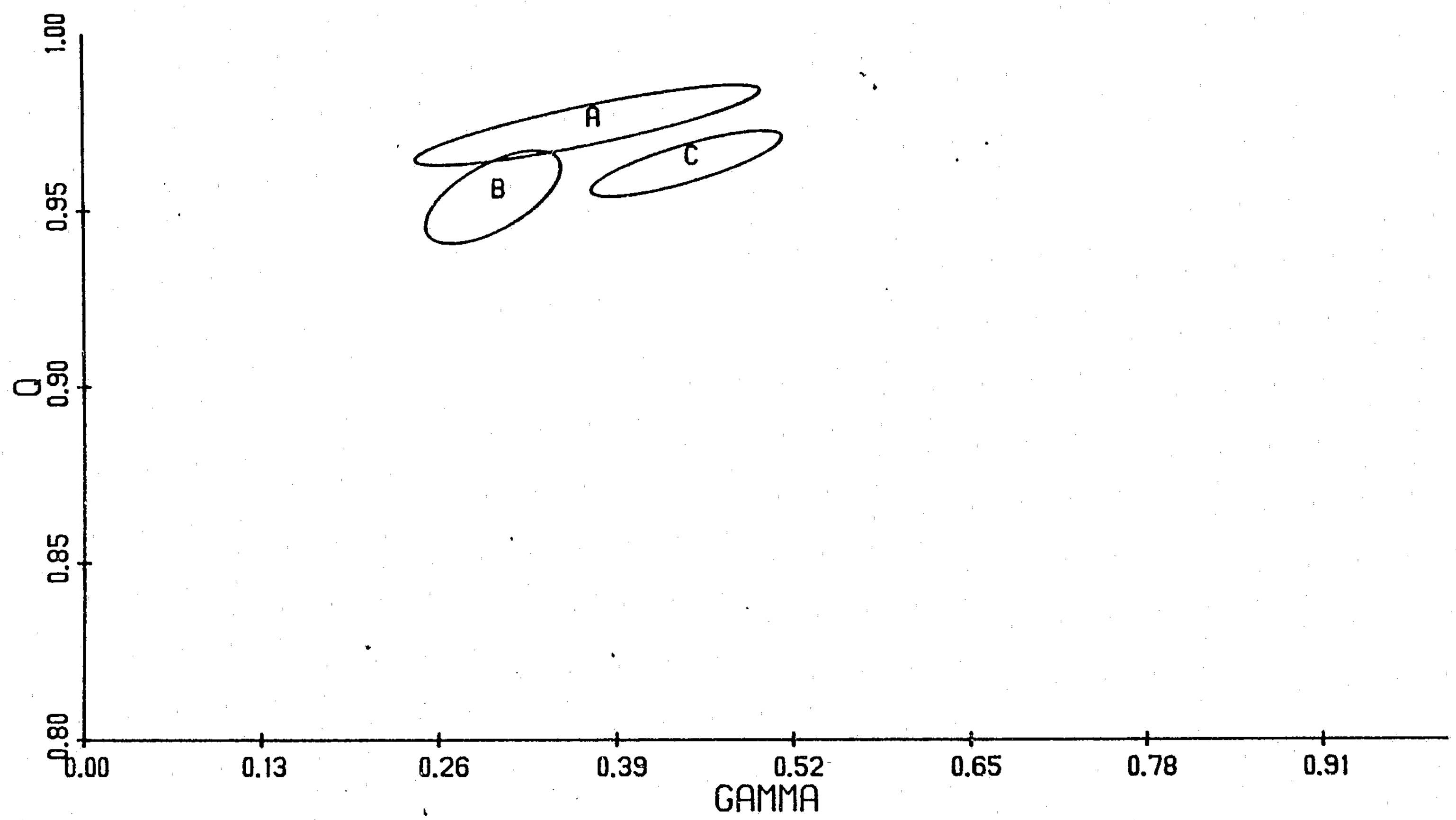


Figure 3h: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by ALCOHOL INVOLVEMENT
A. None
B. Under intoxication at arrest
C. History of alcoholism

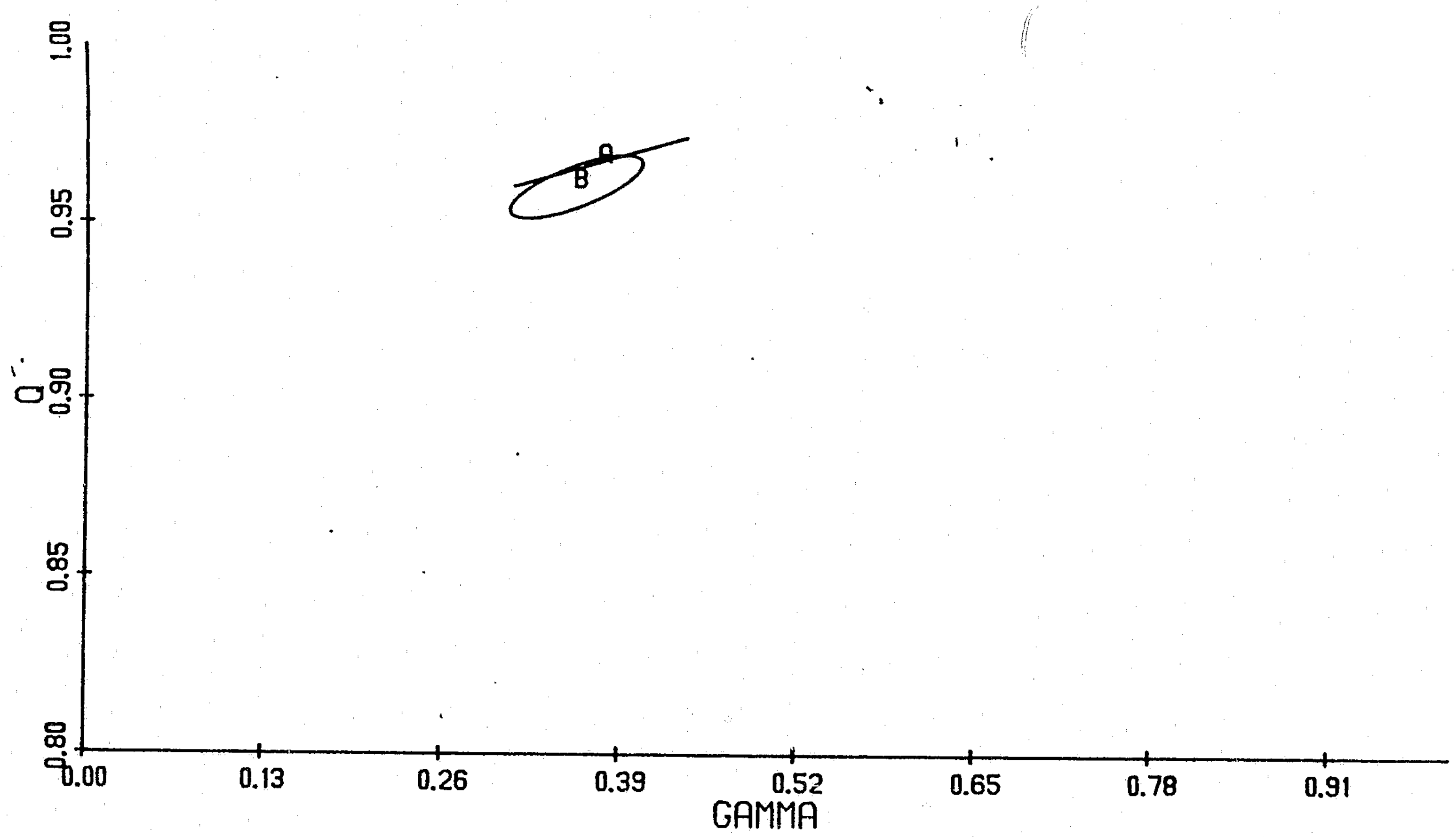


Figure 31: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by DRUG INVOLVEMENT
A. None
B. Some

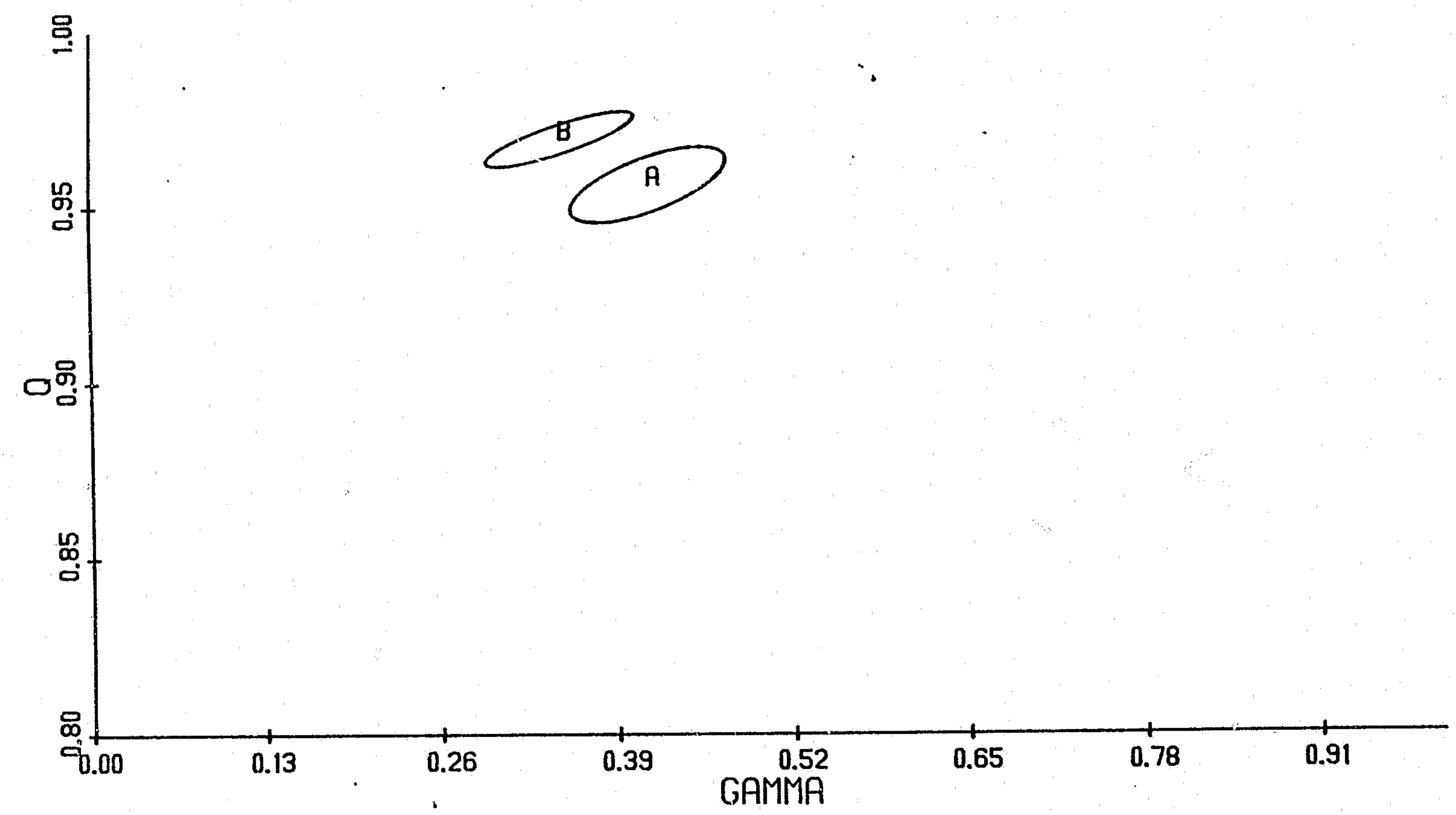


Figure 3j: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by JUVENILE COMMITMENTS
A. One or more
B. None

CONTINUED

3 OF 4

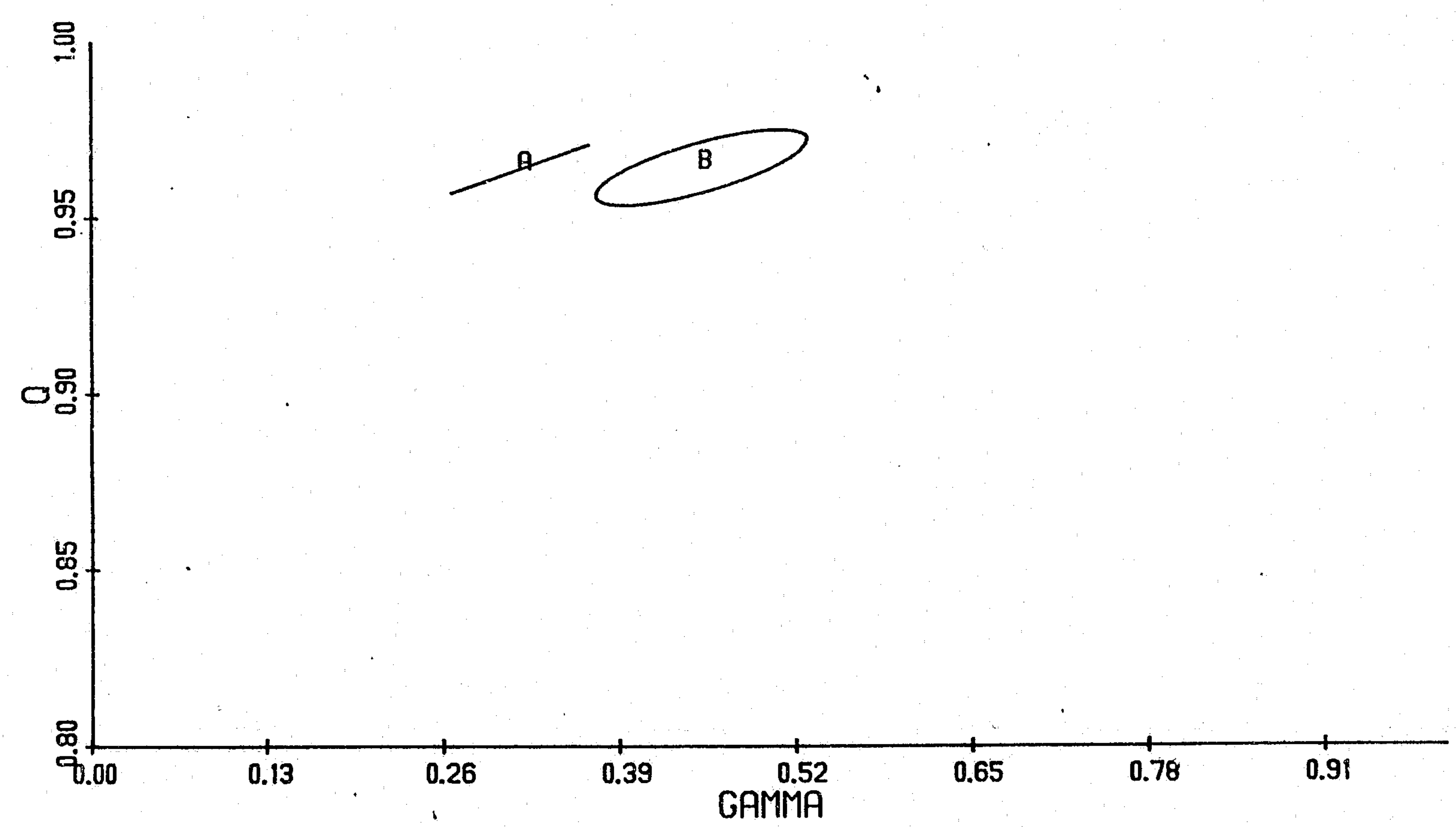


Figure 3k: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by PRIOR PRISON RECORD
A. None
B. One or more

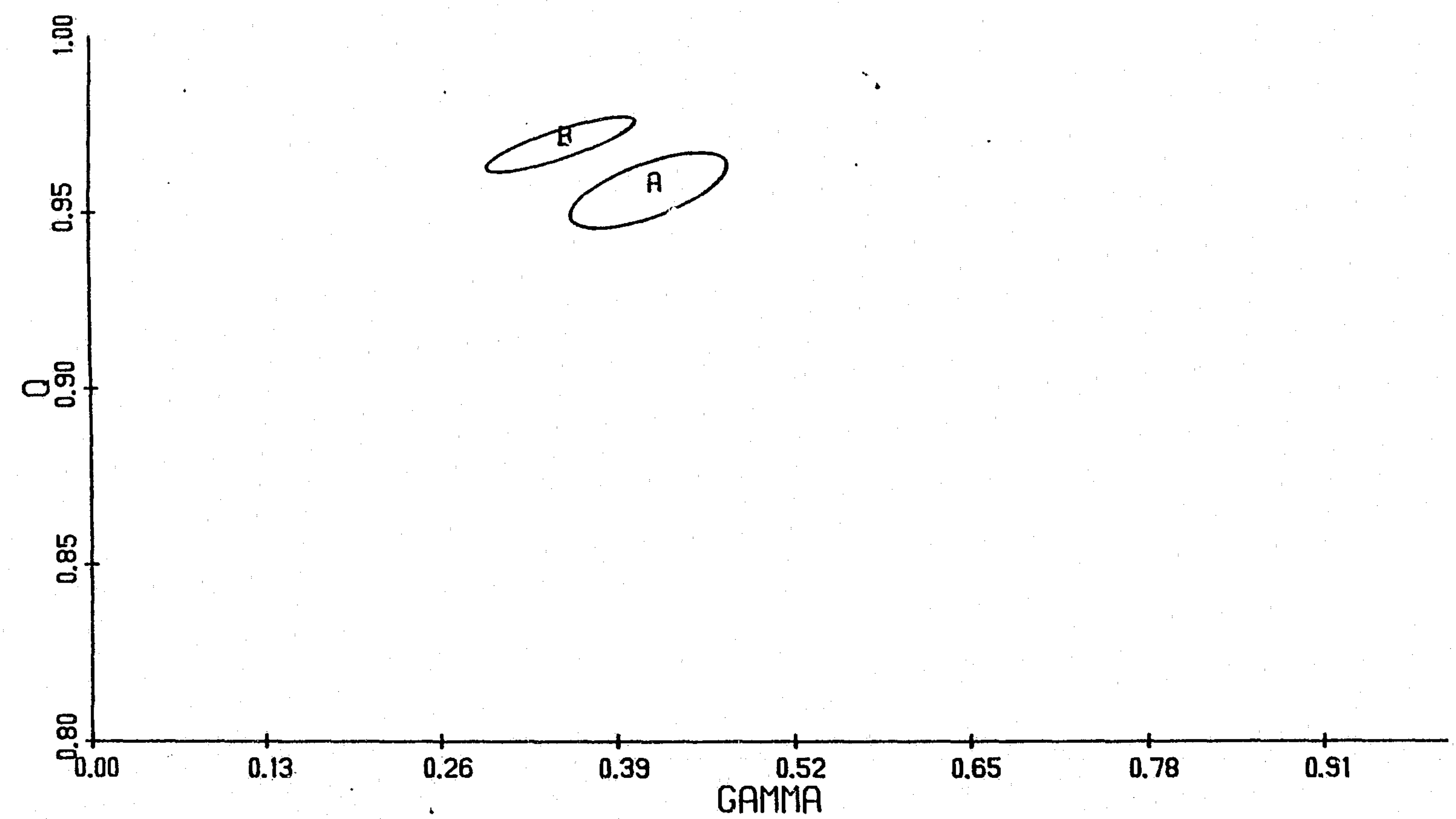


Figure 3j: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by JUVENILE COMMITMENTS
A. One or more
B. None

