MATCHACURVE - 2 FOR ALGEBRAIC TRANSFORMS TO DESCRIBE CURVES OF THE CLASS X^n

Chester E. Jensen and Jack W. Homeyer

INTERMOUNTAIN FOREST AND RANGE EXPERIMENT STATION
Ogden, Utah 84401

USDA Forest Service Research Paper INT-106 1971

USDA Forest Service
Research Paper INT-106
August 1971

MATCHACURVE-2 FOR ALGEBRAIC TRANSFORMS TO DESCRIBE CURVES OF THE CLASS X^n

Chester E. Jensen and Jack W. Homeyer

INTERMOUNTAIN FOREST AND RANGE EXPERIMENT STATION
Forest Service
U.S. Department of Agriculture
Ogden, Utah 84401
Robert W. Harris, Director

THE AUTHORS

CHESTER E. JENSEN is Principal Statistician for the Intermountain Forest and Range Experiment Station. Previously, he held the same position at the Northeastern and Central States Forest Experiment Stations. Mr. Jensen, who holds a master of forestry degree from Michigan State University, received his training in graduate statistics at Iowa State University.

JACK W. HOMEYER is a Statistician working with Jensen at Intermountain Forest and Range Experiment Station. He graduated from Utah State University in 1969 with a bachelor of science degree in mathematics.

CONTENTS

ABSTRACT

Graphed curves of the class, X^n, can be quickly scaled to and compared with the wide array of standard curves presented. If a standard can be found that suitably matches the scaled curve, the algebraic transform specified for the standard can be applied to the X values of a relevant data set and these can be scaled to their corresponding Y values by least squares to arrive at the fitted model.

INTRODUCTION

The analyst often initiates a two-dimensional regression analysis by hand-fitting the expected curve for an XY relation through a set of plotted data points called the "graphed curve" in this paper.

Assume he wants next to find a mathematical form of the independent variable, X, which when scaled to the graph emulates the graphed curve with acceptable accuracy. If the curve is linear, X itself is, of course, the appropriate form. But if the curve is nonlinear, a matching nonlinear form of X should be the object of his search. Once having found a suitable form, the analyst can rescale it to the actual data points of any relevant data set[1] by least squares, and the development of the estimating equation is complete.

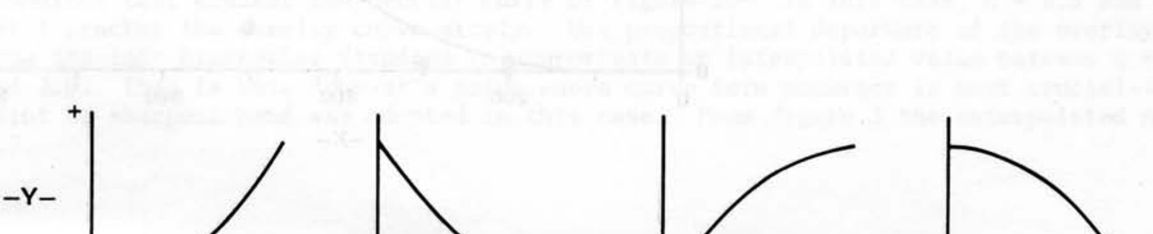
In this paper we have attempted to reduce the effort required to find acceptable transforms of the X^n class. The analyst simply compares a scaled version of his graphed curve to graphed Standards (page 9) and selects the two adjacent ones most nearly like his own in shape, a process hereinafter referred to as "Matchacurve."

Interpolation between the transforms given for these Standards results in identification of the best alternative offered by the system. It is this transform of X that is finally fitted to a relevant data set by least squares.

Three unique sets of standard curves are presented. Each set is based on a selected array of n in X^n, with limits as shown:

Set	*n-array limits*
1	$1.00 \leq n \leq 20.00$
2	$0.10 \leq n \leq 1.00$
3	$-2.00 \leq n \leq -0.01$

Each set of Standards is transformed as required (see Appendix C) to appear in each of the four basic positions in which the analyst's curve can occur in the upper right quadrant so that there are 12 sets of Standards in all--3 sets in each of 4 positions.



[1]Either the data set from which the graphed curve is derived or some other set to which this curve is judged applicable.

The scaling procedure for sets 1 and 2 is shown in the example that follows. Set 3, because it involves reciprocals and because the latter are indeterminant at X = 0, requires that Y_p be determined within a slightly abbreviated X-range as explained in the last paragraph under the section labeled "An Example."

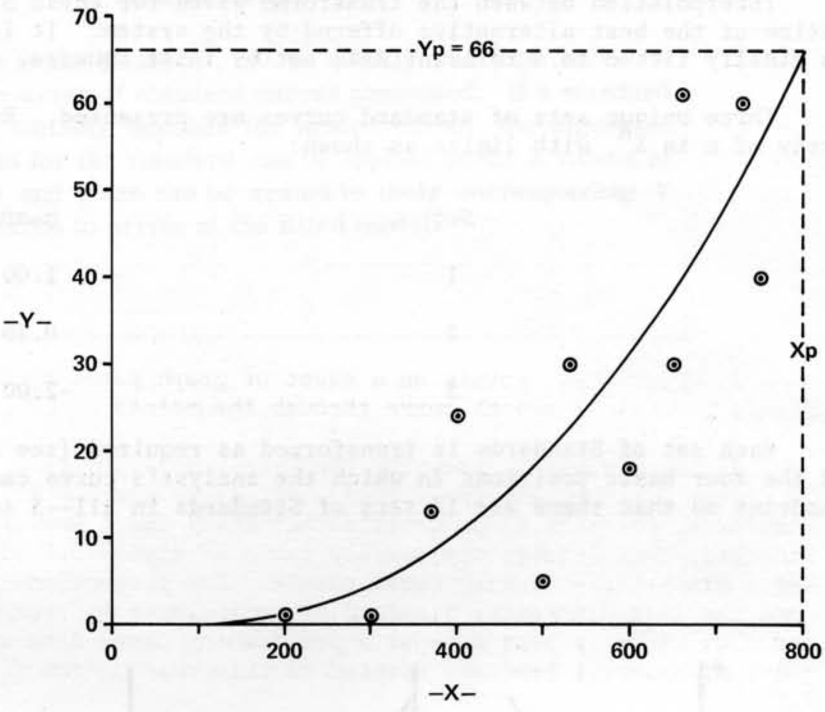
For the sake of efficiency, then, it is suggested that the analyst's curve be scaled first for comparison to sets 1 and 2[2/] ; the curve will be scaled for comparison to set 3 only when no suitably similar form can be found in sets 1 and 2.

