

QUANTILE DATA ANALYSIS OF IMAGE DATA

Emanuel Parzen
Department of Statistics
Texas A&M University
College Station, Texas 77843

ABSTRACT

Quantile data analysis and functional statistical inference methods are introduced and applied to provide representations of spectral data which may lead to simple statistical discriminators effective for the estimation of ground truth from satellite spectral measurements.

To estimate the ground truth of a pixel, we propose to estimate the probability of each possible ground truth, given observed (estimated) quantile-theoretic statistical characteristics of the multi-spectral image data corresponding to the pixel and its neighboring pixels. This paper describes a research strategy for determining which statistical characteristics discriminate best.

Results are reported of quantile data analysis of an extensive collection of training files of image data.

1. Introduction

To conduct research in image analysis, one must define its data, ends, and means.

The data consists of files. An image file consists of measurements taken on a specified date at a specified 5 x 6 nautical mile site on the earth's surface. A site is divided into a rectangular grid of (more than 20,000) surface elements [approximately 1 acre] called pixels. On each pixel, spectral measurements are made by satellite on four (and perhaps seven) channels [of the electromagnetic energy spectrum]. Each spectral measurement is an integer from 0 to 256.

The ends [goals] of image analysis is to estimate ground truth within the pixel; labels for ground truth include alfalfa, corn, soybeans, sugar beets, spring wheat, spring oats, grass, pasture, trees, water.

A file is called a training file if a ground truth record is available; each pixel is divided into six sub-pixels and ground truth is recorded for each sub-pixel.

The means of image analysis are currently under investigation by many investigators. A probability approach considers ground truth as a parameter [denoted θ]. A formal Bayesian statistical solution to the estimation of ground truth from data is to calculate $p(\theta|data)$, the posterior probability distribution of θ [the ground truth parameter] given the data. A formal maximum likelihood solution to the estimation of ground truth from data consists of two steps: (1) calculate the

likelihood function of θ , which equals $p(\text{data}|\theta)$, the conditional distribution of the data given that it is observed from a pixel with ground truth θ , and (2) use optimization algorithms to determine $\hat{\theta}$, the parameter value which maximizes likelihood. The foregoing formal statistical procedures are often described as being theoretically "optimal." But they may not be "good" in practice in the sense of correctly identifying ground truth with high probability.

To obtain high probability of discrimination, we recommend

(1) measuring suitable characteristics of probability models of the data, (2) treating the measured characteristics as new data, and estimating the likelihood function $p(\text{measured characteristics of data}|\theta)$, and (3) determining characteristics whose distributions for different values of θ are as wide apart as possible [the likelihood function is not flat and its optimum is easily determined].

This paper investigates the use of quantile data analysis to obtained measured characteristics of image data which have good power of discrimination between different values of ground truth. Only univariate analysis methods are used on channel 2 and channel 3 spectral observations. Future research will be concerned with bivariate analysis of the joint distribution of channel 2 and channel 3 measurements. Our approach to quantile data analysis strongly recommends that bivariate analysis be built on a foundation of univariate analysis. Therefore the univariate analysis techniques developed in this paper will not be rendered obsolete by the bivariate techniques to be developed in future research.

2. Outline of Quantile-Data Analysis of a Pixel

Let us describe a proposed method of statistical data analysis based on characteristics of the sample quantile functions of batches of measurements. Given a pixel whose ground truth we would like to estimate, let (t_1, t_2) be its coordinates which represent its position within the rectangular grid of pixels into which the scene has been divided.

Define $A_v(t_1, t_2)$, the v -neighborhood of a pixel, to be the set of pixels with coordinates $(t_1 + j_1, t_2 + j_2)$, where $j_1, j_2 = 0, \pm 1, \dots, \pm v$. For example $A_1(t_1, t_2)$ contains 9 pixels, $A_2(t_1, t_2)$ contains 25 pixels, $A_3(t_1, t_2)$ contains 49 pixels.

For $k=2$ and 3, the channel k measurements of the pixels in $A_v(t_1, t_2)$ are collected to form a data batch whose sample quantile function $\tilde{Q}(u)$ is formed. The "measured data characteristics" we associate with a pixel are various characteristics of the sample quantile function of a batch of measurements formed from the pixels surrounding a given pixel. The remainder of this section reviews quantile data analysis and defines the summary statistics that it suggests.

The probability law of a random variable X is usually described by its distribution function $F(x) = \Pr[X \leq x]$, $-\infty < x < \infty$, and probability density function $f(x) = F'(x)$. The quantile approach uses [see, for example, Parzen (1983)]

$$(1) \quad Q(u) = F^{-1}(u) = \inf \{x : F(x) \geq u\} ,$$

$$(2) \quad q(u) = Q'(u) ,$$

$$(3) fQ(u) = f(Q(u)) = \{q(u)\}^{-1}, \text{ and since } \frac{d}{du} F(Q(u)) = 1 \Rightarrow f(Q(u)) \cdot q(u) = 1$$

$$(4) J(u) = -(fQ)'(u)$$

A quick measure of location is the median $Q(0.5)$. A quick index of scale is the interquartile range $Q(0.75) - Q(0.25)$, formed for the quartiles $Q(0.25)$ and $Q(0.75)$.

Quick measures of distributional shape are provided by values (as u tends to 0 and 1) of the informative quantile function [recently introduced by Parzen].

$$IQ(u) = \frac{Q(u) - Q(0.5)}{2\{Q(0.75) - Q(0.25)\}}, \quad 0 \leq u \leq 1.$$

We cannot emphasize how powerful the IQ function appears to be in practice as a tool for the diagnosis of distributional shapes.

The IQ function is independent of location and scale parameters. It is approximately equivalent to normalizing a quantile function to have the properties $Q(0.5) = 0$, $Q'(0.5) = 1$. The IQ graph of the function provides us at a glance with a vague estimate of tail behavior as defined by tail exponents.

A fundamental description of the tail behavior of distributions is provided by the left tail exponent α_0 and the right tail exponent α_1 defined as follows:

$$fQ(u) = u^{\alpha_0} L_0(u) \text{ as } u \rightarrow 0$$

$$fQ(u) = (1-u)^{\alpha_1} L_1(u) \text{ as } u \rightarrow 1$$

where $L_0(u)$ and $L_1(u)$ are slowly varying functions.

A function $L(u)$ is slowly varying as $u \rightarrow 0$ if, for every $y > 0$,

$$\lim_{u \rightarrow 0} \frac{L(yu)}{L(u)} = 1.$$

Tail behavior is defined in terms of a tail exponent as follows:

$\alpha < 1$: short tail

$\alpha = 1$: medium tail

$\alpha > 1$: long tail

Medium tail ($\alpha = 1$) distributions are further classified by the value of

$$h_0 = \lim_{u \rightarrow 0} \frac{f(u)}{u}, \quad h_1 = \lim_{u \rightarrow 1} \frac{f(u)}{1-u};$$

the letter h is suggested by the notion of hazard function. We define

$h = .0$: medium-long tail

$0 < h < \infty$: medium-medium tail

$h = \infty$: medium-short tail

Extensive calculations of informative quantile functions indicate that the value IQ_0 of $IQ(u)$ for u near 0 is a quick indicator of left tail behavior:

$-0.5 \leq IQ_0 < 0$: short left tail,

$-1.0 \leq IQ_0 < -0.5$: medium-short left tail,

$IQ_0 < -1.0$: medium-medium to long left tail.

proof?
→ investigate

Similarly the value IQ_1 of $IQ(u)$ for u near 1 is a quick indicator of right tail behavior:

$0 < IQ_1 \leq 0.5$: short right tail,

$0.5 < IQ_1 < 1.0$: medium-short left tail,

$1.0 < IQ_1$: medium-medium to long right tail

An important family of distributions is the Weibull with shape parameter β . Its quantile function $Q(u)$ is of the form

$$Q(u) = \mu + \sigma Q_0(u)$$

where

$$Q_0(u) = \frac{1}{\beta} \{\log(1-u)^{-1}\}^{\beta}$$

Its density-quantile

$$f_0 Q_0(u) = (1-u) \{\log(1-u)^{-1}\}^{1-\beta}$$

Its right tail exponent is $\alpha = 1.$, and its left tail exponent is $\alpha_0 = 1-\beta$. Insight into the interpretation of informative quantile functions is obtained by computing them for Weibull distributions.

Given data, we distinguish three types of estimators of population parameters, which we call: (1) fully non-parametric, (2) fully parametric, and (3) functional-parametric. Fully non-parametric estimators assume no model, and provide quick estimators. Fully parametric estimators assume a model known up to a finite number of parameters which must be estimated. Functional-parametric estimators are based on methods of functional statistical inference.

A fully non-parametric estimator $\tilde{Q}(u)$ of $Q(j)$, given a sample of n distinct values $x_{1;n} < x_{2;n} < \dots < x_{n;n}$, is defined by (for $j=1, \dots, n$)

$$\tilde{Q}(u) = x_{j;n}, \quad \frac{j-1}{n} < u \leq \frac{j}{n}$$

For a large sample, or for grouped values, we form a histogram before

computing $\hat{Q}(u)$ by linear interpolation at an equi-spaced grid of values kh , $k=1, 2, \dots, [1/h]$ where usually $h = 0.01$.

3. Example and Interpretation of a Quantile Data Analysis

To illustrate the uses of measured data characteristics provided by quantile data analysis, let us consider the analysis of a training file which contains both image data and ground truth data. We search through the ground truth file to see what codes appear more than a few times. The codes found to be present corresponded to the ground truth values listed in Table A. For a ground truth value j , we created a data batch consisting of all the channel 2 values observed in a pixel one of whose sub-pixels had a ground truth equal to the value j . We created a similar data batch of channel 3 observations. Table A lists the sample sizes of the number of observations in these data batches and the medians and interquartile ranges of the channel 2 and channel 3 observations. One immediately sees a pattern which might provide a discrimination statistic Δ to be used in determining ground truth. One might be able to readily distinguish the category "grass, pasture, trees" from "corn, soybeans, sugar beets, spring wheat, spring oats" by the values of

$$\Delta_1 = \text{median (channel 3)} - \text{median (channel 2)}$$

$$\Delta_2 = \log \frac{\text{IQ range (channel 3)}}{\text{IQ range (channel 2)}}$$

The values of these statistics are given in Table A. Note that $\Delta_1 > 2$ for grass, pasture, and trees, and $\Delta_1 < 2$ for crops. Of the crops,

alfalfa is closest in statistical characteristics to grass, pasture, and trees; this conclusion is reached also in Table B.

Table A lists statistics based on comparisons of location and scale estimators; Table B lists discriminators which are based on shape and tail characteristics. We consider the following four characteristics as statistics which might discriminate between (ground truth) distributions:

$$\Delta_3 = \text{MEAN IQ} = \frac{\text{MEAN} - \text{MEDIAN}}{2 \times \text{INTERQUARTILE RANGE}}$$

$$\Delta_4 = \text{STD IQ} = \frac{\text{STANDARD DEVIATION}}{2 \times \text{INTERQUARTILE RANGE}}$$

$$\Delta_5 = \text{IQ}_0 = \text{IQ}(u) \text{ for } u \text{ near 0}$$

$$\Delta_6 = \text{IQ}_1 = \text{IQ}(u) \text{ for } u \text{ near 1}$$

The values of these statistics in this example indicate that trees, pasture, and grass have spectral observations with distributions closer to normal, while crops have spectral observations with distributions further from normal.

It should be strongly emphasized that these empirical patterns found in one file are not intended to be general algorithms applicable to all files. They are presented only as an illustration of the kinds of facts about image data which quantile data analysis proposes to discover through extensive computation on training files.

→ Quantile-based
statistics can be
used as discriminators

TABLE A

	Sample Size	Median Channel 2	Median Channel 3	$\frac{\Delta_1}{\text{Median}(3) - \text{Median}(2)}$
Alfalfa	377	19	20	1
Corn	8,755	15	14	-1
Soybeans	11,000	15	13	-2
Sugar Beets	793	14	12	-2
Spring Wheat	2,296	18	16	-2
Spring Oats	558	18	16	-2
Grass	174	23	26	3
Pasture	-248	-21	28	7
Trees	95	20	24	4

	IQ Range Channel 2	IQ Range Channel 3	$\frac{\Delta_1}{\text{Log IQ}(3) - \text{Log IQ}(2)}$
Alfalfa	9	16.75	.62
Corn	5	6.5	.26
Soybeans	5	8	.47
Sugar Beets	4	4.5	.12
Spring Wheat	6	9	.41
Spring Oats	8	11	.32
Grass	8	12.5	.45
Pasture	5	13	.96
Trees	6	11	.61

TABLE B

	Mean IQ Channel 2	Mean IQ Channel 3	STD IQ Channel 2	STD IQ Channel 3
Alfalfa	-.08	-.07	.32	.27
Trees	-.08	-.01	.38	.35
Pasture	-.04	-.06	.41	.32
Grass	-.01	.02	.36	.34

Spring Wheat	.07	.11	.38	.41
Spring Oats	.09	.12	.36	.35
Sugar Beets	.14	.06	.42	.49
Corn	.17	.10	.44	.51
Soybeans	.17	.13	.46	.41

	IQ ₀ Channel 2	IQ ₀ Channel 3	IQ ₁ Channel 2	IQ ₁ Channel 3
--	------------------------------	------------------------------	------------------------------	------------------------------

Alfalfa	-.34	-.32	1.05	.68
Trees	-.75	-.72	1.0	.68
Pasture	-1.0	-.76	1.1	.65
Grass	-.75	-.72	.81	.68

Spring Wheat	-.58	-.44	2.08	1.66
Spring Oats	-.43	-.36	1.18	1.0
Sugar Beets	-.37	-.44	2.25	2.55
Corn	-.40	-.46	2.8	3.0
Soybeans	-.40	-.31	2.7	1.93

Note: STD IQ = .37 for normal. Above line characteristics close to normal. Below line characteristics far from normal.

4. Quantile Data Analysis of Statistical Characteristics Estimated from Pixel Neighborhoods

A program of fundamental research on the quantile data analysis approach to picture segmentation poses many detailed problems for research. This section gives an example of one sample quantile data analysis. (1) Consider all pixels in a file whose ground truth is a specified crop (spring wheat is considered here). (2) For each such pixel form a 5 by 5 neighborhood of pixels (with the specified pixel at the center). (3) For each neighborhood form a data batch of spectral observations (channels 2 and 3 are considered here). (4) For each data batch, form its sample quantile function and estimate typical univariate quantile theoretic statistical characteristics: median, IQR (interquartile range), mean IQ (mean of informative quantile function), STDIQ (standard deviation of IQ function), IQ(.01), IQ(.99), average log spacings (which is a fully non-parametric estimator of entropy of the IQ function), and $\log \sigma_0$ [where σ_0 is the score deviation, defined as the sum of products of the spacings of the IQ function and a specified density-quantile function $f_0 Q_0(u)$]. The specified density quantile functions that we use are the logistic distribution

$$f_0 Q_0(u) = u(1-u) ,$$

and the Weibull distribution with quantile shape parameter β [we choose $\beta = 0.7$]