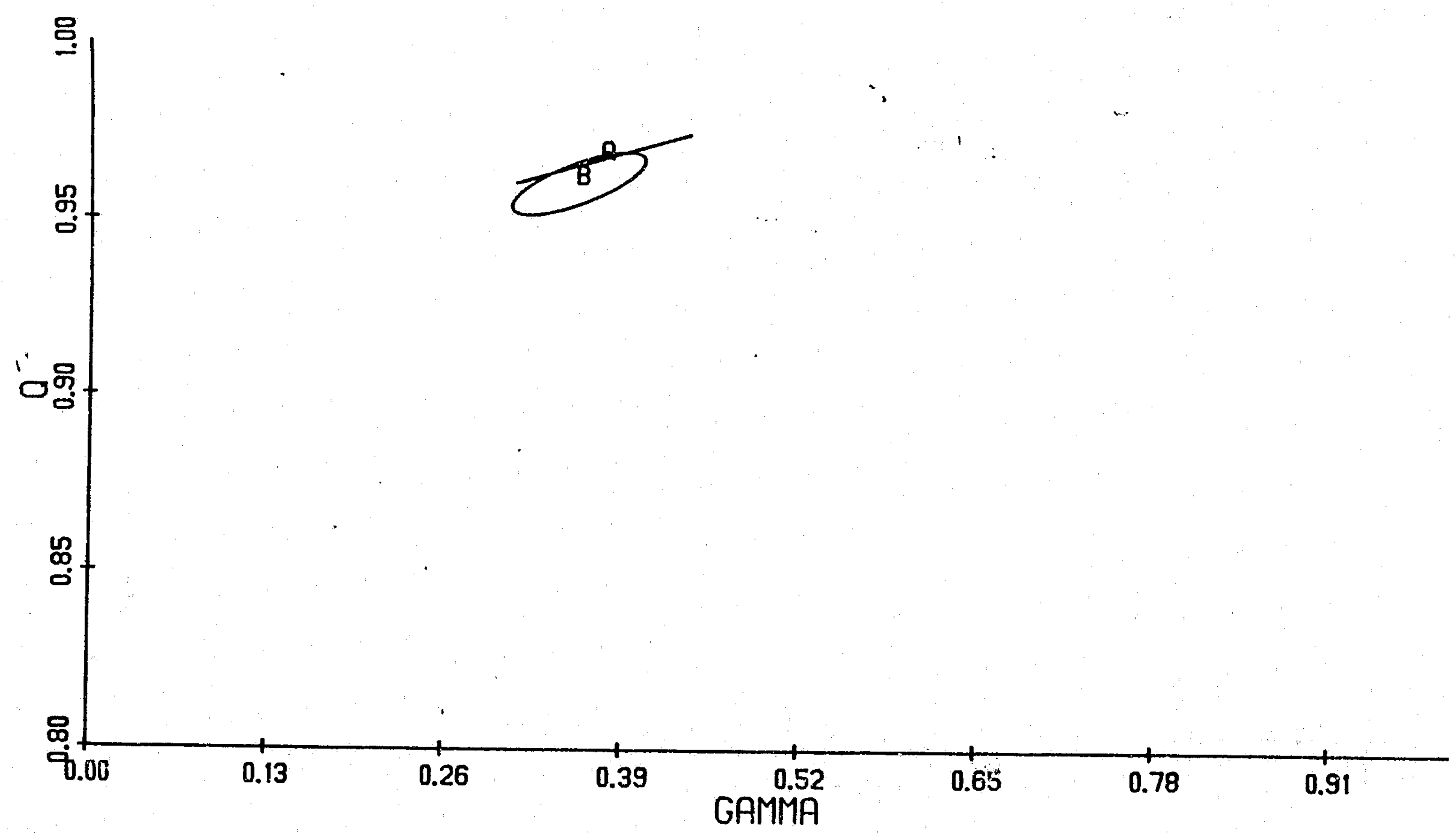


Figure 3i: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by DRUG INVOLVEMENT
A. None
B. Some

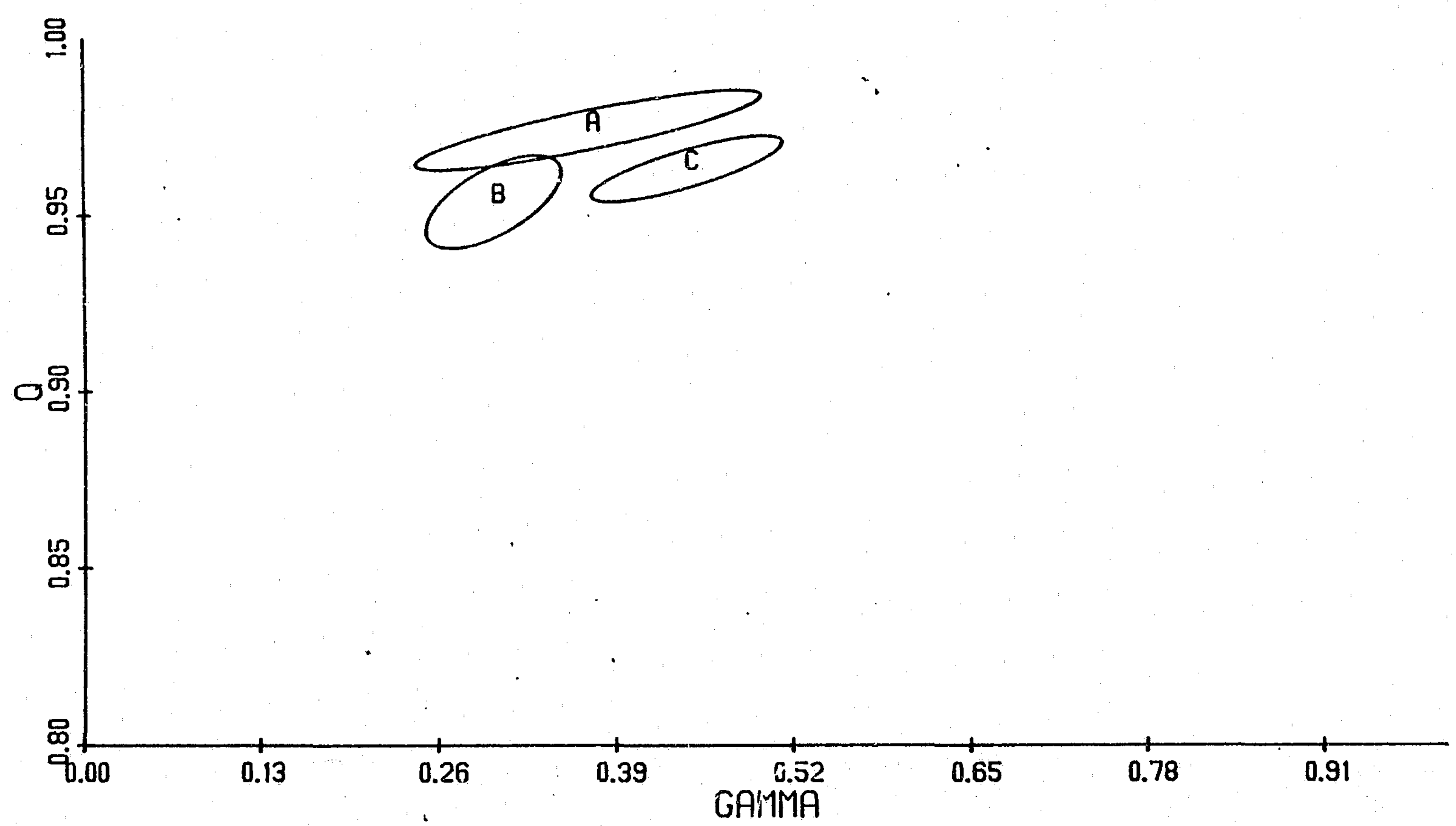


Figure 3h: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by ALCOHOL INVOLVEMENT
A. None
B. Under intoxication at arrest
C. History of alcoholism

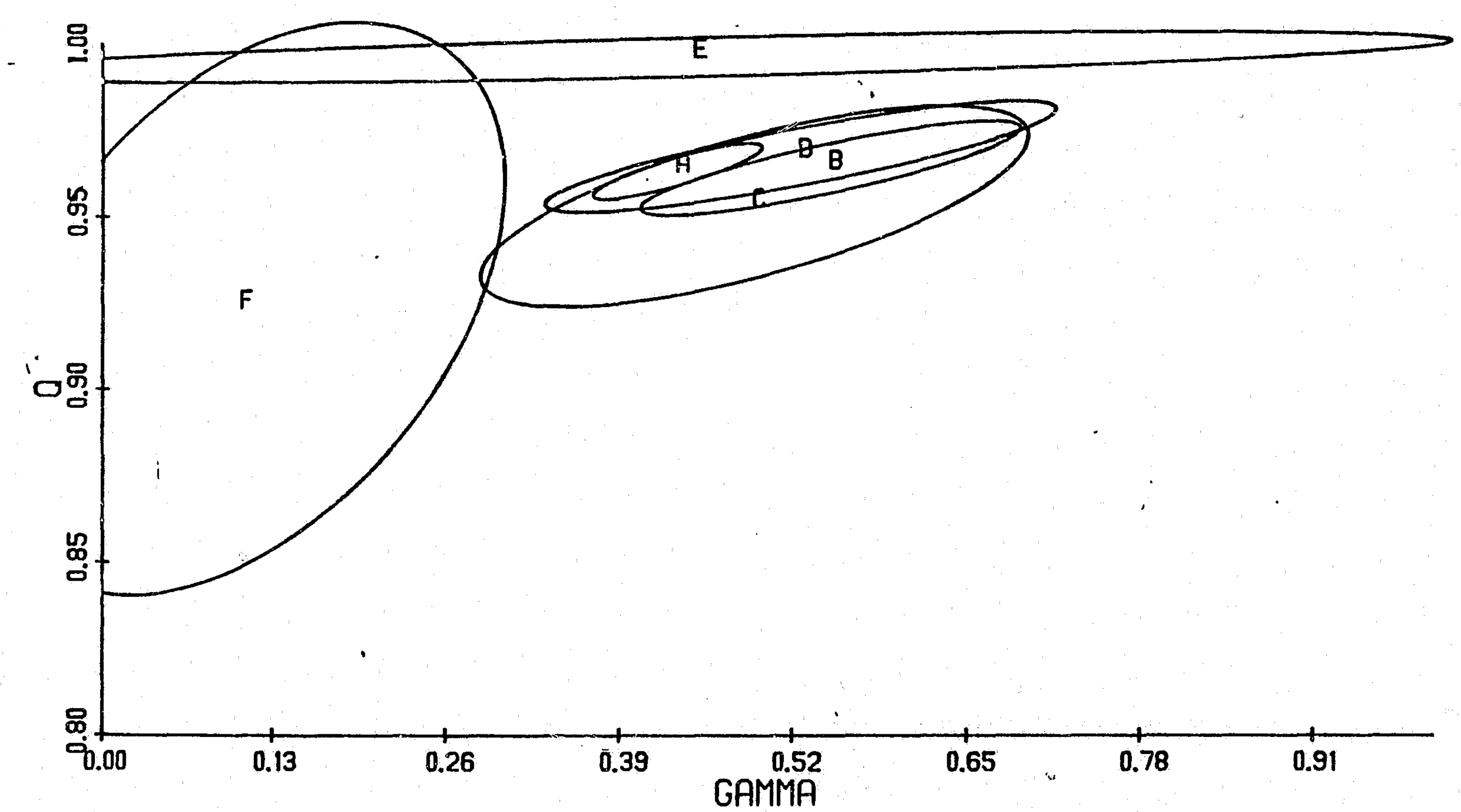


Figure 3g: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by TYPE OF ADMISSION

A. Direct CRT committment
B. Probation revocation
C. Parole revocation - NOA
D. Parole violation
E. Safekeeping or Evaluation
F. Other

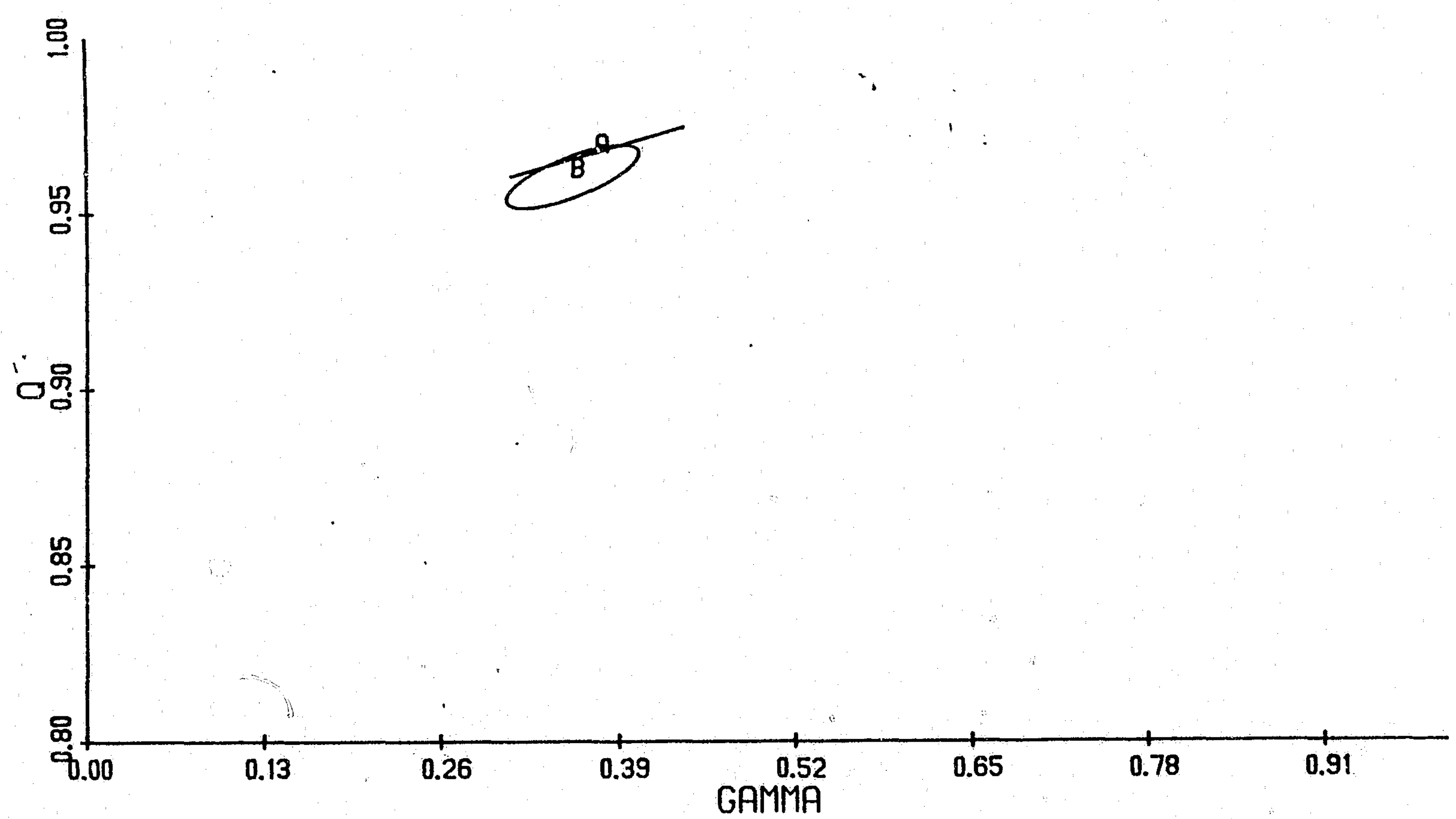


Figure 3i: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by DRUG INVOLVEMENT
A. None
B. Some

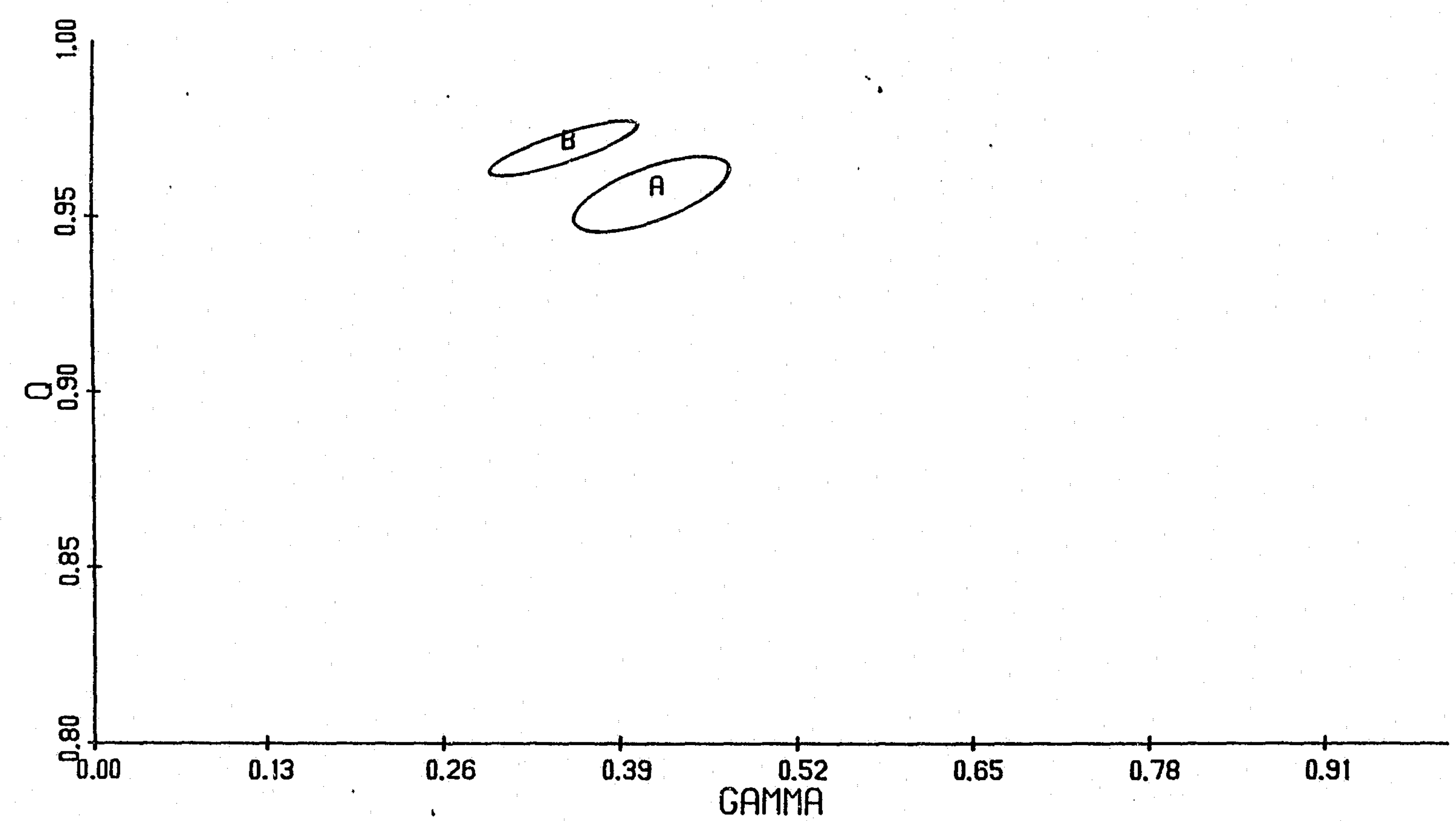


Figure 3j: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by JUVENILE COMMITMENTS
A. One or more
B. None

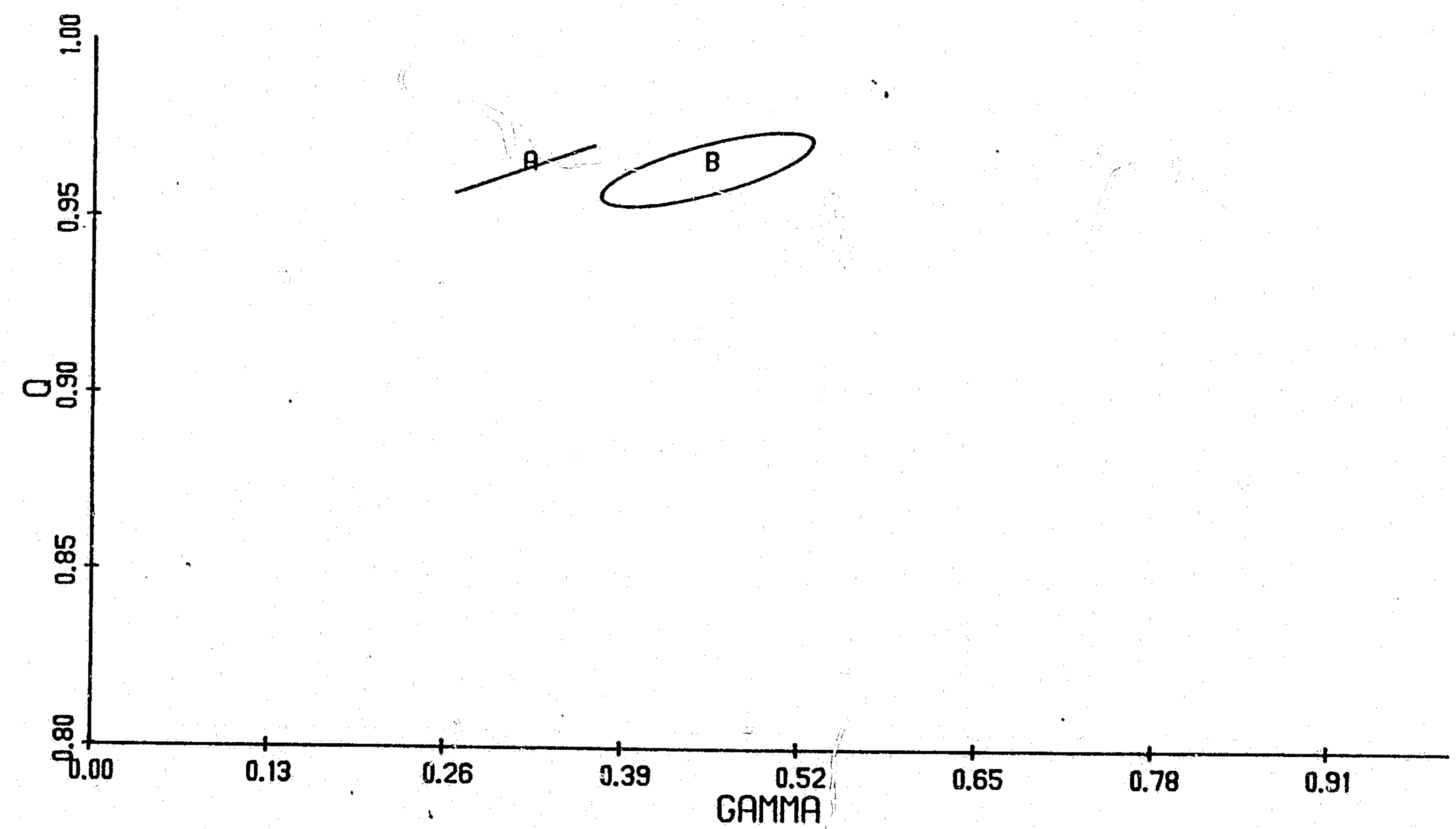


Figure 3k: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by PRIOR PRISON RECORD
A. None
B. One or more

