1 *Department of Statistics, University of North Carolina, Chapel Hill, N. C. 27514.* *The research of this author and the preparation of the manuscript was supported by NSF Grant GP-23520.*

2 *Department of Sociology, University of North Carolina, Chapel Hill, N.C.*

A Time Series Approach to the Life Table

by

Edward J. Wegman[1] *and* Cris R. Kukuk[2]

University of North Carolina at Chapel Hill
Department of Statistics

Institute of Statistics Mimeo Series No. 730

December, 1970

A Time Series Approach to the Life Table

by

Edward J. Wegman[1] *and* Cris R. Kukuk[2]

1. Introduction

"Statistics" was historically the name given to the science which collects facts about the state. One of the most useful and best-known collection of facts about the state is the life table. In this paper, we will present a time series approach to estimating the life table.

There are two general forms of the life table: the cohort life table and the current life table. The former follows the fortunes of a cohort or group of individuals born at the same time throughout their lifetimes until the death of the last-surviving member. The latter version of the life table describes the mortality experiences of an entire population at a fixed point in time.

In either case, ℓ_x is the number of people surviving at age x. The ℓ_x are usually related by the equations

$$\ell_x = p_{0,x}\ell_0 \tag{1.1}$$

or

$$\ell_x = p_{x-1,x}\,\ell_{x-1}. \tag{1.2}$$

In these equations, p_{ij} may be interpreted as the conditional probability

[1] *Department of Statistics, University of North Carolina, Chapel Hill, N. C. 27514.* *The research of this author and the preparation of the manuscript was supported by NSF Grant GP-23520.*

[2] *Department of Sociology, University of North Carolina, Chapel Hill, N.C.*

of an individual surviving to age j given that he was alive at age i. If the function ℓ_x is normalized so that $\sum_x \ell_x = 1$, it is referred to as the population distribution. An extensive discussion of the life table and the related biometric functions may be found in Keyfitz, (1968).

As early as 1907, Lotka had enunciated the idea of the stationary population distribution and suggested the form

$$\text{(1.3)} \qquad \ell(x) = p(x)\lambda e^{-(\lambda-\mu)x}.$$

Here $\ell(x)$ is meant to be a continuous (density) version of ℓ_x, $p(x)$ is the probability of an individual surviving to age x, and λ and μ are respectively the crude birth and death rates. If we assume an individual has been exposed to a constant death rate, μ, throughout his x years, then

$$p(x) = e^{-\mu x}$$

so that (1.3) becomes

$$\ell(x) = \lambda e^{-\lambda x} .$$

In general, this suggests that ℓ_x decreases like a negative exponential. This is approximately true although such a parametric form is generally not satisfactory to explain all details appearing in real data. A description of Lotka's work can be found in Lotka, (1956).

The fact that ℓ_x is a monotone decreasing function is used to advantage by Grenander, (1956). He gives a non-parametric maximum likelihood estimate of ℓ_x. Grenander's type of estimate may be used to estimate any monotone probability density and has generated more work along these lines.

Chiang, (1960) demonstrates that $\{\ell_x, \; x = 0,1,2,\ldots\}$ is a Markov Stochastic Process. We will take advantage of this fact in order to use time series methods to study $\{\ell_x, \; x = 0,1,2,\ldots\}$. If $\{\ell_x, \; x = 0,1,2,\ldots\}$ is a stochastic process, it is reasonable to expect that (1.2) still holds except that the relation is perturbed by an error term, ε_x. Hence we assume

(1.4) $$\ell_x = p_{x-1,x}\,\ell_{x-1} + \varepsilon_x, \quad x = 1,2,3,\ldots$$

where ε_x is a sequence of uncorrelated zero mean random variables and where ε_x is uncorrelated with $\ell_{x'}$, $x' < x$. Let us also, henceforth, write p_x for $p_{x-1,x}$. A process satisfying equation (1.4) is known as a first-order autoregressive process. The relation (1.4) is sometimes referred to as a simple Markov relation. We shall devote the next section to outlining some facts about first order autoregressive processes. In the third section, we shall carry through some computations using the 1961 Indian Census data to illustrate our methods.

2. Autoregressive and Related Processes

In this section, let us write (1.4) as

(2.1) $$X_t = p(t)\,X_{t-1} + \varepsilon_t \quad t = 1,2,\ldots$$

in order to keep the notation more consistent with conventional time series notation. We will give, in this section, the mean and covariance structure of $\{X_t:\ t = 1,2,\ldots\}$ and propose estimates of $p(t)$.

