DOCUMENT RESUME

ED 117 134 TM 004 986

AUTHOR Marco, Gary L.
TITLE Use of the Logistic Model as an Alternative to Linear Interpolation for Computing Percentile Ranks.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RM-75-6
PUB DATE Oct 75
NOTE 17p.

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS *Comparative Analysis; Intervals; *Mathematical Models; *Percentage; *Scores; *Scoring Formulas; Statistical Bias; Test Results
IDENTIFIERS *Logistic Interpolation; Percentile Ranks

ABSTRACT

A method of interpolation has been derived that should be superior to linear interpolation in computing the percentile ranks of test scores for unimodal score distributions. The superiority of the logistic interpolation over the linear interpolation is most noticeable for distributions consisting of only a small number of score intervals (say fewer than 10), particularly distributions that are relatively unskewed. Logistic interpolation thus should be useful in practical situations in which percentile ranks of number right scores must be estimated from very coarse groupings. The logistic method may also be applied to distributions of formula scores. However, the method should probably not be used for unsmoothed distributions of formula scores unless it is desired to smooth out the peaks and valleys that result from rounding scores to integer values. The usefulness of logistic interpolation in computing percentile ranks for test score distributions is illustrated using three score distributions for item analysis samples from 1974 Law School Admissions Tests (LSAT). (Author/BJG)

RM-75-6

RESEARCH MEMORANDUM

USE OF THE LOGISTIC MODEL AS AN ALTERNATIVE TO LINEAR INTERPOLATION FOR COMPUTING PERCENTILE RANKS

Gary L. Marco

This Memorandum is for interoffice use. It is not to be cited as a published report without the specific permission of the author.

2

Educational Testing Service
Princeton, New Jersey
October 1975

Use of the Logistic Model as an Alternative to Linear Interpolation for Computing Percentile Ranks

The usual method of computing percentile ranks for test scores is linear interpolation. For this method it is assumed that the scores are distributed uniformly throughout the score interval. One formula for computing percentile ranks in this way is

$$PR(x) = p_1 + \frac{x - x_1}{x_u - x_1}(p_u - p_1) \qquad (1)$$

where x is the score for which the percentile rank is to be computed,

x_1 is the lower theoretical limit of the score interval,

x_u is the upper theoretical limit of the score interval,

p_1 is the percentage of the cases scoring below the lowest score in the interval, and

p_u is the percentage of the cases scoring below the lowest score in the next higher interval.

In this paper we consider the logistic distribution function as an alternative to (1). Formula (1) is known to yield percentile ranks that are too low above the mode and too high below the mode of a continuous unimodal distribution. The reason for this bias is that above the mode the score density is greater in the lower part of the interval than in the upper part of the interval, and, conversely, below the mode the score density is greater in the upper part of the interval. If the score interval in which interpolation is used is small, the assumption of a uniform distribution of scores in the interval results in negligible bias, and linear interpolation and the alternative proposed here will give nearly identical results. However, there are other instances in which the logistic interpolation yields results superior to linear interpolation. These are illustrated in the second section of the paper.

Theoretical Considerations

The logistic cumulative distribution function, which very nearly coincides with the normal ogive model (Lord and Novick, 1968), has a closed mathematical form that makes it convenient to apply. This function is

$$p = L(y) = 1/(1 + e^{-y}) , - \infty < y < \infty . \quad (2)$$

The inverse function is

$$y = L^{-1}(p) = \ln[p/(1 - p)] , 0 < p < 1 . \quad (3)$$

If we substitute $p = p_u$ and $p = p_1$ in (3), we can determine the corresponding y_u and y_1. Then we can find the y in this interval that corresponds to x in (1). Thus,

$$y_u = L^{-1}(p) = \ln[p_u/(1 - p_u)] , \quad (4)$$

$$y_1 = L^{-1}(p) = \ln[p_1/(1 - p_1)] , \text{ and} \quad (5)$$

$$y = y_1 + \frac{x - x_1}{x_u - x_1}(y_u - y_1) . \quad (6)$$

The formula for computing the percentile rank for score x is then given by (2).

Since $\ln[p/(1 - p)]$ is undefined for $p = 1$ and $p = 0$, which correspond to the upper and lower theoretical limits of the highest and lowest obtained score intervals, respectively, we resort to a practical rule suggested by M. S. Bartlett (1947, p. 46) in these instances, namely, substituting $1/4n$ for 0 and $(1 - 1/4n)$ for 1.00. Other working values, such as $1/2n$ and $(1 - 1/2n)$, might also be used (Berkson, 1955). We can then use equations (2) through (6) as before.