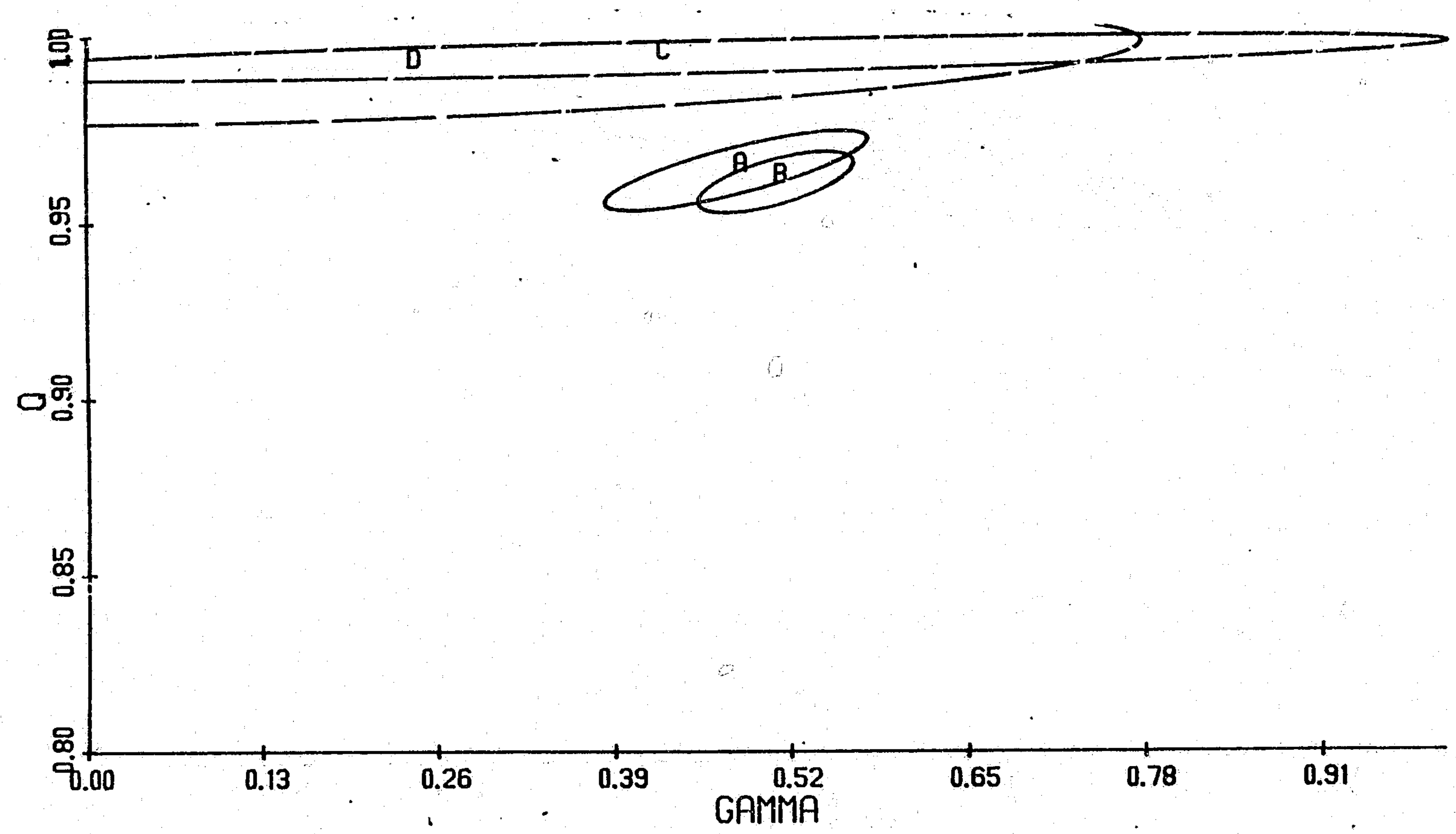


Figure 31: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by TYPE OF RELEASE
A. Expiration of sentence
B. Parole
C. Safekeeping or evaluation
D. Other

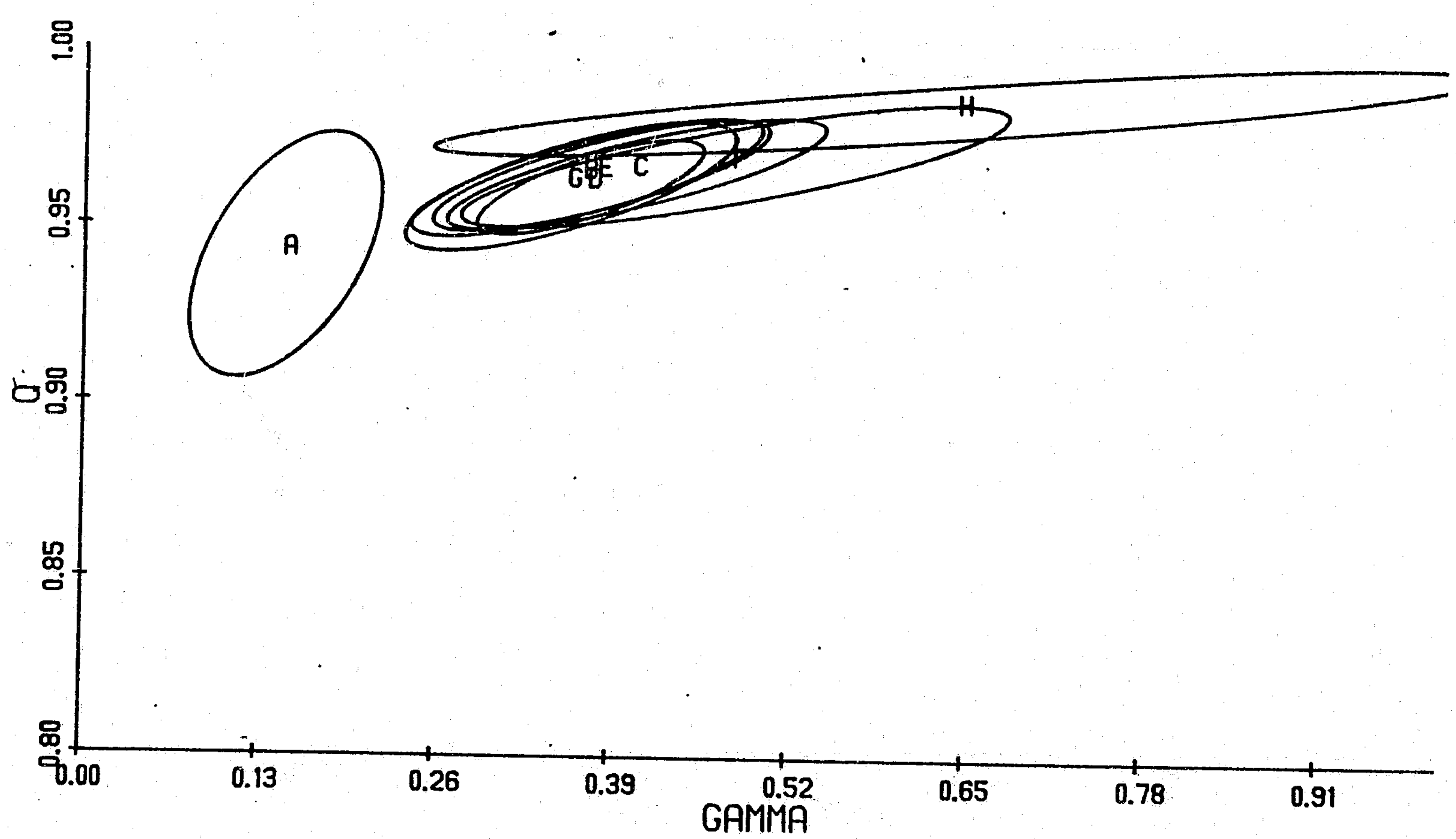


Figure 3m: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by AGE AT RELEASE

A. 19 or less
B. 20-21
C. 22-23
D. 24-26
E. 27-29
F. 30-35
G. 36-46
H. 47 and over

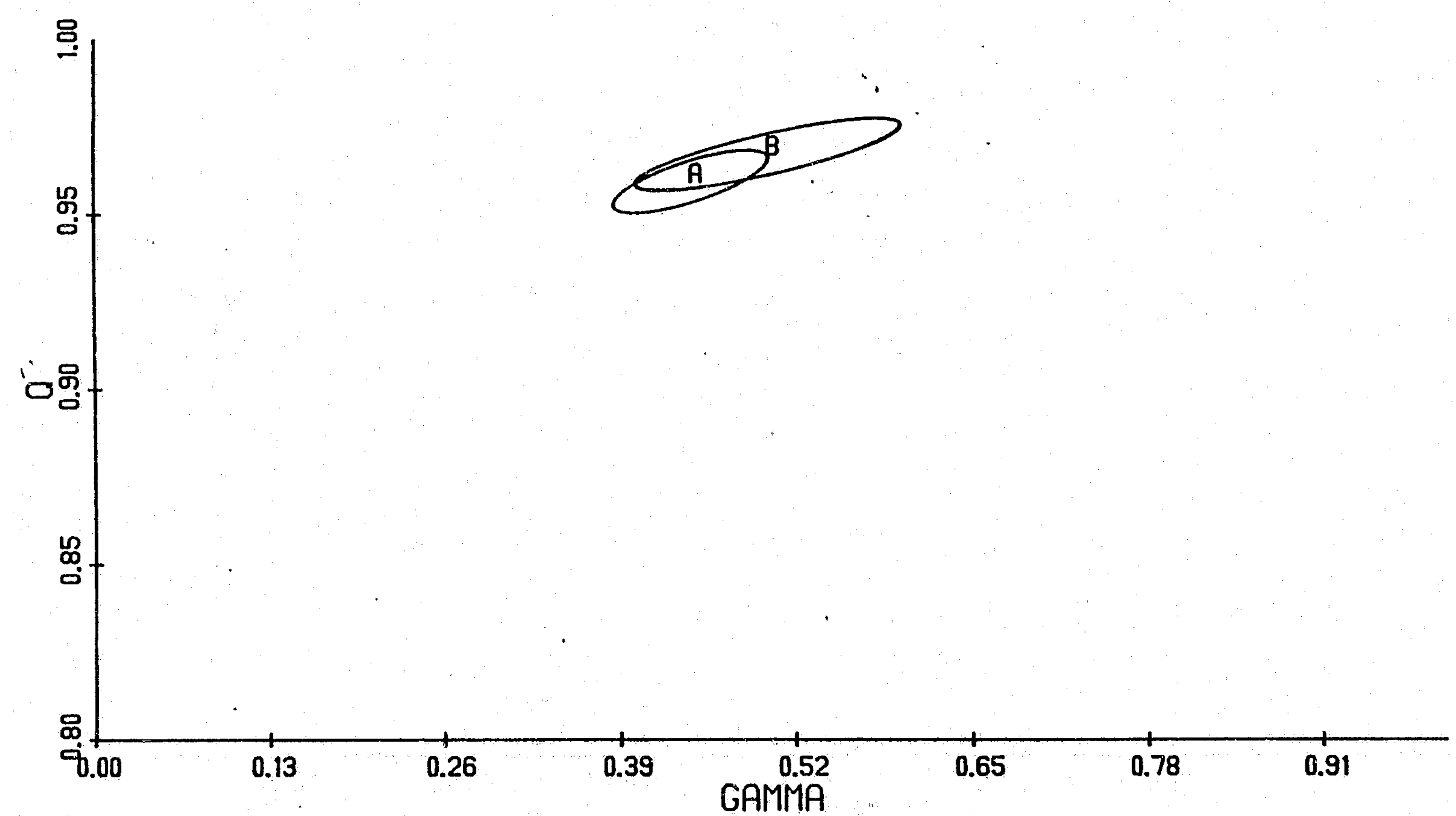


Figure 3n: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by OCCUPATION AT ADMISSION
A. None or unskilled
B. Skilled or higher

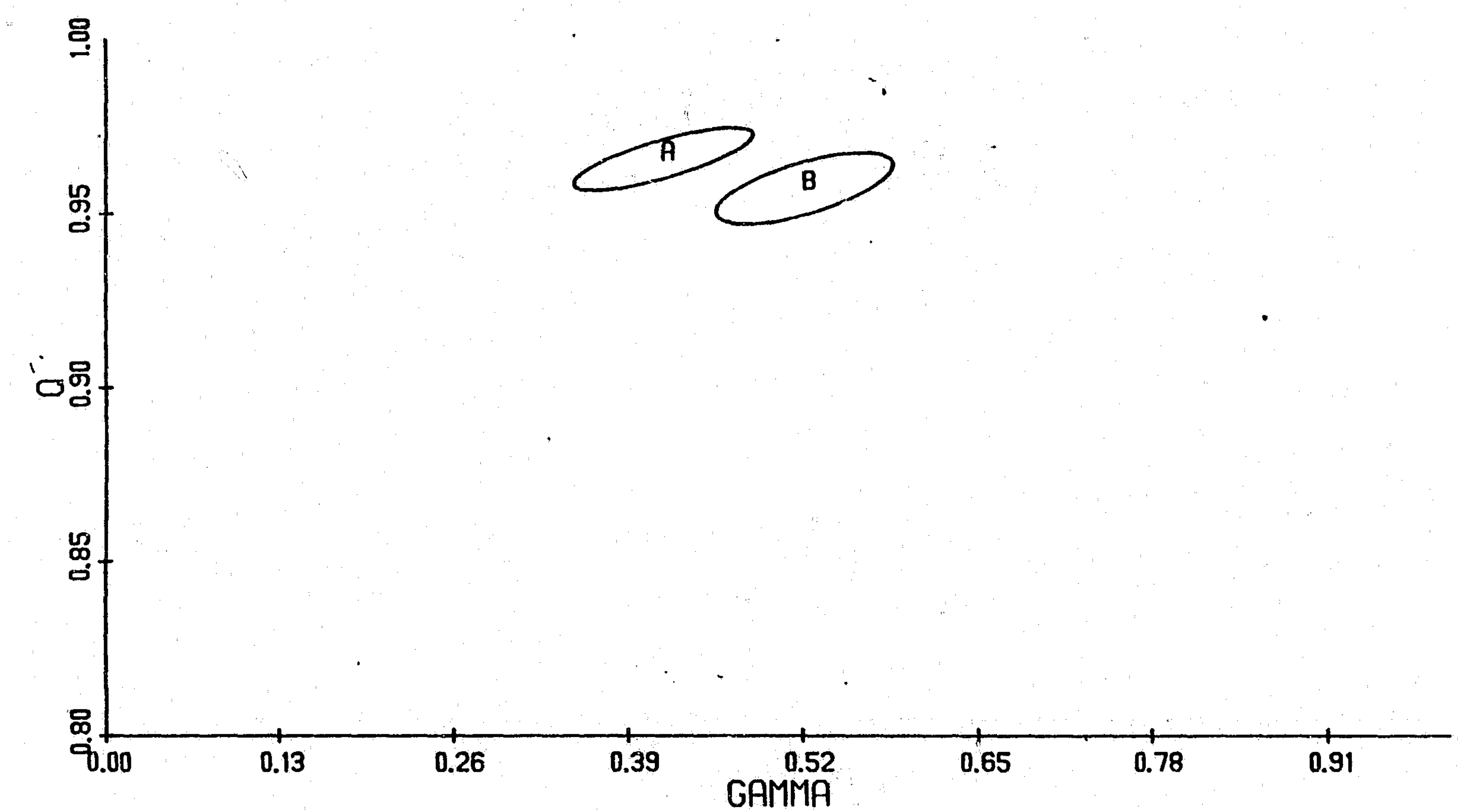


Figure 2o: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by PRIOR JUVENILE ARRESTS
A. None
B. One or more

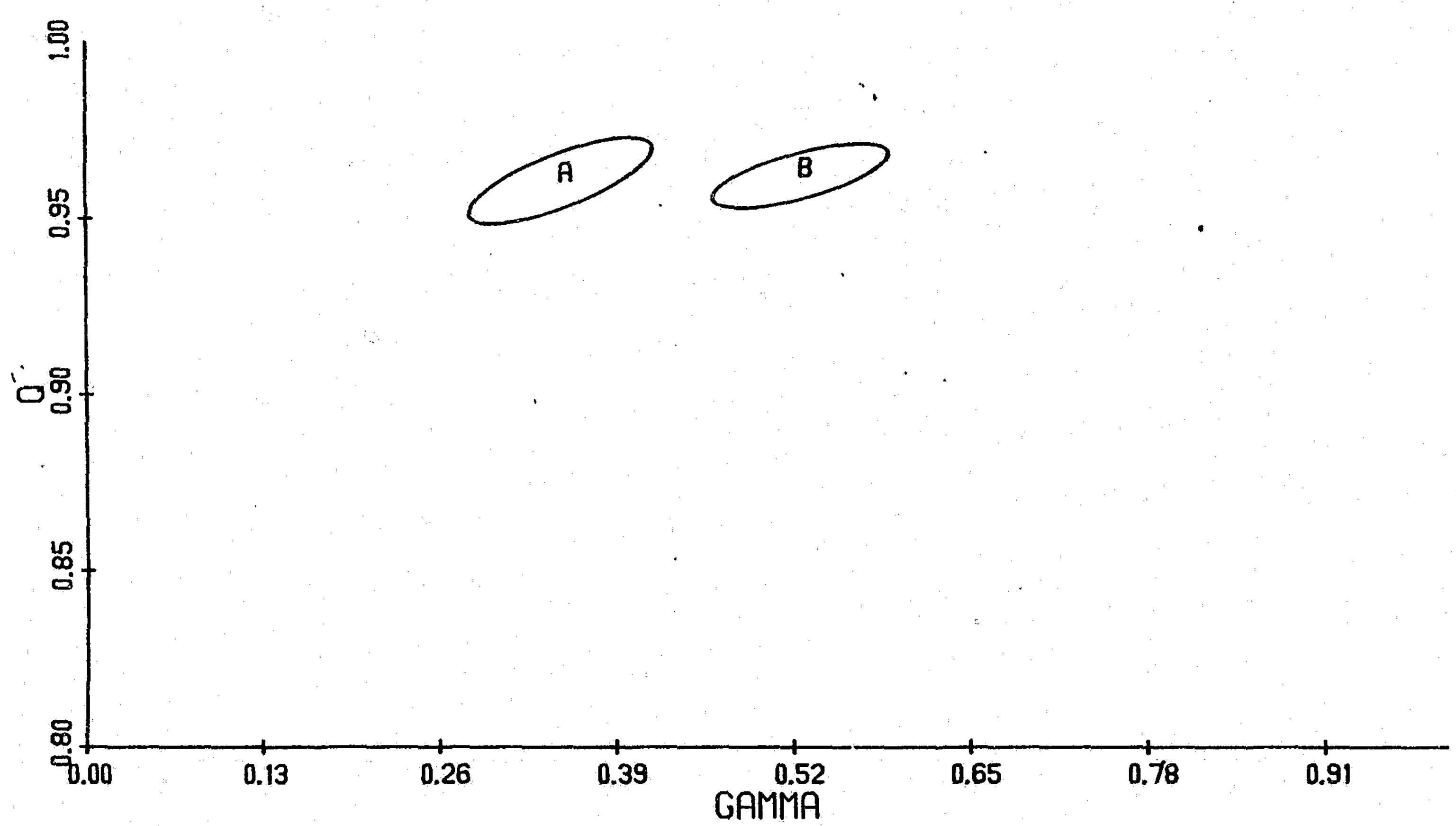


Figure 2p: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by PRIOR ARRESTS
A. None
B. One or more