$$f_0 Q_0(u) = (1-u) \{-\log (1-u)\}^{1-\beta}$$

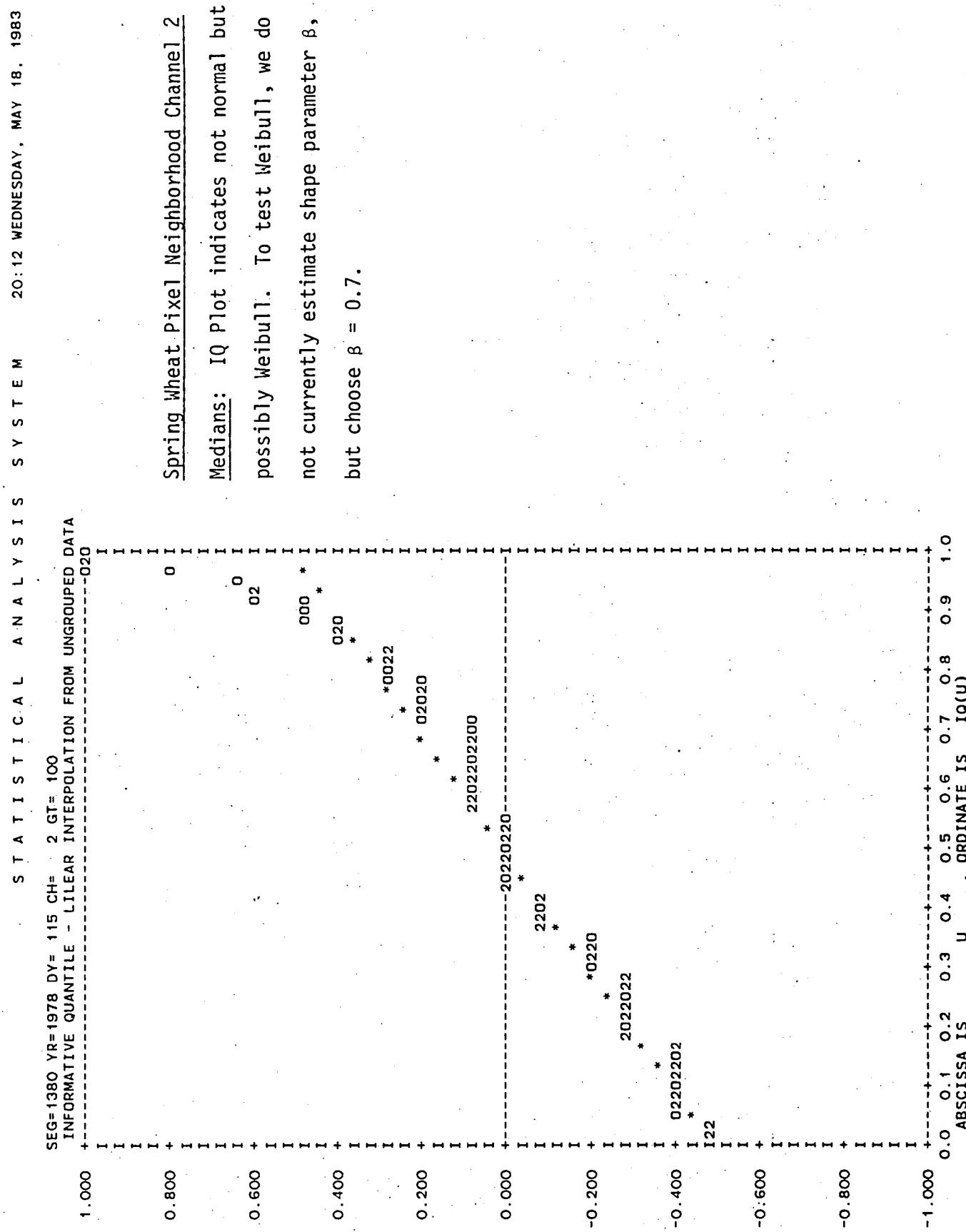
Step (5) is to form, for each statistical characteristic, a data batch of the several thousand estimates of that characteristic corresponding to the pixels in the training file with the specified ground truth [here, spring wheat (code 100)]. Step (6) is to do a one-sample quantile data analysis of this data batch. These analyses are presented in detail for the following estimators: median channel 2, mean IQ channel 2, median channel 3, mean IQ channel 3.

The following table lists some basic summary measures for a one-sample statistical analysis of a data batch of statistical characteristics of Spring Wheat pixel neighborhoods:

	Median Channel 2	Median Channel 3	Mean IQ Channel 2	Mean IQ Channel 3
Median	18	16	.02	.04
IQR	5	7.75	.14	.13
Mean IQ	.04	.09	.03	.01
Std IQ	.41	.36	.44	.49
Av. Log Spacings	-.68	-.59	.43	.43
σ_0 Weibull	.67	.61	.78	.80
σ_0 Logistic	.34	.20	.22	.22

We give for these variables: (1) printer plots of the informative quantile functions; (2) estimated density quantile functions, computed by the method of autoregressive density estimation; using the logistic and Weibull distributions as bases; and (3) diagnostic distribution functions (to be compared with the uniform) that help us decide which autoregressive order to accept as providing a "parsimonious" estimator.

haven't read

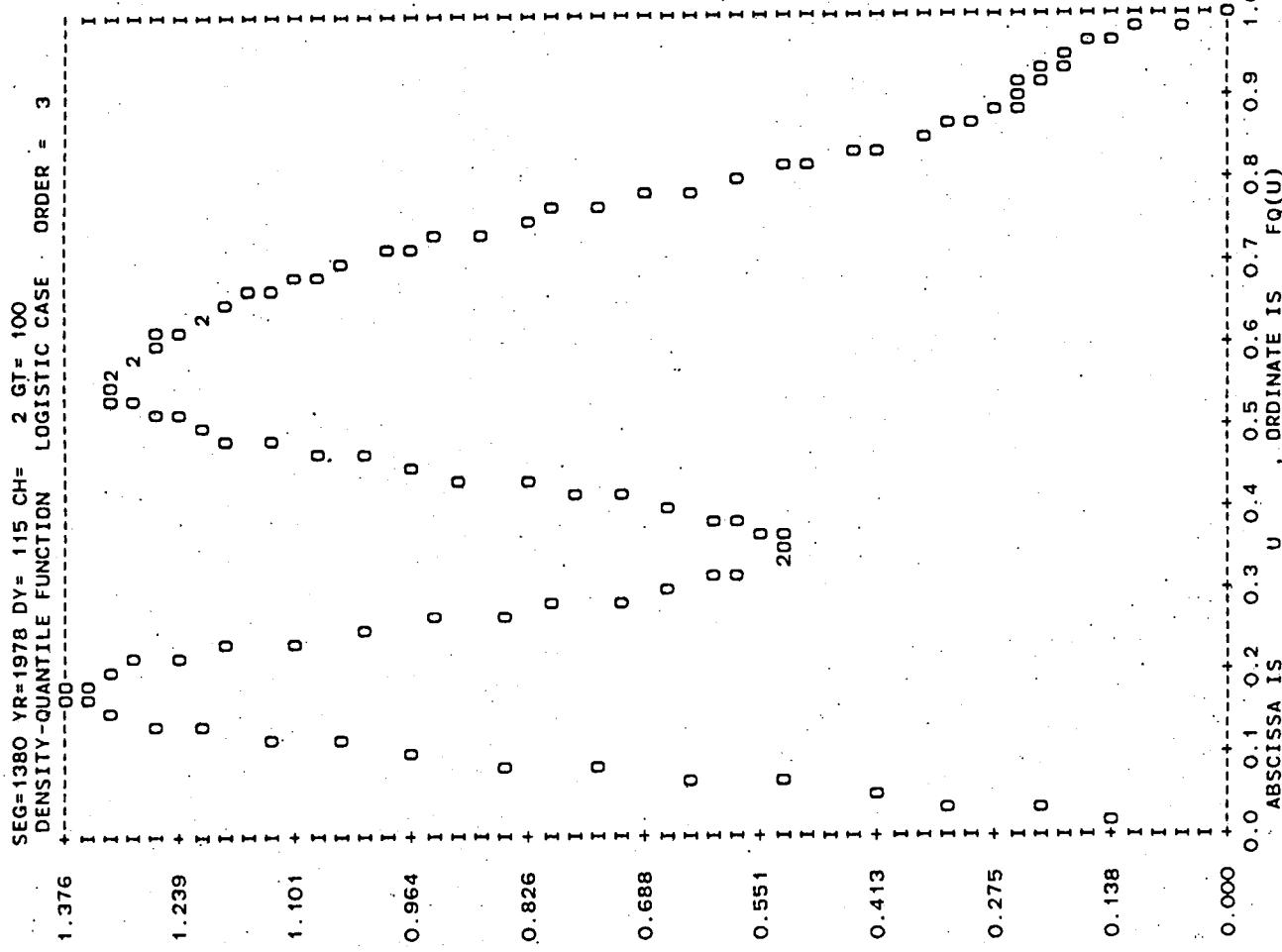


STATISTICAL ANALYSIS SYSTEM

20:12 WEDNESDAY, MAY 18, 1983

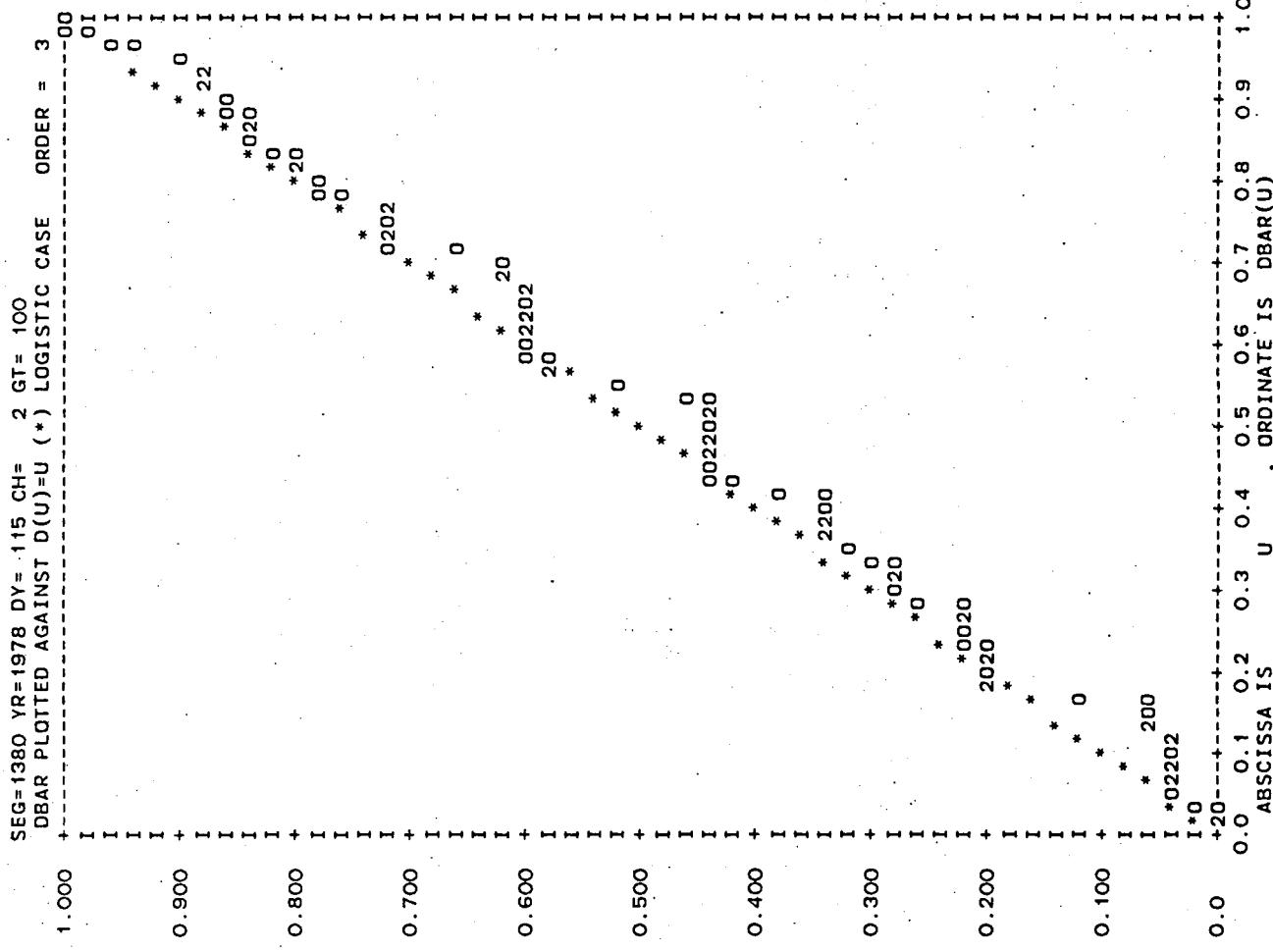
12

206

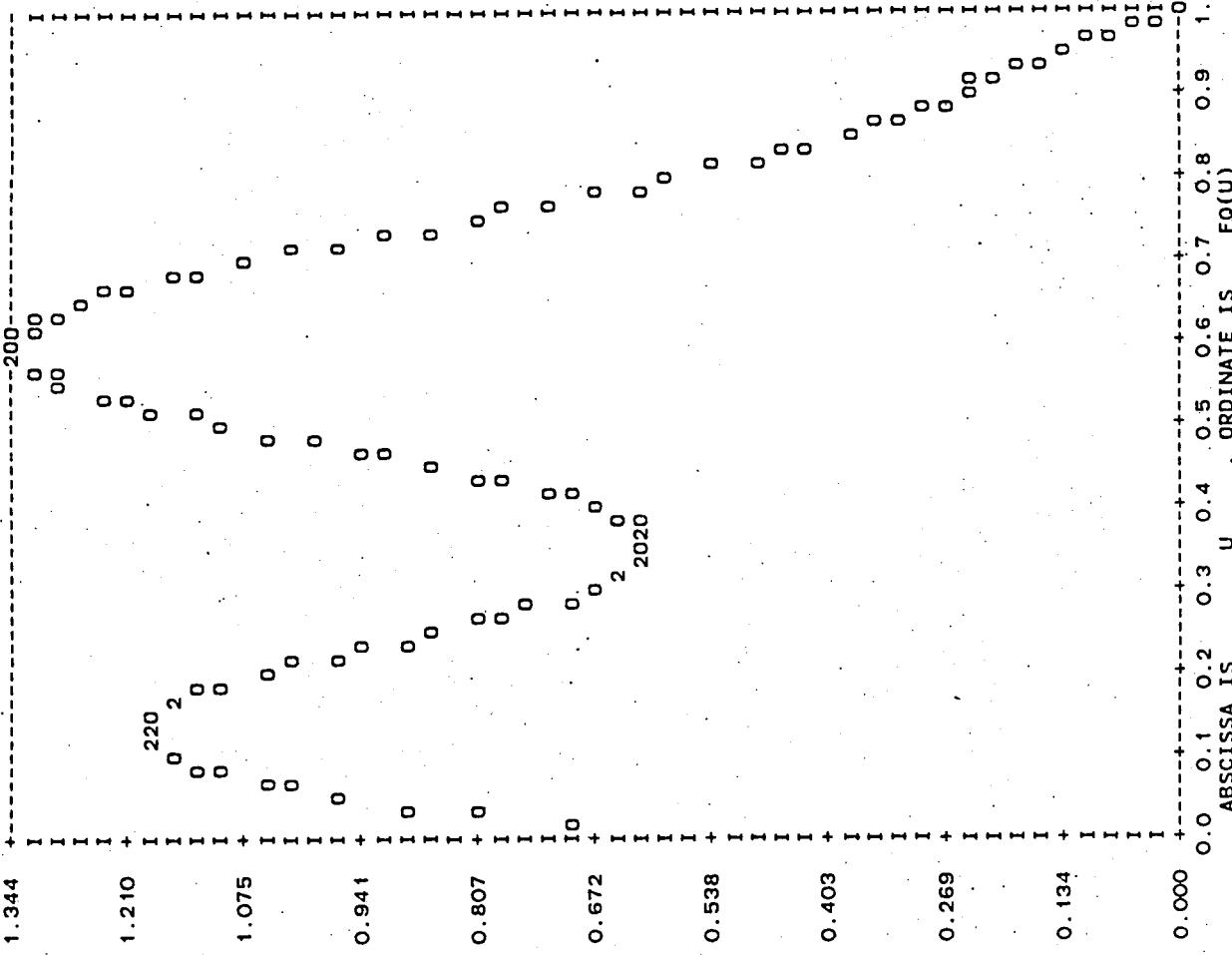
Spring Wheat Pixel Neighborhood Channel 2Medians: Autoregressive density quantile estimator (with logistic base and order 3)

indicates bimodal density.