Where the analyst's curve lies either partly or wholly in some quadrant other than "upper right" (i.e., not oriented as the curves above), adjust either the X- or Y-scale, or both, to put the curve on the same basis as the standards before applying Matchacurve --see Appendix A.

AN EXAMPLE

This example applies as shown to sets 1 and 2.

Figure 1.--Here, the expected curve form has been hand-fitted through plotted data points. Plus and minus departures are balanced approximately.



Let's say that an analyst wants to generate an equation for Y from the data set plotted in figure 1. From previous knowledge of the relation between X and Y, he expects the curve to be concave upward with larger values of X. The apparent data trend supports his expectation and he hand-fits such a curve through the data, also shown in

---

[2/]This scaling procedure is identical to that employed for the 1970 companion paper for sigmoids, Matchacurve-1, so the same scaled curve can be compared to the sigmoid standards, if pertinent.

figure 1. This curve then, is to be emulated using the Matchacurve process detailed in the steps below:

1. Let X_p be the value of X at or near the largest X-value in the data set. Let Y_p be the value of Y at X_p. Determine the values of X_p (=800) and Y_p (=66) as shown in figure 1. Note that for sets 1 and 2, Y_p is measured at X = 0 or at X_p depending on position--see table 1. Select representative point coordinates from the smoothed curve.

Five points have been chosen for this example.[3/] Call the array of X values X_i, and the array of Y values, Y_i.

Point coordinates from the smoothed curve.

X_i	0	200	400	600	800
Y_i	0	1.3	10.6	31.4	66

2. Scale the X_i to X_i/X_p and the Y_i to Y_i/Y_p.

Scaled coordinates

$X_i/800$	0.00	0.25	0.50	0.75	1.00
$Y_i/66$	0.00	0.02	0.16	0.48	1.00

3. Plot these points on a sheet of graph paper at the *exact* scale[4/] shown in figure 2. Draw a smooth curve through the points.

4. From table 1, identify the "position"--position A, here--of the scaled curve in figure 2. Then use the scaled curve as an overlay for Set 1 and Set 2 of the Standards, position A; be certain that the X and Y axes are matched exactly. Find the Standards that bracket the overlay curve of figure 3.[5/] In this case, n = 2.5 and 3.0 of Set 1 bracket the overlay curve nicely. Use proportional departure of the overlay curve from the left bracketing Standard to approximate an interpolated value between n = 2.5 and 3.0. This is best done at a point where curve form accuracy is most crucial--the point of sharpest bend was adopted in this case. From figure 3 the interpolated n is 2.7.

[3/] Use fewer points where the desired sensitivity is less, more for greater sensitivity.

[4/] This is the same scale used for the Standards. Note that the odd X scale was necessitated by publication limitations on paper size.

[5/] A light source behind the graphs that are being compared will assist in the matching process.

Table 1.--*Information on the Standards (page 9)*

Analyst's curve position	Set	Exponent range	Y_p measured at X =	X transform[1/] to be fitted by least squares	Sign of $\hat{B}_1$ under least squares fit
A 	[2/]1	$1.0 \leq n \leq 20.0$	X_p	$(X)^n$	+
	[2/]2	$0.1 \leq n \leq 1.0$	X_p	$(X_p-X)^n$	-
	[2/]3	$-2.00 \leq n \leq -0.01$	X_p	$(X_p-X)^n$	+
B 	1	)	0	$(X_p-X)^n$	+
	2	) As above	0	$(X)^n$	-
	3	)	$0.01X_p$	$(X)^n$	+
C 	1	)	X_p	$(X_p-X)^n$	-
	2	) As above	X_p	$(X)^n$	+
	3	)	X_p	$(X)^n$	-
D 	1	)	0	$(X)^n$	-
	2	) As above	0	$(X_p-X)^n$	+
	3	)	$0.01X_p$	$(X_p-X)^n$	-

[1/]To achieve the specified positions for some of the sets, the X-scale must be reversed before applying the exponents as indicated here.

[2/]CAUTION: Limits of use for Matchacurve transforms are as follows:

Sets 1 and 2, $0 \leq X \leq X_p$

Set 3, $0.01X_p \leq X \leq X_p$

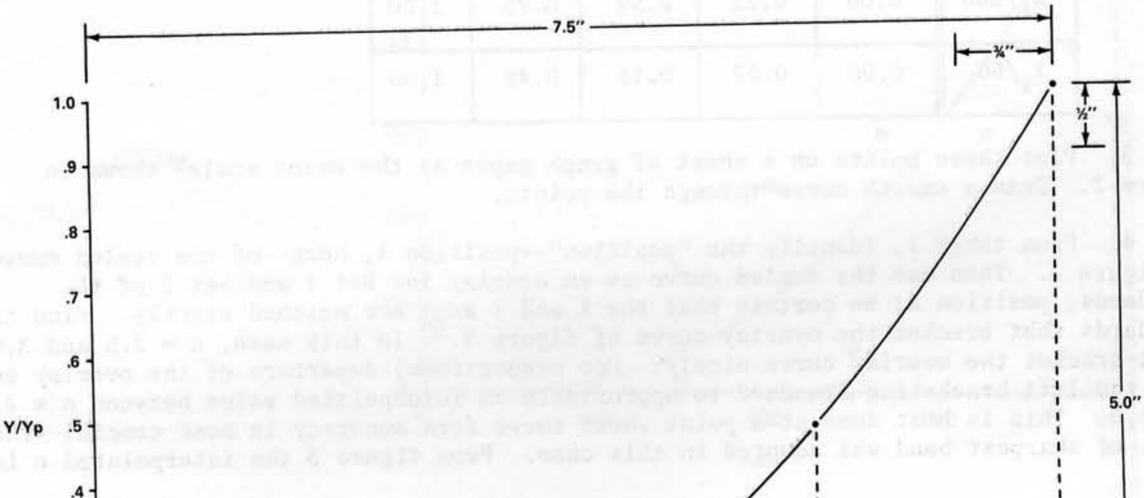


Figure 2.--Here, the graphed curve has been scaled to the Standards.