We shall denote by EX_t the expected value of X_t. From (2.1), it is easy to see

$$EX_t = p(t)EX_{t-1}, \quad t = 1,2,\ldots\ .$$

Using this formula recursively, we obtain

(2.2) $$EX_t = \left[\prod_{j=1}^{t} p(j)\right] \cdot EX_0 \quad t = 1,2,\ldots$$

If X_0 is given as an initial condition, the conditional expectation of X_t given X_0 is

$$E(X_t|X_0) = \left[\prod_{j=1}^{t} p(j)\right] \cdot X_0, \qquad t = 1,2,\ldots$$

Again using (2.1), we may study the covariance structure. Multiplying both sides of (2.1) by X_{t-s}, we have

$$X_t\, X_{t-s} = p(t)\, X_{t-1}\, X_{t-s} + \varepsilon_t\, X_{t-s}$$

so that by taking the expectation, we have

$$E(X_t\, X_{t-s}) = p(t)\, E(X_{t-1}\, X_{t-s}).$$

Subtracting the product of the appropriate means,

$$\text{(2.3)} \qquad cov(X_t, X_{t-s}) = p(t)\, cov(X_{t-1}, X_{t-s}).$$

Applying (2.3) recursively,

$$\text{(2.4)} \qquad cov(X_t, X_{t-s}) = p(t)\, p(t-1) \ldots p(t-s+1)\, var(X_{t-s}).$$

Thus, the system of covariances, depends on the variances. Applying (2.1) recursively leads to

$$X_t = \left[\prod_{j=1}^{t} p(j)\right] X_0 + \sum_{j=1}^{t-1} p(t) \ldots p(t-j+1)\varepsilon_{t-j} + \varepsilon_t.$$

Subtracting (2.2) from this yields

$$(X_t - EX_t) = \left[\prod_{j=1}^{t} p(j)\right] \cdot (X_0 - EX_0) + \sum_{j=1}^{t-1} p(t) \ldots p(t-j+1)\varepsilon_{t-j} + \varepsilon_t\ .$$

Squaring, taking expected values and recalling ε_t is uncorrelated with $\varepsilon_{t'}$, $t \neq t'$ and with $X_{t'}$, $t' < t$, we have

$$\text{(2.5)} \qquad var(X_t) = \left[\prod_{j=1}^{t} p(j)\right]^2 var(X_0) + \left[\sigma^2_t + \sum_{j=1}^{t-1} [p(t) \ldots p(t-j+1)]^2\, \sigma^2_{t-j}\right].$$

Here, $\sigma^2_j = var(\varepsilon_j)$. In the special case $p(t)$ is a constant, say p, and $var(\varepsilon_j)$ is a constant, say σ^2, (2.5) becomes

$$var(X_t) = p^{2t}\, var(X_0) + \sigma^2 \sum_{j=0}^{t-1} p^{2j} .$$

Using the formula for a finite geometric sum

$$var(X_t) = p^{2t}\, var(X_0) + \frac{\sigma^2(1-p^{2t})}{1-p^2} .$$

If X_0 is known, then $var(X_0) = 0$ and the conditional variance is

$$var(X_t|X_0) = \frac{\sigma^2(1-p^{2t})}{1-p^2} .$$

Finally, the conditional covariance is

$$cov(X_t, X_{t-s}|X_0) = \frac{p^s\sigma^2(1-p^{2t})}{1-p^2} .$$

In certain cases, particularly in connection with census data, the errors appear not to have a constant variance, but rather to be proportional, roughly speaking, to population size. That is to say, there appears to be a constant percentage error. Thus, it is reasonable to let $\varepsilon_t = X_{t-1}\cdot\eta_t$ where η_t is uncorrelated with $\eta_{t'}$, $t \neq t'$ and with $X_{t'}$, $t' < t$ and where η_t has zero mean and variance σ^2. In this case, (2.1) becomes

(2.6) $$X_t = p(t)\, X_{t-1} + \eta_t\, X_{t-1}.$$

For an equation of the form (2.6), (2.2) and (2.4) are still true. However, (2.5) is no longer true. Solving (2.6) recursively leads to

$$X_t = \prod_{j=1}^{t} (p(j)+\eta_j)X_0.$$

Squaring and taking expected values,

$$EX_t^2 = \prod_{j=1}^{t} (p^2(j)+\sigma^2) \cdot EX_0^2 .$$

From the second moments and equation (2.2), the variance of X_t may be calculated,

$$var(X_t) = \left[\prod_{j=1}^{t}(p^2(j)+\sigma^2)\right] EX_0^2 - \left[\prod_{j=1}^{t} p^2(j)\right] \{EX_0\}^2.$$

If $p(t)$ is a constant,

$$var(X_t) = (p^2+\sigma^2)^t \; EX_0^2 - p^{2t}[EX_0]^2.$$

If X_0 is known,

$$var(X_t|X_0) = X_0^2[(p^2+\sigma^2)^t - p^{2t}].$$

In both (2.1) and (2.6), it is of considerable interest to estimate $p(t)$. If we parameterize $p(t)$ as an ℓ-th degree polynomial, we may use standard least squares procedures to determine the coefficients. In (2.1), let us assume ε_t has a constant variance and let us also assume we have n observations $x_0,\ldots,x_{n-1}$. We then wish to minimize

$$\sum_{t=1}^{n-1} \left(x_t - (\sum_{j=0}^{\ell} a_j t^j) x_{t-1}\right)^2 .$$

Taking the partial derivative with respect to a_k and equating to zero, we have

$$\text{(2.7)} \qquad \sum_{t=1}^{n-1} t^k \, x_t \, x_{t-1} = \sum_{j=0}^{\ell} a_j \Big(\sum_{t=1}^{n-1} t^{j+k} x_{t-1}^2\Big), \quad k = 0,\ldots,\ell.$$

Solving (2.7) simultaneously gives an estimate of a_j, $j = 0,\ldots,\ell$ and, hence, an estimate of $p(t)$. If $p(t)$ is assumed to be a constant, then the estimate of $a_0 = p$ is $\hat{p} = [\sum_{t=1}^{n-1} x_t x_{t-1}]\cdot[\sum_{t=1}^{n-1} x_{t-1}^2]^{-1}$.