Application to Empirical Distributions

To illustrate the usefulness of logistic interpolation in computing percentile ranks for test score distributions, we use three score distributions for item-analysis samples from the February and December 1974 administrations of the Law School Admissions Test (LSAT): one negatively skewed, one positively skewed, and one relatively unskewed. These three distributions are based on the number-right scores from the 35-item Data Interpretation section of the February 1974 LSAT and the 30-item Data Interpretation and Sentence Correction sections of the December 1974 LSAT. Table 1 gives the one-point interval distributions and summary statistics for all three scores. Skewness was computed by the formula

$$s = [\Sigma(x - \mu_x)^3]/n\ \sigma_x^{\ 3} .$$

Percentile ranks were computed by both linear and logistic interpolation for all scores in the obtained score range for one-point (0, 1, 2, ...), three-point (0-2, 3-5, ...), and six-point (0-5, 6-11, ...) interval distributions.

Insert Table 1 about here

The results are shown in Table 2 and Figures 1 through 6. Since the percentile ranks derived by the two methods for one-point interval distributions were virtually undistinguishable, only the percentile ranks computed by linear interpolation are given in the figures and were the base against which the other percentile ranks were compared.

Insert Table 2 and Figures 1-6 about here

It is clear from the figures that when 10 or more score intervals are used, there is little to gain from using the logistic method. However, for the six-point interval distributions (six categories) the logistic method yields percentile ranks that are much closer to the baseline. The logistic method works exceedingly well for a relatively unskewed distribution (see Figure 2). Figures 4 and 6 show that the percentile ranks for scores at the high end of the score range are too high when the distribution is negatively skewed. Correspondingly, the percentile ranks are too low for scores at the low end of the score range when the distribution is positively skewed. A study of the figures suggests that an average of the logistic and linear results might correct this bias.

Conclusion

A method of interpolation has been derived that should be superior to linear interpolation in computing the percentile ranks of test scores for unimodel score distributions. We have seen that the superiority of the logistic interpolation over the linear interpolation is most noticeable for distributions consisting of only a small number of score intervals (say fewer than 10), particularly distributions that are relatively unskewed. Logistic interpolation thus should be useful in practical situations in which percentile ranks of number right scores must be estimated from very coarse groupings.

The logistic method may also be applied to distributions of formula scores. However, the method should probably not be used for unsmoothed distributions of formula scores unless it is desired to smooth out the peaks and valleys that result from rounding scores to integer values.

References

Bartlett, M. S. The use of transformations. Biometrics, 1947, 3, 39-52.

Berkson, J. Maximum likelihood and minimum χ^2 estimates of the logistic function. Journal of the American Statistical Association, 1955, 59, 130-162.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Table 1

Distribution of Section Scores from

the February 1974 and December 1974 Administrations

of the Law School Admissions Test

Score	Data Interpretation (February 1974)		Data Interpretation (December 1974)		Sentence Correction (December 1974)	
	Frequency	Percent Below	Frequency	Percent Below	Frequency	Percent Below
34	1	99.9				
33	1	99.9				
32	10	99.3				
31	13	98.5				
30	16	97.5	3	99.8	1	99.9
29	25	95.9	5	99.4	8	99.5
28	22	94.6	5	99.0	20	98.3
27	28	92.9	13	98.1	38	96.0
26	49	89.8	14	97.0	60	92.4
25	53	86.6	19	95.6	71	88.1
24	68	82.4	29	93.5	95	82.5
23	74	77.8	31	91.2	116	75.5
22	84	72.7	45	87.8	140	67.1
21	111	65.8	37	85.1	142	58.6
20	112	59.0	53	81.1	146	49.9
19	147	49.9	70	75.9	143	41.3
18	104	43.5	94	68.9	137	33.1
17	106	37.0	94	61.9	111	26.5
16	116	29.8	99	54.6	111	19.8
15	108	23.2	118	45.8	89	14.5
14	92	17.5	102	38.2	86	9.3
13	66	13.5	102	30.6	44	6.7
12	55	10.1	97	23.4	28	5.0
11	44	7.4	80	17.5	27	3.4
10	41	4.9	74	12.0	22	2.1
9	31	3.0	63	7.3	13	1.3
8	18	1.8	48	3.7	8	0.8
7	9	1.3	27	1.7	2	0.7
6	8	0.8	13	0.7	5	0.4
5	10	0.2	5	0.4	2	0.3
4	1	0.1	2	0.2	4	0.1
3	2	0.0	1	0.1	0	0.1
2			2	0.0	0	0.1
1					0	0.1
0					1	0.0
N	1625		1345		1670	
Mean	18.43		15.29		19.27	
S.D.	5.40		4.92		4.46	
Skewness	0.07		0.35		-0.38	
Median	18.51		14.98		19.51	