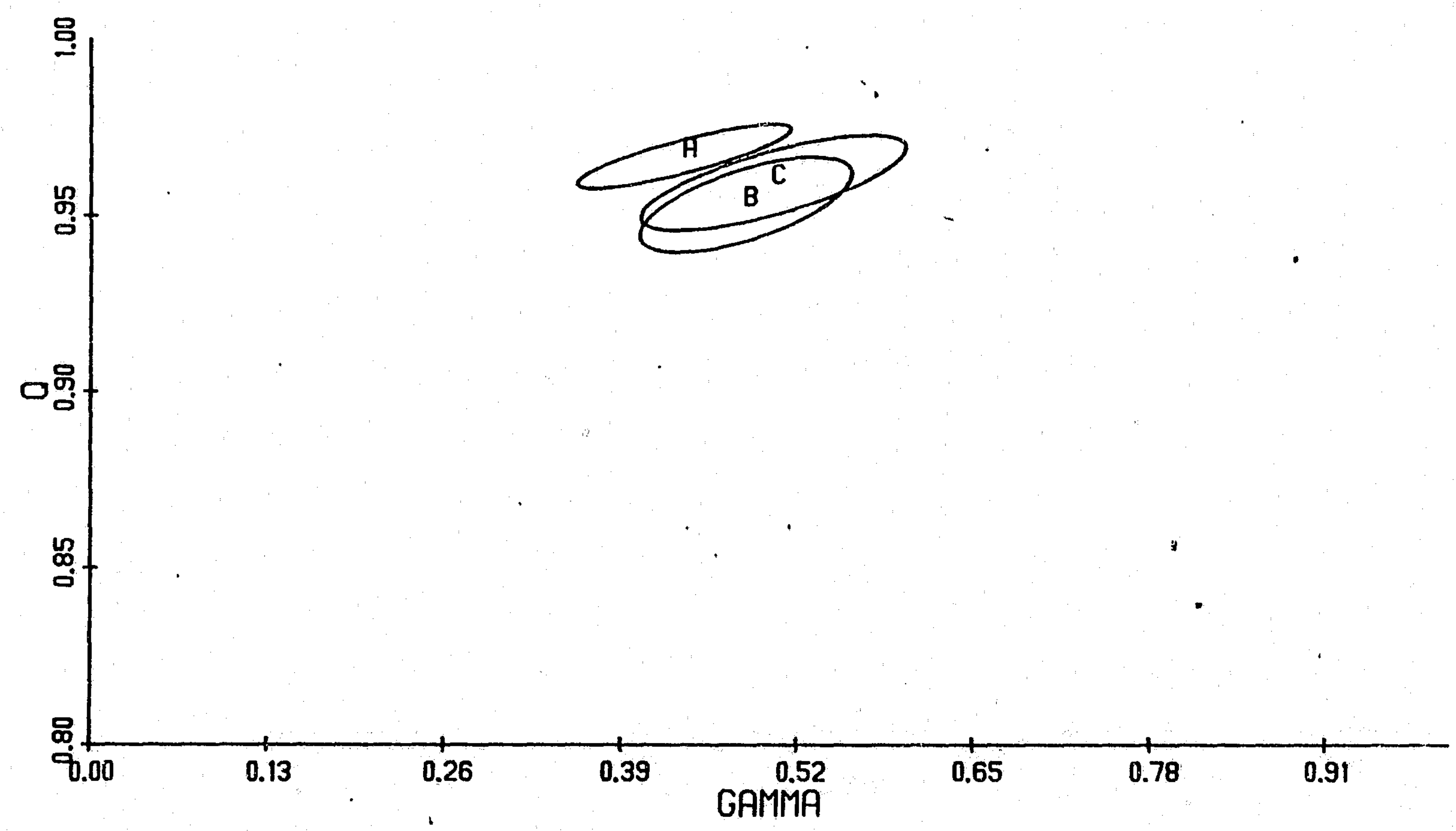


Figure 2q: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by PRIOR FELONY CONVICTIONS

A. None
B. One
C. Two or more

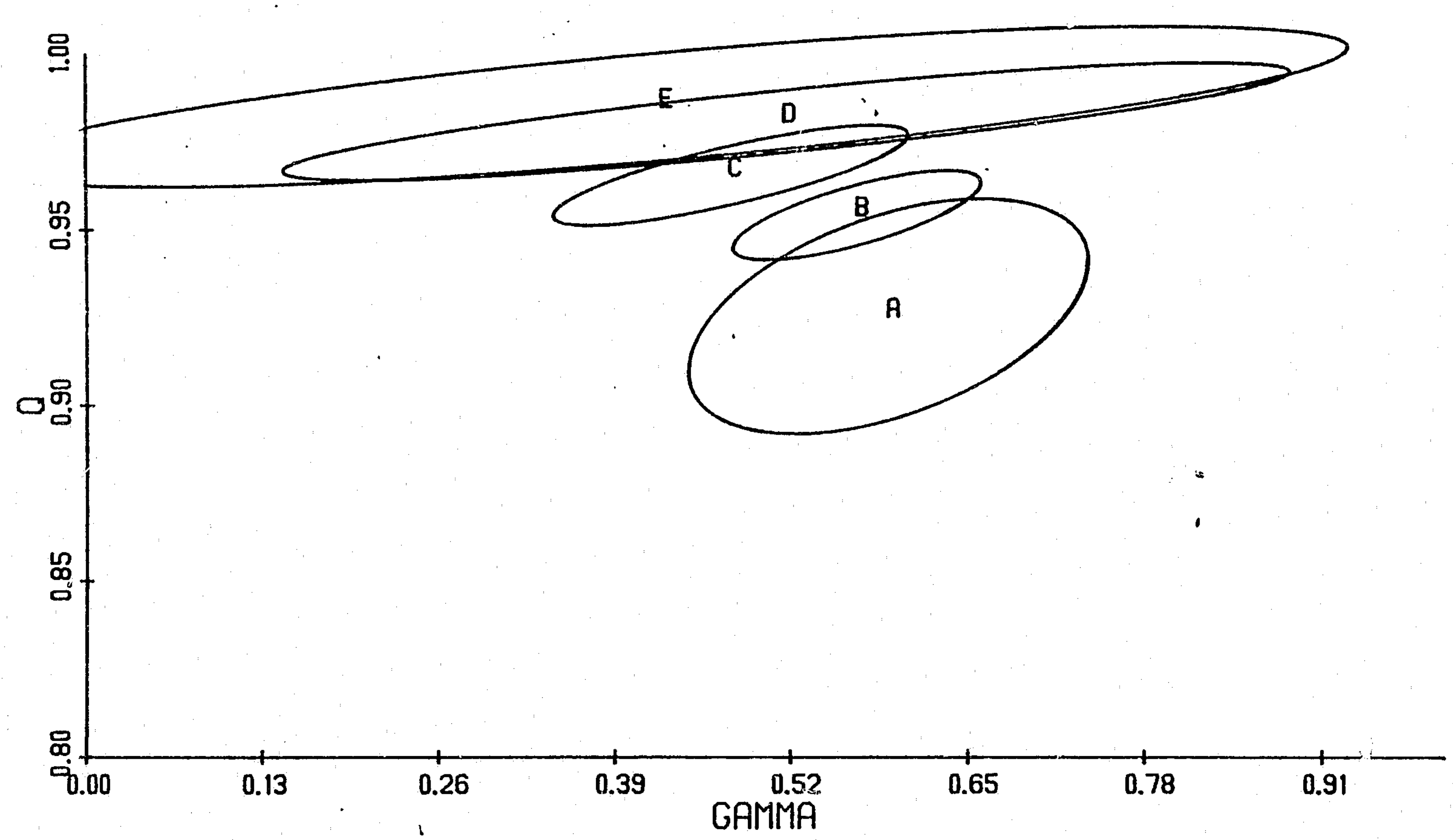


Figure 2r: Iowa Cohort
Bayes 90% Confidence Regions for estimates of γ and q by PAROLE RISK AT ADMISSION

A. Ultra High
B. High
C. Medium
D. Low
E. Nil

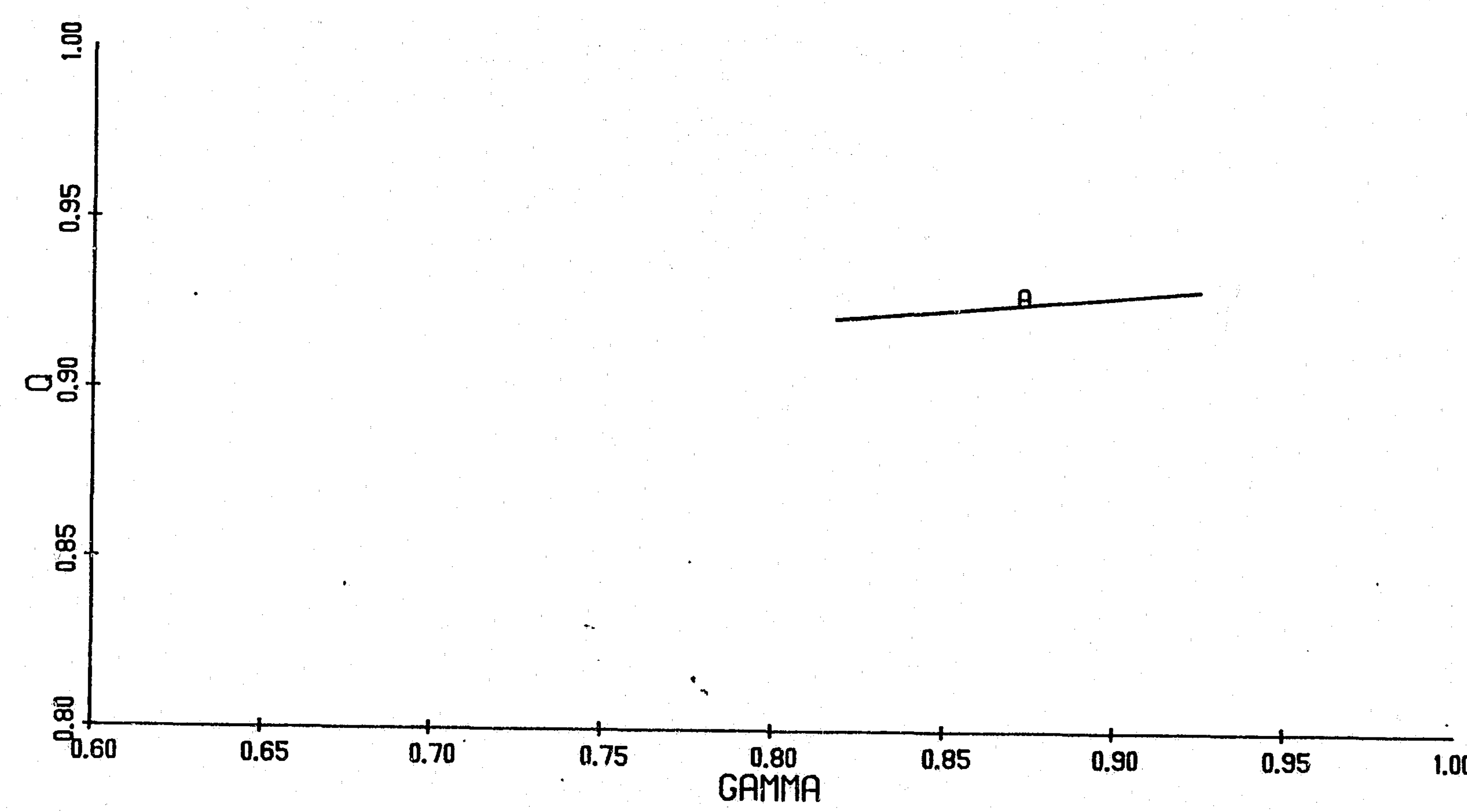


Figure 4a: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q.

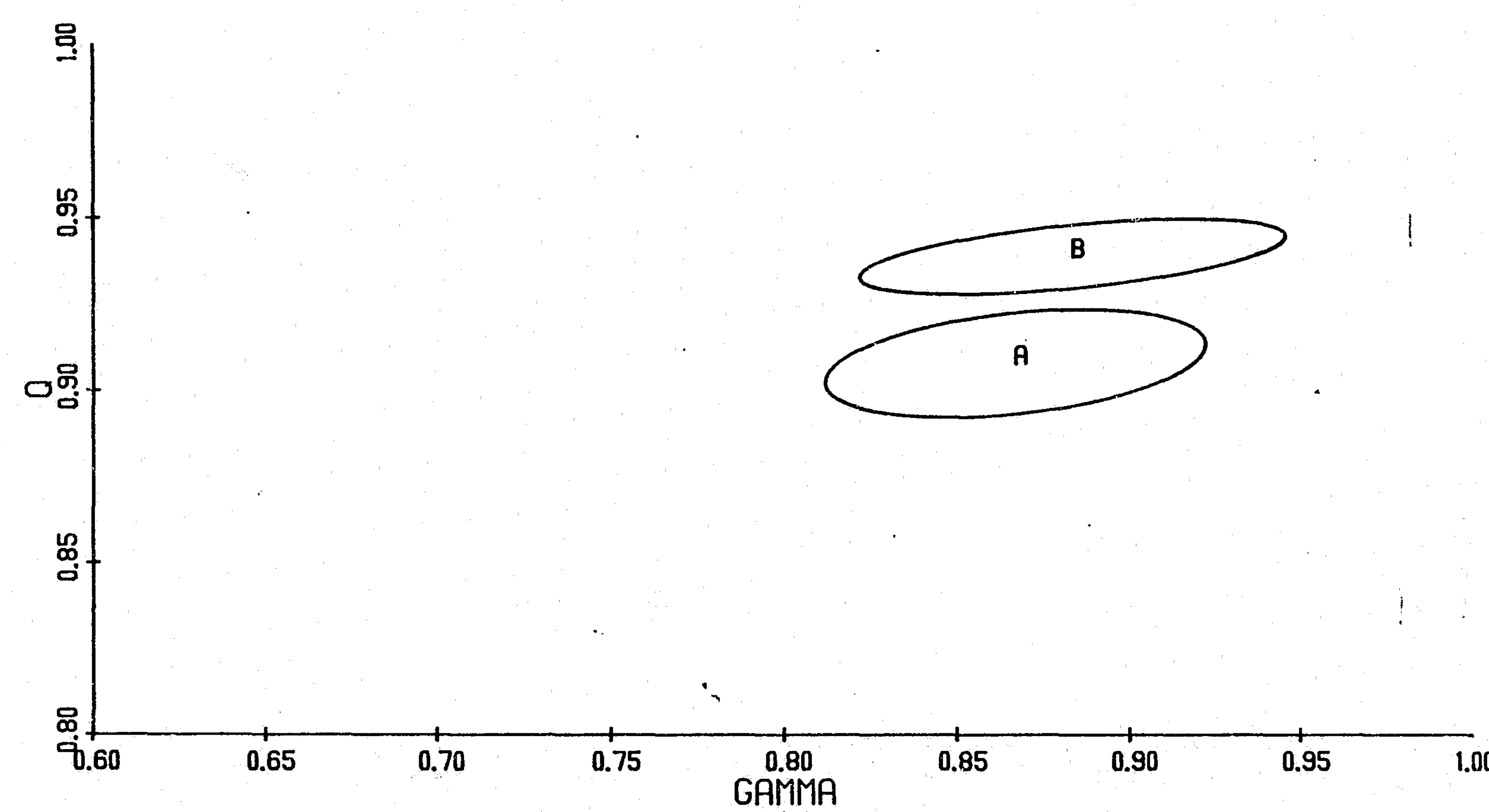


Figure 4b: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by RACE
A. White
B. Black or Indian

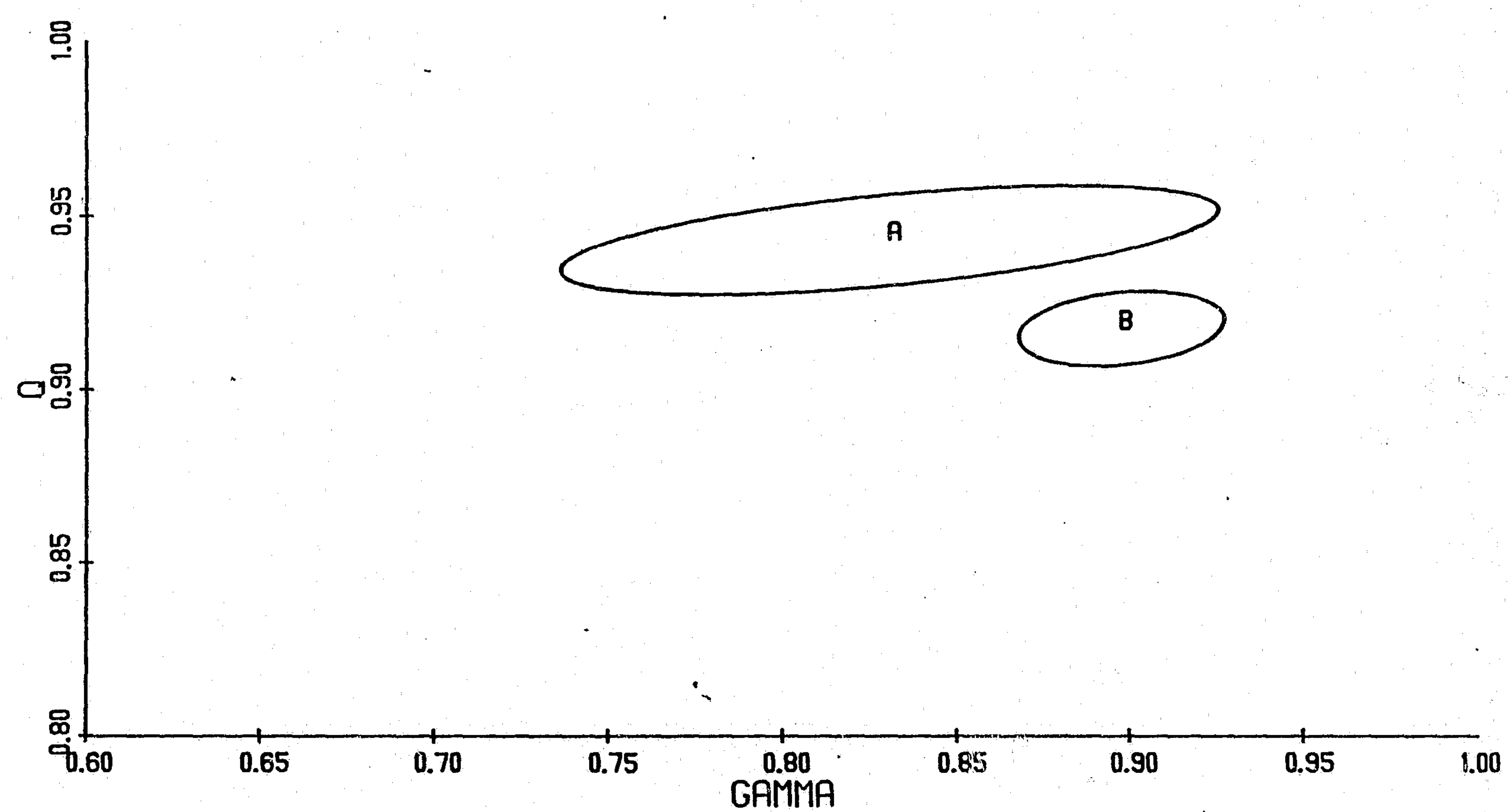


Figure 4c: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by TYPE OF RELEASE
A. Other
B. Unconditional