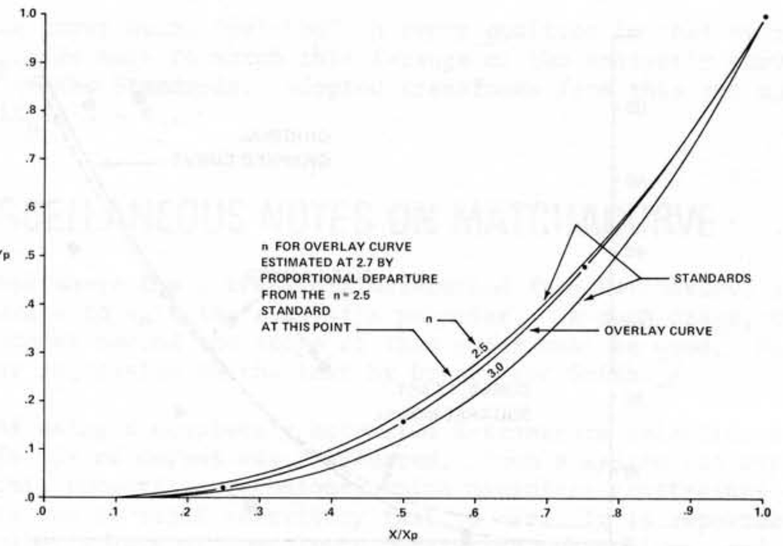


Figure 3.--The Standards best matched to the original curve bracket the overlay curve.

5. Substitute Y for Y/Y_p in the final, simple linear model to be fitted to a relevant data set (the original set is used here) by least squares:

$$Y = B_0 + B_1 X^{2.7}$$

Then the estimated B_0, B_1 -- or $\hat{B}_0$, $\hat{B}_1$ -- will scale the X_i transforms to the Y_i.

It is generally preferable to use a computer for transforming the X_i and applying the least squares fitting process and, in this case, the fitted model is:[6]

$$\hat{Y} = 2.06 + (0.8932 \times 10^{-6}) X^{2.7}, \; 0 \leq X \leq X_p$$

The curve for this model is compared to the original graphed curve in figure 4. Differences indicate the extent to which the hand-fitted curve failed to follow the least squares path. We assume here that the least squares curve is the most desirable alternative *within the X-range of the data.*

[6] Using logarithm tables for the $X^{2.7}$ transforms and a desk calculator for the fitting process, the more laborious hand calculations would be as shown in Appendix B.

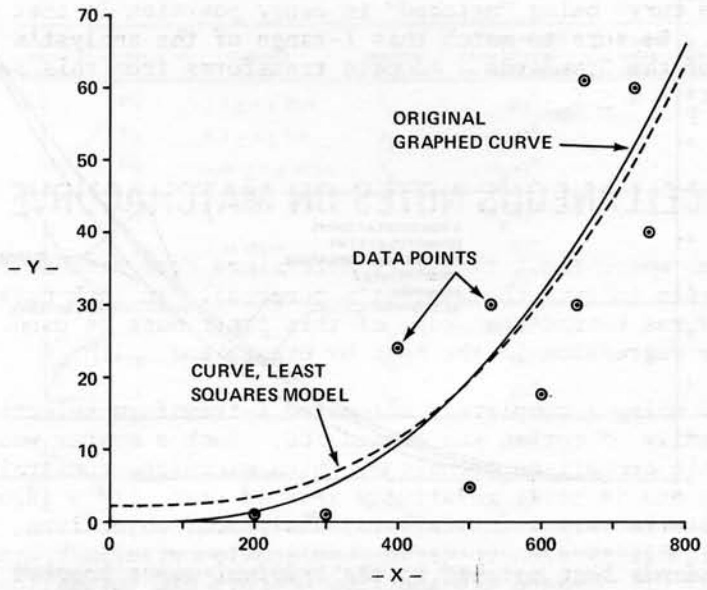
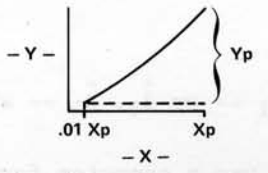
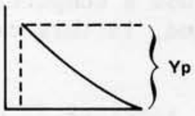
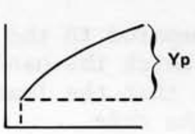
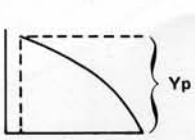


Figure 4.--The fitted model has been plotted over the original graphed form.

The foregoing procedures apply in detail to set 3 with the exception that Y_p is measured as shown in the graphs below:

Position	*Analyst's curve*	Y_p
A		$Y_{X_p} - Y_{0.01X_p}$ *
B		$Y_{0.01X_p}$
C		$Y_{X_p} - Y_{0.01X_p}$
D		$Y_{0.01X_p}$

---

*Y_{X_p} = Y at X_p, $Y_{0.01X_p}$ = Y at $0.01X_p$

The portion of the curve being "matched" in every position is that over the X-range from $0.01X_p$ through X_p. Be sure to match this X-range of the analyst's curve with the corresponding X-range of the Standards. Adopted transforms from this set apply *only* within the limits $0.01X_p \le X \le X_p$.