Table 2

Percentile Ranks for Selected Scores on LSAT Sections

by Size of Score Interval and Type of Interpolation

Score	One-Point Intervals Linear	One-Point Intervals Logistic	Three-Point Intervals Linear	Three-Point Intervals Logistic	Six-Point Intervals Linear	Six-Point Intervals Logistic
			Data Interpretation (February 1974)			
28	95.26	95.31	95.17	95.73	93.71	95.80
22	75.26	75.35	74.12	75.03	72.68	74.88
16	33.42	33.32	33.35	32.54	35.15	32.24
10	6.12	6.00	6.52	5.52	7.77	5.49
4	0.15	0.15	0.40	0.11	0.40	0.11
			Data Interpretation (December 1974)			
28	99.22	99.24	98.92	99.34	98.20	99.47
22	89.48	89.60	89.26	90.02	87.32	89.97
16	58.25	58.30	57.36	57.79	57.55	57.47
10	14.72	14.51	15.35	13.42	17.75	10.79
4	0.30	0.29	0.45	0.33	0.46	0.19
			Sentence Correction (December 1974)			
28	98.86	99.03	97.96	99.50	95.57	99.74
22	71.32	71.50	70.54	72.07	70.12	72.81
16	23.14	22.98	23.80	22.46	26.09	22.07
10	2.75	2.68	3.17	2.59	3.88	2.74
4	0.18	0.13	0.24	0.16	0.31	0.18

Figure Captions

Figure 1. Comparison of linear and logistic interpolation for February 1974 LSAT Data Interpretation scores grouped into three-point intervals.

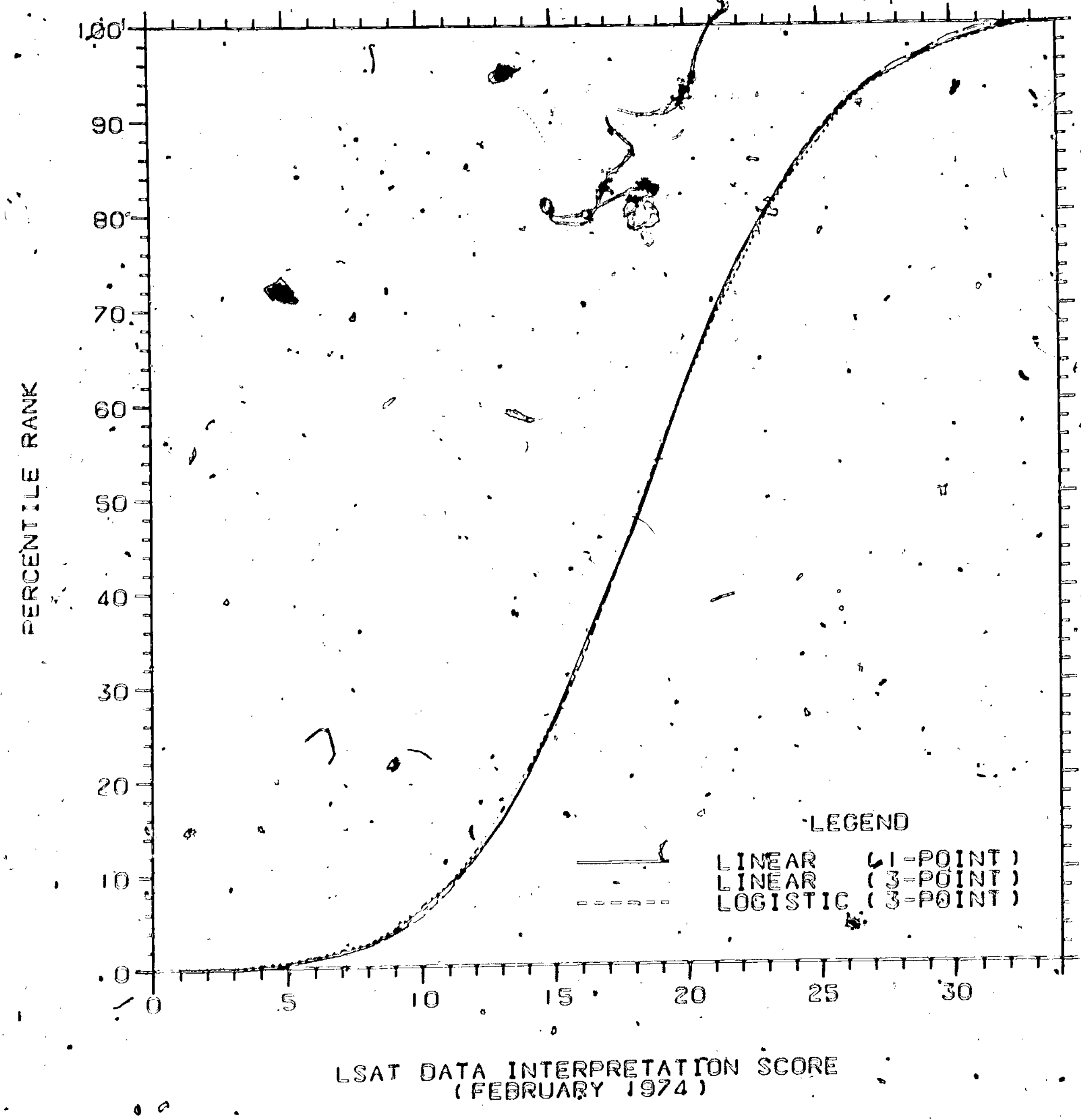
Figure 2. Comparison of linear and logistic interpolation for February 1974 LSAT Data Interpretation scores grouped into six-point intervals.