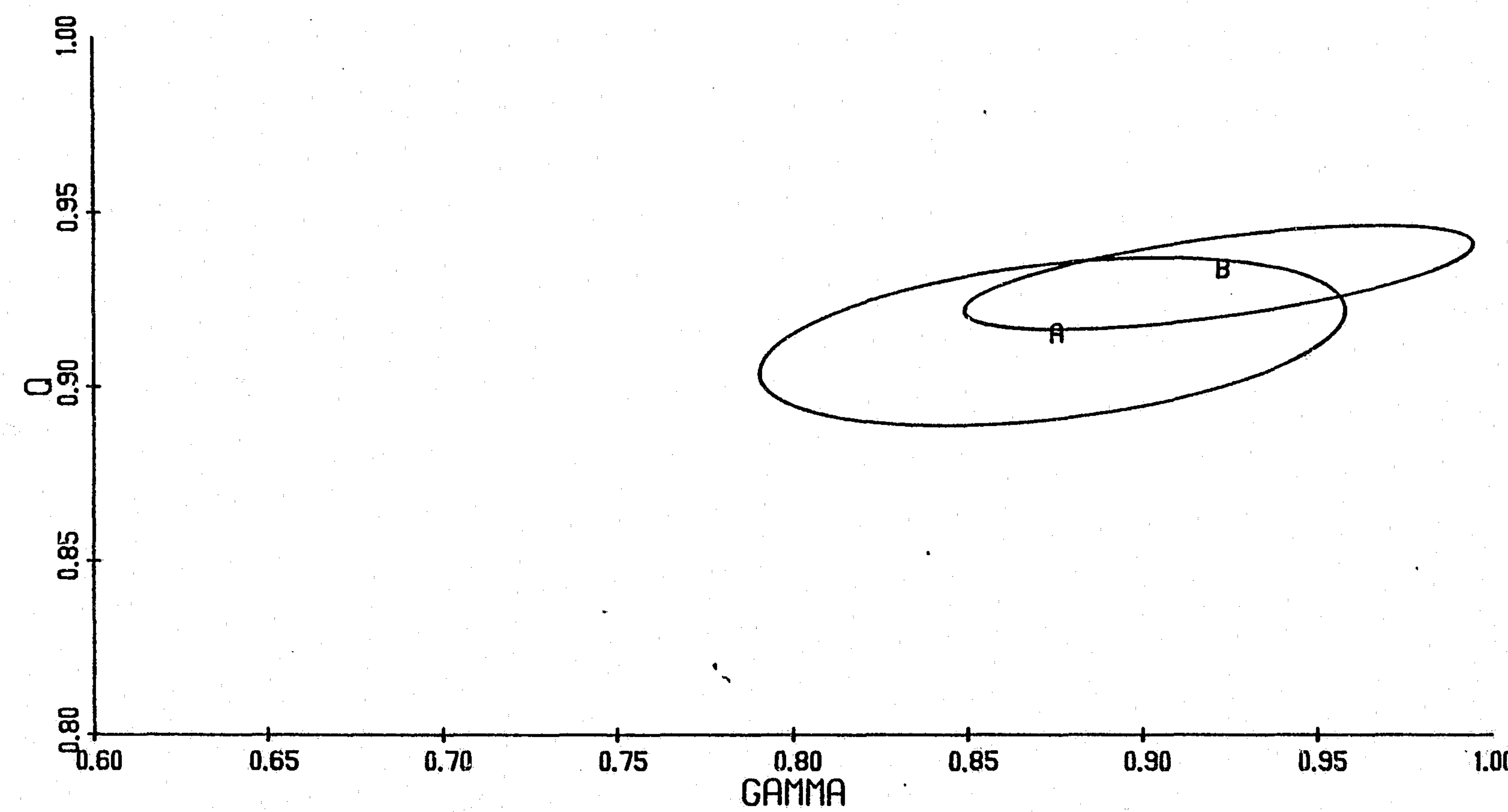


Figure 4d: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by IQ
  A. 100 or less
  B. Greater than 100

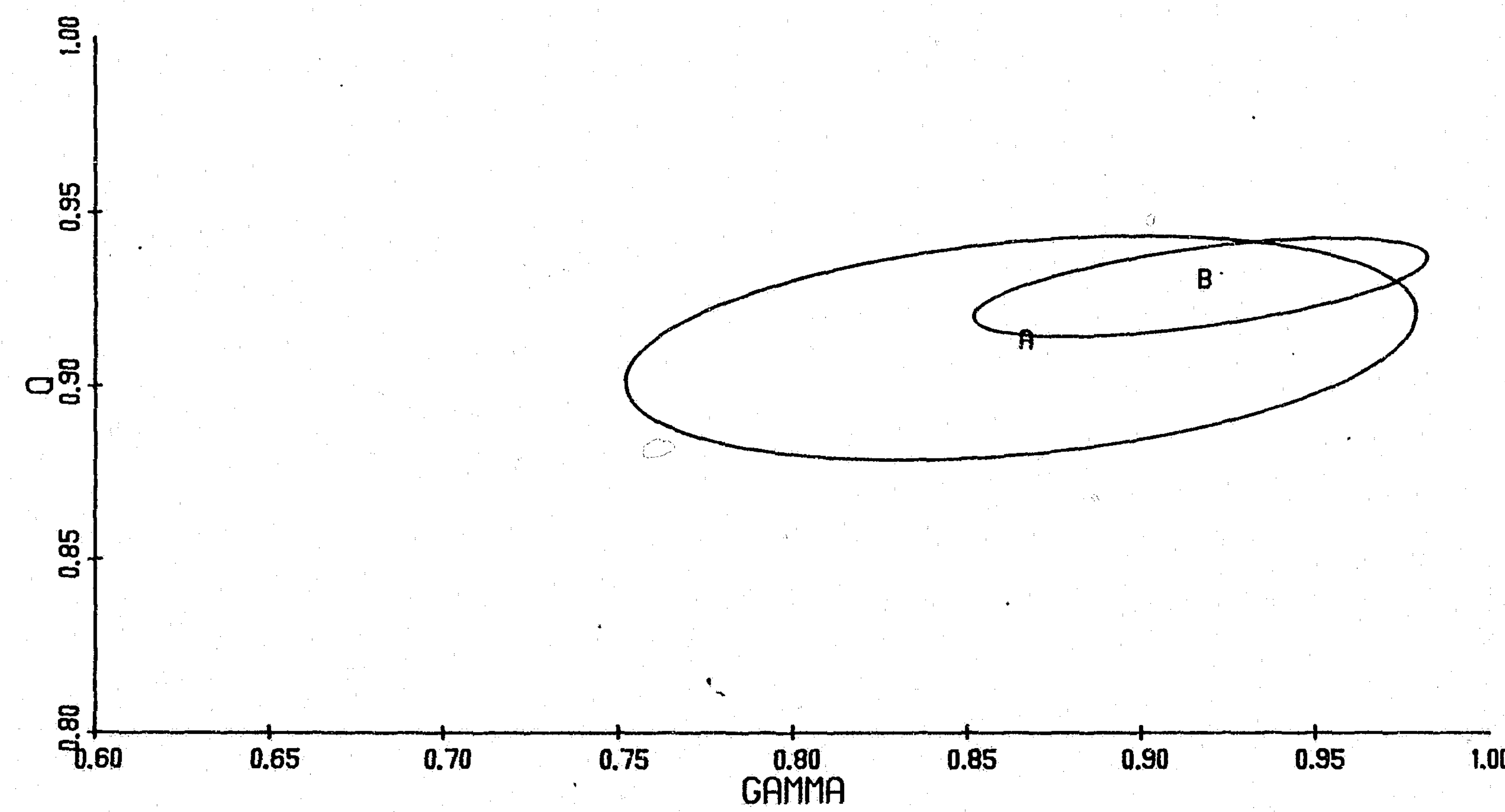


Figure 4e: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by SCHOOL ACHIEVEMENT
  A. 8 years or less
  B. 9 years or more

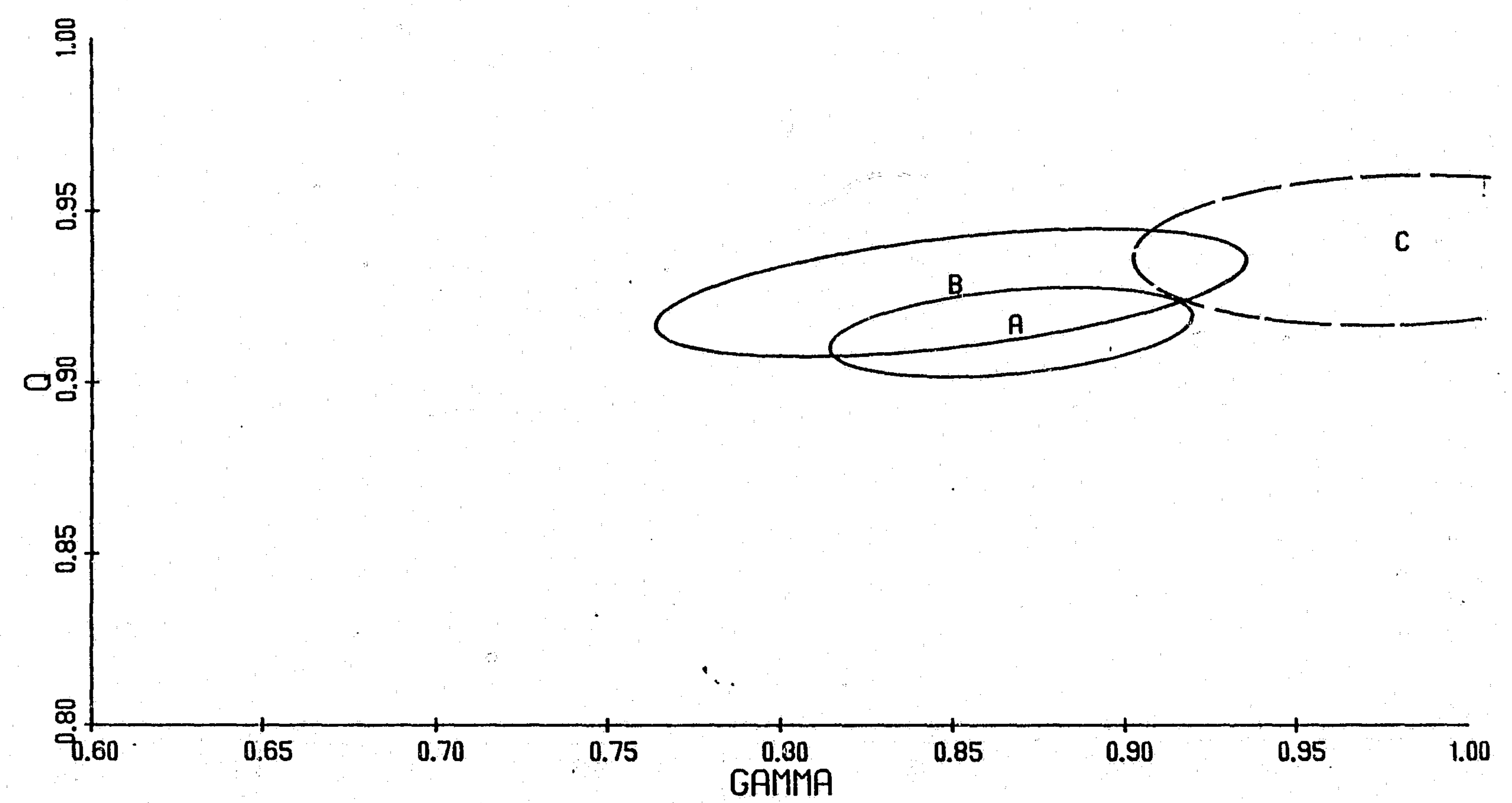


Figure 4f: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by
WORK STABILITY 5 YEARS PRIOR TO SAMPLE TERM
A. Other
B. Two or fewer job changes
C. Student

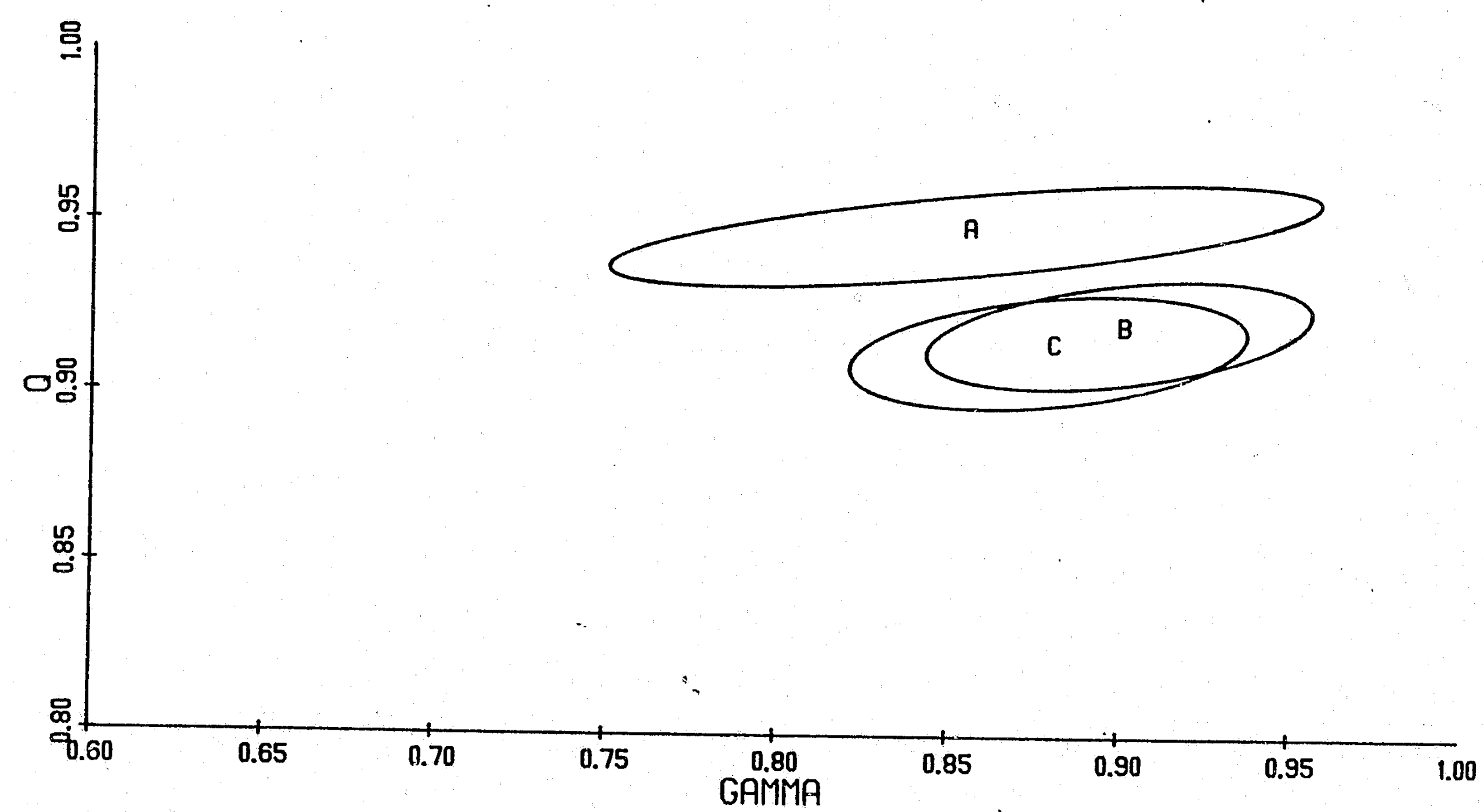


Figure 4g: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by PRIOR ARRESTS
A. None
B. One or two
C. Three or more

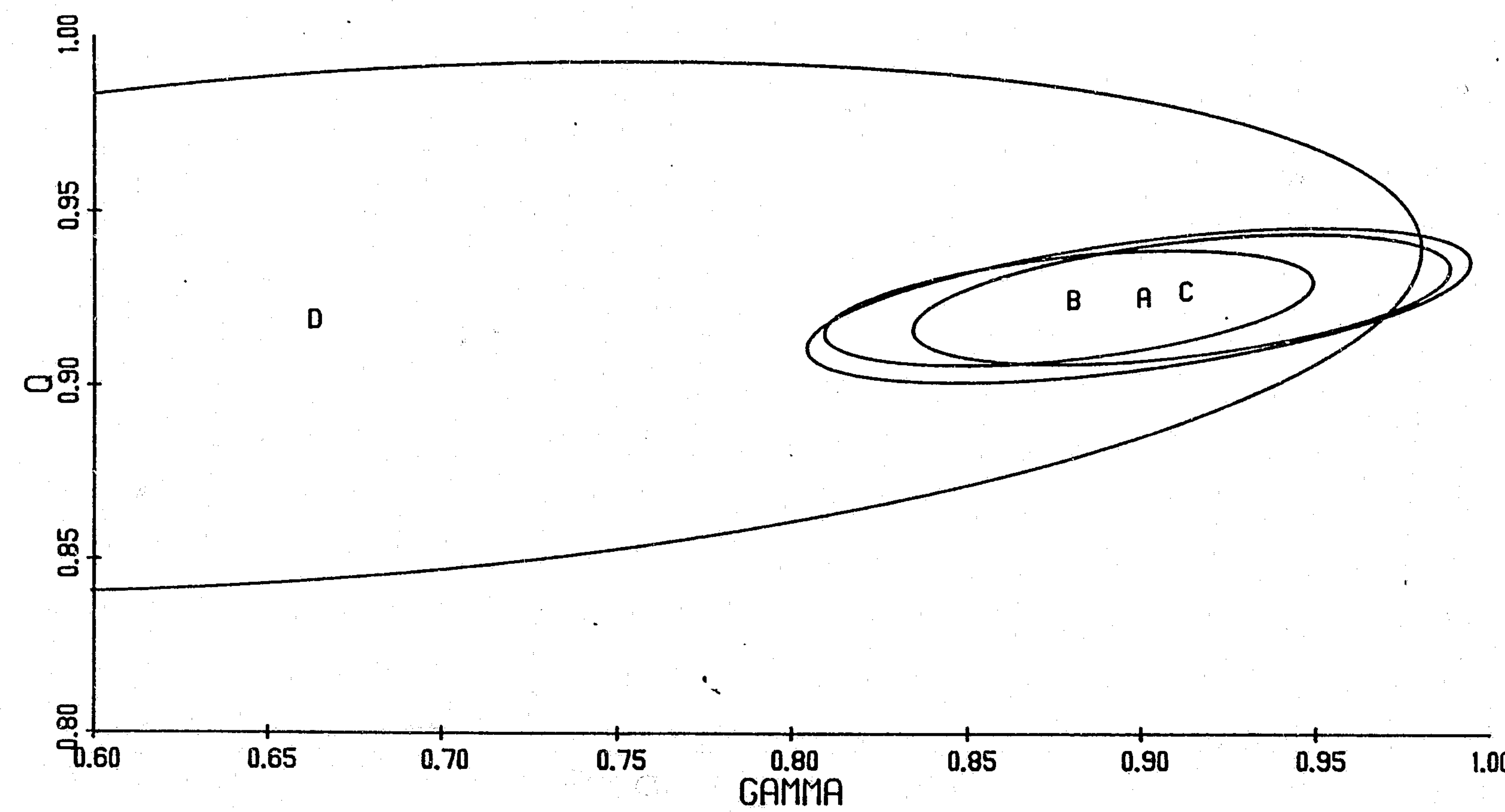


Figure 4h: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by MARITAL STATUS AT INTERVIEW