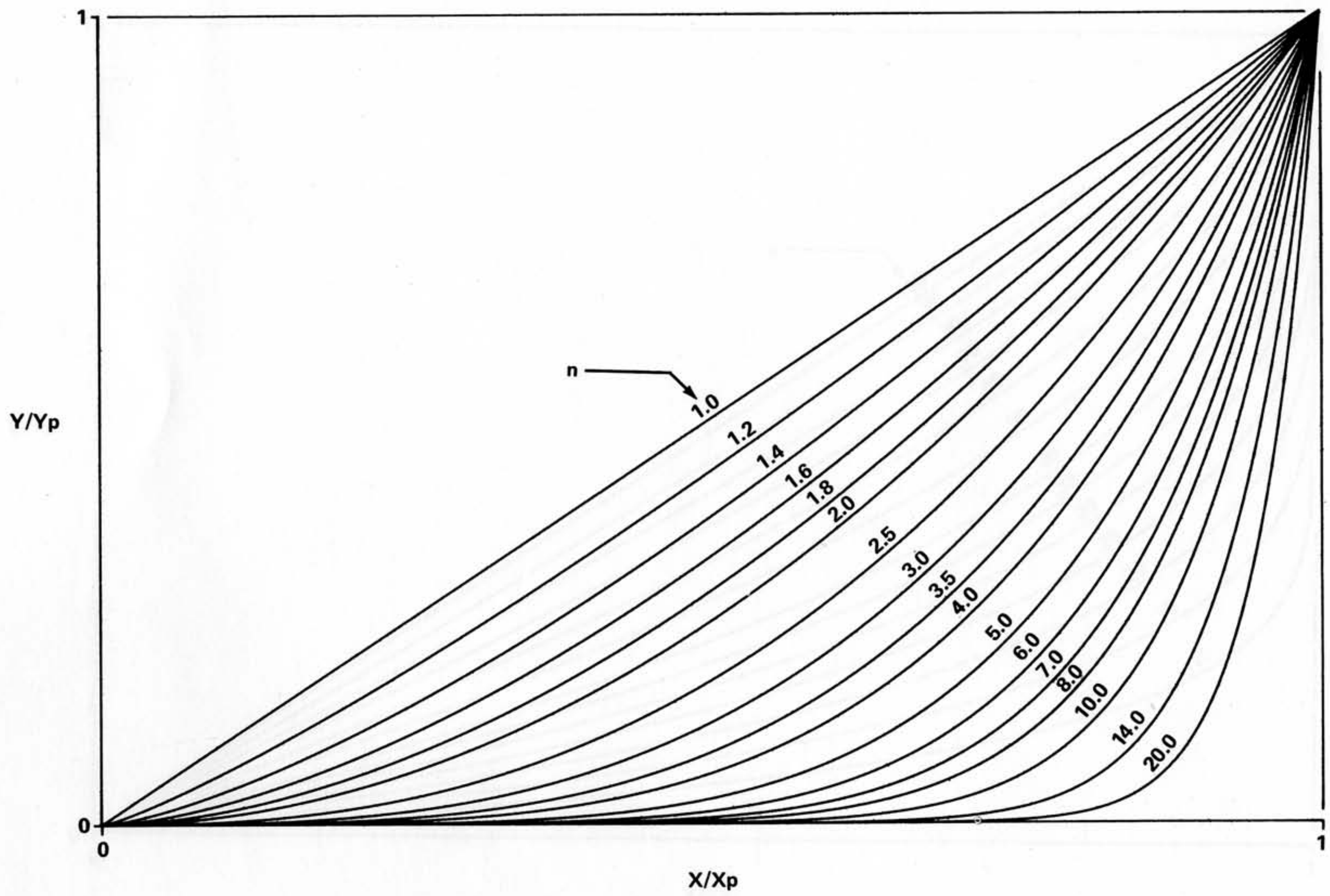
MISCELLANEOUS NOTES ON MATCHACURVE

- There will be cases where the X transform determined from Matchacurve is not sufficiently accurate to suit the analyst's purposes. In such cases, curve-form description procedures beyond the scope of this paper must be used. For example, refer to nonlinear regression in the text by Draper and Smith.[7/]

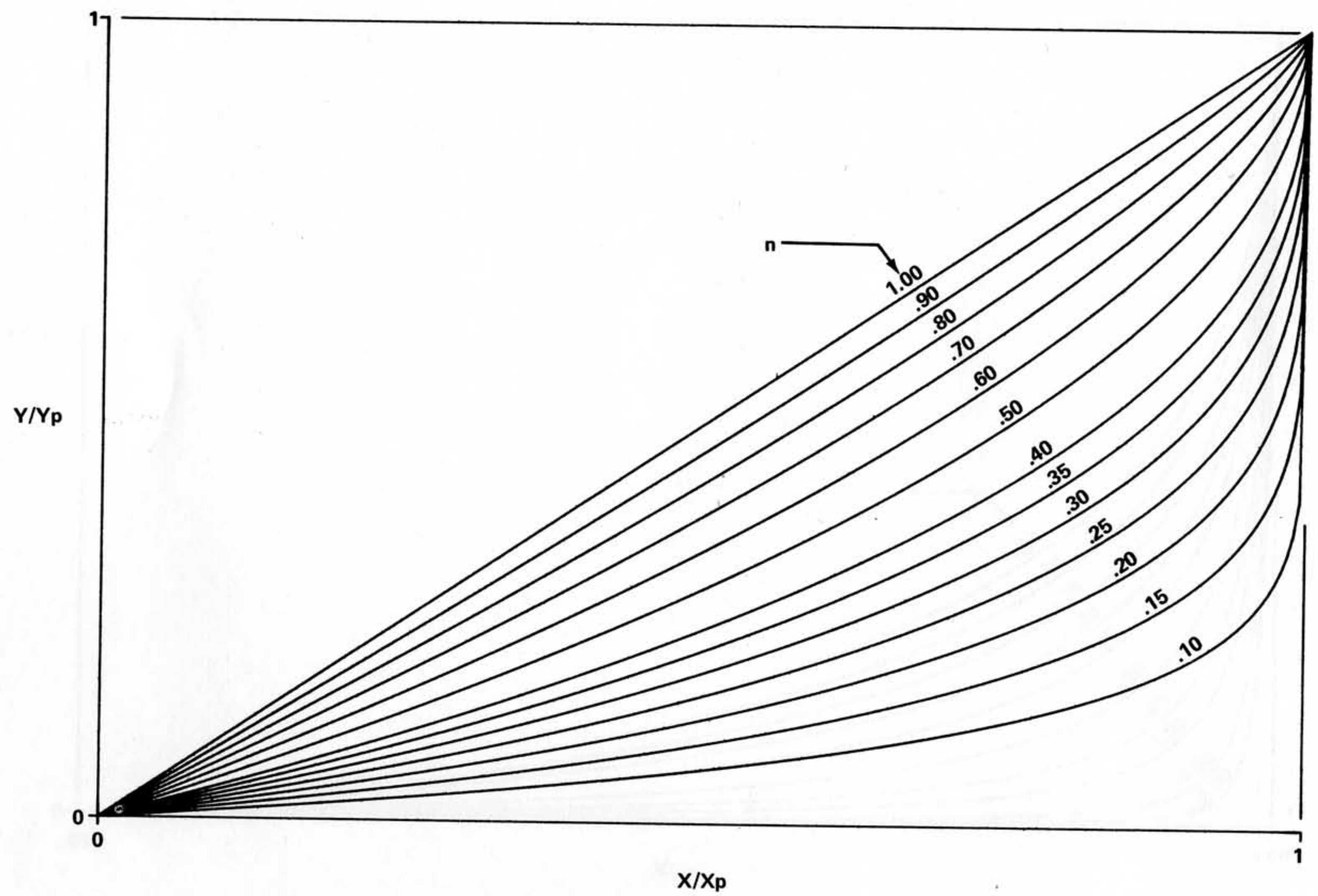
- The alternative of using a completely automated X-transform selection system from the Matchacurve family of curves was considered. Such a system was bypassed in favor of the graphic comparison technique, which minimizes constraints on curve selection criteria and is still relatively fast to use. It is important to note that selection criteria vary with analysts, analytical objectives, and data characteristics . Our inability even to define applicable criteria commonly used (aside from least squares) discouraged the adoption of complete automation here.