Figure 3. Comparison of linear and logistic interpolation for December 1974 Data Interpretation scores grouped into three-point intervals.

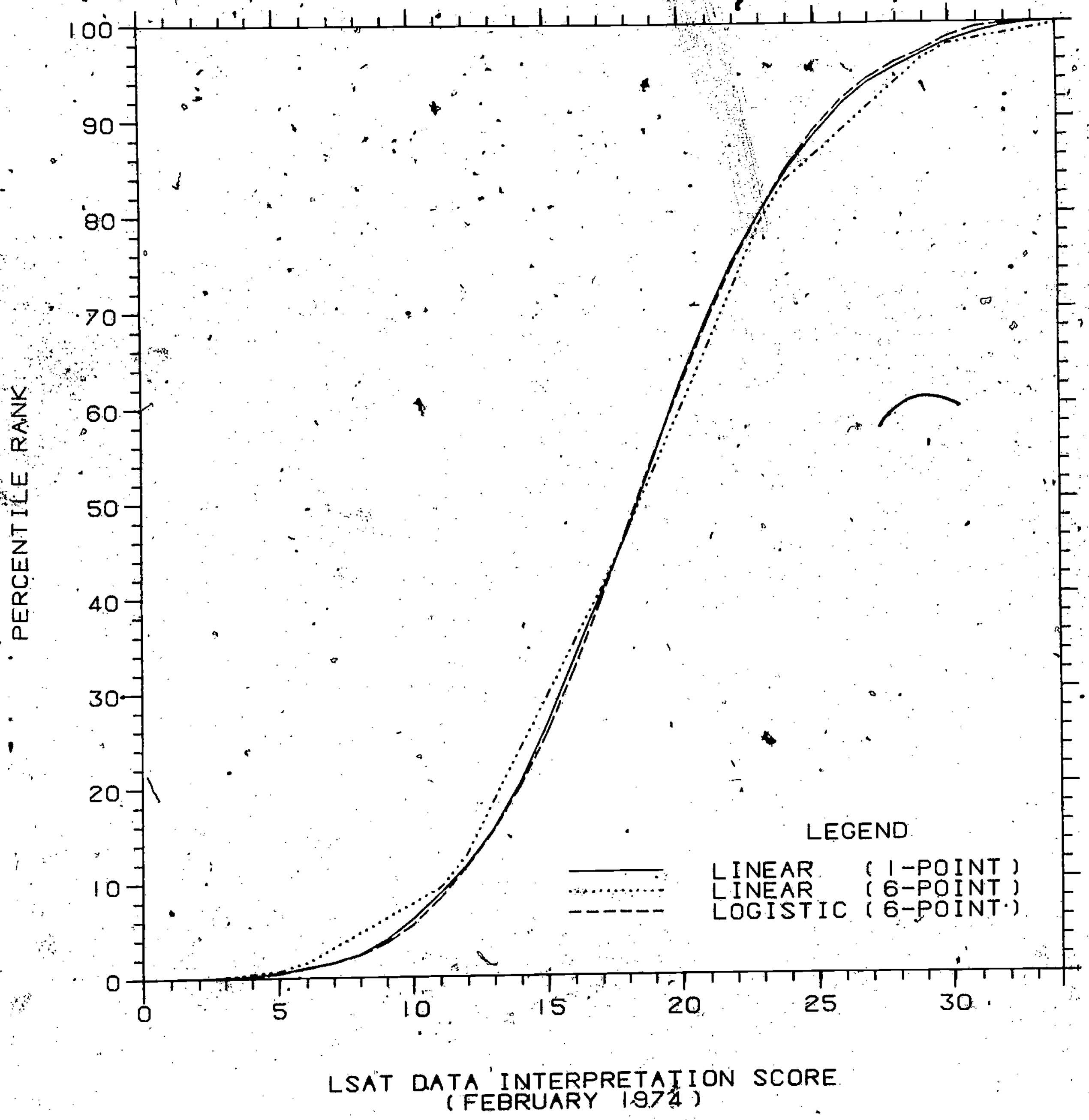
Figure 4. Comparison of linear and logistic interpolation for December 1974 Data Interpretation scores grouped into six-point intervals.