A. Single
B. Married
C. Divorced
D. Other

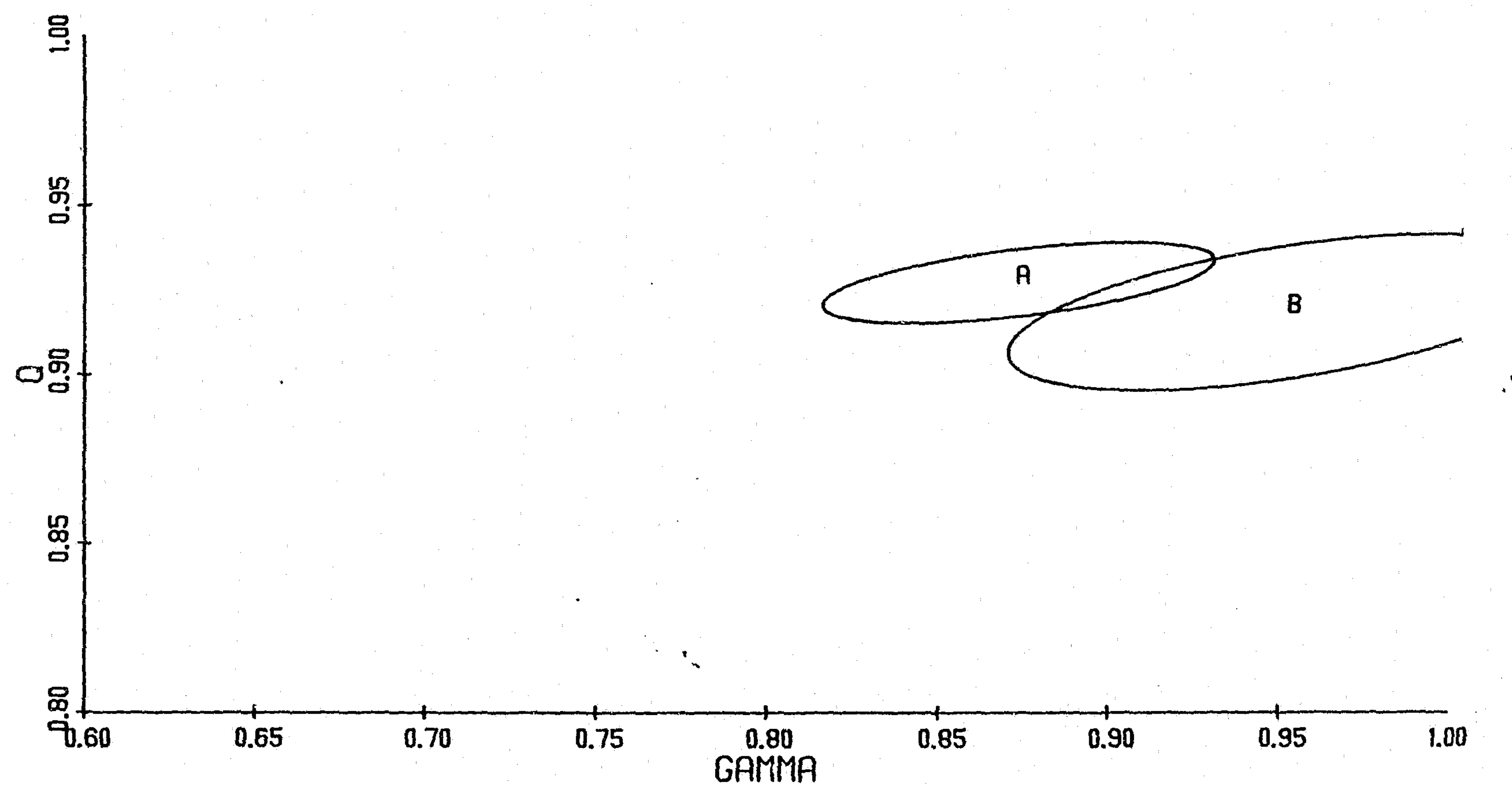


Figure 4i: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by EMPLOYMENT STATUS AT INTERVIEW
A. No Job
B. Other

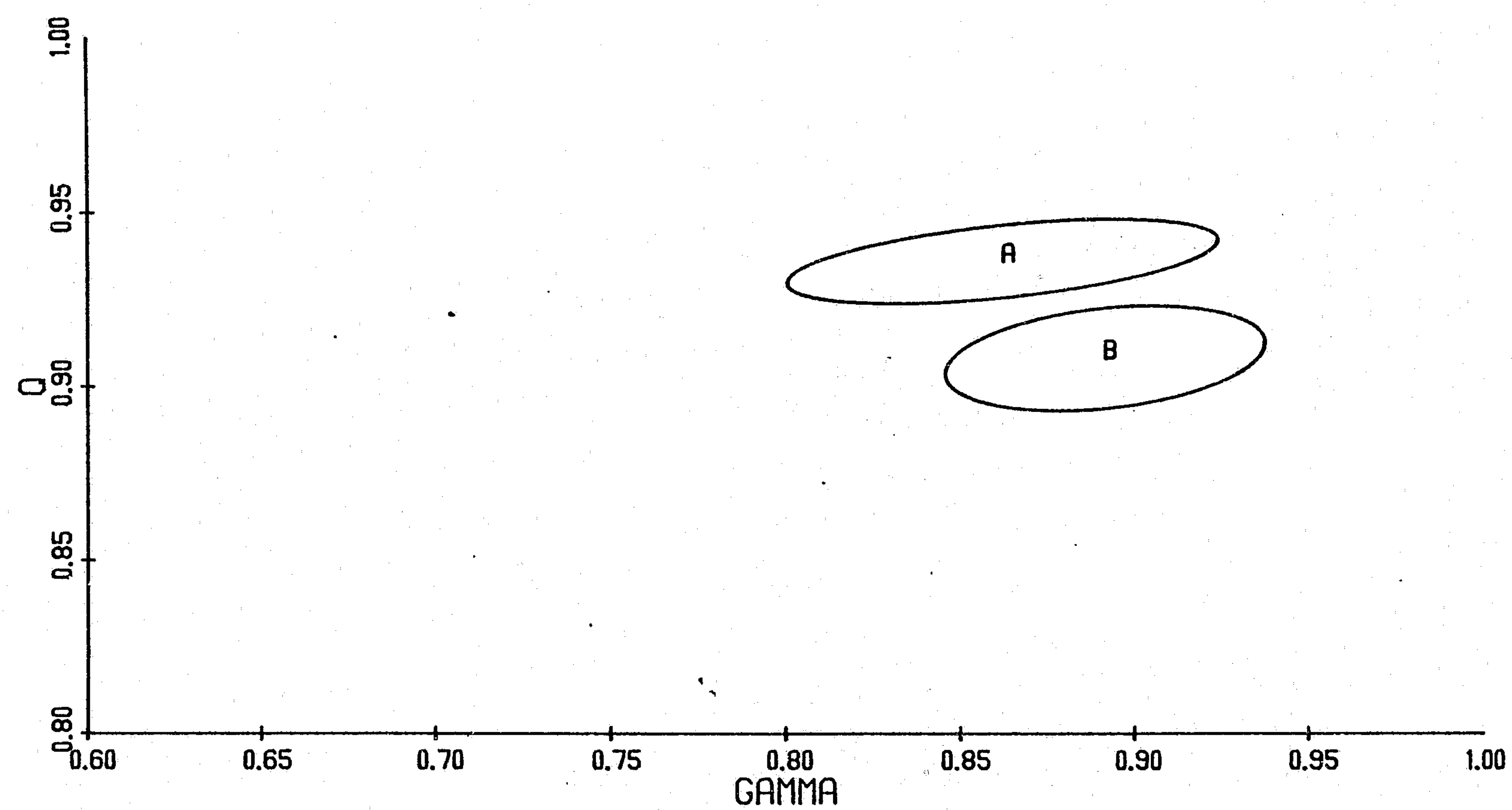


Figure 4j: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by DRINKING PROBLEM
A. None
B. Some

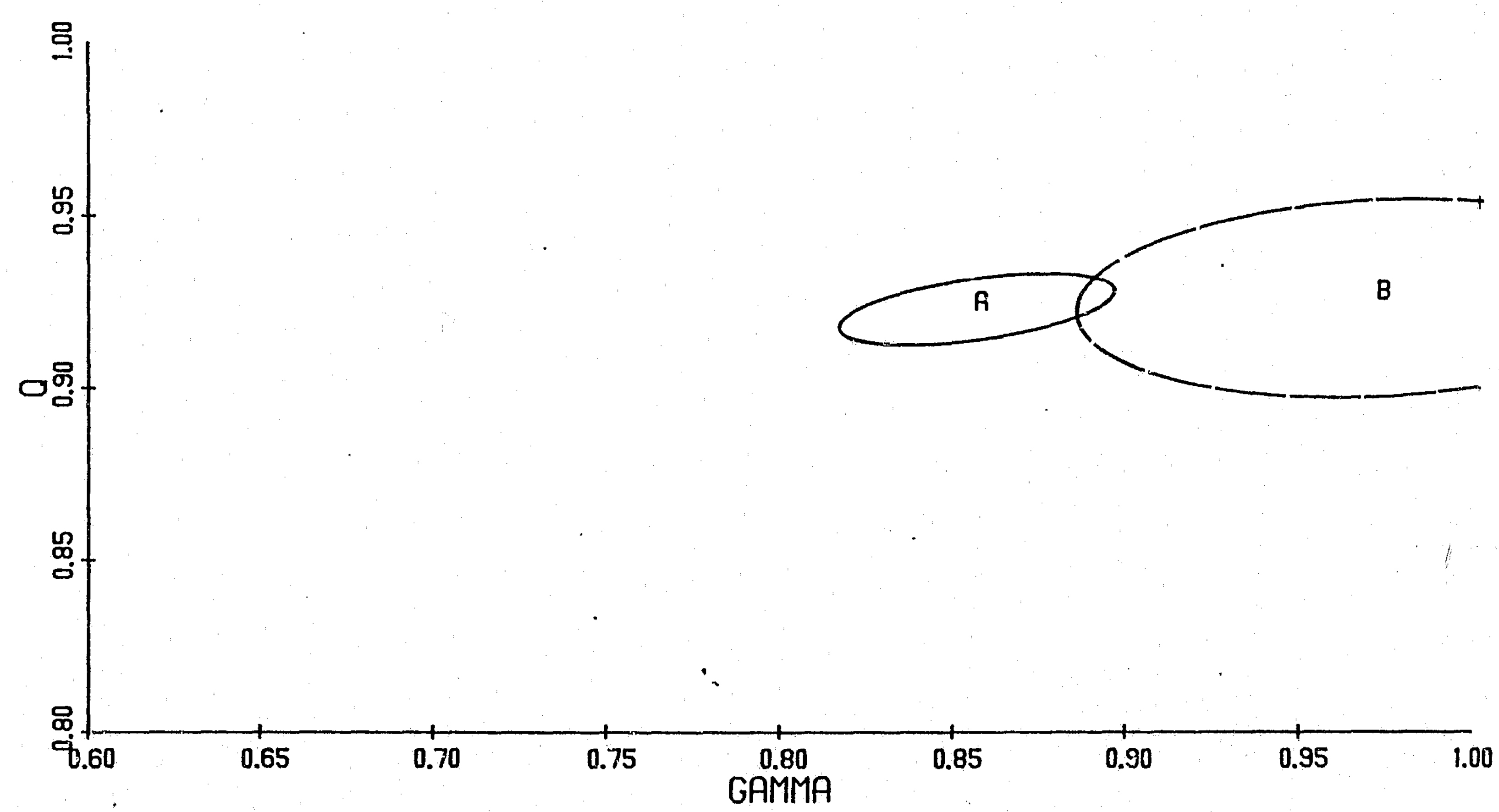


Figure 4k: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by DRUG USE
A. None
B. Some

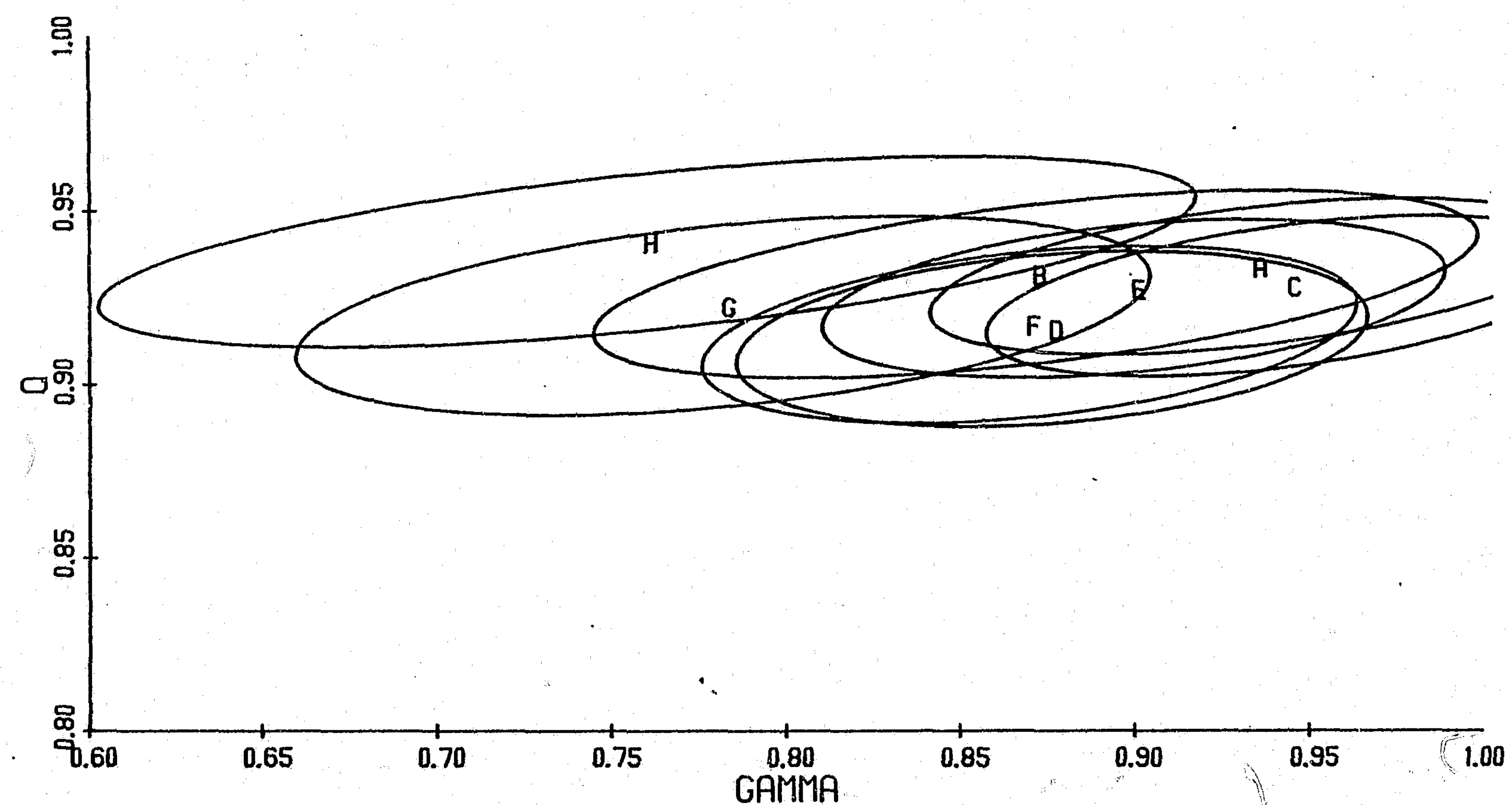


Figure 41: North Carolina Cohort (rearrest)
Bayes 90% Confidence Regions for estimates of γ and q by AGE AT RELEASE

- A. twenty or less
- B. 21, 22
- C. 22-24.5
- D. 24.5-28
- E. 28-34
- F. 34-40
- G. 40-47
- H. Greater than 47.

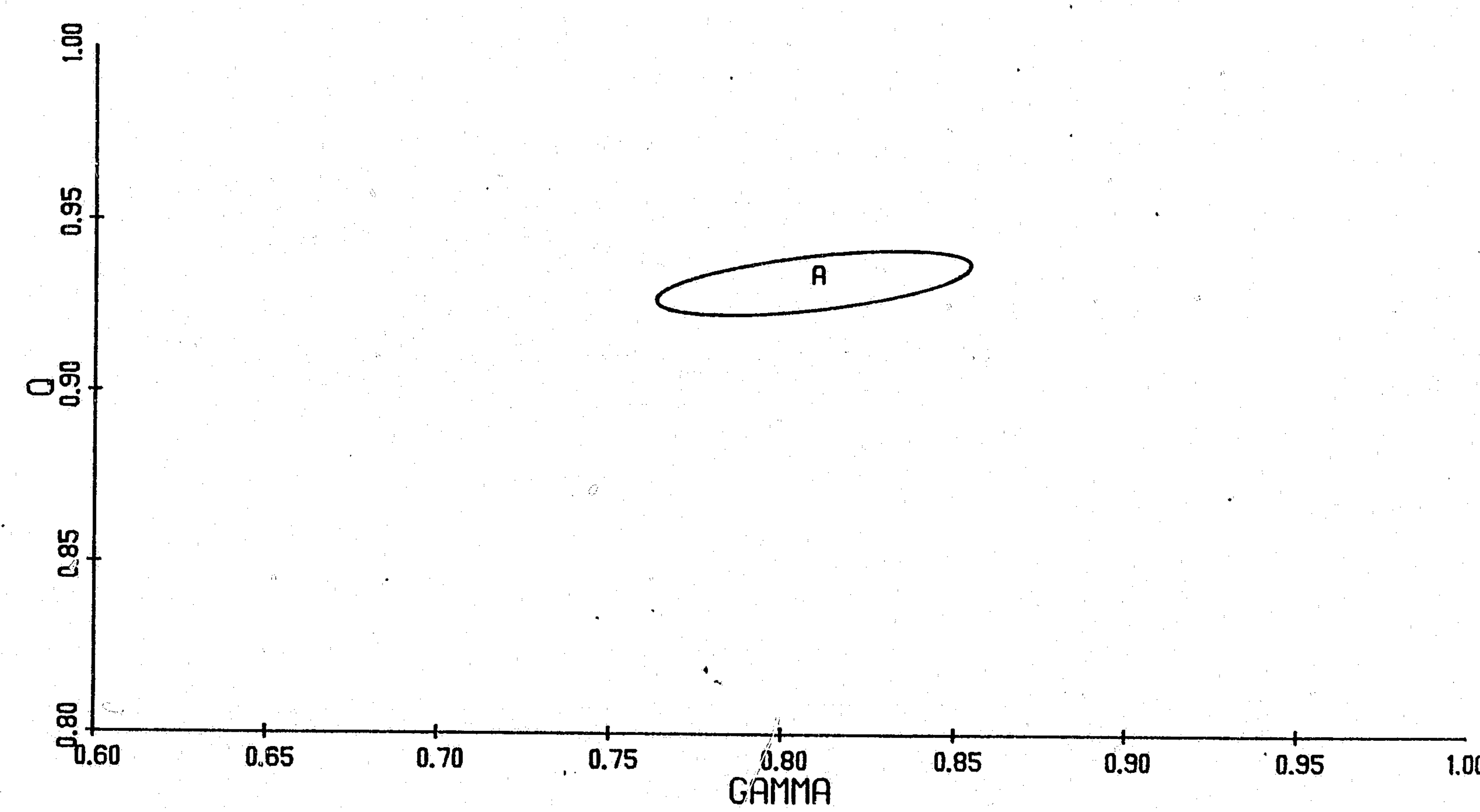


Figure 5a: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q.