Let us rewrite (2.6) as

$$X_t X_{t-1}^{-1} = p(t) + n_t.$$

Here we wish to minimize

$$\sum_{t=1}^{n-1} (x_t x_{t-1}^{-1} - (\sum_{j=0}^{\ell} a_j t^{\ell}))^2.$$

Taking partial derivatives with respect to a_k and equating to zero, we have

$$(2.8) \qquad \sum_{t=1}^{n-1} x_t x_{t-1}^{-1} t^k = \sum_{j=0}^{\ell} a_j \sum_{t=1}^{n-1} t^{j+k} .$$

Again, solving (2.8) simultaneously gives the estimate of $p(t)$. In the special case $p(t)$ is constant, the estimate of $a_0 = p$ is $\hat{p} = (n-1)^{-1} \cdot \sum_{t=1}^{n} X_t X_{t-1}^{-1}$. An interesting feature of $\hat{p}$ is that it is an unbiased estimate of p. In addition, $var(\hat{p}) = (n-1)^{-1} \cdot \sigma^2$ where σ^2 is the variance of n_t.

We close this section by noting that we describe the fitting of a polynomial $p(t) = \sum_{j=0}^{\ell} a_j t^j$. The set of functions $\{t^j\}$ are linearly independent but not orthonormal. It is frequently advisable to investigate the fitting of an orthonormal series, rather than one which is merely linearly independent.

3. Analysis of the Indian Population Count by Age

The data to which we shall refer in this section is taken from papers published in India on the 1961 Indian census. See Census of India, (1963). We shall present our technique for arriving at the population distribution, ℓ_x, but first we would like to mention the procedure followed by the Indian Census officials.

The population count of India by the stated age of the respondant is found in Table 1. Perhaps the most noticeable facet of this census data is the marked preference shown for the ages divisible by 5. This age heaping effect is well-known and represents a considerable distortion of the true situation. To eliminate this peaking, the Indian census officials smoothed their data by an eleven term moving average. They next formed 5 year totals (with some special adjustments for the very young and the very old), smoothed the result with a weighted 3 term moving average and finally interpolated by the well-known fifth difference osculatory interpolation formulae due to Kozakiewicz.

We have several objections to these procedures. First of all, from the point of view of time series an eleven term moving average is a low pass filter, filtering out variations with period smaller than 11 years. Since we are primarily interested in filtering the peaks which occur at 5 year intervals, a five term moving average is sufficient. (Clearly, one wants to use the filter which least changes the data, but still removes the unwanted periodicities.) A second objection is that the ordinary arithmetic moving average and the procedure of forming 5 years totals are both procedures which tend to reshape the data so that the altered data will have a tendency to cluster along a straight line. In Section 1, we observed, in the case of a stationary population, the population distribution tends to decrease as a negative exponential, hence, very much in a non-linear fashion. (In the case of a nonstationary population where the total population size increases, the population distribution will drop off even more quickly than in the stationary case.) In view of this situation, the data analyst should avoid as much as possible procedures which linearize the data.* We believe that the two

* *We have experimented with 11- and 5- term moving averages, using these to filter a curve known to decrease exponentially. The filtered data, in both cases, fell above the original curve. The percentage difference was respectively about 8% and about 1% for the 11 term and the 5 term filters.*

TABLE 1

Population count (1961) of India by stated age of respondant

0	13,551,710	26	5,472,520	51	960,486
1	11,673,441	27	3,986,209	52	1,970,895
2	13,427,444	28	7,616,137	53	800,709
3	14,129,333	29	2,376,492	54	996,797
4	13,242,893	30	18,049,763	55	6,217,166
5	15,227,496	31	1,992,606	56	1,181,324
6	13,966,354	32	6,205,031	57	664,442
7	12,268,094	33	2,116,072	58	1,198,366
8	13,655,781	34	2,432,703	59	551,415
9	9,471,091	35	14,523,328	60	8,723,146
10	14,366,679	36	3,626,705	61	626,729
11	7,115,477	37	1,920,928	62	1,022,540
12	12,729,548	38	3,822,719	63	392,757
13	6,922,327	39	1,532,454	64	453,118
14	8,094,839	40	15,496,709	65	3,493,498
15	8,403,095	41	1,350,982	66	358,707
16	8,184,236	42	3,334,978	67	306,016
17	4,641,026	43	1,287,810	68	460,037
18	10,225,422	44	1,352,650	69	223,171
19	4,370,035	45	11,580,622	70+	8,601,562
20	13,596,192	46	1,643,541		
21	4,536,574	47	1,207,143		
22	9,314,331	48	2,596,110		
23	4,427,801	49	1,002,262		
24	5,395,409	50	12,352,529		
25	17,076,873				

procedures mentioned above have distorted the shape of their curve sufficiently so that their estimate of ℓ_x is virtually linear from age 20 to age 70. See for example Census of India, (1963, page 4).