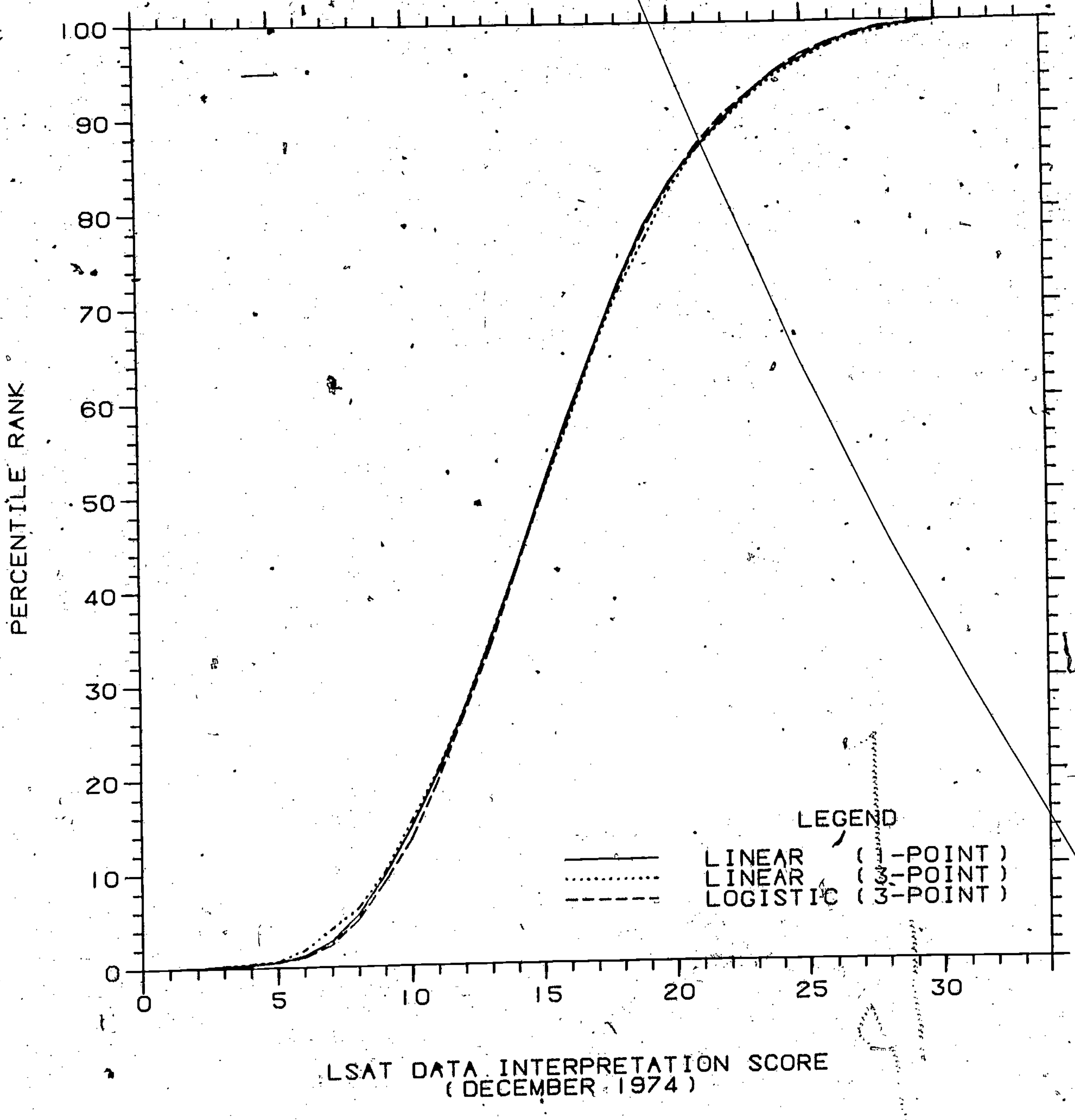
Figure 5. Comparison of linear and logistic interpolation for December 1974 Sentence Correction scores grouped into three-point intervals.