STATISTICAL ANALYSIS SYSTEM 20:12 WEDNESDAY, MAY 18, 1983 11



STATISTICAL ANALYSIS SYSTEM

SEG=1380 YR=1978 DY= 115 CH= 2 GT= 100
DENSITY-QUANTILE FUNCTION WEIBULL CASE ORDER = 2

208

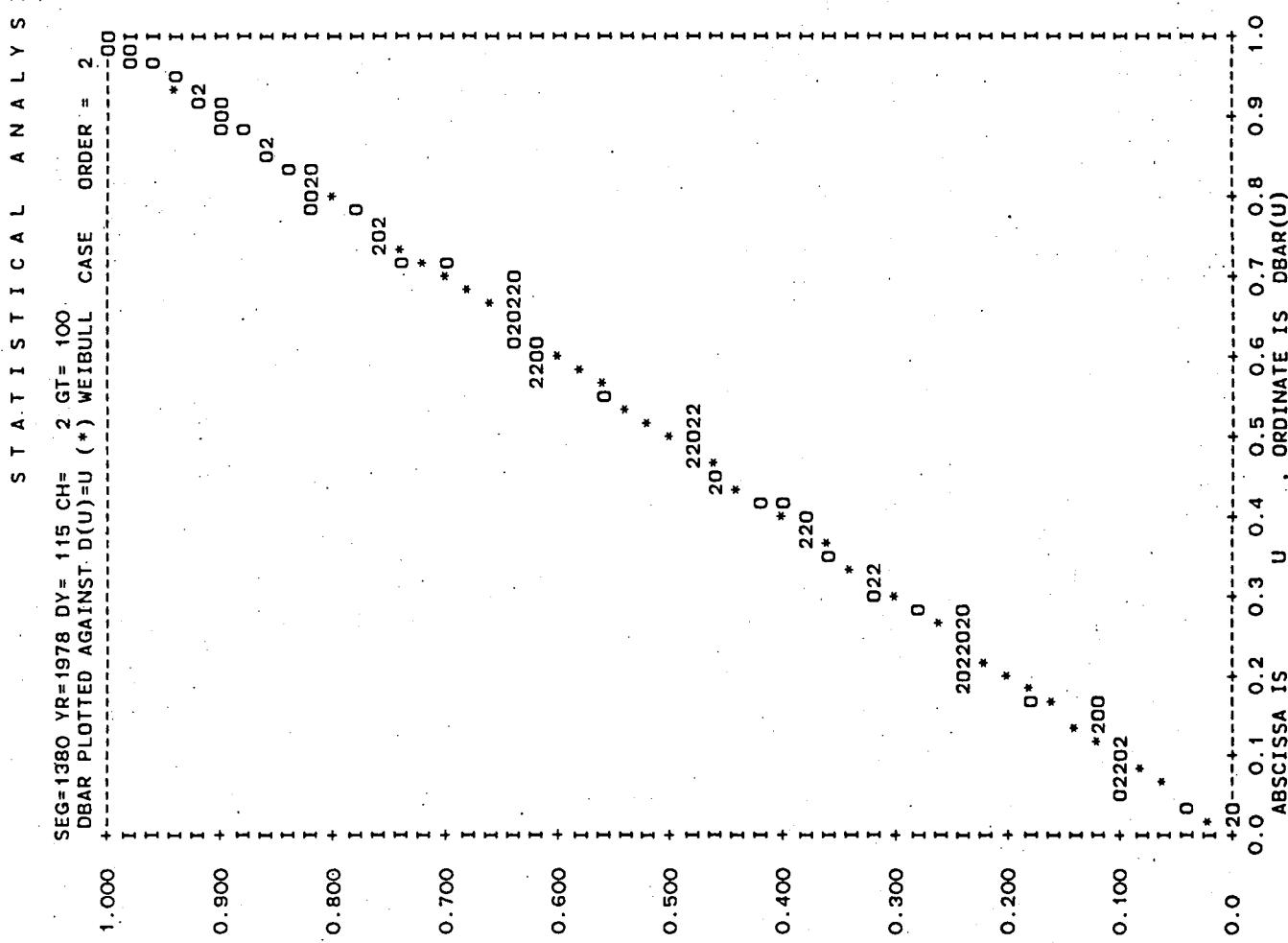
Spring Wheat Pixel Neighborhood Channel 2Medians: Autoregressive density quantile

analysis (with Weibull shape parameter .7

base and order 2) indicates bimodal

density.

23:45 WEDNESDAY, MAY 18, 1983 10



Spring Wheat Pixel Neighborhoods Channel 1 2

Medians: Diagnostic of fit of AR density
 quantile estimator (with Weibull shape
 parameter .7 base and order 2).

STATISTICAL ANALYSIS SYSTEM

20:12 WEDNESDAY, MAY 18, 1983

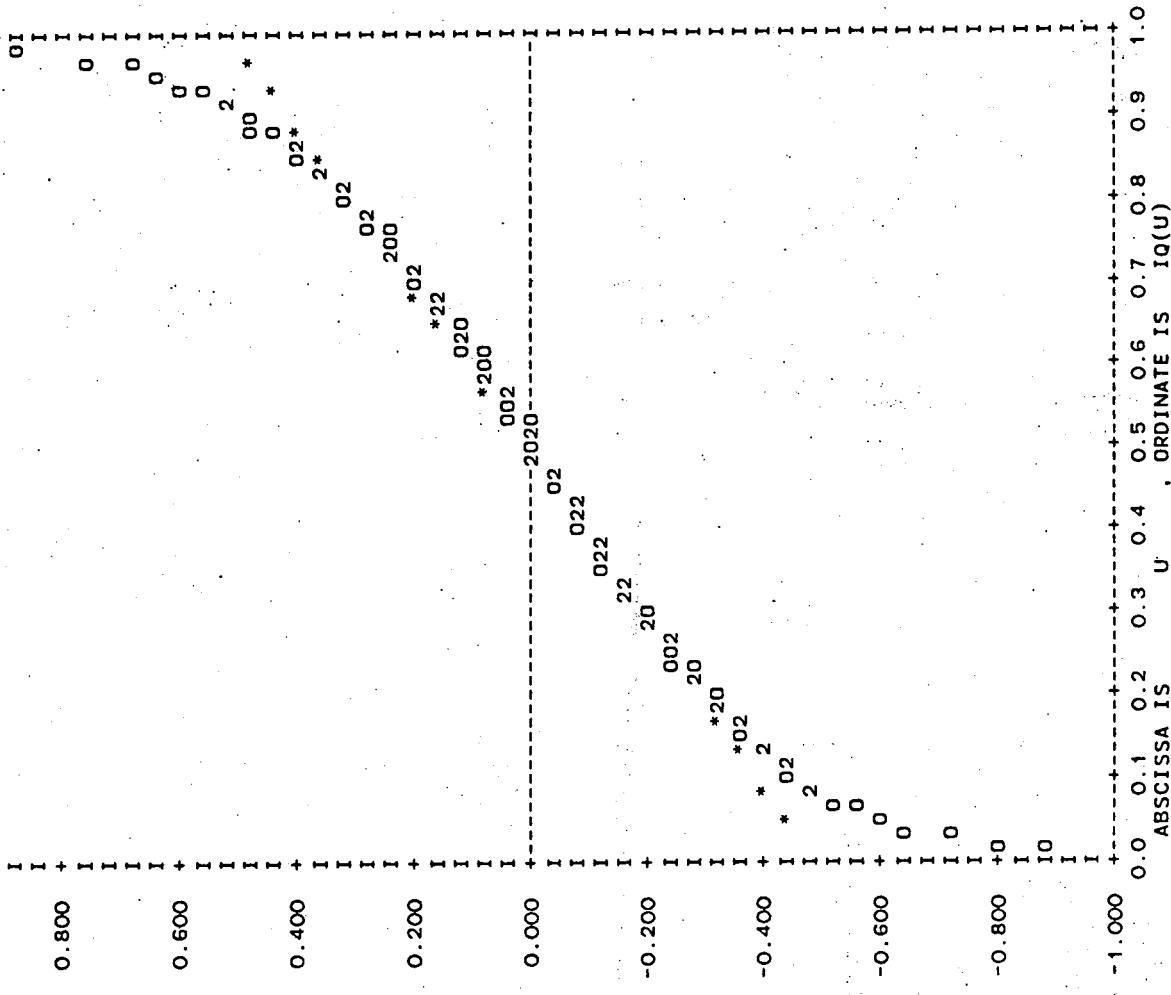
37

SEG=1380 YR=1978 DY= 115 CH= 2 GT= 100
 INFORMATIVE QUANTILE - LINEAR INTERPOLATION FROM UNGROUPED DATA

210

Spring Wheat Pixel Neighborhood Channel 2

Mean IQ: IQ plot indicates almost perfect normality.



AUTOREGRESSIVE PARAMETRIC SELECT ANALYSIS

Mean IQ Logistic Base

SQUARED MODULUS OF FOURIER COEFFICIENTS

```

PHI2( 1) = 0.000999923
PHI2( 2) = 0.000039880 *
PHI2( 3) = 0.000049510
PHI2( 4) = 0.000095499
PHI2( 5) = 0.000067322 *

```

AUTOREGRESSIVE PARAMETRIC SELECT ANALYSIS

Mean IQ Weibull Case

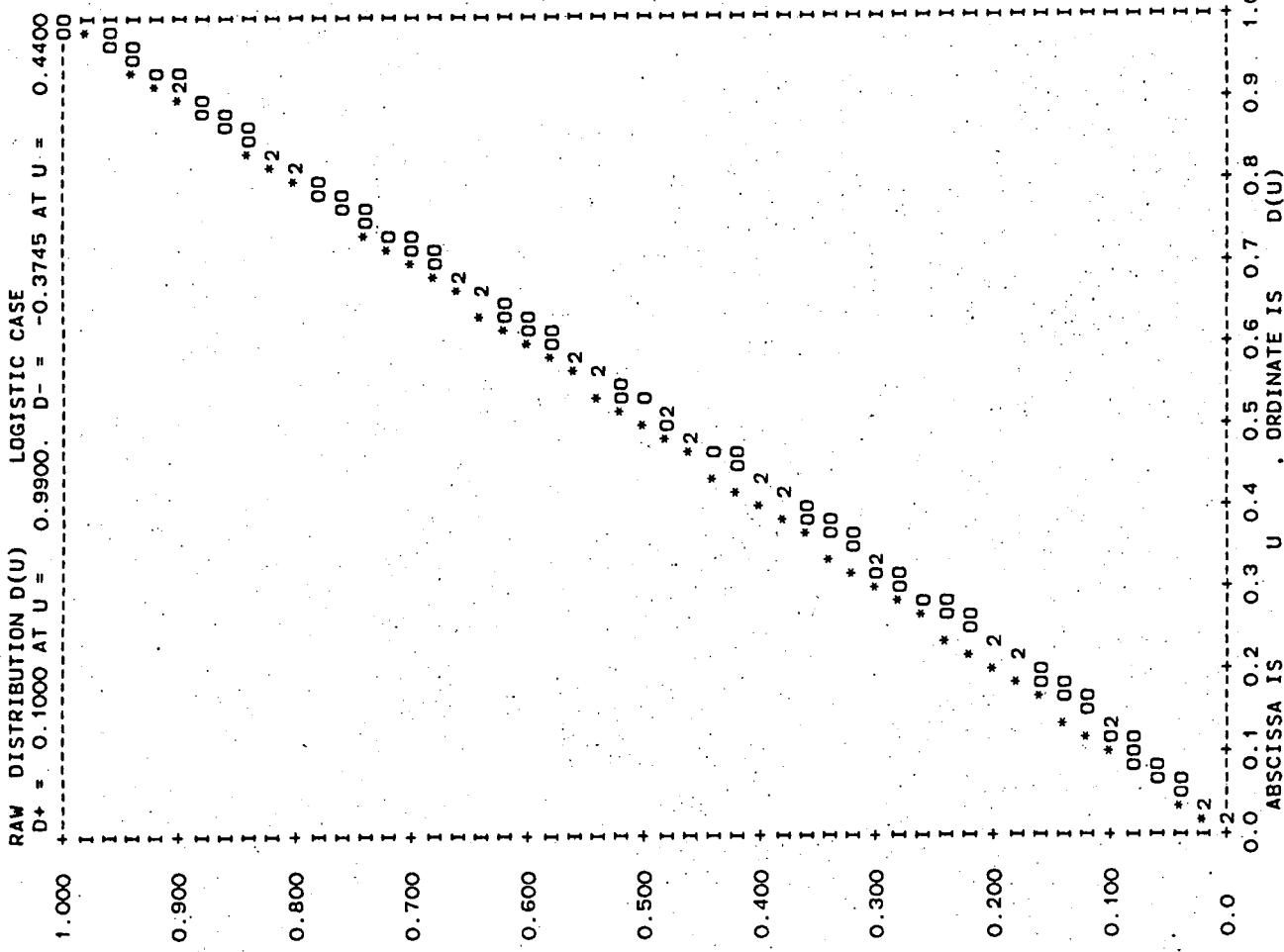
SQUARED MODULUS OF FOURIER COEFFICIENTS

```

PHI2( 1) = 0.07646000
PHI2( 2) = 0.03759194 *
PHI2( 3) = 0.02402839 *
PHI2( 4) = 0.02089854 *
PHI2( 5) = 0.01483078 *

```

Spring Wheat Pixel Neighborhood Channel 2 Mean IQ: Pseudo-correlations square nodulus (phi 2) accept logistic distribution, reject Weibull distribution fit.

Spring Wheat Pixel Neighborhood Channel 2

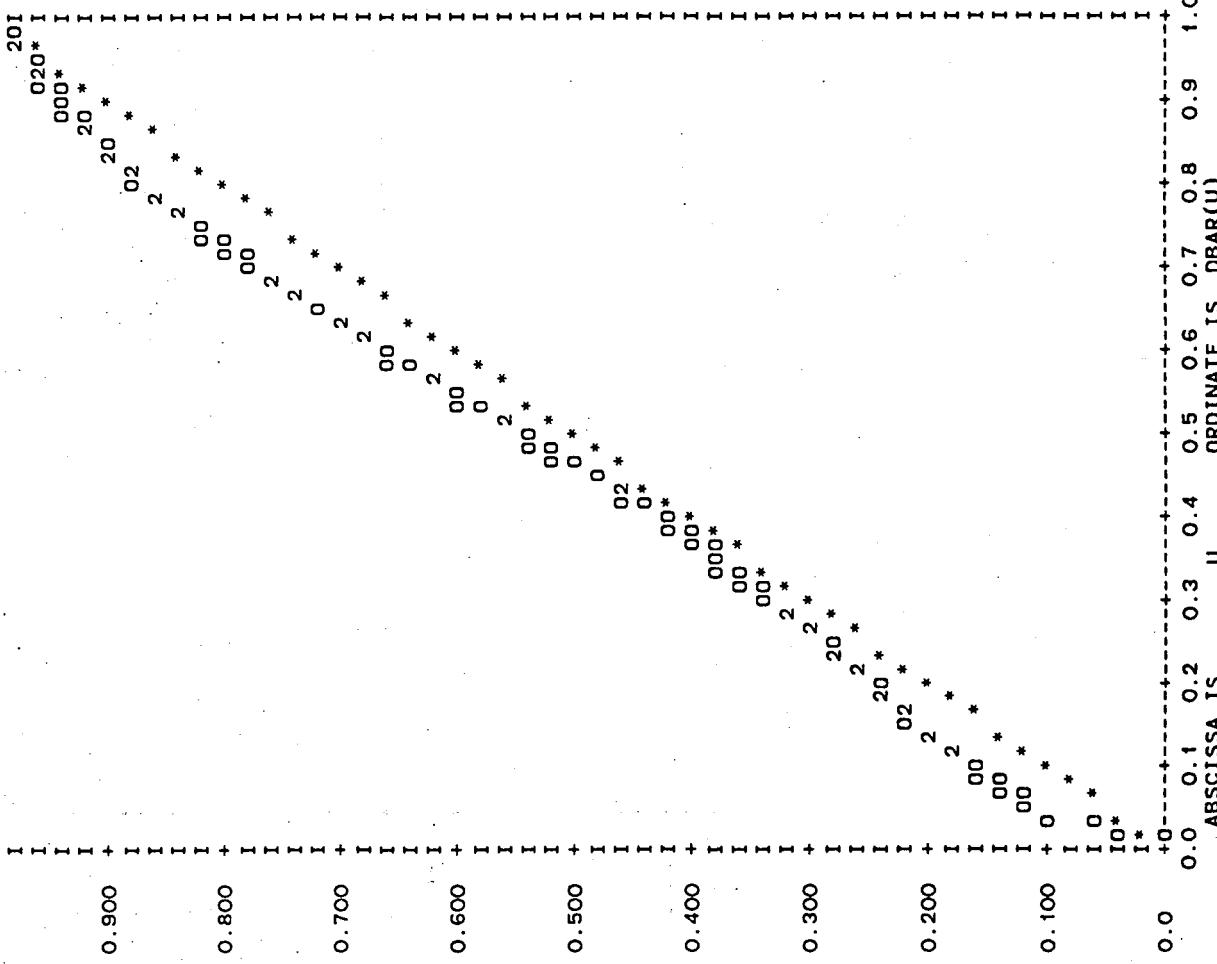
Mean IQ: Cumulative weighted spacings

$D(u)$ plot indicates accept fit of logistic distribution.