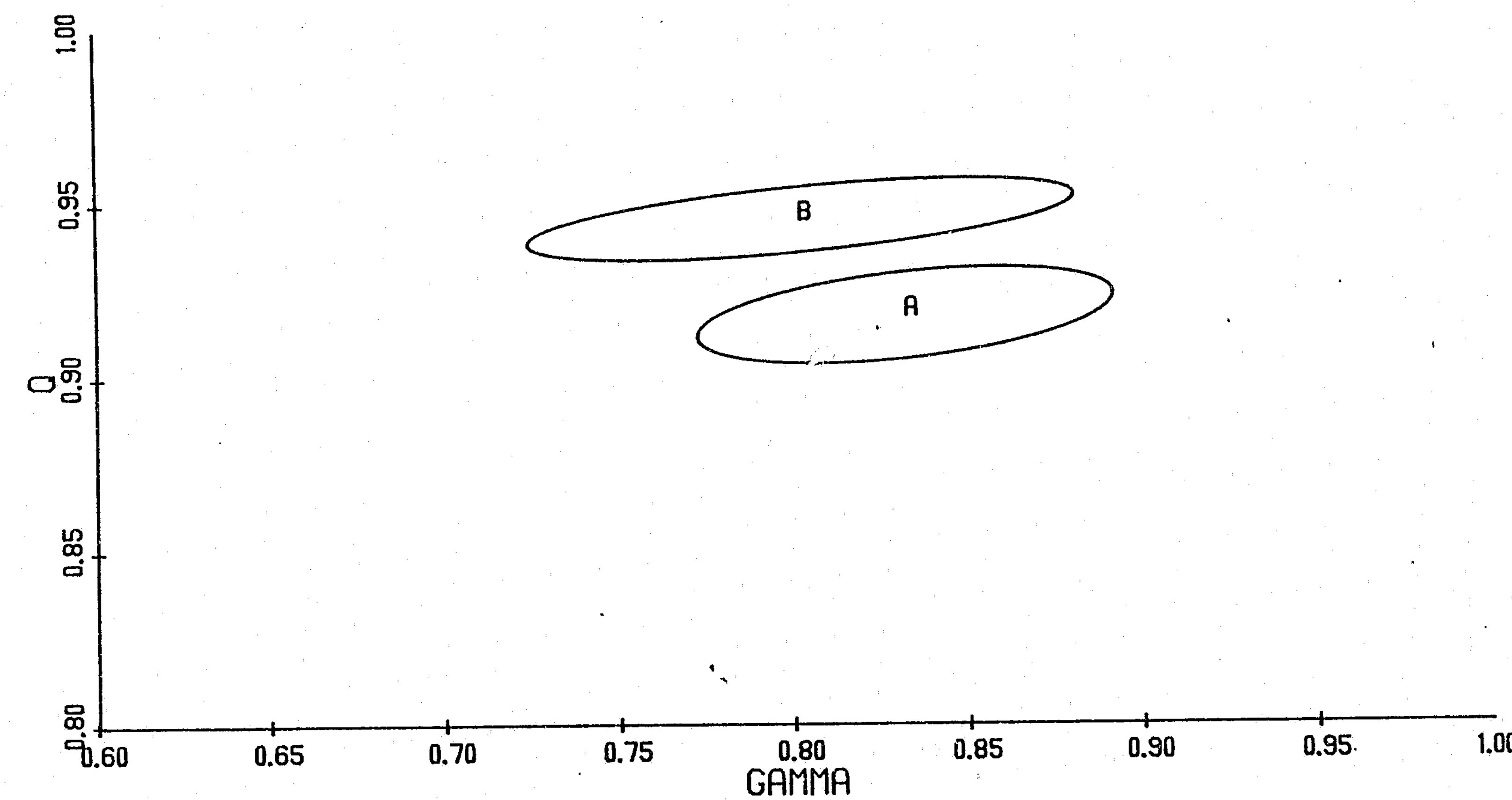


Figure 5b: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by RACE
A. White
B. Black or Indian

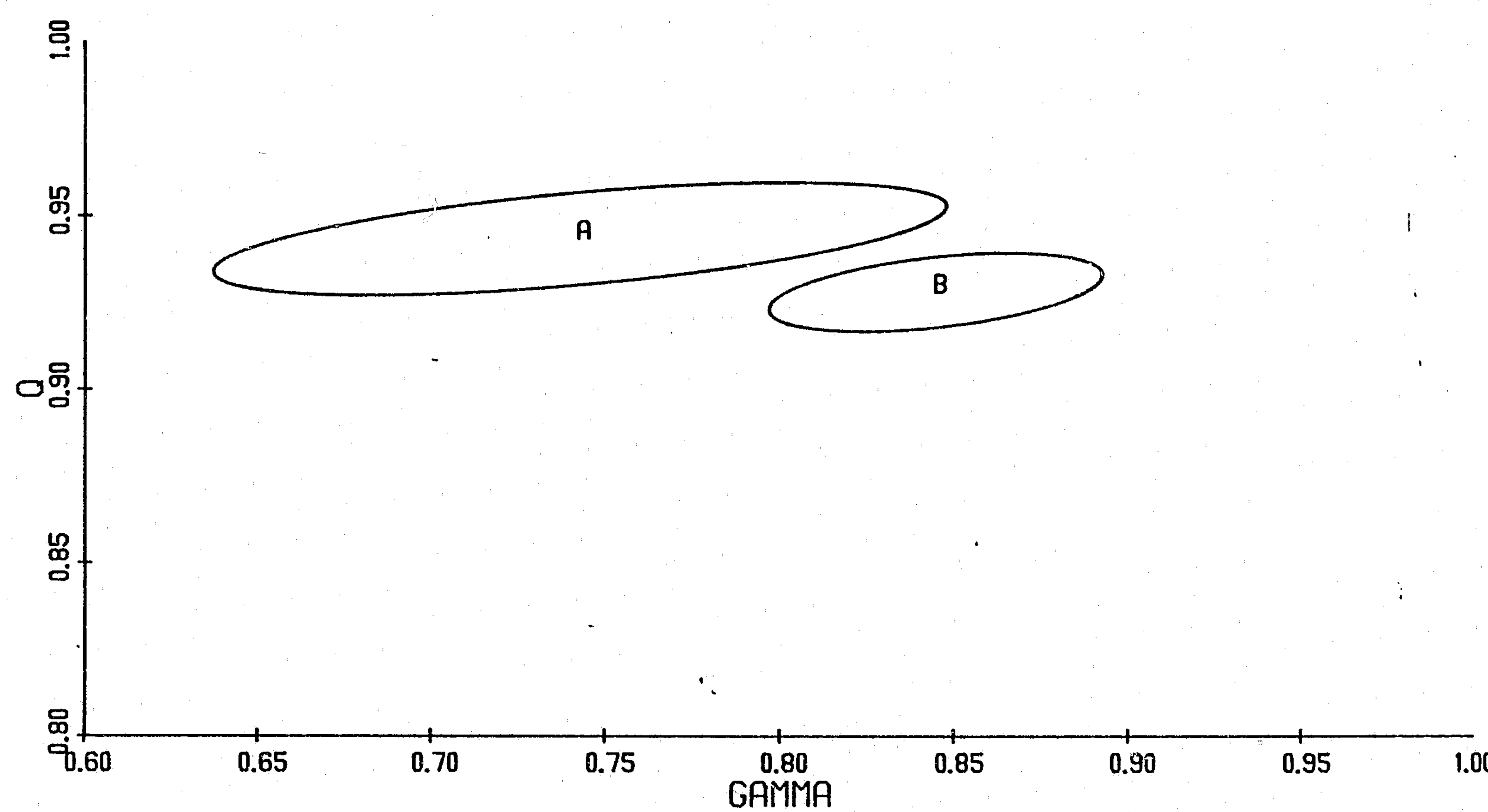


Figure 5c: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by TYPE OF RELEASE
A. Other
B. Unconditional

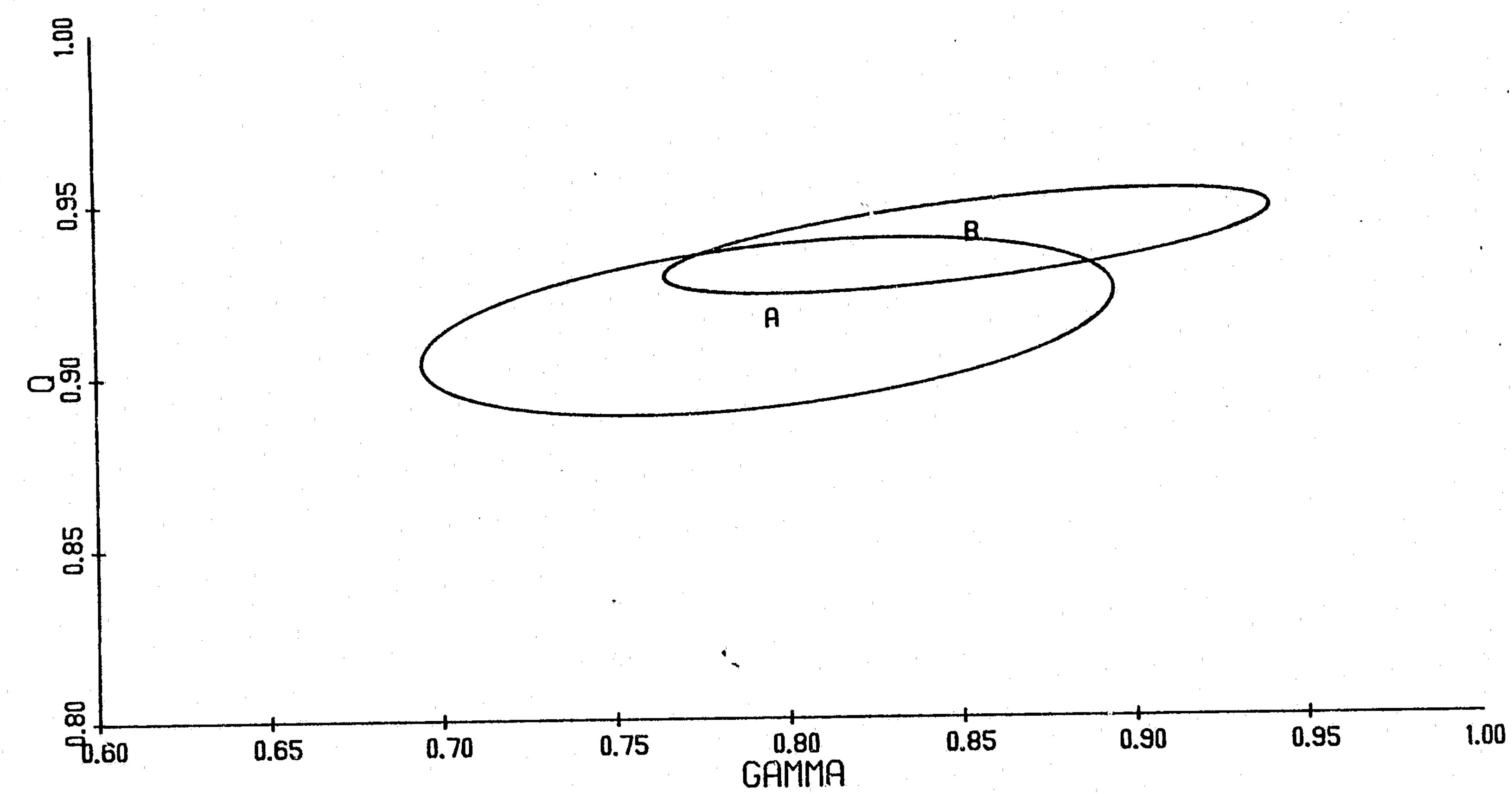


Figure 5d: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by IQ
A. 100 or less
B. Greater than 100

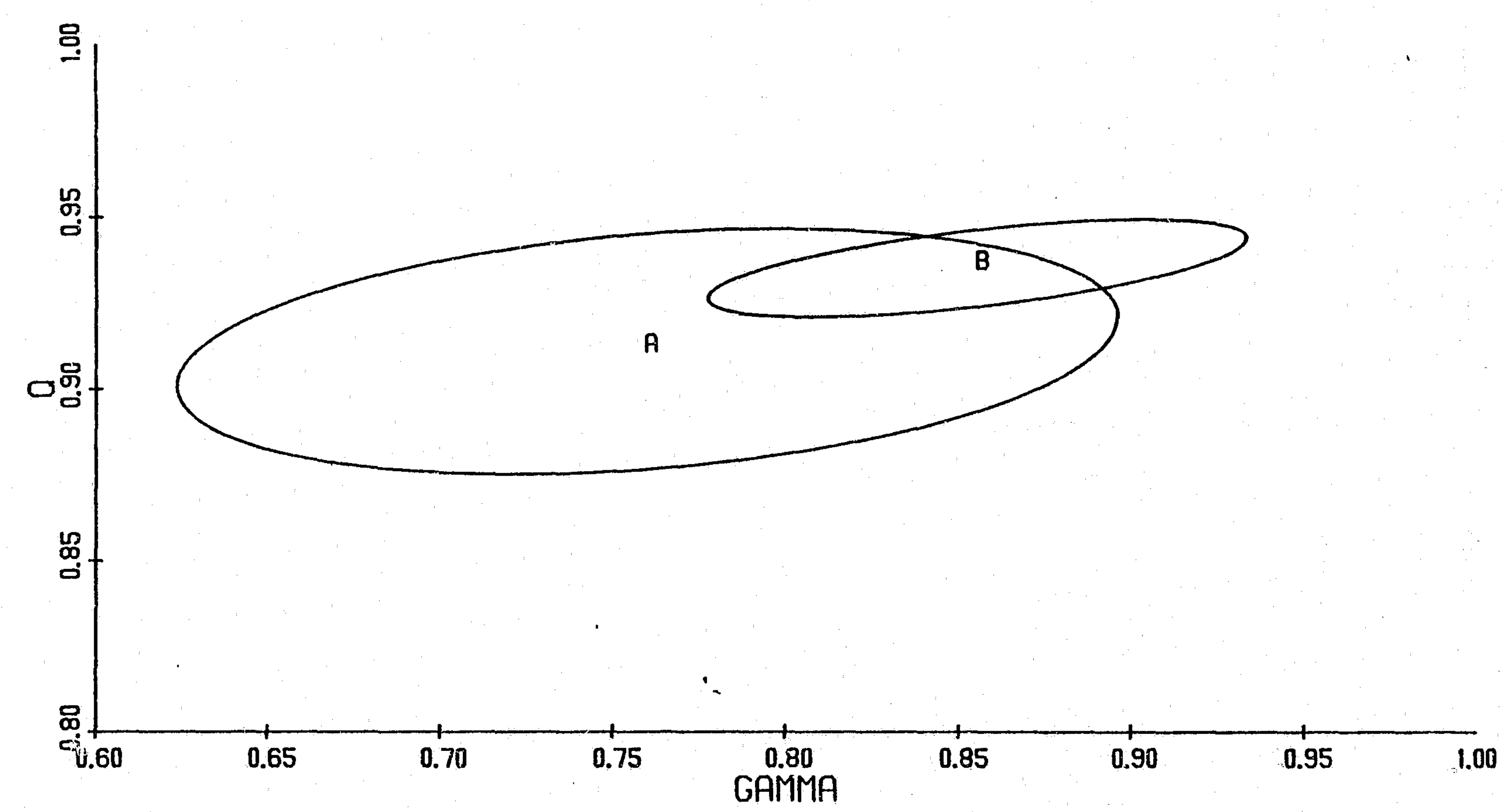


Figure 5e: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by SCHOOL ACHIEVEMENT
A. 8 years or less
B. 9 years or more

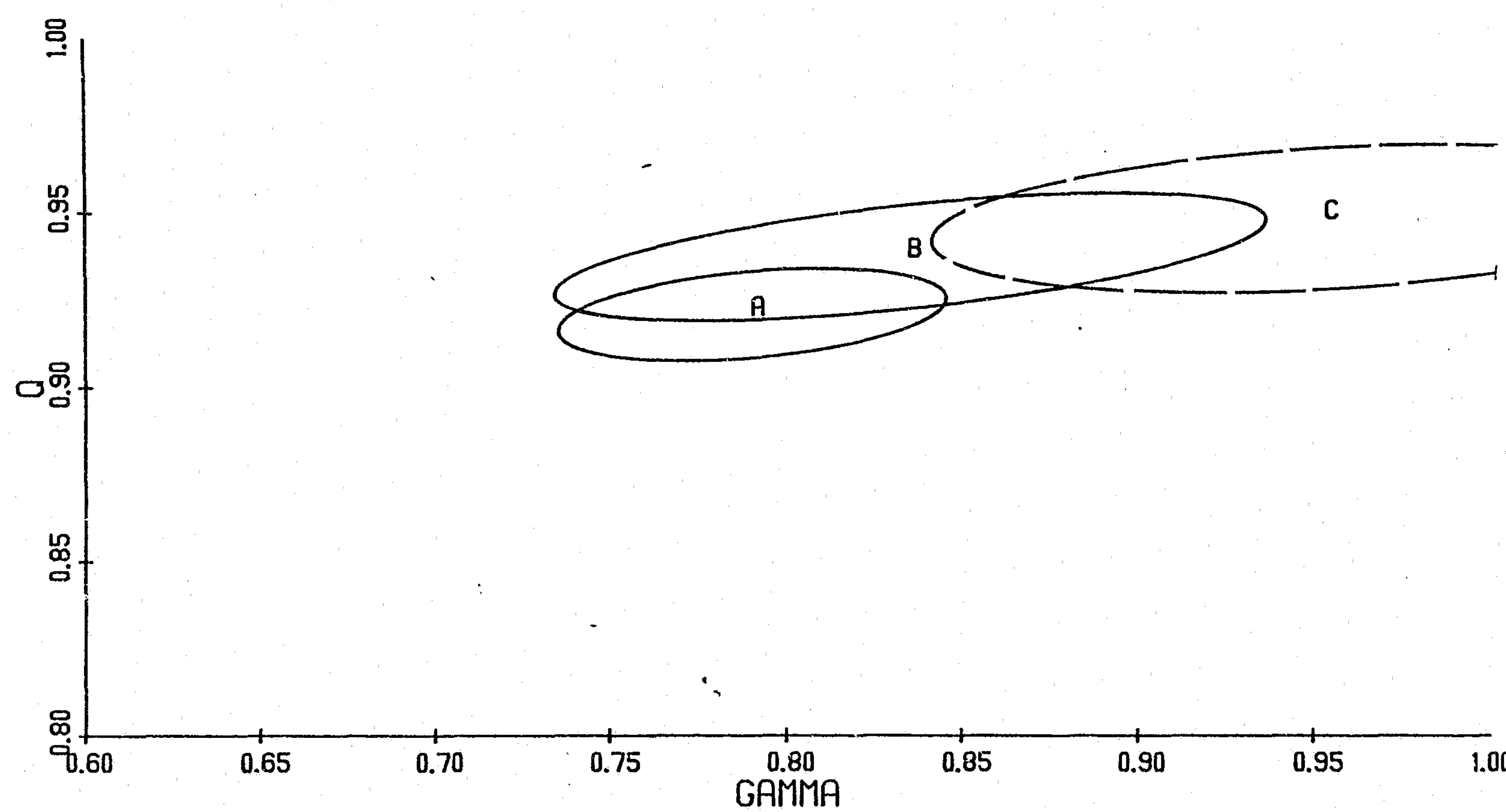


Figure 5f: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by
WORK STABILITY 5 YEARS PRIOR TO SAMPLE TERM
A. Other
B. Two or fewer job changes
C. Student

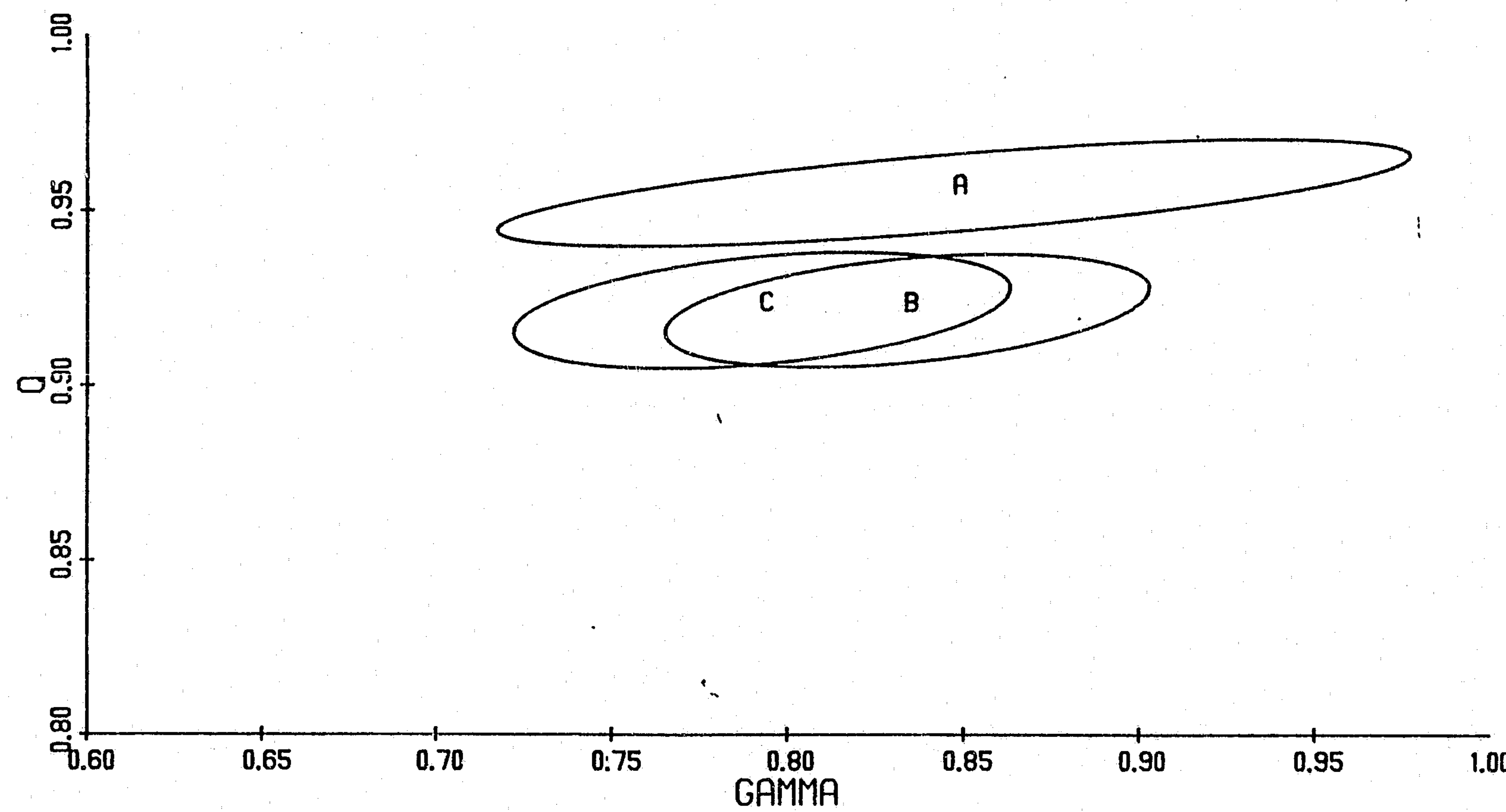


Figure 5g: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by PRIOR ARRESTS
A. None
B. One or two
C. Three or more