Figure 6. Comparison of linear and logistic interpolation for December 1974 Sentence Correction scores grouped into six-point intervals.

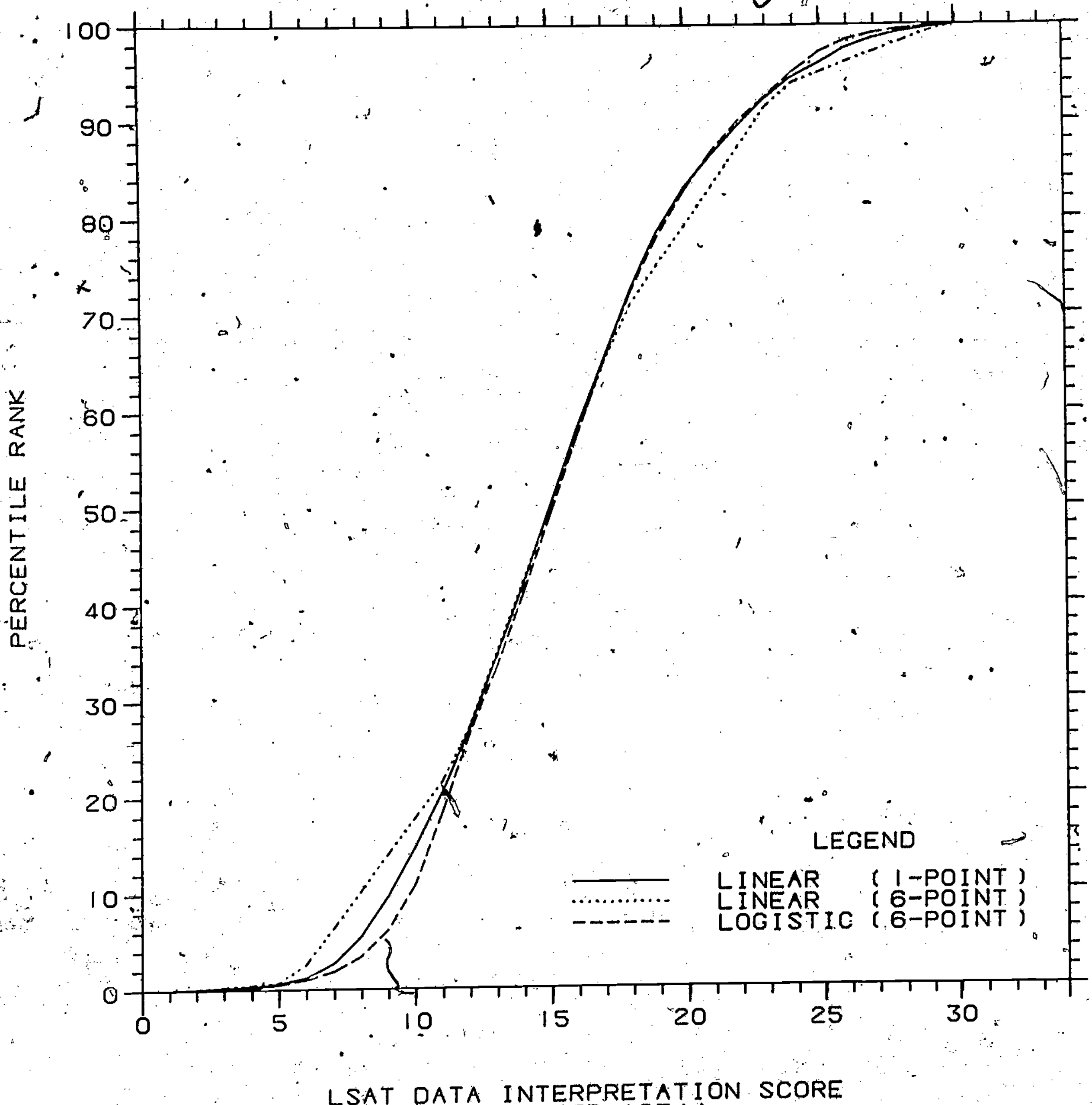


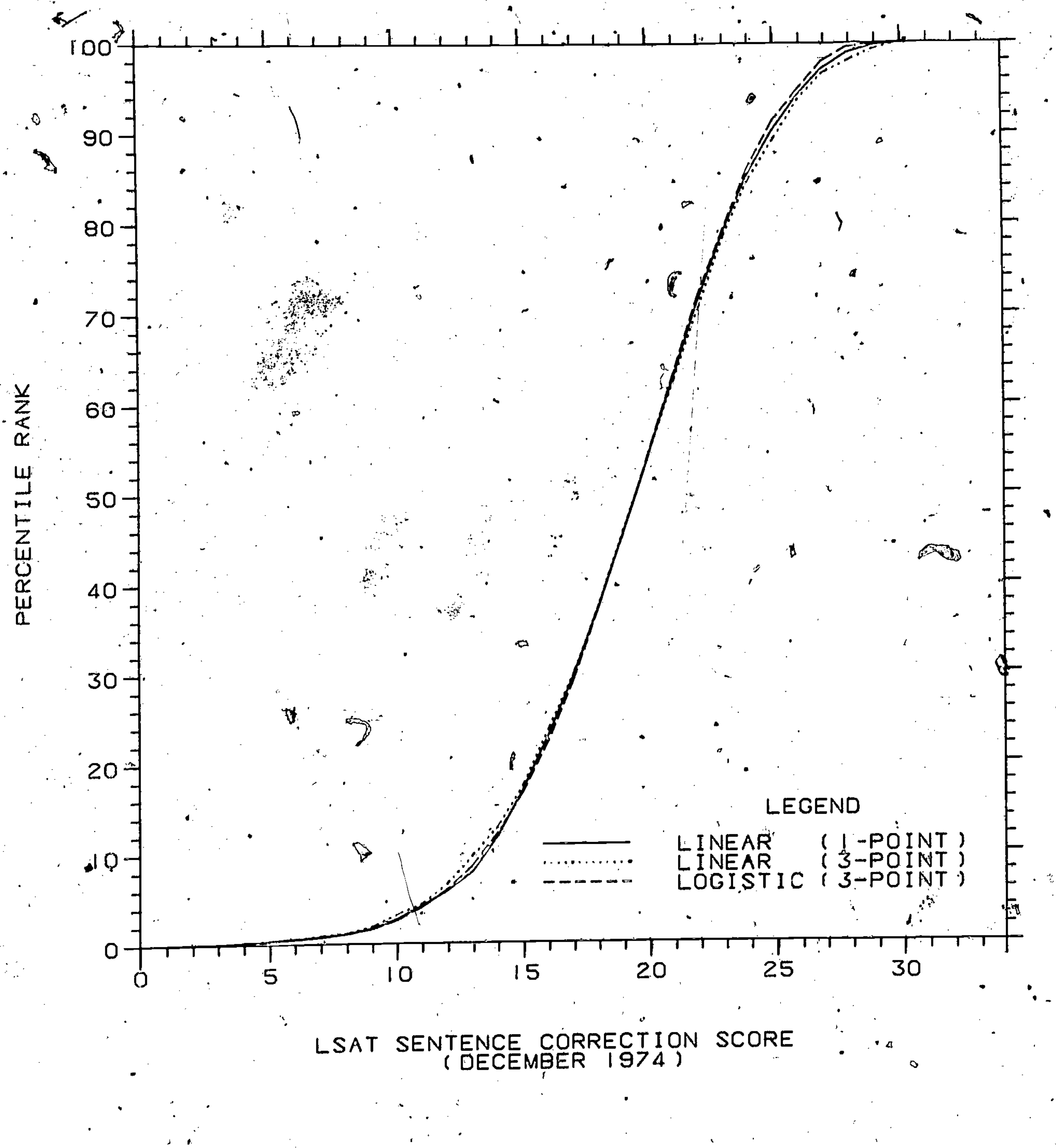
PERCENTILE RANK
100
90
80
70
60
50
40
30
20
10
0
0
5
10
15
20
25
30
LEGEND
LINEAR (1-POINT)
LINEAR (3-POINT)
LOGISTIC (3-POINT)
LSAT DATA INTERPRETATION SCORE
(FEBRUARY 1974)


PERCENTILE RANK
100
90
80
70
60
50
40
30
20
10
0
0
5
10
15
20
25
30
LEGEND
LINEAR (1-POINT)
LINEAR (6-POINT)
LOGISTIC (6-POINT)
LSAT DATA INTERPRETATION SCORE
(FEBRUARY 1974)