[7/] N. R. Draper and H. Smith. Applied regression analysis. New York: Wiley and Sons, Inc. 1968.

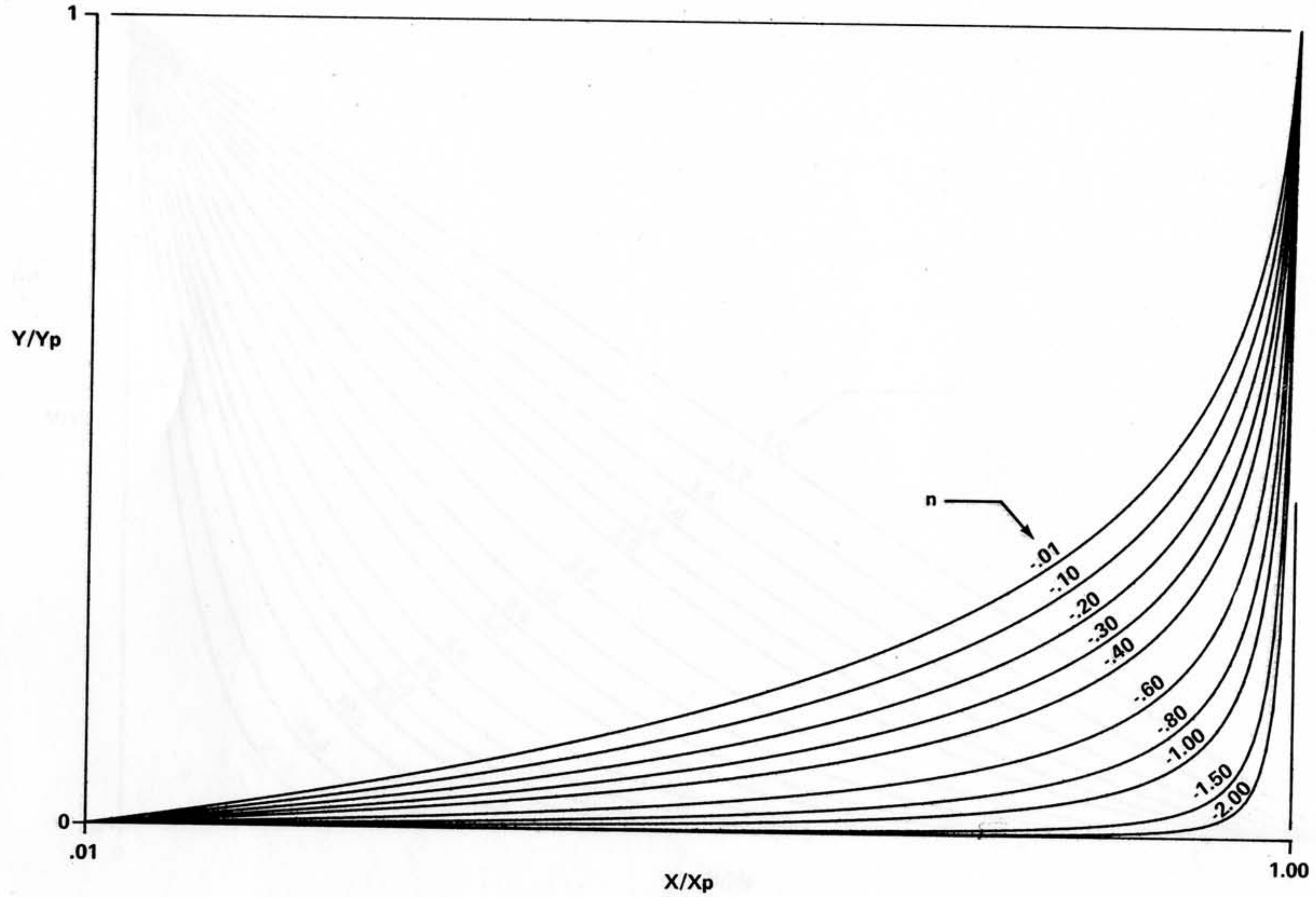
STANDARDS



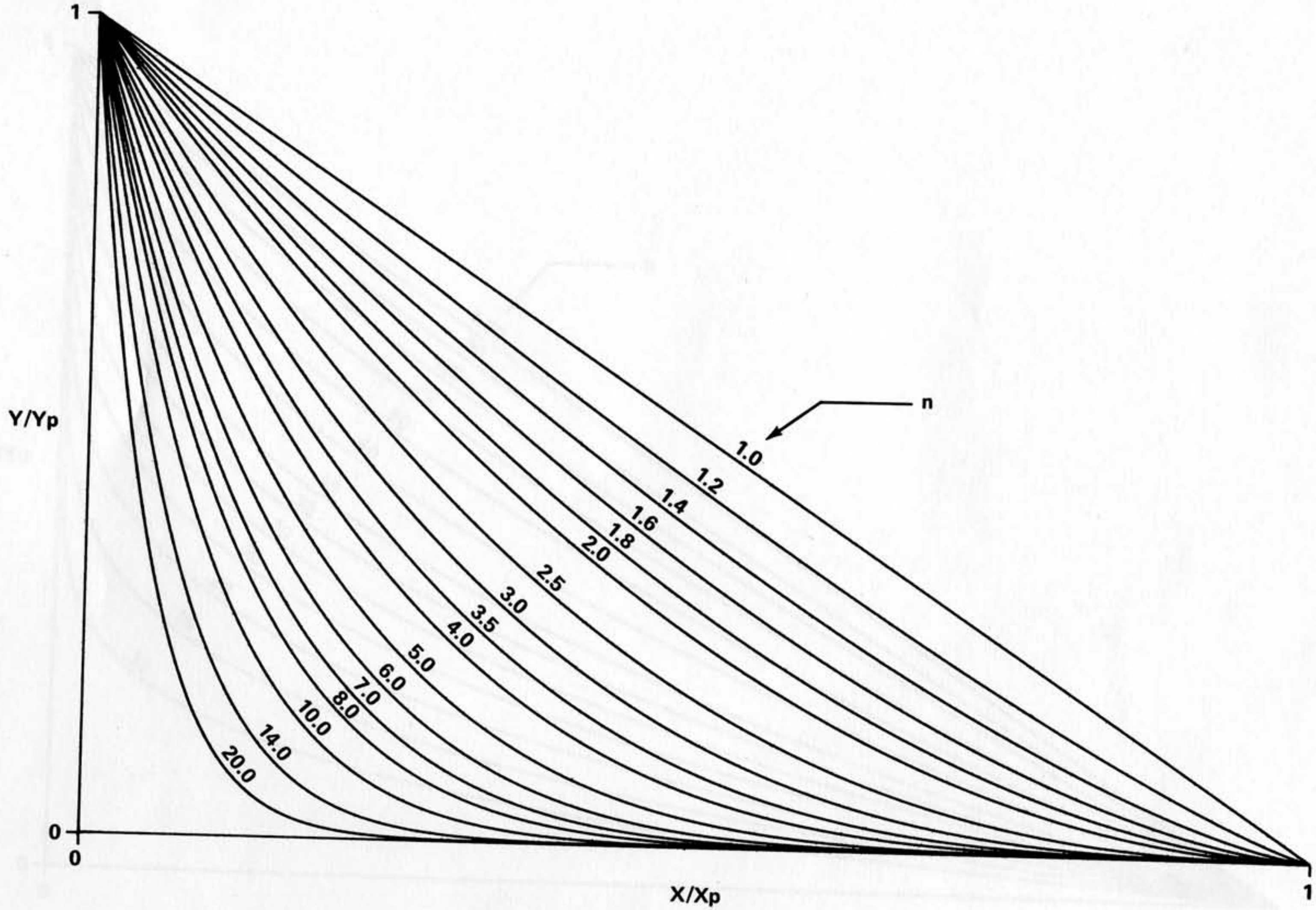
A-1.--Standards for position A, set 1--(X-transform