STATISTICAL ANALYSIS SYSTEM

23:45 WEDNESDAY, MAY 18, 1983

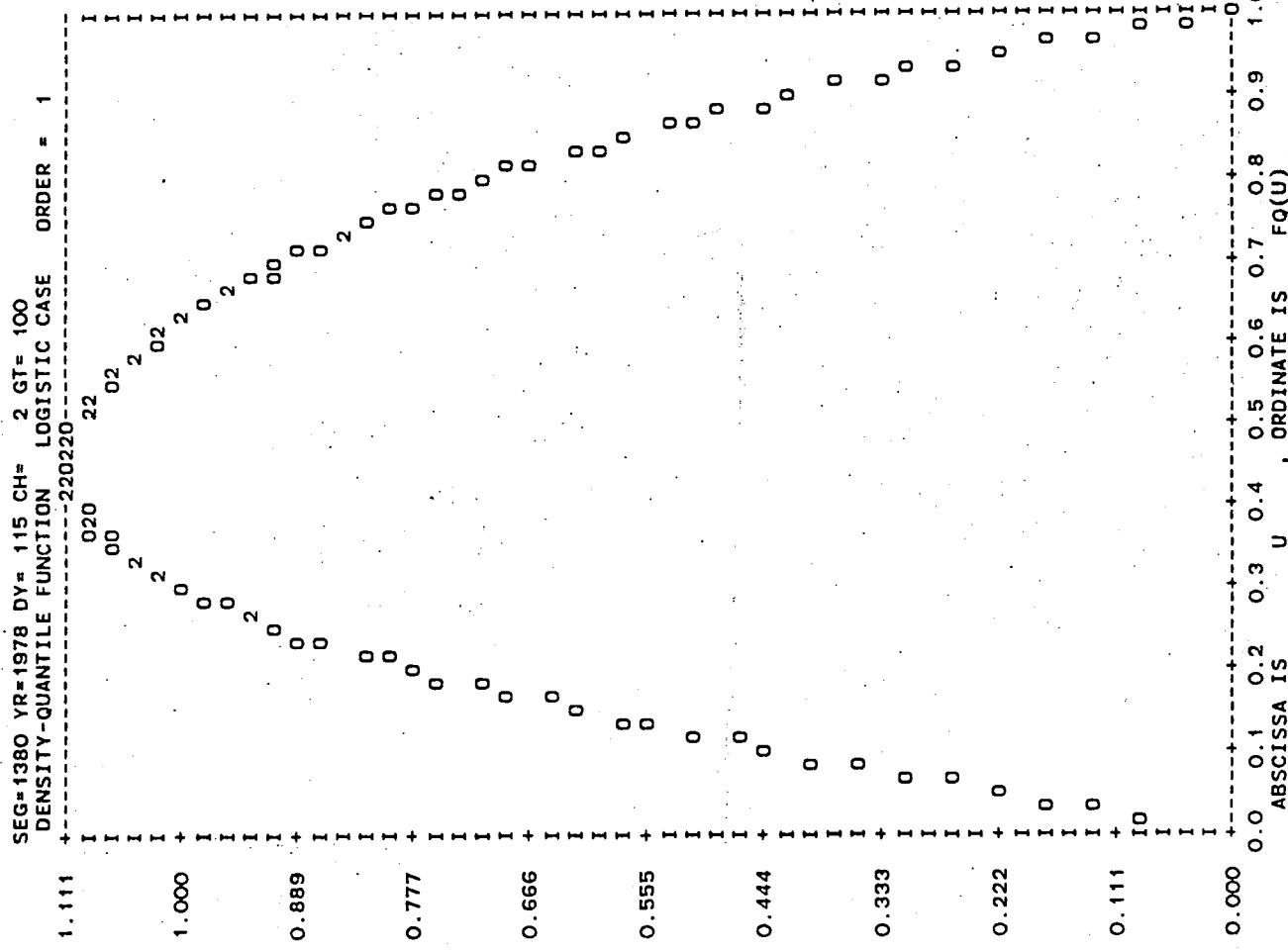
SEG=1380 YR=1978 DY= 115 CH= 2 GT= 100
DBAR PLOTTED AGAINST D(U)=U (*) WEIBULL CASE ORDER = 1



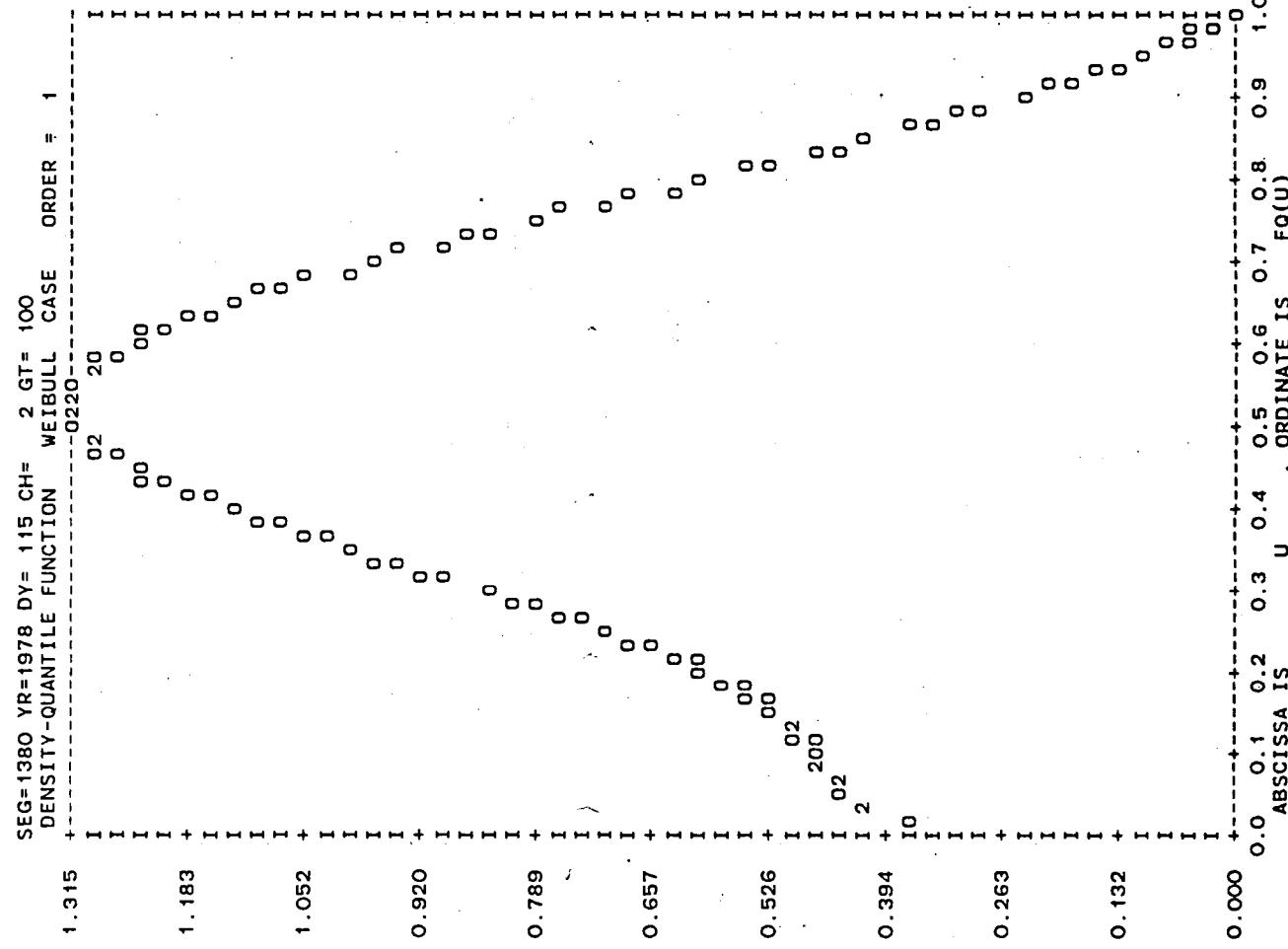
STATISTICAL ANALYSIS SYSTEM

20:12 WEDNESDAY, MAY 18, 1983

42



STATISTICAL ANALYSIS SYSTEM 23:45 WEDNESDAY, MAY 18, 1983 42



STATISTICAL ANALYSIS SYSTEM

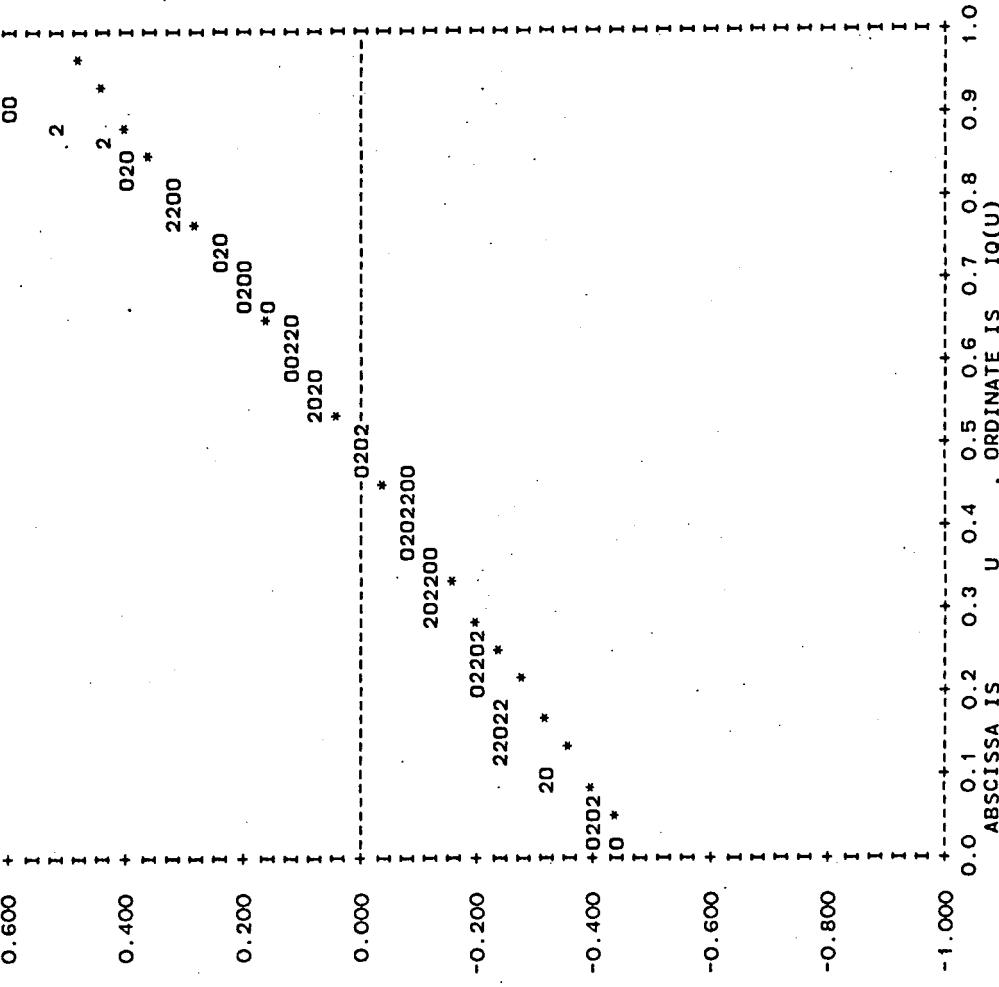
23:45 WEDNESDAY, MAY 18, 1983 139

SEG=1380 YR=1978 DY= 115 CH= 3 GT= 100
 INFORMATIVE QUANTILE - LINEAR INTERPOLATION FROM UNGROUPED DATA

1.000 +
 0.800 +
 0.600 +
 0.400 +
 0.200 +
 0.000 +
 -0.200 +
 -0.400 +
 -0.600 +
 -0.800 +
 -1.000 +

Spring Wheat Pixel Neighborhood Channel 3Medians: IQ plot indicates not normal

but possibly Weibull.



STATISTICAL ANALYSIS SYSTEM 20:12 WEDNESDAY, MAY 18, 1983 142

AUTOREGRESSIVE PARAMETRIC SELECT ANALYSIS

SQUARED MODULUS OF FOURIER COEFFICIENTS

```
.....  
PHI2( 1) = 0.0388847  
PHI2( 2) = 0.00517420 *  
PHI2( 3) = 0.01129055 *  
PHI2( 4) = 0.00169468 *  
PHI2( 5) = 0.01800444 *
```

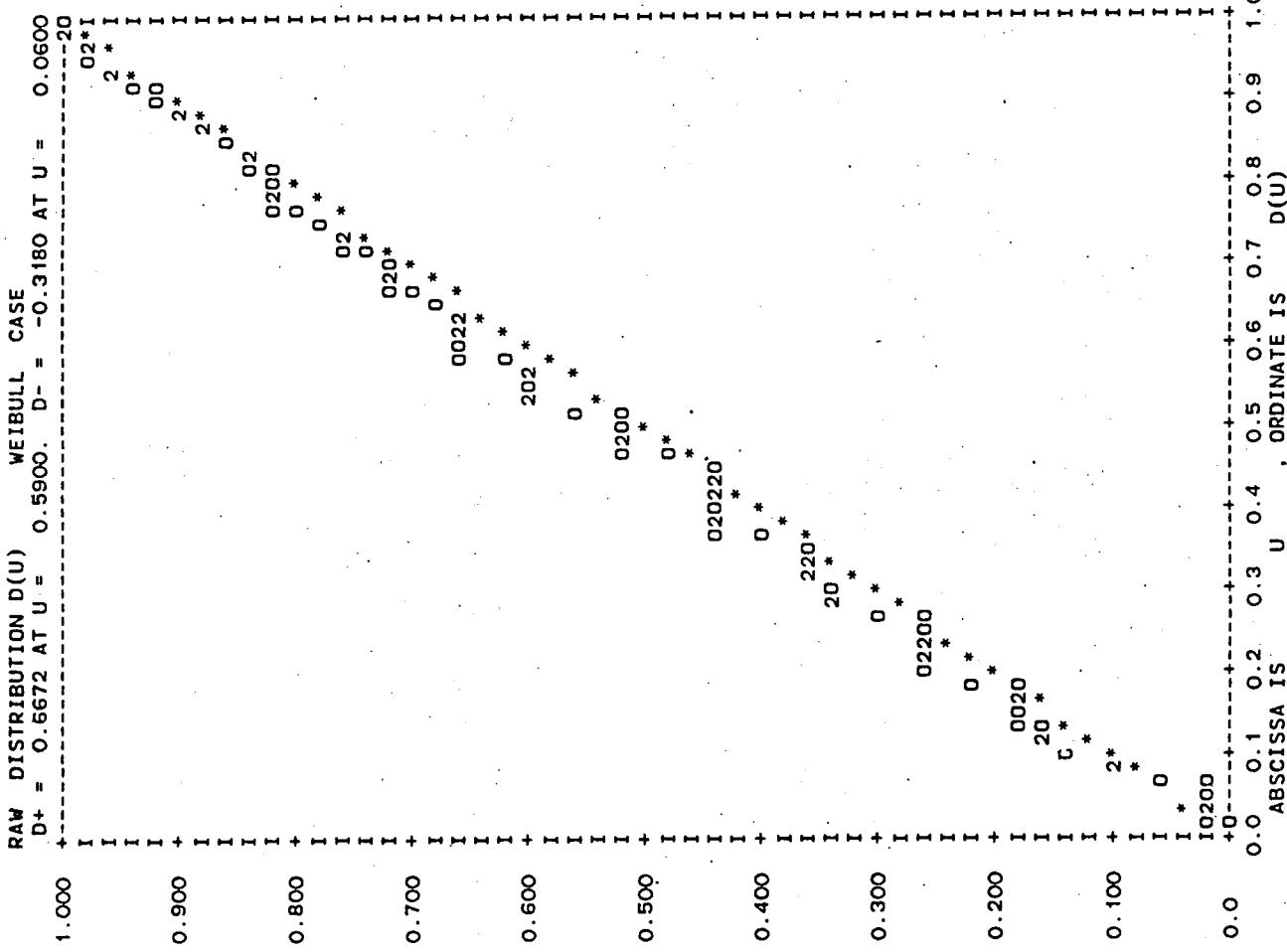
AUTOREGRESSIVE PARAMETRIC SELECT ANALYSIS

SQUARED MODULUS OF FOURIER COEFFICIENTS

```
.....  
PHI2( 1) = 0.00177358 *  
PHI2( 2) = 0.00257716 *  
PHI2( 3) = 0.00229294 *  
PHI2( 4) = 0.00134046 *  
PHI2( 5) = 0.01826305 *
```