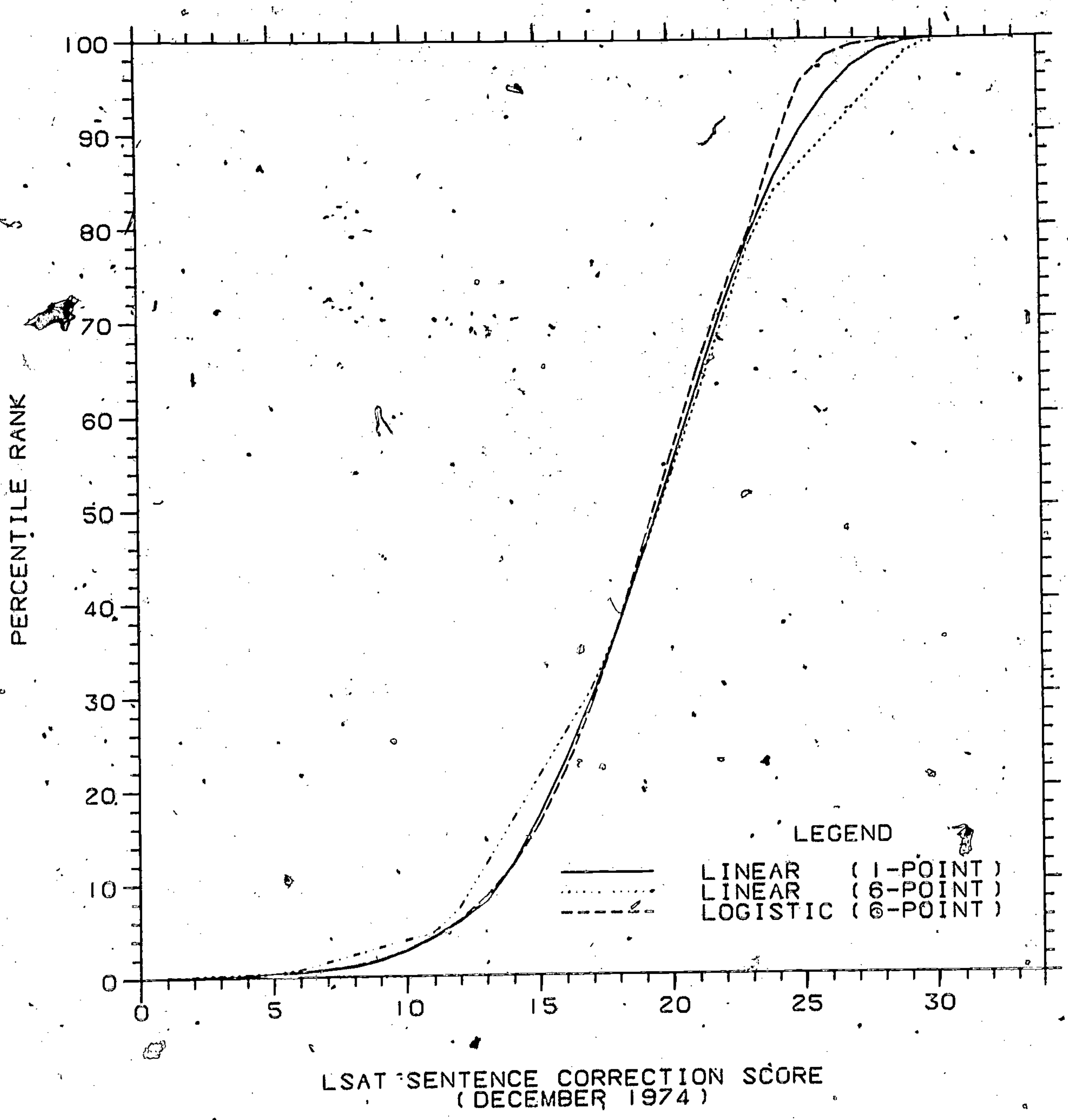


100
90
80
70
60
50
40
30
20
10
0
PERCENTILE RANK
0
5
10
15
20
25
30
LSAT DATA INTERPRETATION SCORE
(DECEMBER 1974)
LEGEND
LINEAR (1-POINT)
LINEAR (3-POINT)
LOGISTIC (3-POINT)


PERCENTILE RANK
100
90
80
70
60
50
40
30
20
10
0
0
5
10
15
20
25
30
LEGEND
LINEAR (1-POINT)
LINEAR (6-POINT)
LOGISTIC (6-POINT)
LSAT DATA INTERPRETATION SCORE
(DECEMBER 1974)



100
90
80
70
60
50
40
30
20
10
0
PERCENTILE RANK
0
5
10
15
20
25
30
LEGEND
LINEAR (1-POINT)
LINEAR (3-POINT)
LOGISTIC (3-POINT)
LSAT SENTENCE CORRECTION SCORE
(DECEMBER 1974)


100
90
80
70
60
50
40
30
20
10
0
PERCENTILE RANK
0
5
10
15
20
25
30
LEGEND
LINEAR (1-POINT)
LINEAR (6-POINT)
LOGISTIC (6-POINT)
LSAT SENTENCE CORRECTION SCORE
(DECEMBER 1974)