to be fitted by least squares = $(X)^n$, $0 \le X \le X_p$*)*



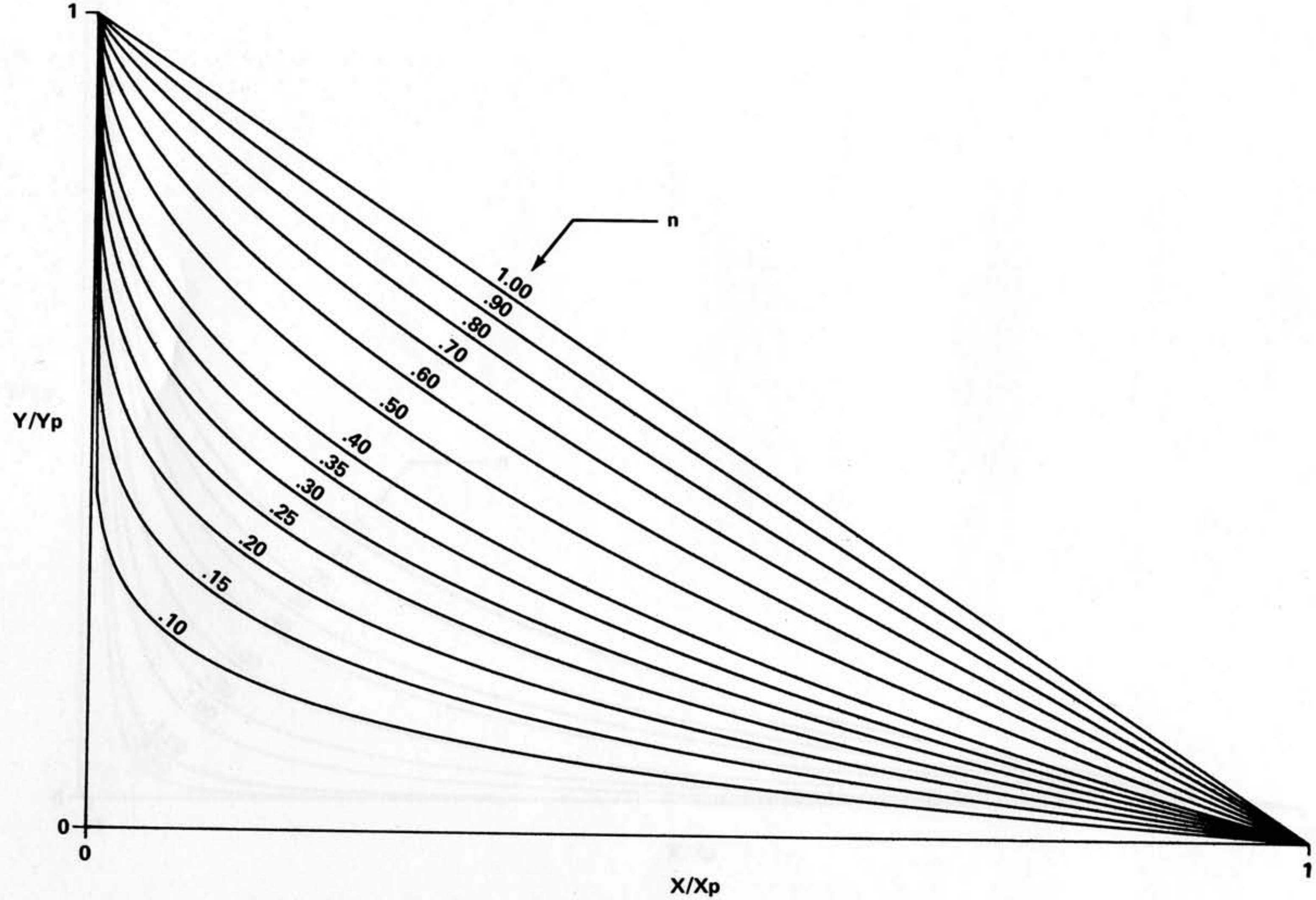
A-2.--Standards for position A, set 2--(X-transform to be fitted by least squares = $(X_p - X)^n$*,* $0 \le X \le X_p$*)*