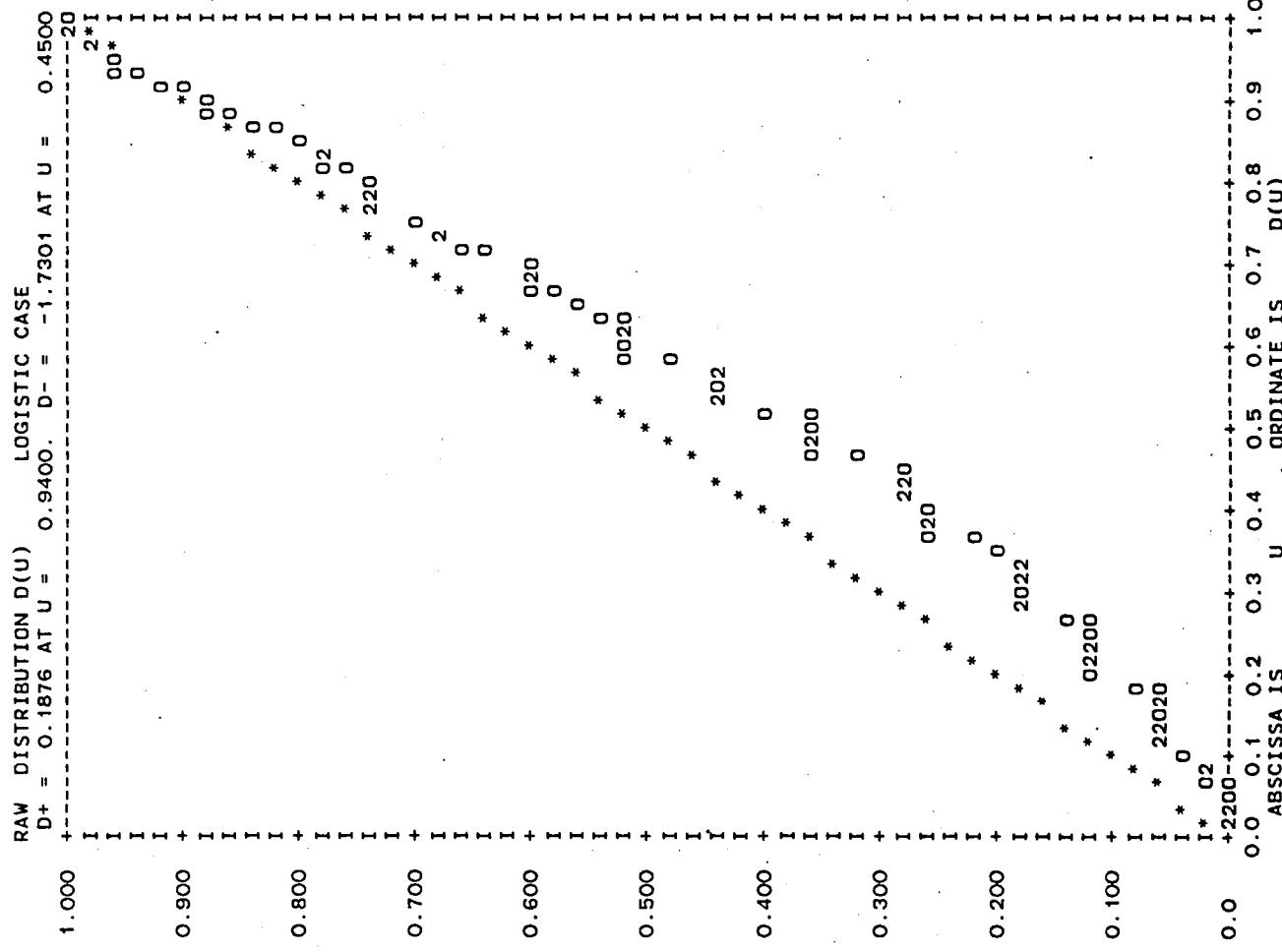
Median Channel 3 Logistic

Spring Wheat Pixel Neighborhood Channel 3 Medians: Pseudo-correlations square modulus (phi 2) accept Weibull distribution, reject logistic distribution fit.

Spring Wheat Pixel Neighborhood Channel 3Medians: Cumulative weighted spacings

$\bar{D}(u)$ plot indicates accept fit of Weibull distribution (shape parameter $\beta = 0.7$).

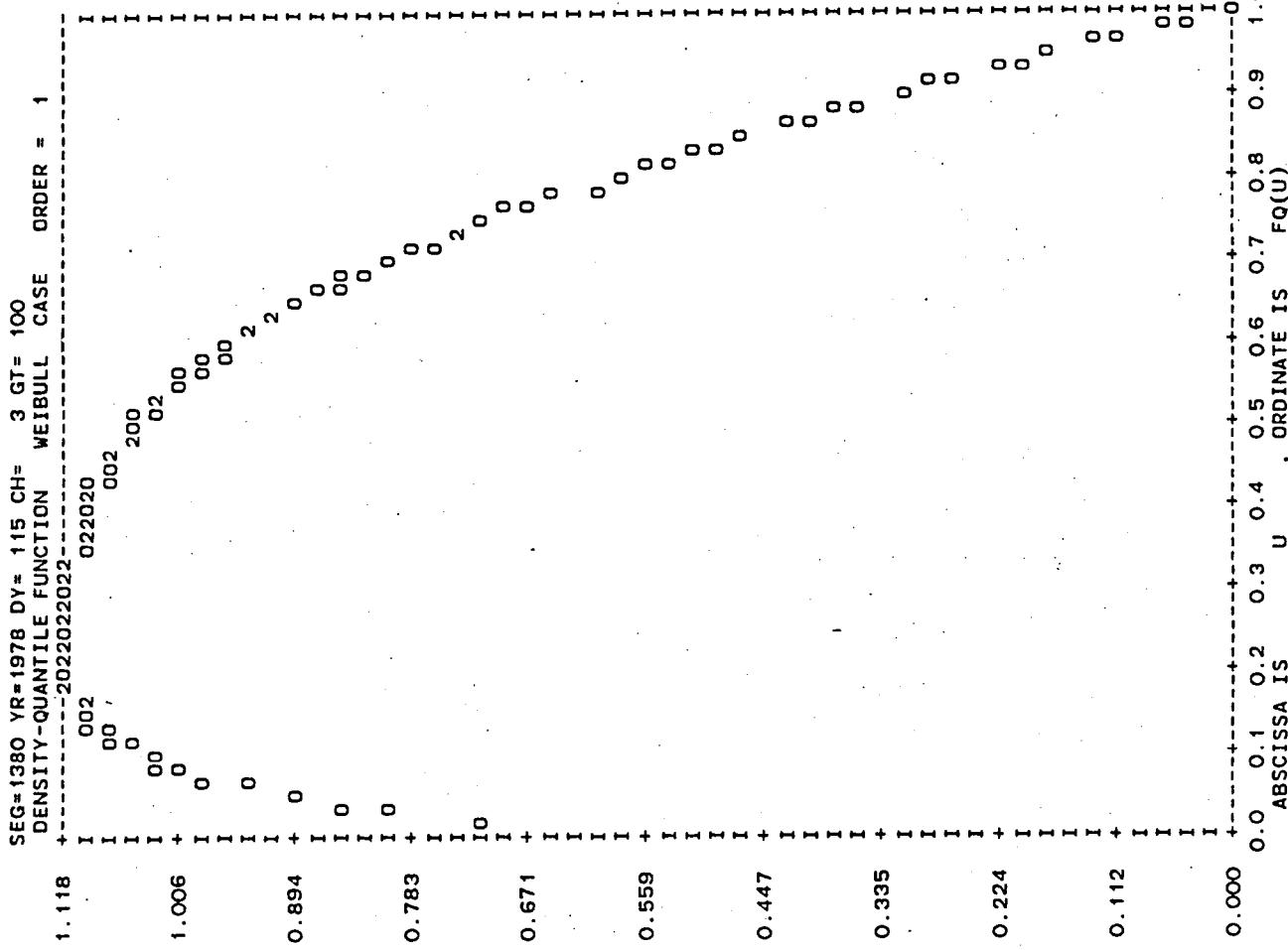
STATISTICAL ANALYSIS SYSTEM 20:12 WEDNESDAY, MAY 18, 1983 141



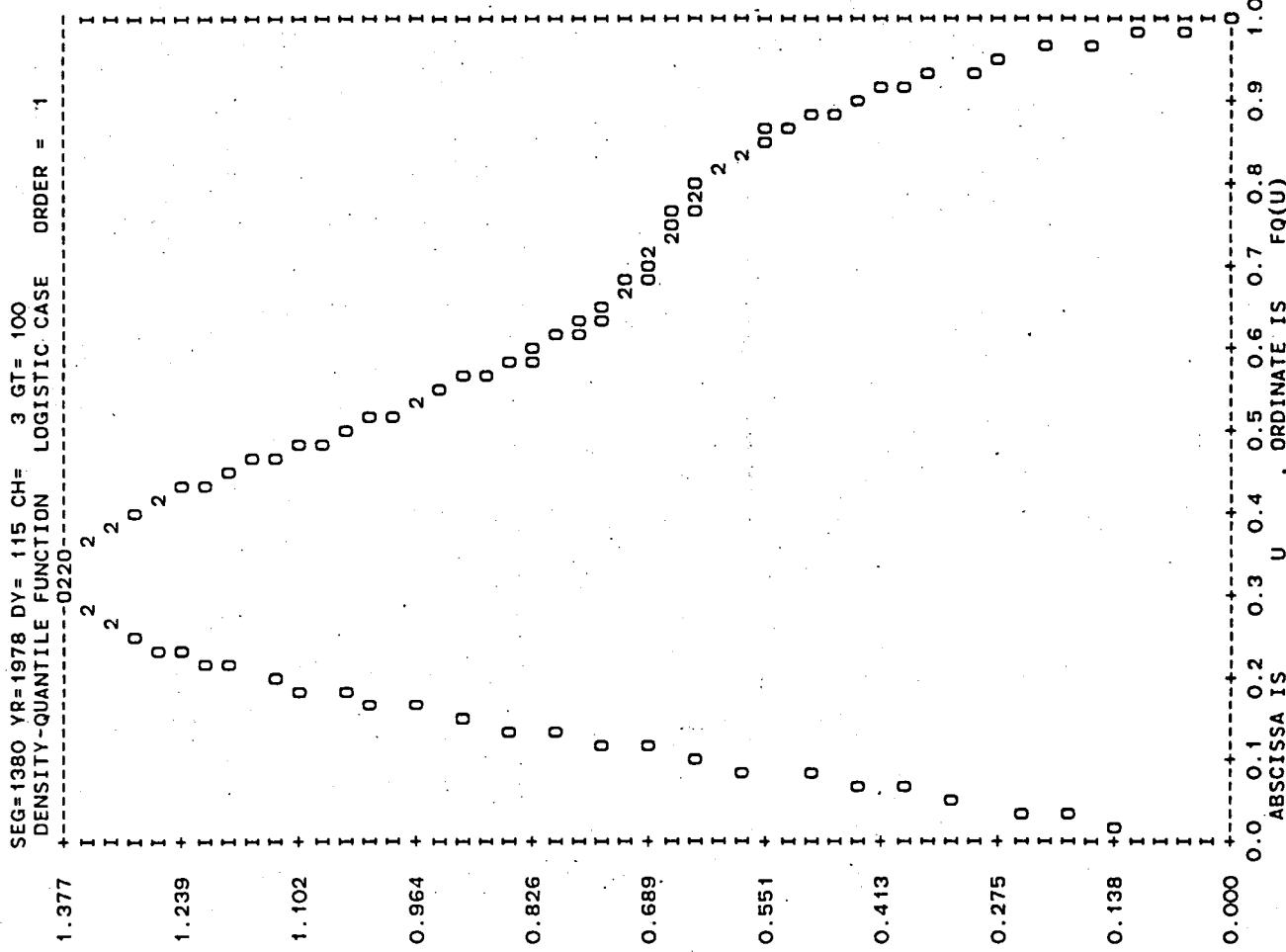
Medians: Cumulative weighted spacings

$\tilde{D}(u)$ plot indicates reject fit of logistic distribution.

220



STATISTICAL ANALYSIS SYSTEM 20:12 WEDNESDAY, MAY 18, 1983 144



Spring Wheat Pixel Neighborhood Channel 3
Medians: Autoregressive density-quantile
estimator (with logistic base and order 1).

STATISTICAL ANALYSIS SYSTEM

20:12 WEDNESDAY, MAY 18, 1983 153

SUMMARY OF AR PARAMETRIC SELECT ANALYSIS FOR LOGISTIC CASE

ORDER	RES_VAR	LOG(RES_VAR)	AIC	CAT
1	0.96161	-0.03914	-0.01914	-1.01922
2	0.94997	-0.05133	-0.01133	-1.01100
3	0.93683	-0.06525	-0.00525	-1.00444
4	0.92880	-0.07387	0.00613	-0.99229
5	0.91700	-0.08665	0.01335	-0.98433

OPTIMAL ORDER BY CAT CRITERION IS 1 MAXIMUM ORDER CHECKED IS 5

OPTIMAL ORDER BY AIC CRITERION IS 1 MAXIMUM ORDER CHECKED IS 5

SUMMARY OF AR PARAMETRIC SELECT ANALYSIS FOR WEIBULL CASE

ORDER	RES_VAR	LOG(RES_VAR)	AIC	CAT
1	0.99823	-0.00177	0.01823	-0.98184
2	0.99556	-0.00445	0.03555	-0.96461
3	0.99366	-0.00636	0.05364	-0.94667
4	0.99218	-0.00785	0.07215	-0.92837
5	0.97480	-0.02552	0.07448	-0.92561