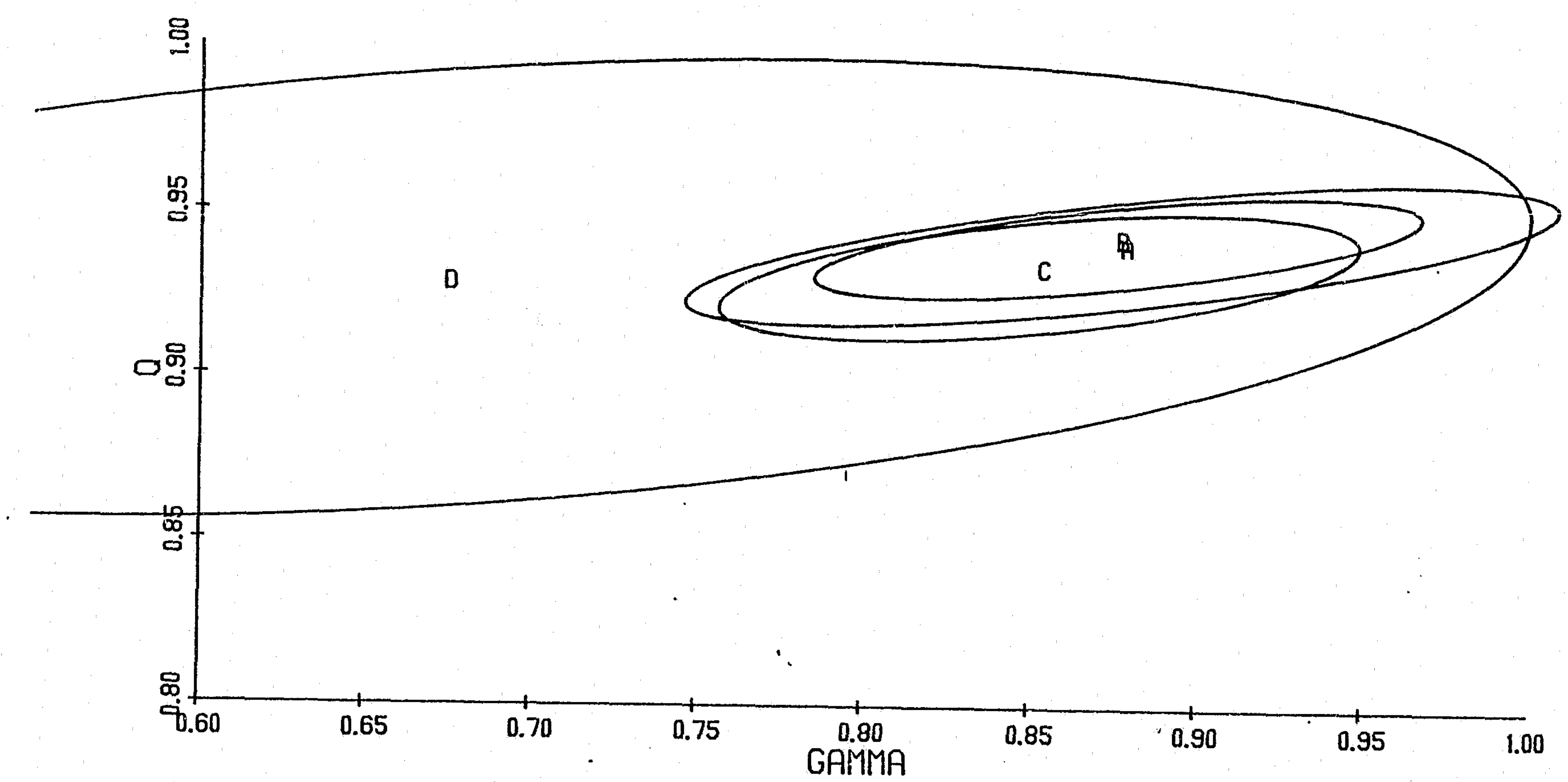


Figure 5h: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by MARITAL STATUS AT INTERVIEW

A. Single
B. Married
C. Divorced
D. Other

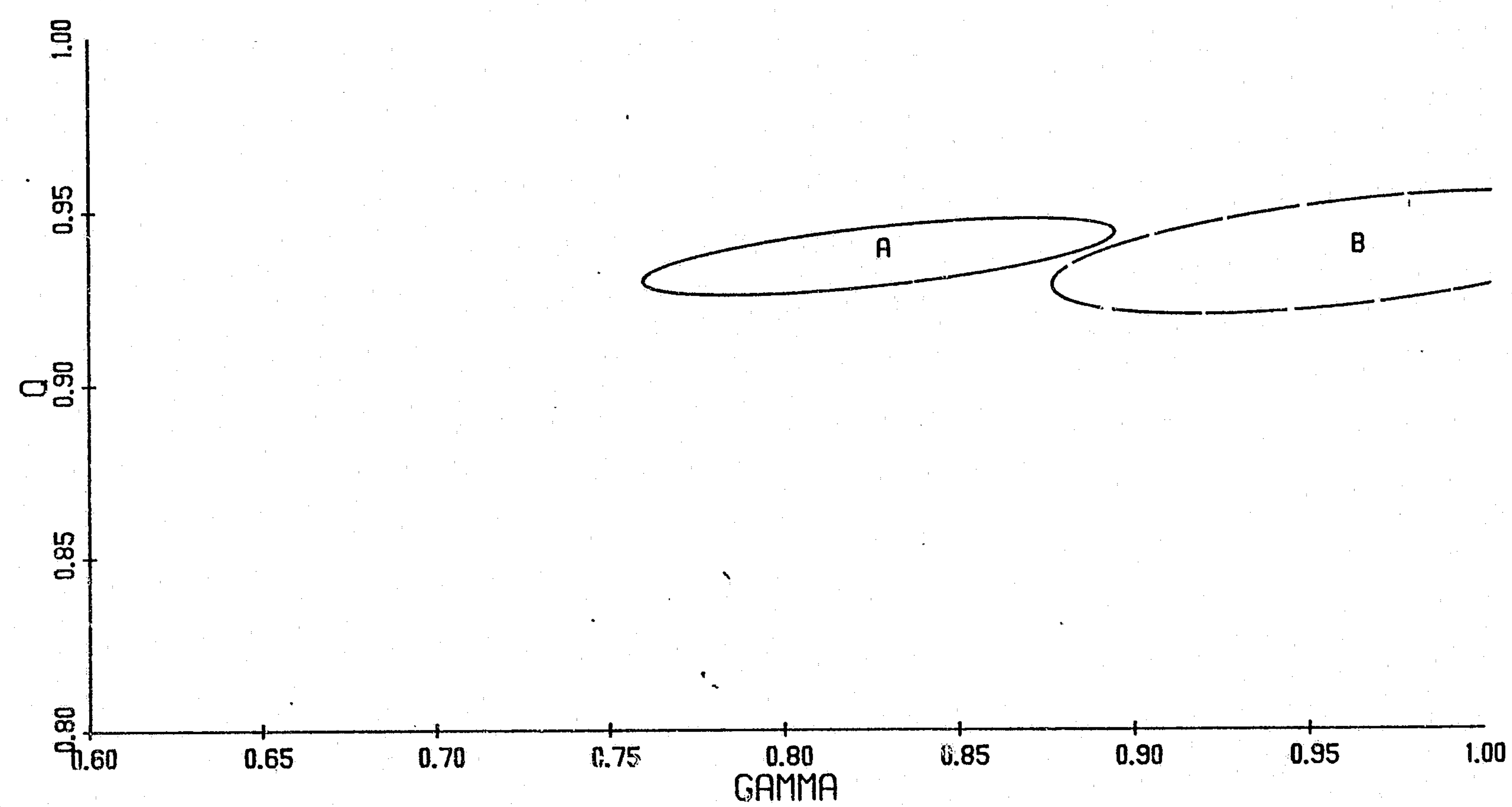


Figure 5i: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by EMPLOYMENT STATUS AT INTERVIEW
A. No Job
B. Other

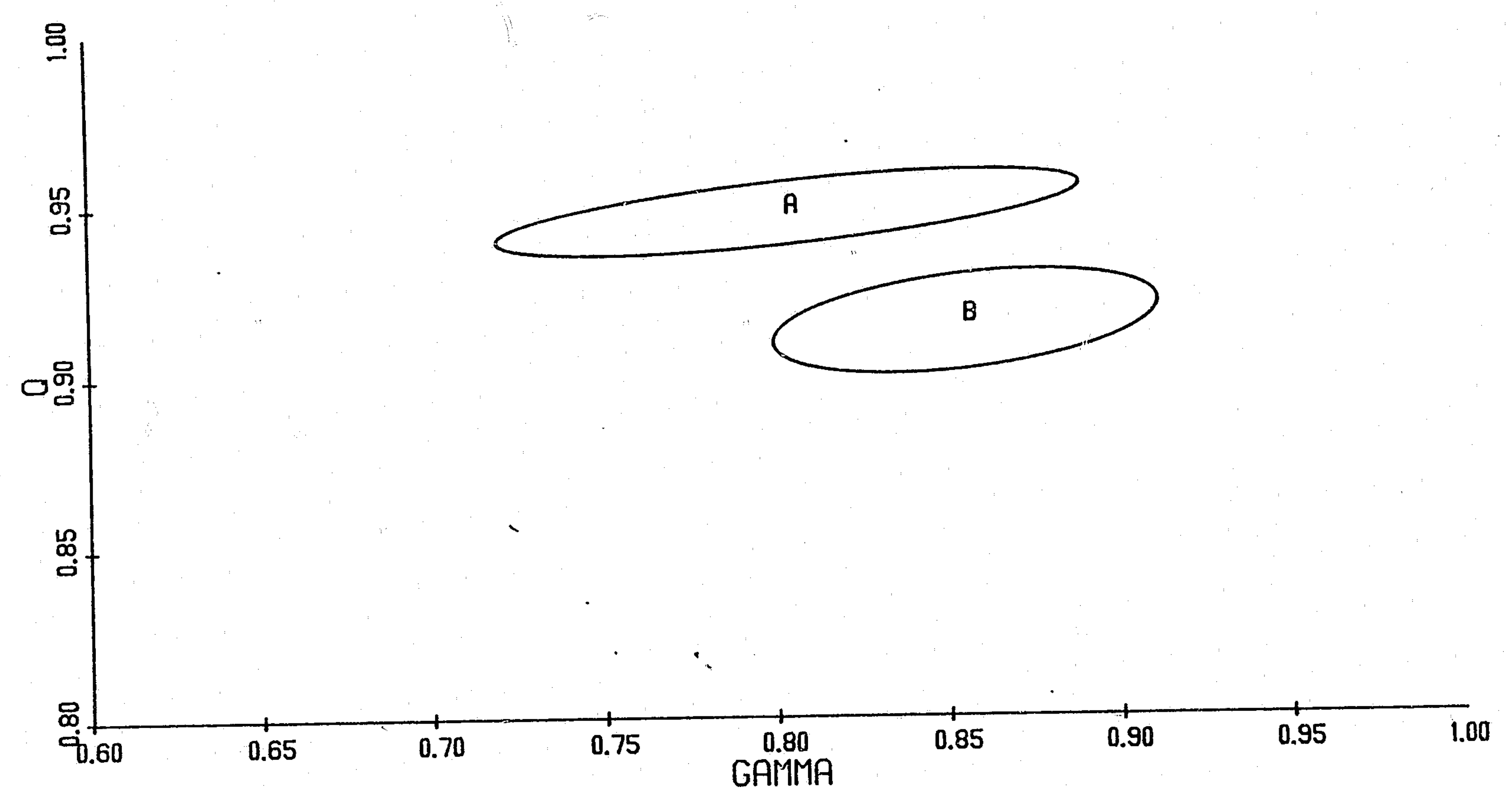


Figure 5j: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by DRINKING PROBLEM

A. None
B. Some

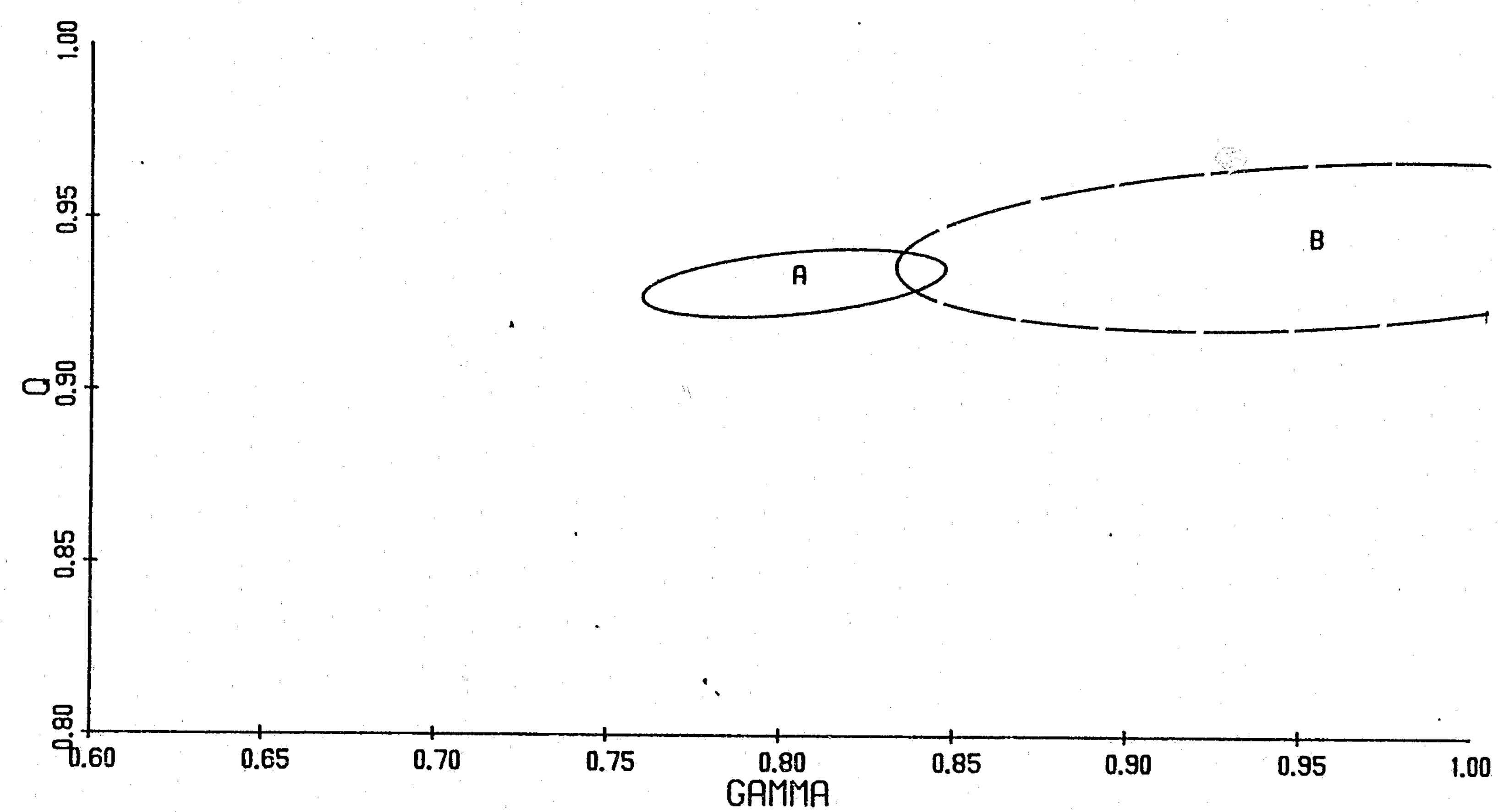


Figure 5k: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by DRUG USE
A. None
B. Some

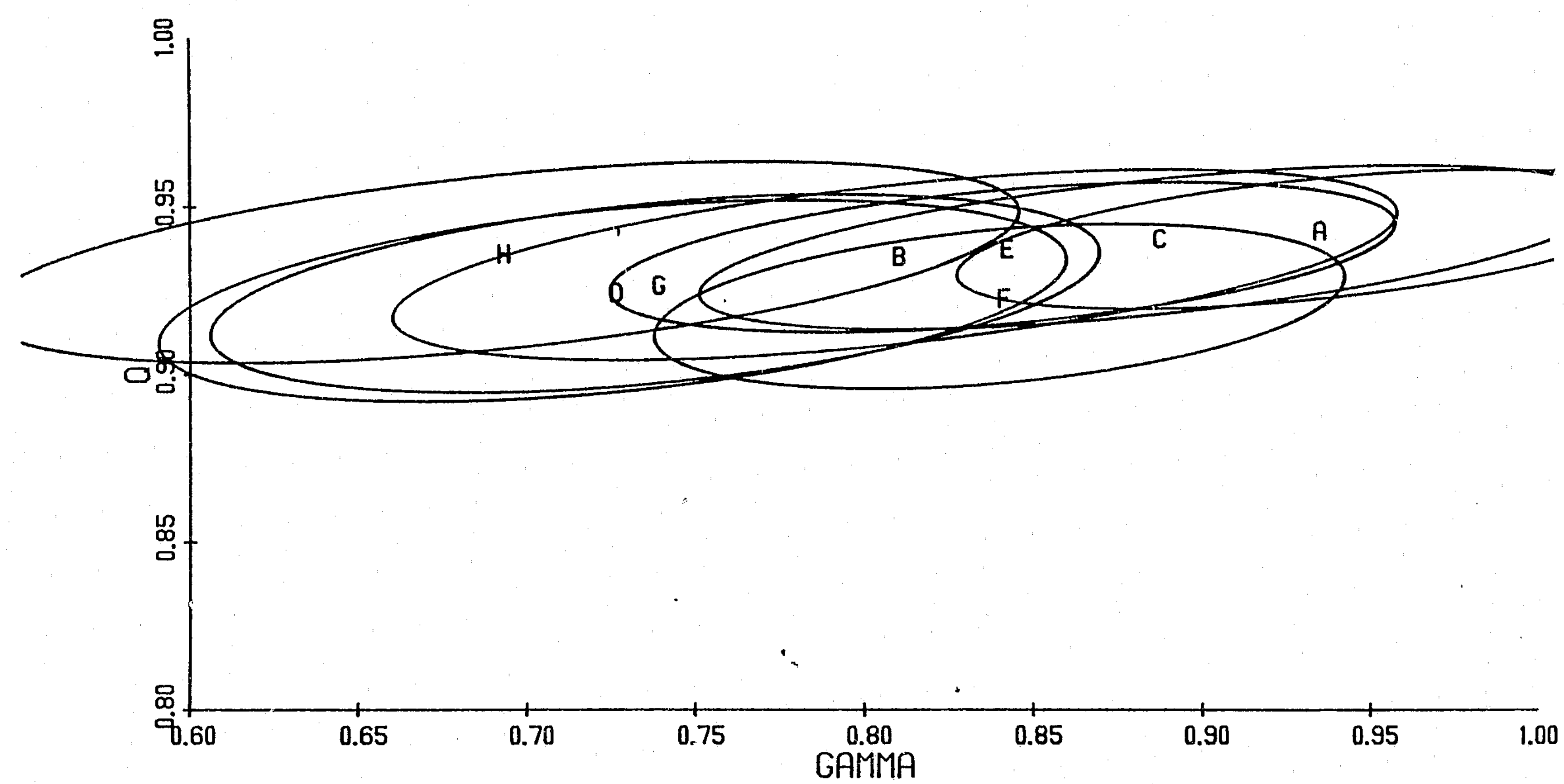


Figure 51: North Carolina Cohort (reconviction)
Bayes 90% Confidence Regions for estimates of γ and q by AGE AT RELEASE

A. twenty or less
B. 21, 22
C. 22-24.5
D. 24.5-28
E. 28-34
F. 34-40
G. 40-47
H. Greater than 47

The following pages (appendix I: 389-397) contain material protected by the Copyright Act of 1976 (17 U.S.C.): Book Reviews from Crime and Delinquency

END