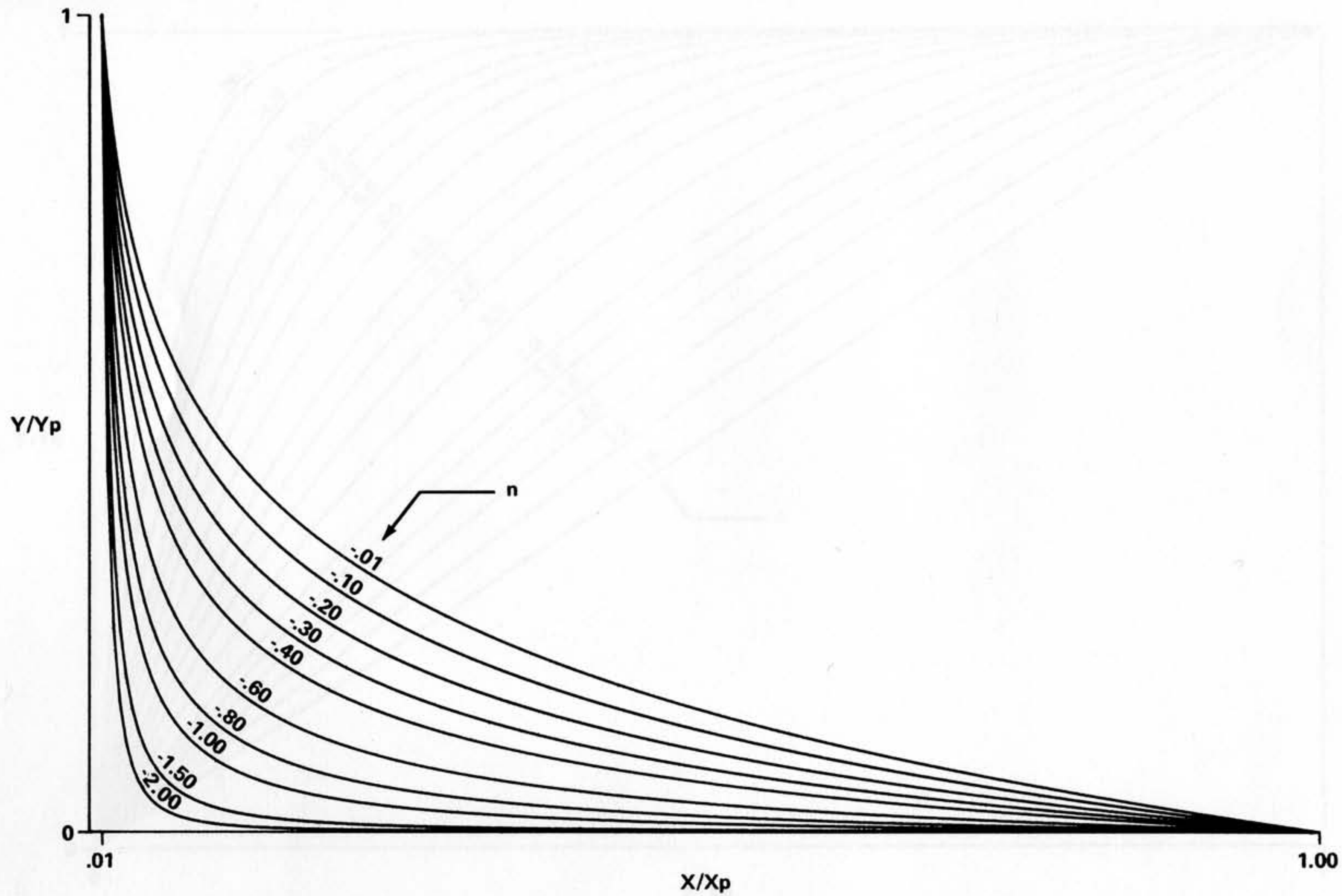
A-3.--Standards for position A, set 3--(X-transform to be fitted by least squares = $(X_p - X)^n$, $.01X_p \le X \le X_p$*)*