OPTIMAL ORDER BY CAT CRITERION IS 0 MAXIMUM ORDER CHECKED IS 5

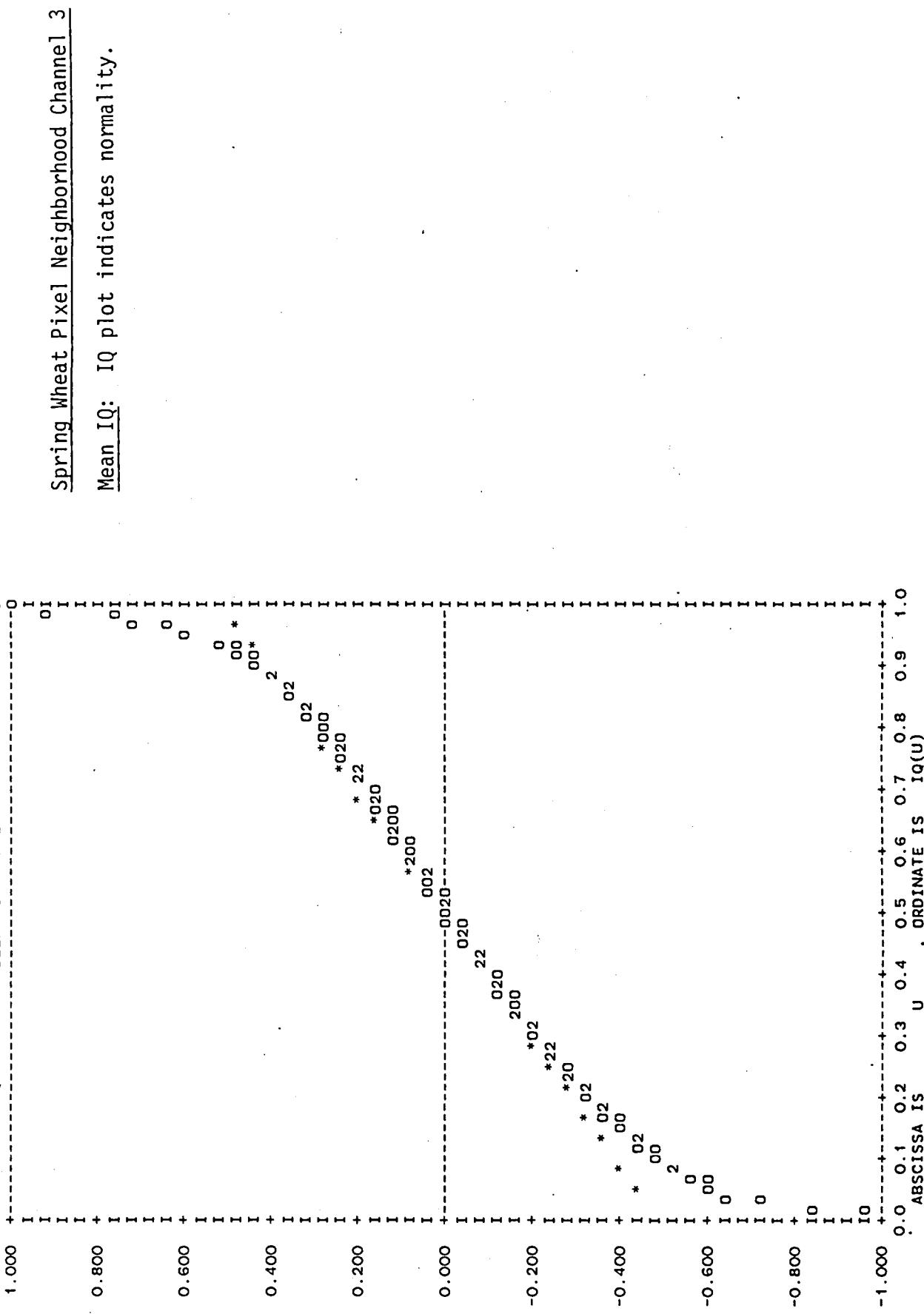
OPTIMAL ORDER BY AIC CRITERION IS 0 MAXIMUM ORDER CHECKED IS 5

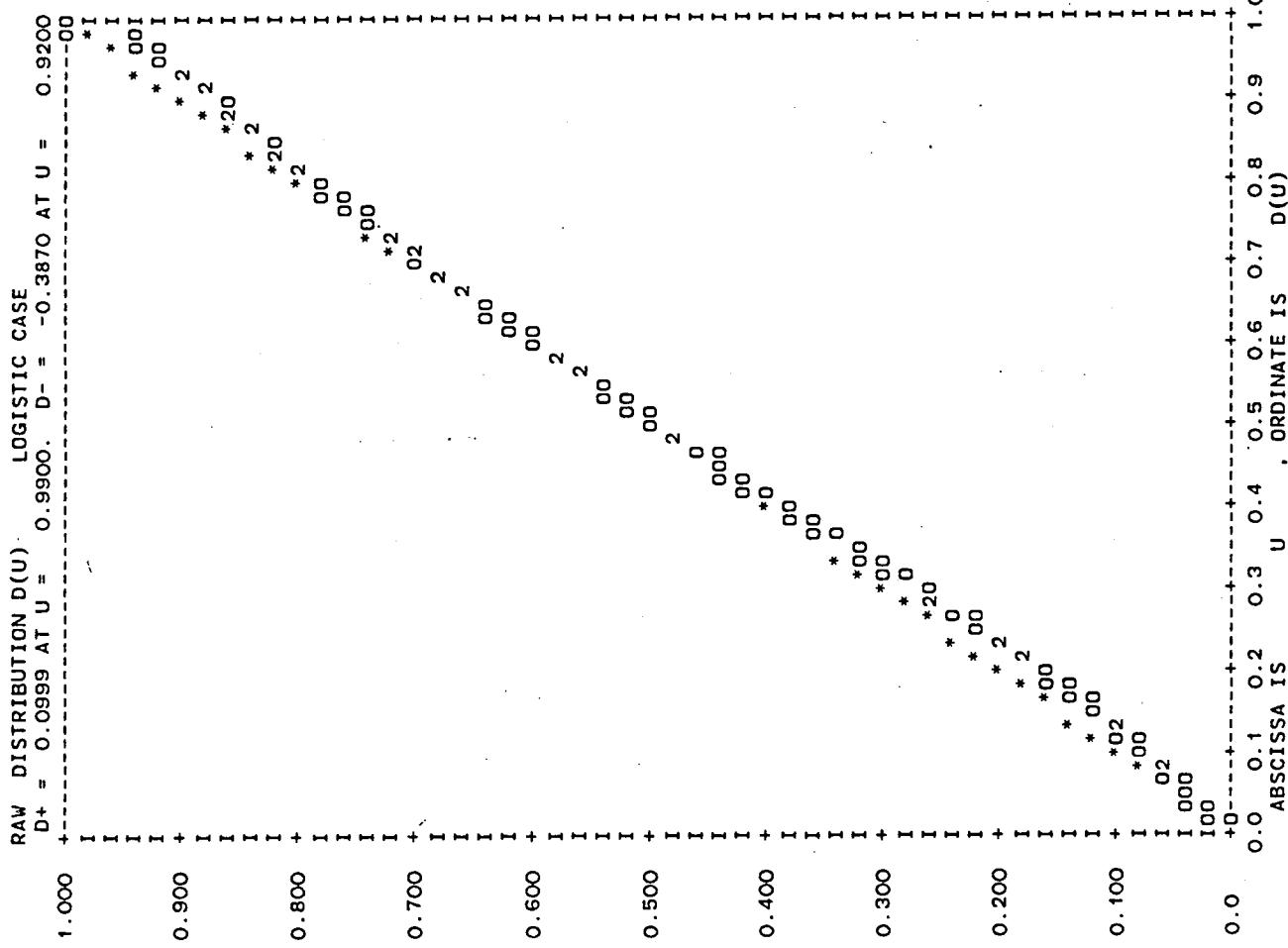
Spring Wheat Pixel Neighborhood Channel 3 Medians: AIC AR order

determining analysis.

STATISTICAL ANALYSIS SYSTEM 23:45 WEDNESDAY, MAY 18, 1983 173

SEG=1380 YR=1978 DY= 115 CH= 3 GT= 100
INFORMATIVE QUANTILE - LINEAR INTERPOLATION FROM UNGROUPED DATA



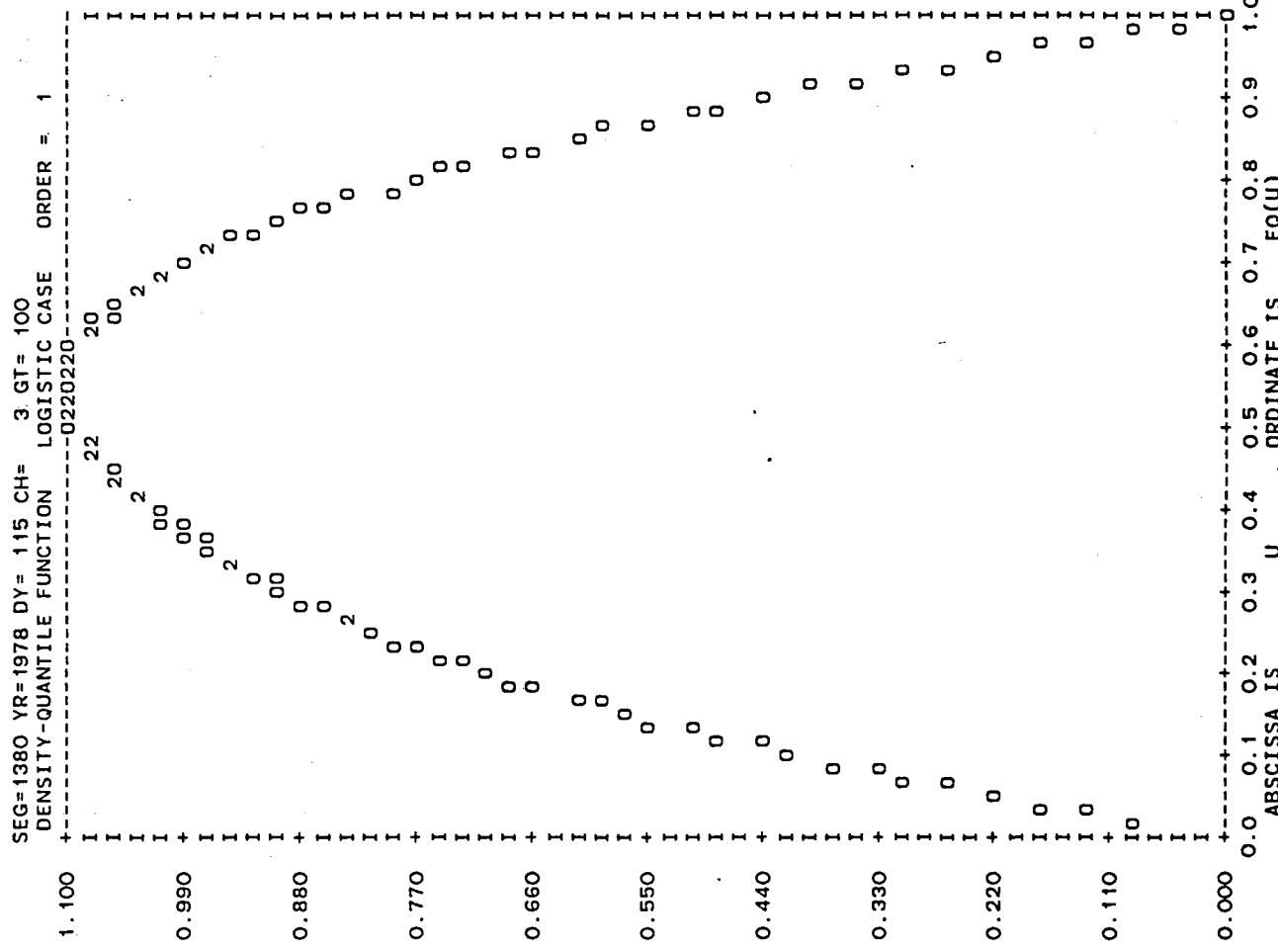
Spring Wheat Pixel Neighborhood Channel 3

Mean IQ: Cumulative weighted spacings D(u)

plot indicates accept fit of logistic distribution.

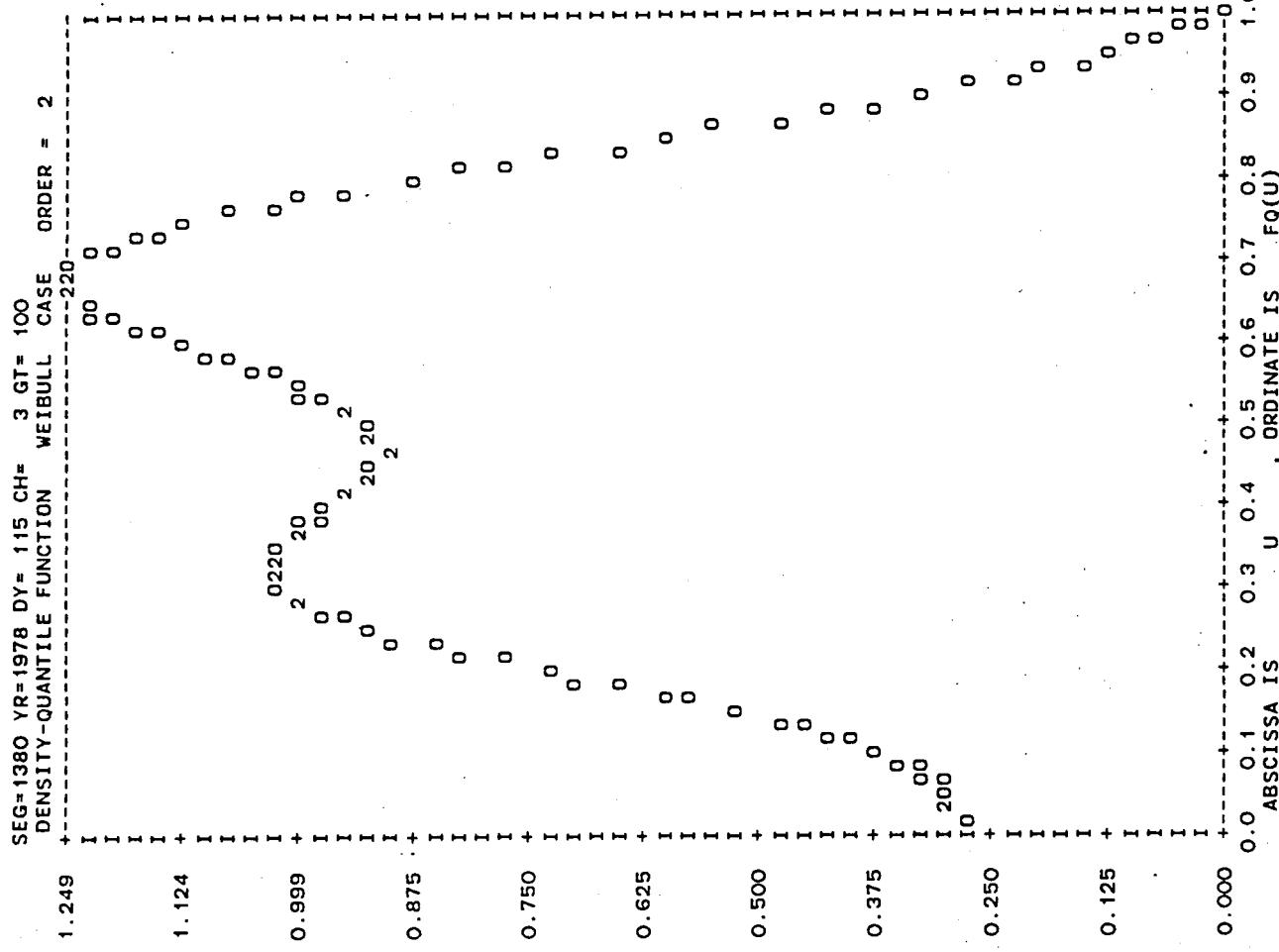
STATISTICAL ANALYSIS SYSTEM

20:12 WEDNESDAY, MAY 18, 1983 178



Mean IQ: Autoregressive density-quantile estimator (with logistic base and order 1)

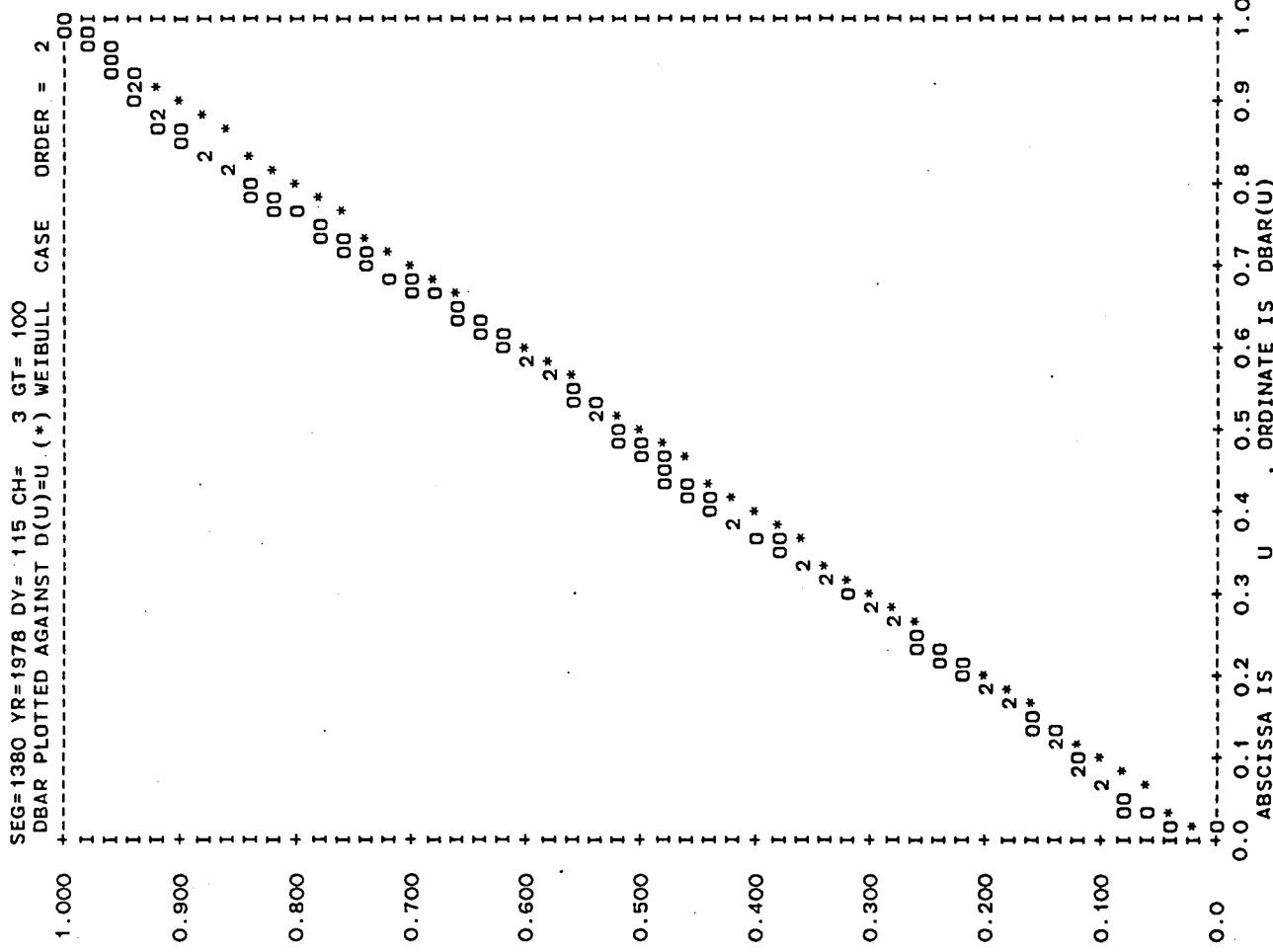
indicates normal-like density.