If the moving average is modified by forming a geometric moving average, i.e.

$$Y_t = (X_{t-2} \cdot X_{t-1} \cdot X_t \cdot X_{t+1} \cdot X_{t+2})^{\frac{1}{5}}$$

instead of the arithmetic moving average

$$Y_t = (X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2}) \div 5,$$

two advantages result. First, if $\{X_t: t = 0,1,2,\ldots\}$ truely is an exponential curve, then $\{Y_t: t = 2,3,\ldots\}$ will be the same exponential curve, and second, it is well-known that geometric averages are less affected by extreme values than arithmetic averages, hence data filtered geometrically is smoother than data filtered through a corresponding arithmetic filter. We have used both the geometric and arithmetic filter and the results may be found in Figures 1 and 2 respectively. Both have a similar shape, but the curve of the arithmetically filtered data tends to lie above the curve of the geometrically filtered data.

Before we move on to the fitting of an autoregressive process, we should point out that the procedures followed by the Indian census officials will tend to keep the total population count constant throughout the whole graduation process whereas the geometric filter most assuredly will not. However, we believe it is the shape of the curve that is of interest and not the total population count.

We should now like to estimate ℓ_x based on the filtered data. (We shall illustrate the results for the geometrically filtered data only, but we shall give, in Table 2, results for the arithmetically filtered data also.)

Recall that our autoregressive model is

$$\ell_x = p_x \ell_{x-1} + \varepsilon_x, \quad x = 1,2,\ldots .$$

To give us an idea of the form of p_x, we can plot ℓ_x verses ℓ_{x-1}. See Figure 3. It is evident from Figure 3, that p_x is probably a constant and that the errors ε_x are fairly small. Fitting higher degree polynomials to p_x confirms that p_x is effectively a constant.**

We fitted models both of the form (2.1) with ε_x having constant variance σ^2 and of the form (2.6). An analysis of the residuals indicates that the model of (2.6) is more appropriate. Figure 4 is the plot of the residuals for the fitting of this model. If one assumes that the errors are normally distributed†, then one may compute the marginal likelihood product as a function of p. The ratio of this likelihood product as a function of p to the likelihood product for the least squares (maximum likelihood) estimate for model (2.6) is given in Figure 5. The sharpness of the plot of this ratio lends a good deal of confidence to this model and to the estimate of p, $\hat{p}$. The results of both models using data filtered both ways is summarized in Table 2.

Recall from (2.2), we have

$$E\ell_x = p^x E\ell_0.$$

** *That p_x should be a constant in a human population is indeed remarkable. In developed countries, one expects the probability of surviving to the next year to be constantly decreasing. Perhaps, this effect is due to relatively harsh conditions in India, where only the very hardy can survive to an old age. Being hardy, the probability of dying in the next year is not significantly different from that of the generally less hardy younger generations.*

† *We constructed a histogram of the residuals and it appears as though the errors are probably* not *normally distributed, rather tending more towards a lighter tailed distribution. Nonetheless, the likelihood ratio will be at least approximately correct. In the case that we fitted (2.6) to the data filtered by the arithmetic filter, the normal assumption is much more believable.*

TABLE 2

	Model (2.1) ϵ_x has variance σ^2	Model (2.6)
Geometric Filter	Estimated p_x = .9626476654 Estimated σ^2 = 19420364.46 Ratio of the Sums of Squares = 135.44	Estimated p_x = .9571073391 Estimated variance of η_x = .7596421223 Ratio of the Sums of Squares = 78.18
Arithmetic Filter	Estimated p_x = .9702683796 Estimated σ^2 = 21267288.58 Ratio of the Sums of Squares = 156.32	Estimated p_x = .9673863102 Estimated variance of η_x = .7538528904 Ratio of the Sums of Squares = 81.69

Thus $E\ell_x \propto p^x$. If we normalize so that $\sum_x E\ell_x = 1$, we have the expected population distribution. Using our estimate $\hat{\beta}$ from (2.6) and the geometrically filtered data, we may plot the estimated expected population distribution. This is done in Figure 6 along with the original data similarly normalized.

A direct statistical comparison of our methods with commonly used procedures is difficult to make since many commonly used procedures tend to arise from considerations other than statistical ones. We believe, however, that the statistical evidence gives a good deal of confidence in our procedures, particularly because of the relatively close agreement found in Table 2 in spite of the variations in the procedures. This statistical evidence together with the relative simplicity of our procedure and the previously mentioned objections to other procedures, certainly indicates that the time series approach merits consideration.

References

Census of India. 1963. 1961 Census-Age Tables. Paper No. 2 of 1963.

Chiang, C. L. 1960a. "A stochastic study of the life table and its applications: I. Probability distributions of the biometric functions." *Biometrics, 16*, 618-635.

_____ 1960b. "A stochastic study of the life table and its applications: II. Sample variance of the observed expectation of life and other biometric functions." *Human Biology, 32*, 221-238.

_____ 1961. "A stochastic study of the life table and its applications: III. The follow-up study with the consideration of competing risks." *Biometrics, 17*, 57-78.