Mean IQ: Autoregressive density-quantile estimator (with Weibull base and order 2) indicates a density not in accord with logistic analysis, thus casting doubt on current reliability of AR order determining techniques.

STATISTICAL ANALYSIS SYSTEM

23:45 WEDNESDAY, MAY 18, 1983 179



Mean IQ: Diagnostic of fit of AR density-

quantile estimator (with Weibull base
and order 2) indicates that it "overfits"
and might generate spurious modes in the
density.

STATISTICAL ANALYSIS SYSTEM

23:45 WEDNESDAY, MAY 18, 1983 187

SUMMARY OF AR PARAMETRIC SELECT ANALYSIS FOR WEIBULL CASE

ORDER	RES_VAR	LOG(RES_VAR)	AIC	CAT
1	0.90931	-0.09507	-0.07507	-1.07785
2	0.88209	-0.12546	-0.08546	-1.08899
3	0.86845	-0.14104	-0.08104	-1.08376
4	0.86084	-0.14984	-0.06984	-1.07086
5	0.85631	-0.15512	-0.05512	-1.05400

OPTIMAL ORDER BY CAT CRITERION IS 2 MAXIMUM ORDER CHECKED IS 5

OPTIMAL ORDER BY AIC CRITERION IS 2 MAXIMUM ORDER CHECKED IS 5

SUMMARY OF AR PARAMETRIC SELECT ANALYSIS FOR LOGISTIC CASE

ORDER	RES_VAR	LOG(RES_VAR)	AIC	CAT
1	0.99882	-0.00118	0.01882	-0.98126
2	0.99655	-0.00346	0.03654	-0.96365
3	0.99392	-0.00610	0.05390	-0.94643
4	0.99086	-0.00918	0.07082	-0.92966
5	0.98756	-0.01251	0.08749	-0.91315

OPTIMAL ORDER BY CAT CRITERION IS 0 MAXIMUM ORDER CHECKED IS 5

OPTIMAL ORDER BY AIC CRITERION IS 0 MAXIMUM ORDER CHECKED IS 5

Spring Wheat Pixel Neighborhood Channel 3 Mean IQ: AIC AR order
determining analysis.

Appendix: Quantile and FUN.STAT Data Analysis

This appendix presents some of the new characterizations of probability laws which are being developed under the names of quantile data analysis, and functional statistical inference analysis.

Estimators of these characteristics are currently available for one sample and two samples, univariate and bivariate [Parzen (1979), (1983), Woodfield (1982)].

These methods seem to have much potential to contribute to the solution of the problem of digital image representation: the determination and modeling of basic characteristics or features of the digital image which can be incorporated into the process of identifying classes and attributes in a scene. They provide new approaches to determining scene probability density functions and class conditional density functions of digital image data in order to understand spectral characteristics and extract desired information. They can provide data representations which reduce the dimensions of multivariate image data while preserving information pertaining to scene classes and attributes.

A. One Sample: Univariate

Let X be continuous random variable of which we observe a random sample. To estimate distribution function $F_X(x) = \Pr[X \leq x]$ and probability density $f(x) - F'(x)$, we estimate: quantile function $Q_X(u) = F_X^{-1}(u)$; quantile density $q_X(u) = Q'_X(u)$; density quantile $f_{Q_X}(u) = f_X(Q_X(u))$. A quantile data analysis of the random sample

1. Forms sample distribution function $\tilde{F}_X(x)$, sample quantile function $\tilde{Q}_X(u)$, sample quantile density $\tilde{q}(u)$ at a grid of values of u in $0 < u < 1$.

2. Plots sample version of informative quantile function

$$IQ(u) = \frac{Q(u) - Q(0.5)}{2\{Q(0.75) - Q(0.25)\}}$$

whose values as u tends to 0 and 1 indicates the tail exponents of the probability law of X .

3. Determines standard distribution functions $F_0(x)$ to test

$$H_0: F(x) = F_0\left(\frac{x-\mu}{\sigma}\right) \text{ or } Q(u) = \mu + \sigma Q_0(u)$$

for location and scale parameters μ and σ to be estimated. A test of H_0 which does not require estimation of μ and σ can be based on [Parzen (1979)]

$$\tilde{d}(u) = f_0 Q_0(u) \tilde{q}(u) \div \tilde{\sigma}_0 ,$$

$$\tilde{\sigma}_0 = \int_0^1 f_0 Q_0(t) \tilde{q}(t) dt$$

which estimate respectively

$$d(u) = f_0 Q_0(u) q(u) \div \sigma_0 ,$$

$$\sigma_0 = \int_0^1 f_0 Q_0(t) q(t) dt.$$

4. Form successive autoregressive estimators

$$\hat{d}_m(u) = \hat{K}_m \left| 1 + \hat{\alpha}_m(1) e^{2\pi i u} + \dots + \hat{\alpha}_m(m) e^{2\pi i um} \right|^{-2}$$

whose negentropy

$$\hat{H}_m = \int_0^1 -\log \hat{d}_m(u) du = -\log \hat{K}_m$$

is used to determine optimal orders \hat{m} . Note that \hat{H}_m estimates the entropy difference

$$\Delta = \{\log \sigma_0 - \int_0^1 \log f_0 Q_0(u) du\} - \{- \int_0^1 \log fQ(u) du\}$$

5. Estimate $fQ(u)$ by

$$\hat{f}Q_m(u) = f_0 Q_0(u) \div \tilde{\sigma}_0 \hat{d}_m(u)$$

where m is chosen equal to an optimal order \hat{m} .

B. Two Sample: Univariate

Let X and Y be continuous random variables with random samples X_1, \dots, X_m and Y_1, \dots, Y_n respectively, and with respective distribution functions $F(x) = \Pr[X \leq x]$, $G(x) = \Pr[Y \leq x]$. The pooled sample $X_1, \dots, X_m, Y_1, \dots, Y_n$ can be regarded as a random sample from the distribution function

$$H(x) = \lambda F(x) + (1-\lambda) G(x), \quad \lambda = \frac{m}{m+n} .$$

To test the hypotheses of equality of distributions, $H_0: F(x) = G(x) = H(x)$, it is customary in non-parametric statistics to introduce

$$D_X(u) = F H^{-1}(u), \quad D_Y(u) = G H^{-1}(u)$$

with densities [equivalent to likelihood ratios]

$$d_X(u) = \frac{f H^{-1}(u)}{f H^{-1}(u)}, \quad d_Y(u) = \frac{g H^{-1}(u)}{h H^{-1}(u)} .$$

Note that $h H^{-1}(u) = \lambda f H^{-1}(u) + (1-\lambda) g H^{-1}(u)$; therefore

$$d_X(u) = \left\{ \lambda + (1-\lambda) \frac{g H^{-1}(u)}{f H^{-1}(u)} \right\}^{-1} .$$

Parzen (1983) shows that all conventional two-sample nonparametric test procedures are functionals of the following raw estimator of $D_X(u)$:

$$\tilde{D}_X(u) = \{\tilde{H} \tilde{F}_X^{-1}\}^{-1}(u)$$

from which one can form "pseudo-correlations" $\tilde{\rho}(v)$ and linear rank statistics $\Delta(J)$ with score function $J(u)$,

$$\tilde{\rho}(v) = \int_0^1 e^{2\pi i uv} d\tilde{D}_X(u), \quad \Delta(J) = \int_0^1 J(u) d\tilde{D}_X(u),$$

and autoregressive estimators $\hat{d}_{X,m}(u)$ of $d_X(u)$.

When one observes several variables $x^{(1)}, x^{(2)}, \dots, x^{(j)}, \dots$ one estimates functionals of $D_j(u) = F_{X^{(j)}}(H^{-1}(u))$ or $D_{jk}(u) = F_{X^{(j)}}(F_{X^{(k)}}^{-1}(u))$.

C. One Sample: Bivariate

Let (X_1, X_2) be jointly continuous random variables with distribution function $F_{X_1, X_2}(x_1, x_2) = \Pr[X_1 \leq x_1, X_2 \leq x_2]$ and density $f_{X_1, X_2}(x_1, x_2)$. The joint density quantile function is defined by

$$f_{Q_{X_1, X_2}}(u_1, u_2) = f_{X_1, X_2}(Q_{X_1}(u_1), Q_{X_2}(u_2))$$

To estimate f_Q we define

$$D_{X_1, X_2}(u_1, u_2) = F_{X_1, X_2}(Q_{X_1}(u_1), Q_{X_2}(u_2))$$

which is the distribution function of $U_1 = F_{X_1}(X_1)$, $U_2 = F_{X_2}(X_2)$; it has density

$$d_{X_1, X_2}(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} D(u_1, u_2)$$

satisfying

$$fQ_{X_1, X_2}(u_1, u_2) = fQ_{X_1}(u_1) fQ_{X_2}(u_2) d_{X_1, X_2}(u_1, u_2).$$

To estimate d_{X_1, X_2} from a random sample $(X_1^{(j)}, X_2^{(j)}), j=1, \dots, n$,
form

$$\tilde{d}_{X_1, X_2} = \tilde{F}_{X_1, X_2}(\tilde{Q}_{X_1}(u_1), \tilde{Q}_{X_2}(u_2))$$

and a raw estimator $\hat{d}_{X_1, X_2}(u_1, u_2)$. We smooth $\log \tilde{d}_{X_1, X_2}(u_1, u_2)$ by a
smooth estimator $\log \hat{d}_{X_1, X_2}(u_1, u_2)$ minimizing a criterion similar to

$$\sum_{j=1}^n |\log d[U_1^{(j)}, U_2^{(j)}] - \log d_m[U_1^{(j)}, U_2^{(j)}]|^2$$

where $\log d_m(u_1, u_2)$ has the parametric representation

$$\log d_m(u_1, u_2) = \sum_{v_1, v_2} \theta_{v_1, v_2} \exp i(u_1 v_1 + u_2 v_2) - \psi(\theta_{v_1, v_2});$$

where the summation is over $v_1, v_2 = 0, \pm 1, \dots, \pm m$, and $\psi(\theta_{v_1, v_2})$ is an
integrating factor to make $d_m(u_1, u_2)$ a probability density. The
foregoing estimators have been implemented in T. J. Woodfield [1982].
The problem of choosing a best value of the order m is approached by
evaluating the entropy of d_m .

D. Two Samples: Bivariate

Let (X_1, X_2) and (Y_1, Y_2) be random vectors with respective distribution functions $F(x_1, x_2)$ and $G(y_1, y_2)$, and respective random samples

$$(X_1^{(j)}, X_2^{(j)}), \quad j=1, \dots, m \quad \text{and} \quad (Y_1^{(k)}, Y_2^{(k)}), \quad k=1, 2, \dots, n.$$

Let $H(x_1, x_2)$ denote the distribution function of the pooled random sample, with marginal distribution functions $H_1(x_1)$ and $H_2(x_2)$. Define

$$D_1(u_1, u_2) = F(H_1^{-1}(u_1), H_2^{-1}(u_2)) ,$$

$$D_2(u_1, u_2) = F(H_1^{-1}(u_1), H_2^{-1}(u_2)) .$$

From $D_1(u_1, u_2)$ and $D_2(u_1, u_2)$ one can form raw estimators $d_1(u_1, u_2)$ and $d_2(u_1, u_2)$ of the densities

$$d_1(u_1, u_2) = \frac{f(H_1^{-1}(u_1), H_2^{-1}(u_2))}{h_1 H_1^{-1}(u_1) h_2 H_2^{-1}(u_2)} ,$$

$$d_2(u_1, u_2) = \frac{g(H_1^{-1}(u_1), H_2^{-1}(u_2))}{h_1 H_1^{-1}(u_1) h_2 H_2^{-1}(u_2)} .$$

Therefore

$$\begin{aligned} & \log d_1(u_1, u_2) - \log d_2(u_1, u_2) \\ &= \log f(H_1^{-1}(u_1), H_2^{-1}(u_2)) - \log g(H_1^{-1}(u_1), H_2^{-1}(u_2)) . \end{aligned}$$

The likelihood ratio $f(x_1, x_2)/g(x_1, x_2)$ can be effectively estimated by estimating $\log d_1(u_1, u_2) - \log d_2(u_1, u_2)$. It seems most natural to

estimate [using exponential model representations]

$$\log d_1(u_1, u_2) - \log d_{11}(u_1) - \log d_{12}(u_2)$$

where $d_{11}(u_1)$ and $d_{12}(u_2)$ are the marginal densities of $d_1(u_1, u_2)$ which are estimated separately by methods of two samples: univariate.

The final output are contour plots of the classification statistic

$$L(x_1, x_2) = \log f(x_1, x_2) - \log g(x_1, x_2) .$$

A point (x_1, x_2) is classified in population 1 or 2 by whether $L(x_1, x_2)$ exceeds a threshold which depends on the prior probabilities and loss function.

Appendix: Exploratory Quantile Data Analysis
of Training Files

The basic tool for determining statistical characteristics that are good discriminators is to determine (for each file, ground truth, and channel) a data batch of measurements in the specified channel on all pixels with the specified ground truth. The statistical characteristics of these data batches are summarized (as on the attached pages) and studied to determine patterns which can discriminate between different ground truths. The file numbers are those used in the Fundamental Research Data Base [see Guseman (1983)].

STATISTICAL ANALYSIS SYSTEM

21:09 TUESDAY, MAY 17, 1983

441

G	S	M	H	P	L	Q	Q	Q	I	—Q	—Q	M	E	S	T	I	I	Q	—S	—P	L	O	S	G	L	O	S	G	L	O	G	S	—S	O	—W	M	A	I	G	—W	S				
F	C	I	H	L	A	R	U	T	N	—T	—R	E	N	—E	—T	N	D	—A	—T	—I	—Q	—G	—O	—G	—G	—G	—G																		
I	A	R	U	T	N	—T	—R	—E	—E	—T	—R	—E	—E	—T	—T	—R	—R	—Q	—Q																										
E	N	—E	—T	—R	—E	—E	—T	—R	—E	—E	—T	—R	—E	—E	—T	—T	—R	—R	—Q																										
B	N	E	T	R	U	—T	—R	—E	—E	—T	—R	—E	—E	—T	—T	—R	—R	—Q																											
S	O	L	H	S	O	L	H	S	O	L	H	S	O	L	H	S	O	L	H	S	O	L	H	S	O	L	H	S	O	L	H	S	O	L	H	S	O	L	H						
1	1	2	90	377	15.000	19	24.0000	9.0000	0.08564	0.322475	-0.3456	1.05556	-1.3536	-1.5800	-0.90686	0.50636	-0.68050			
2	1	2	92	8755	14.000	15	19.0000	5.0000	0.17327	0.442017	-0.4000	2.80000	0.1066	-0.5821	-0.90812	0.80297	-0.21943			
3	1	2	97	10573	14.000	15	19.0000	5.0000	0.17437	0.463493	-0.4000	2.70000	-1.3959	-1.8841	-0.90812	0.65714	-0.41986					
4	1	2	98	793	13.000	14	17.0000	4.0000	0.14415	0.416479	-0.3750	2.25000	-1.1693	-1.5917	-0.90686	0.61600	-0.48451					
5	1	2	100	2296	15.000	18	21.0000	6.0000	0.07333	0.448731	-0.5833	2.08333	0.0933	-0.6163	-0.90686	0.82098	-0.19726					
6	1	2	104	558	15.000	18	23.0000	8.0000	0.09492	0.382529	-0.4375	1.18750	-0.2036	-2.4581	-0.90812	0.58059	-0.54370					
7	1	2	111	174	19.000	23	27.0000	8.0000	-0.01331	0.326849	-0.7500	0.81250	-5.1738	-5.5843	-0.90812	0.60793	-0.49770					
8	1	2	113	248	19.000	21	24.0000	5.0000	-0.00419	0.407626	-1.0000	1.10000	-1.5537	-2.3137	-0.90812	0.86235	-0.14809					
9	1	2	114	95	18.000	20	24.0000	6.0000	0.08244	0.381941	-0.7500	1.00000	-1.8184	-2.3698	-0.90812	0.69993	-0.35677						
10	1	2	164	2980	15.000	19	22.0000	7.0000	0.00592	0.373707	-0.5714	1.50000	-0.0626	-0.6454	-0.90812	0.72225	-0.32539						
11	1	2	242	1326	15.000	19	21.0000	6.0000	-0.00299	0.3737563	-0.6667	1.58333	-0.0152	-0.6651	-0.90812	0.77239	-0.25827						
12	1	3	90	377	14.000	20	30.7500	16.7500	0.07147	0.274792	-0.3284	0.68657	-1.0524	-1.1649	-0.90686	0.45186	-0.79437					
13	1	3	92	8755	11.500	14	18.0000	6.5000	0.09520	0.506777	-0.4615	3.00000	0.1017	-0.5798	-0.90812	0.79720	-0.22665					
14	1	3	97	10573	10.000	13	18.0000	8.0000	0.13187	0.407418	-0.3125	1.93750	-0.1479	-0.6297	-0.90812	0.65293	-0.42629					
15	1	3	98	793	9.500	12	14.0000	4.5000	0.06506	0.492558	-0.4444	2.55556	-0.6937	-1.3238	-0.90812	0.75730	-0.27800					
16	1	3	100	2296	12.500	16	21.5000	9.0000	0.11656	0.407630	-0.4444	1.66667	-0.5910	-1.0575	-0.90812	0.64299	-0.44162					
17	1	3	104	558	13.000	16	24.0000	11.0000	0.11668	0.345959	-0.36336	1.00000	-0.8114	-1.1307	-0.90812	0.55493	-0.58891					
18	1	3	111	174	20.500	26	33.0000	12.5000	0.02218	0.339094	-0.7200	0.68000	-3.7113	-4.1667	-0.90812	0.63589	-0.45273					
19	1	3	113	248	20.000	28	33.0000	13.0000	-0.05665	0.319116	-0.7692	0.65385	-1.3456	-1.8080	-0.90812	0.64040	-0.44567					
20	1	3	114	95	18.000	24	29.0000	11.0000	-0.01359	0.347879	-0.7273	0.68182	-4.1139	-4.6001	-0.90812	0.65577	-0.42194					
21	1	3	164	2980	13.000	18	24.0000	11.0000	0.05420	0.367705	-0.4545	1.40909	-0.5273	-0.9115	-0.90812	0.59221	-0.52390					
22	1	3	242	1326	14.500	20	26.0000	11.5000	0.03849	0.350681	-0.5217	1.43478	-2.1322	-2.4893	-0.90812	0.57637	-0.55101					
23	6	2	19	84	26.000	27	29.0000	3.0000	0.10974	0.422269	-0.83333	1.33333	-7.7867	-8.4229	-0.90686	0.76287	-0.27066					
24	6	2	20	84	26.000	27	28.9375	4.9125	-0.02021	0.313054	-0.5089	1.01781	-7.0616	-7.3181	-0.90812	0.52187	-0.65034					
25	6	2	21	68	26.780	27	29.0000	2.2000	0.17336	0.514212	-1.7500	1.50000	-7.3150	-8.3163	-0.90812	0.88522	-0.12192					
26	6	2	22	138	26.000	28	30.0000	4.0000	0.05138	0.400085	-0.7500	1.12500	-7.2153	-7.9296	-0.90812	0.82389	-0.19372					
27	6	2	24	75	29.000	30	33.0000	4.0000	0.08499	0.314726	-1.0000	0.75000	-7.9774	-8.4502	-0.90812	0.64713	-0.43521					
28	6	2	25	98	31.880	33	36.0000	4.1200	0.02669	0.411211	-1.3349	0.84951	-7.2732	-8.1135	-0.90812	0.93442	-0.06783					
29	6	2	26	59	28.700	29	32.0000	3.3000	0.08760	0.436342	-1.0606	1.06059	-6.8373	-7.6219	-0.90812	0.88522	-0.12192					
30	6	2	27	66	29.000	29	31.0000	2.0000	0.17336	0.514212	-1.6667	1.6667	-7.3150	-8.3163	-0.90812	0.9776	-0.09327					
31	6	2	29	90	26.920	29	33.0000	4.0000	0.03415	0.538090	-1.1250	1.62500	-4.4185	-5.5421	-0.90812	1.24																													

STATISTICAL ANALYSIS SYSTEM

21:09 TUESDAY, MAY 17, 1983 443

OBS	FILE_NO	CHANNEL	GR_TRUTH	SMPL_SIZ	Q_25	Q_50	Q_75	I_Q_R	MEAN_IQ
99	11	3	14	6.8	41.0000	44.5000	48.7499	7.7499	0.00934
100	11	3	15	154	41.0000	42.0000	44.0000	3.0000	-0.01168
101	11	3	16	176	35.0000	37.0000	39.0000	4.0000	0.04833
102	11	3	20	89	35.0000	37.0000	39.0000	4.0000	0.08762
103	11	3	21	61	36.0000	37.0000	39.0000	3.0000	0.10095
104	11	3	22	187	36.0000	38.0000	39.0000	3.0000	0.10243
105	11	3	23	276	41.0000	42.0000	44.0000	3.0000	0.07719
106	11	3	27	184	37.0000	38.0000	39.0000	2.0000	-0.05223
107	11	3	80	327	38.0000	39.0000	41.0000	3.0000	0.09934
108	11	3	90	137	38.0000	41.0000	44.0000	6.0000	-0.08112
109	11	3	99	1268	42.0000	45.0000	47.0000	5.0000	0.00174
110	11	3	100	1258	34.0000	35.0000	37.0000	3.0000	0.11312
111	11	3	101	908	37.0000	39.0000	44.0000	7.0000	0.07129
112	11	3	104	145	22.0000	29.0000	37.0000	15.0000	0.02661
113	11	3	105	1215	38.0000	39.0000	44.0000	6.0000	0.19637
114	11	3	107	12998	35.0000	37.0000	41.0000	6.0000	0.14851
115	11	3	175	804	38.0000	39.0000	42.0000	4.0000	0.15476
116	11	3	176	75	39.0000	42.0000	44.5699	5.5699	0.01077
117	11	3	179	248	41.0000	44.0000	45.0000	4.0000	-0.13473
118	11	3	240	813	37.0000	41.0000	49.8197	12.8197	0.12388
119	2	2	90	377	18.0000	22.0000	26.7500	8.7500	0.01643
120	2	2	92	8755	17.0000	19.0000	21.0000	4.0000	0.08750
121	2	2	97	10573	18.0000	20.0000	24.0000	6.0000	0.10750
122	2	2	98	793	19.0000	20.0000	23.0000	4.0000	0.21854
123	2	2	100	2296	15.0000	18.0000	21.0000	6.0000	0.07270
124	2	2	104	558	14.0000	15.0000	18.0000	4.0000	0.15617
<hr/>									
OBS	STD_IQ	IQ_01	IQ_99	LOG_SPC	LOG_WSPC	LOG_FQOO	SIGMA_O	LOG_SIGO	LG_SO_WS
99	0.235690	-0.5484	0.35484	-8.2709	-8.3490	-0.90812	0.43605	-0.83000	
100	0.388739	-1.0000	0.83333	-8.6831	-9.3572	-0.90686	0.79239	-0.23270	
101	0.485654	-0.1250	1.25000	-1.5900	-2.5285	-0.90812	1.03079	0.03032	
102	0.389348	-0.1250	1.25000	-7.2055	-7.6596	-0.90812	0.63511	-0.45395	
103	0.412752	-0.1667	1.16667	-7.0332	-7.6502	-0.90812	0.74746	-0.29108	
104	0.758390	-0.33333	4.50000	-6.4167	-7.3487	-0.90812	1.02421	0.02393	
105	0.435878	-0.83333	1.33333	-8.1175	-8.8666	-0.90686	0.85400	-0.15782	
106	0.529777	-0.5000	1.75000	-7.8936	-8.9375	-0.90812	1.14551	0.13585	
107	0.438788	-0.5000	1.66667	-2.0614	-2.8461	-0.90812	0.88397	-0.12333	
108	0.449681	-0.4167	0.50000	-2.1215	-3.1952	-0.90812	1.18011	0.16561	
109	0.453383	-0.9000	2.00000	0.1700	-0.7820	-0.90812	1.04488	0.04391	
110	0.683917	0.1667	2.00000	-1.096	-2.6160	-0.90812	1.83730	0.60830	
111	0.377377	-0.2143	0.92857	-1.1503	-2.0485	-0.90812	0.99014	-0.00991	
112	0.278329	0.23333	0.63333	-1.4291	-1.5725	-0.90812	0.46546	-0.76473	
113	0.587574	-0.2500	4.83333	0.025	-0.7074	-0.90812	0.82017	0.19824	
114	0.566747	-0.0833	4.58333	0.080	-0.6664	-0.90812	0.79156	-0.23375	
115	0.4227600	-0.3750	1.75000	0.0988	-0.7925	-0.90812	0.98332	-0.01682	
116	0.286119	-0.5386	0.89768	-7.8258	-8.0574	-0.90686	0.50897	-0.67536	
117	0.441134	-1.0000	0.87500	-7.2207	-7.9961	-0.90812	0.87576	-0.13266	
118	0.415452	-0.1950	1.67711	-0.6044	-1.1408	-0.90812	0.68958	-0.37167	
119	0.343387	-0.5269	1.20000	-0.9547	-1.3491	-0.59900	0.51249	0.83658	
120	0.572547	-0.6250	3.37500	0.3307	-0.6199	-0.90812	1.04465	0.04368	
121	0.442594	-0.5000	2.33333	0.075	-0.6228	-0.90812	0.83336	-0.20658	
122	0.712776	-0.7500	4.25000	-0.3916	-1.2954	-0.90812	0.99698	-0.00303	
123	0.457327	-0.5025	2.33333	-0.1113	-0.7367	-0.90812	0.75462	-0.28154	
124	0.513247	-0.5000	3.12500	-1.7848	-2.2800	-0.66252	0.66252	-0.41170	

STATISTICAL ANALYSIS SYSTEM 21:09 TUESDAY, MAY 17, 1983 445

S	M	S	L	L	L	L	L	L	L	L	G	G	S	O	S	O	L	G	
F	C	I	H	T	R	E	N	U	N	E	T	S	O	L	S	S	0	0	
G	R	I	A	L	N	E	N	U	N	E	T	1	2	5	0	5	7	51	
3	90	377	47	0.000	51	0.000	57	0.000	10	0.000	0.11187	0.54211	-0.9720	1.8000	-1.0773	-1	9210	-	
75	4	92	8755	49	0.000	54	0.000	59	0.000	10	0.000	0.00794	0.40551	-1.2000	1.7500	-0.8324	-	0.945	-0.06316
75	3	97	10573	0	0.000	0	0.000	0	0.000	0	0.000	0.54000	5.40000	0.00000	54	0.6519	-0.2058	-	0.77583
76	4	98	793	61	0.000	71	0.000	77	0.000	16	0.000	-0.04775	0.33029	-0.8060	0.7463	-0.9628	-1	4863	-0.05657
77	4	98	2296	43	0.000	49	0.000	56	0.000	13	0.000	0.03077	0.39067	-0.8077	1.5385	0.0063	-0.6597	-	0.77583
78	4	100	558	40	0.000	47	0.000	53	0.000	13	0.000	0.01238	0.39469	-0.6538	1.8077	-0.3133	-0.8201	-	0.670
79	4	104	174	46	0.000	55	0.000	62	0.000	15	0.000	0.00690	0.32112	-0.6349	0.7302	-3.7986	-4	1619	-0.40002
80	4	111	248	44	0.000	49	0.000	53	0.000	9	0.000	-0.0196	0.37798	-0.8333	1.0000	-1.3375	-0.9590	-	0.510
81	4	113	97	10573	12	0.000	12	0.000	15	0.000	0.303403	0.50460	-2.1875	1.0625	-0.9529	-2	2132	-0.424	
82	4	114	2980	47	0.000	51	0.000	55	0.000	8	0.000	-0.03403	0.42414	-1.0913	1.5833	-1.2025	-1.9647	-	1.424
83	4	164	1326	49	0.000	53	0.000	59	0.000	12	0.000	0.01909	0.42414	-1.0913	1.5833	-1.2025	-1.9647	-	1.424
84	4	242	377	14	0.000	16	0.000	24	0.000	10	0.000	0.16075	0.37199	-0.2500	1.3500	-1.4246	-1	6591	-0.42011
85	5	90	8755	12	0.000	14	0.000	15	0.000	2	0.000	0.23200	0.45455	-0.4299	0.89847	-1.3389	-0.9590	-	0.575
86	5	92	97	10573	12	0.000	12	0.000	15	0.000	0.30333	0.84649	-0.4167	0.5833	0.5960	-0	4769	-0.00215	
87	5	98	248	18	0.000	21	0.000	23	0.000	5	0.000	-0.00767	0.33992	-0.7826	0.9565	-1.2237	-1	7680	-0.34102
88	5	98	793	15	0.000	16	0.000	18	0.000	3	0.000	0.19833	0.65293	-0.8333	3.6667	0.5786	-0	5688	-0.64293
89	5	100	2296	20	0.000	26	0.000	32	0.000	12	0.000	0.02053	0.36533	-0.5833	1.1667	-0.0841	-0	6151	-0.64293
90	5	104	558	20	0.000	29	0.000	37	0.000	16	0.000	-0.00049	0.32287	-0.5405	0.7568	-1.3346	-1	6888	-0.55274
91	5	111	174	15	0.000	18	0.000	21	0.000	6	0.000	0.03371	0.35199	-0.4571	1.6953	-0.8105	-6	0.0374	-0.98671
92	5	113	248	18	0.000	21	0.000	23	0.000	5	0.000	-0.00767	0.33992	-0.7826	0.9565	-1.2237	-1	7680	-0.67999
93	5	114	95	14	0.000	17	0.000	20	0.000	6	0.000	0.06910	0.47639	-0.5833	1.8333	-1.3445	-1	8854	-0.67999
94	5	164	2980	13	0.000	15	0.000	19	0.000	6	0.000	0.16106	0.53115	-0.3333	3.0000	-0.1000	-0	7010	-0.55745
95	5	242	1326	14	0.000	18	0.000	22	0.000	8	0.000	0.06258	0.39451	-0.4375	1.6250	-1.9339	-2	3006	-0.53745
96	5	90	377	49	0.000	55	0.000	62	0.000	13	0.000	0.03732	0.33783	-0.4804	1.1388	-4.9206	-5	2071	-0.63624
97	5	92	8755	50	0.000	54	0.000	58	0.000	8	0.000	0.06437	0.47052	-1.1250	2.1875	-0.1278	-0	7421	-1.40558
98	5	97	10573	60	0.000	68	0.000	74	0.000	14	0.000	-0.1621	0.33677	-0.7586	1.0000	-0.1971	-0	7710	-0.33296
99	5	98	793	65	0.000	74	0.000	77	0.000	12	0.000	-0.10873	0.35418	-0.4086	0.5825	-0.6155	-1	3129	-0.30590
00	5	100	2296	49	0.000	53	0.000	58	0.000	9	0.000	0.05001	0.46307	-1.0000	1.9444	0.1266	-0	7220	-0.54022
01	5	104	558	47	0.000	53	0.000	57	0.000	9	0.000	0.1209	0.42672	-0.8050	1.4591	-4.8370	-5	4721	-0.64022
02	5	111	174	45	0.000	54	0.000	61	0.000	16	0.000	-0.00587	0.28341	-0.5113	0.6015	-4.4886	-4	7421	-0.64022
03	5	113	248	47	0.000	50	0.000	55	0.000	8	0.000	0.09465	0.47385	-1.5331	1.5000	-1.3288	-2	3599	-0.132
04	5	114	98	44	0.000	49	0.000	54	0.000	10	0.000	0.11776	0.45430	-1.1000	1.3000	-3.4513	-4	2417	-0.12421
05	5	164	2980	50	0.000	55	0.000	62	0.000	12	0.000	0.05976	0.39738	-0.9167	1.4167	-0.5358	-1	2137	-0.12421
06	5	242	1326	50	0.000	54	0.000	61	0.000	11	0.000	0.08167	0.440925	-0.7447	1.4681	-0.7698	-1	3693	-0.30730

References

- Guseman, L. F. (1983) Proceedings of the NASA/MPRIA Workshop: Math/Stat. Appendix: Fundamental Research Data Base.
- Parzen, E. (1979) Nonparametric Statistical Data Modeling. Journal of the American Statistical Association, 74, 105-131.
- Parzen, E. (1983) FUNSTAT Quantile Approach to Two Sample Statistical Data Analysis. The Canadian Journal of Statistics.
- Woodfield, Terry Joe (1982) Statistical Modeling of Bivariate Data. Ph.D. Thesis Texas A&M Department of Statistics.