Grenander, Ulf. 1956. "On the theory of mortality measurement, Part II." *Skan. Aktuarietidskr., 39*, 125-153.

Keyfitz, N. 1968. *Introduction to the Mathematics of Population.* Addison-Wesley. Reading, Mass.

Lotka, A. J. 1956. *Elements of Mathematical Biology.* Dover. New York.

APPENDIX

Legends for Figures

Figure 1. Data in Table 1 filtered by a 5 term geometric filter. Scale on vertical axis is $x10^5$.

Figure 2. Data in Table 1 filtered by a 5 term arithmetic filter. Scale on vertical axis is $x10^5$.

Figure 3. Scatter diagram (ℓ_x vs. ℓ_{x-1}) using data filtered by a five term geometric filter. Scale on both axes is $x10^5$.

Figure 4. Residuals using model of (2.6) fitted to data filtered by geometric filter. The arithmetic filter yields similar results.

Figure 5. Ratio of marginal likelihood products using model of (2.6) fitted to data filtered by geometric filter. The arithmetic filter again yields similar results.

Figure 6. Original data and expected curve from the model of (2.6) using data filtered geometrically. Both are normalized so that the total population is 1. Scale on the vertical axis is $x10^{-2}$.

FIGURE 1

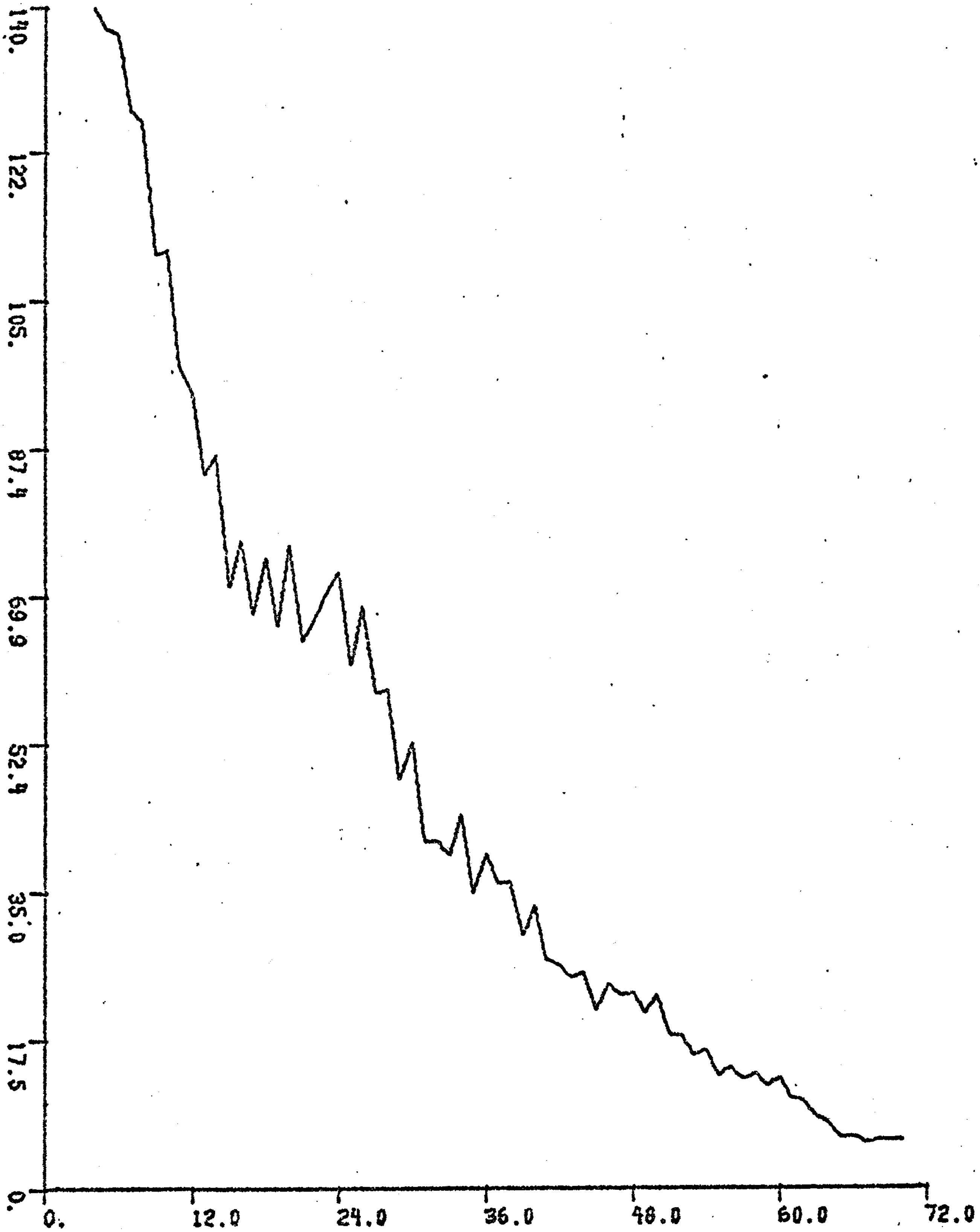

FIGURE 2

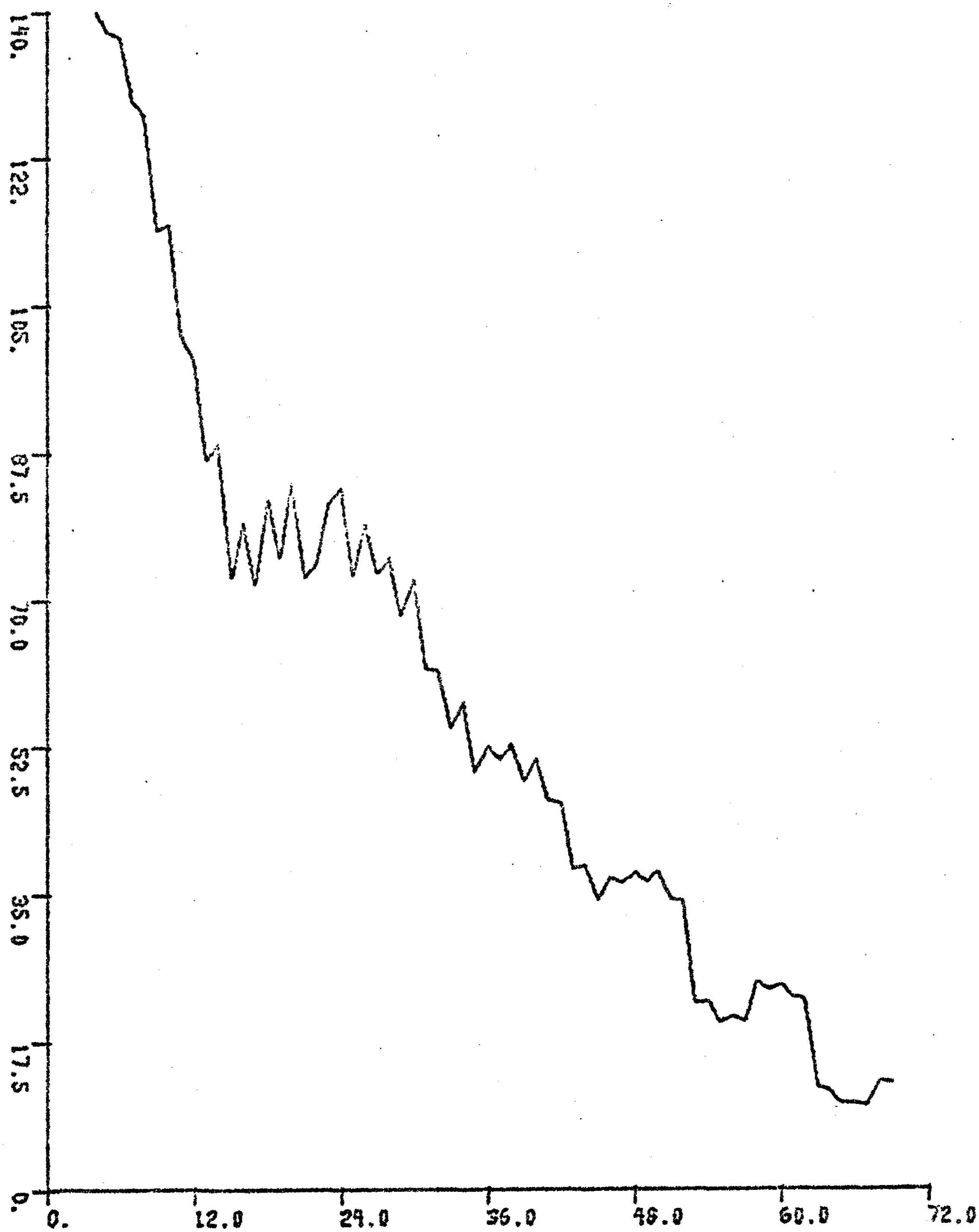

FIGURE 3

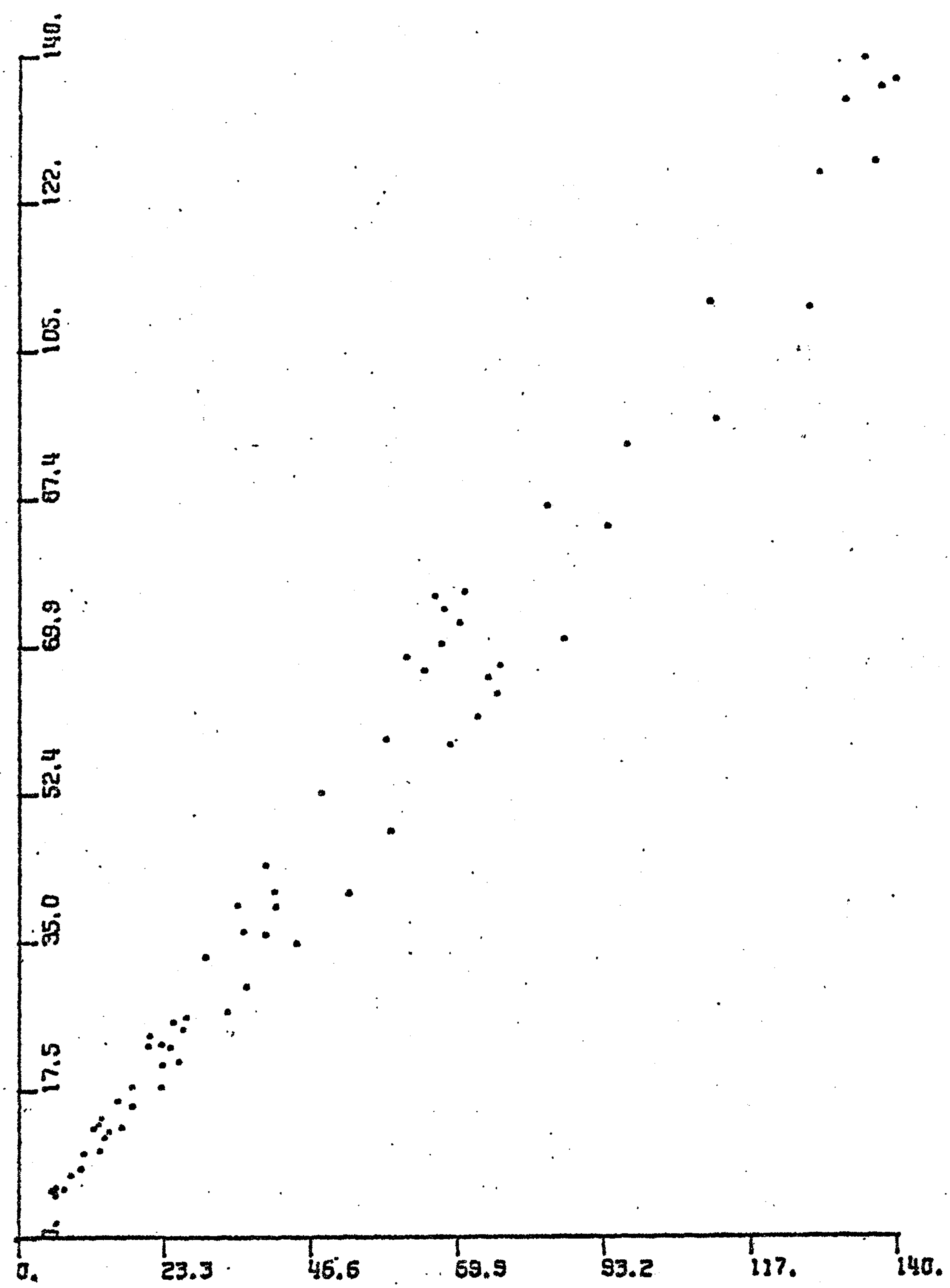

FIGURE 4

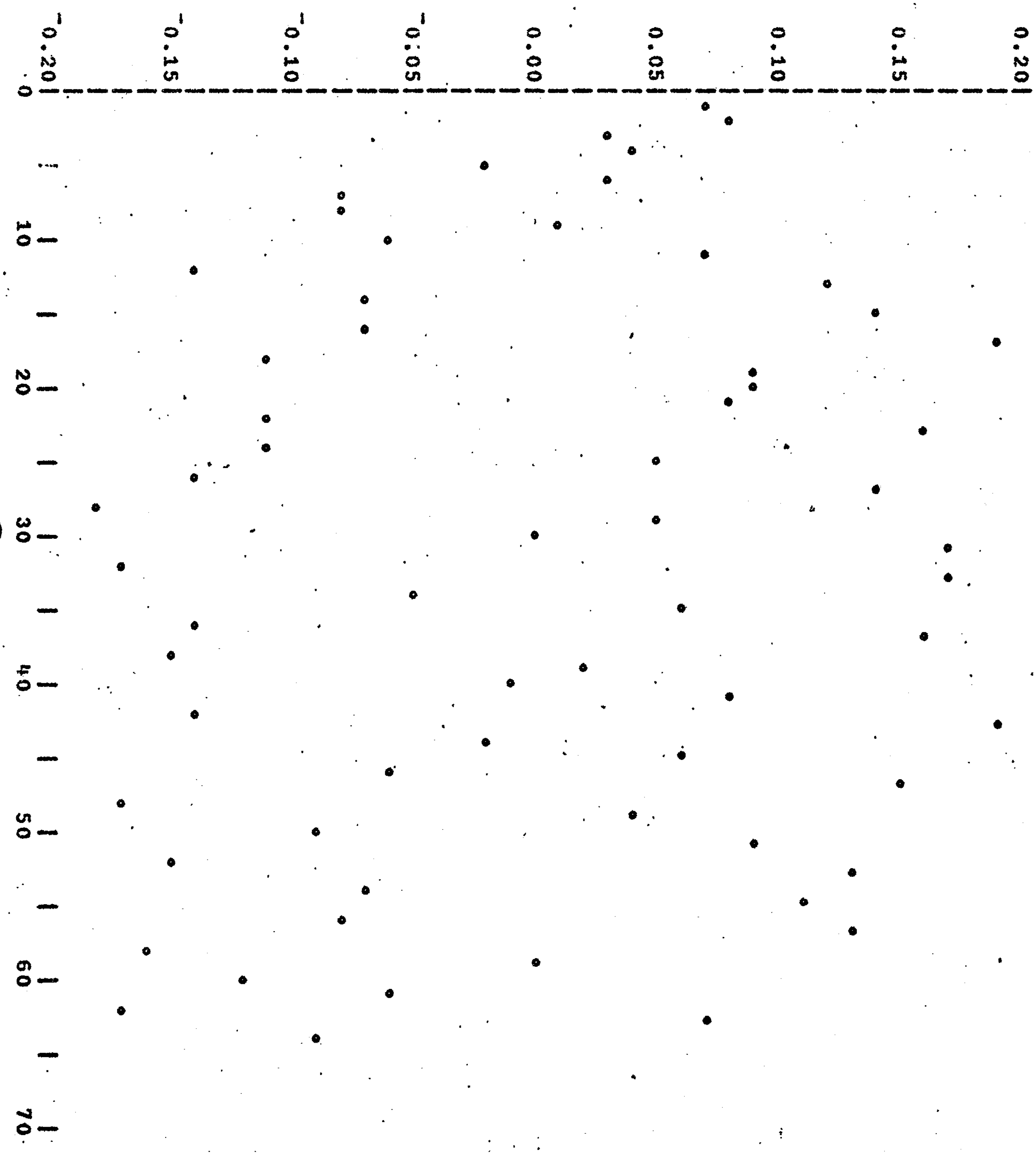

FIGURE 5

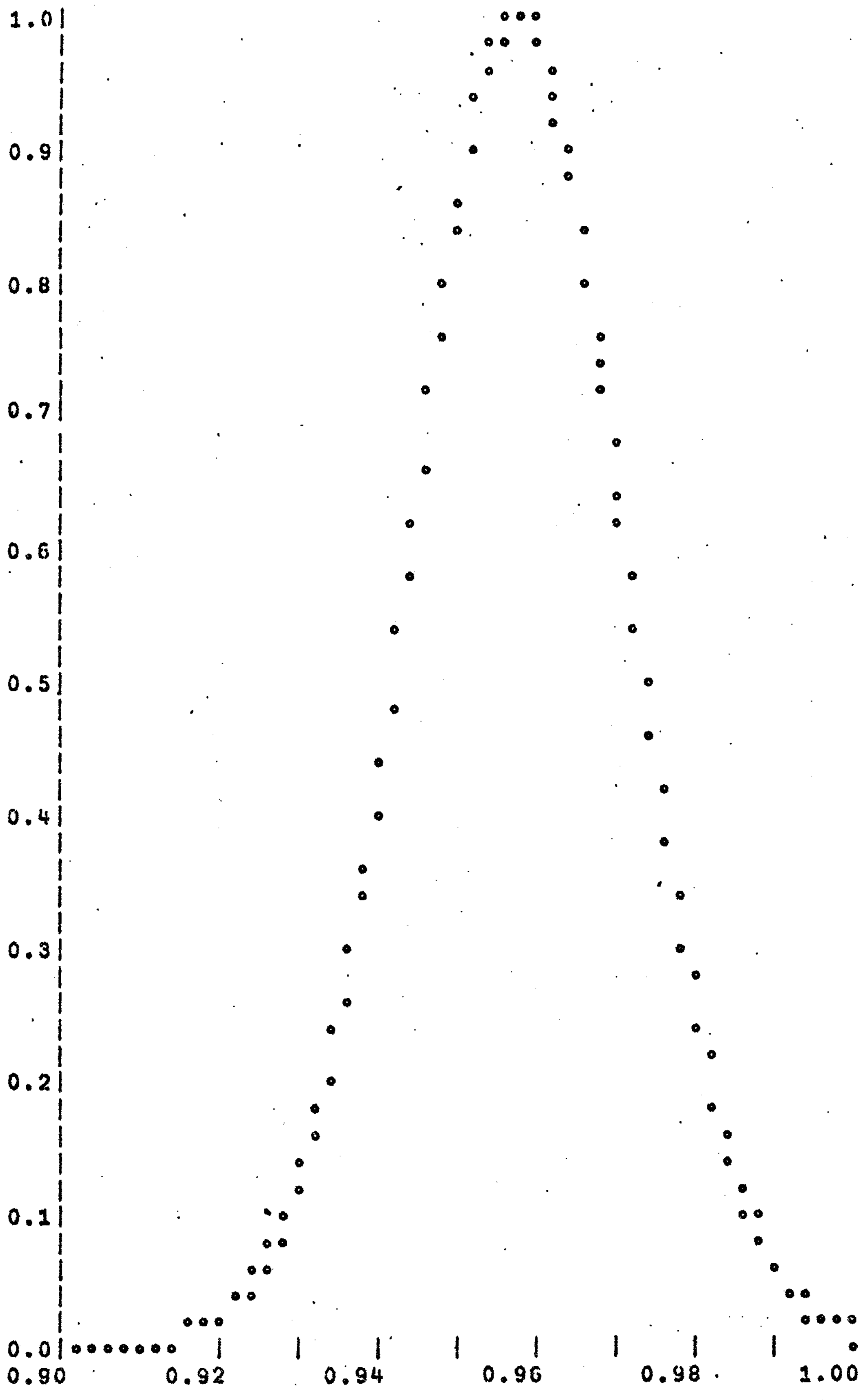

FIGURE 6