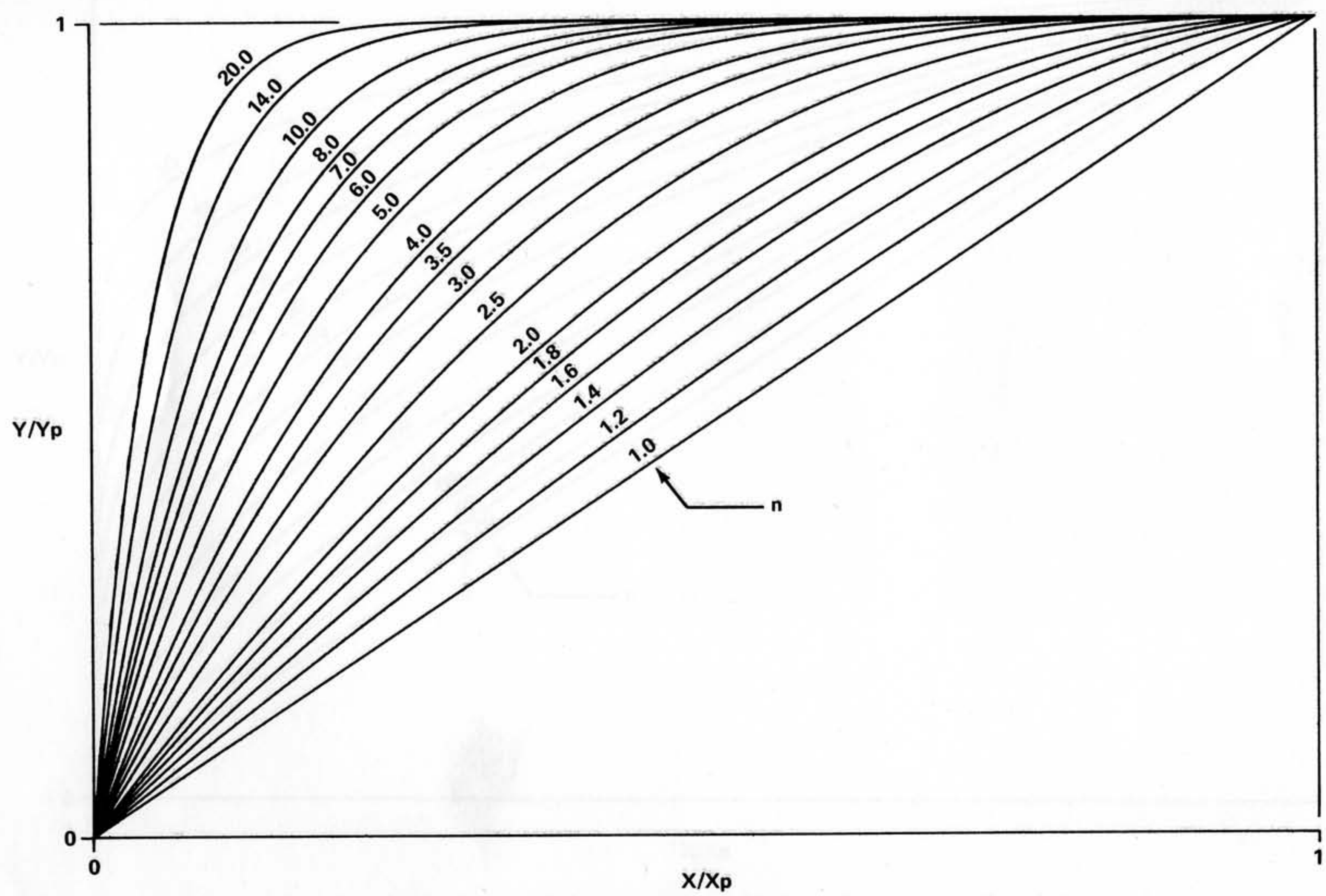
B-1.--Standards for position B, set 1--(X-transform to be fitted by least squares = $(X_p - X)^n$, $0 \le X \le X_p$*)*



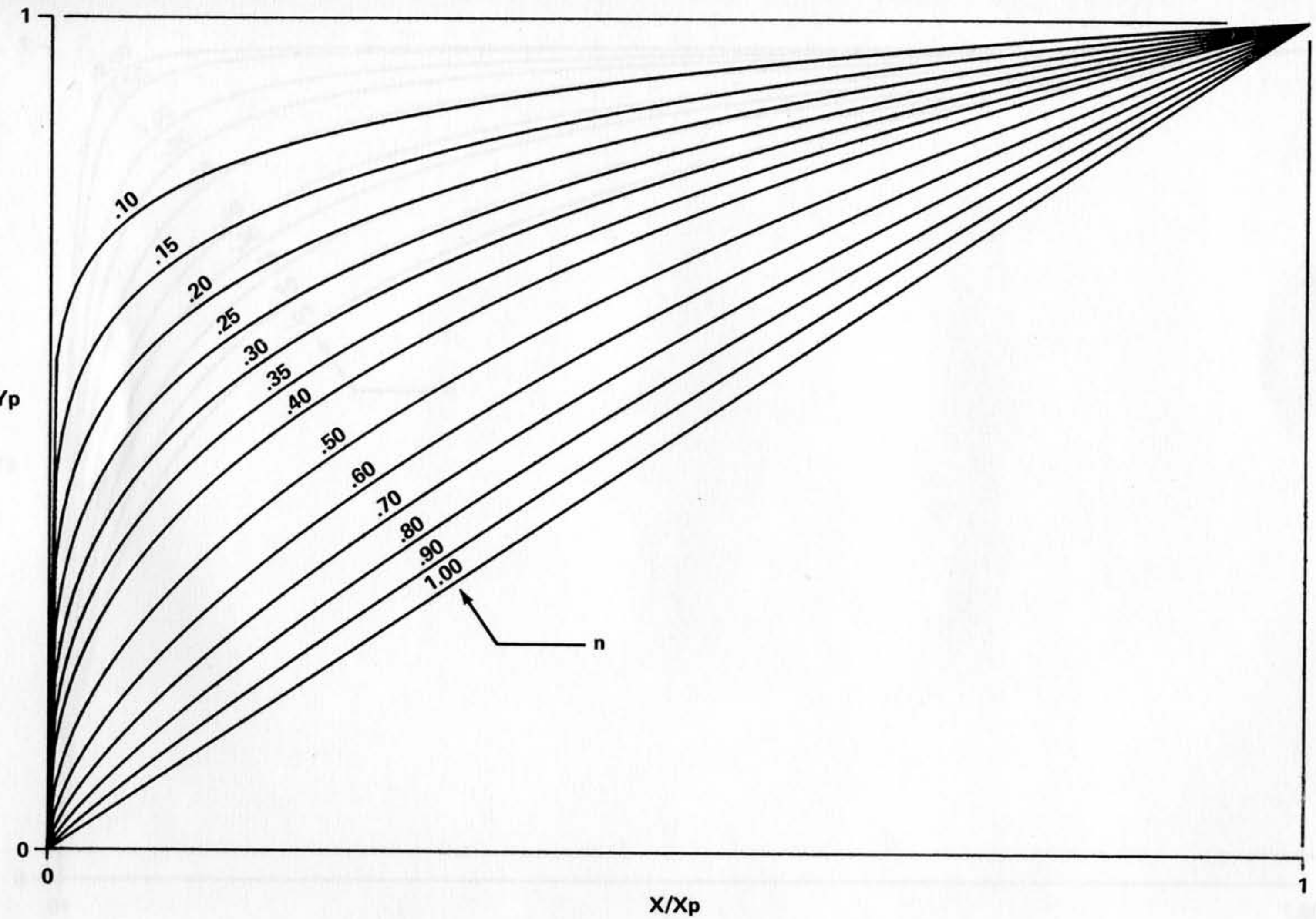
B-2.--Standards for position B, set 2--(X-transform to be fitted by least squares = $(X)^n$*,* $0 \leq X \leq X_p$*)*



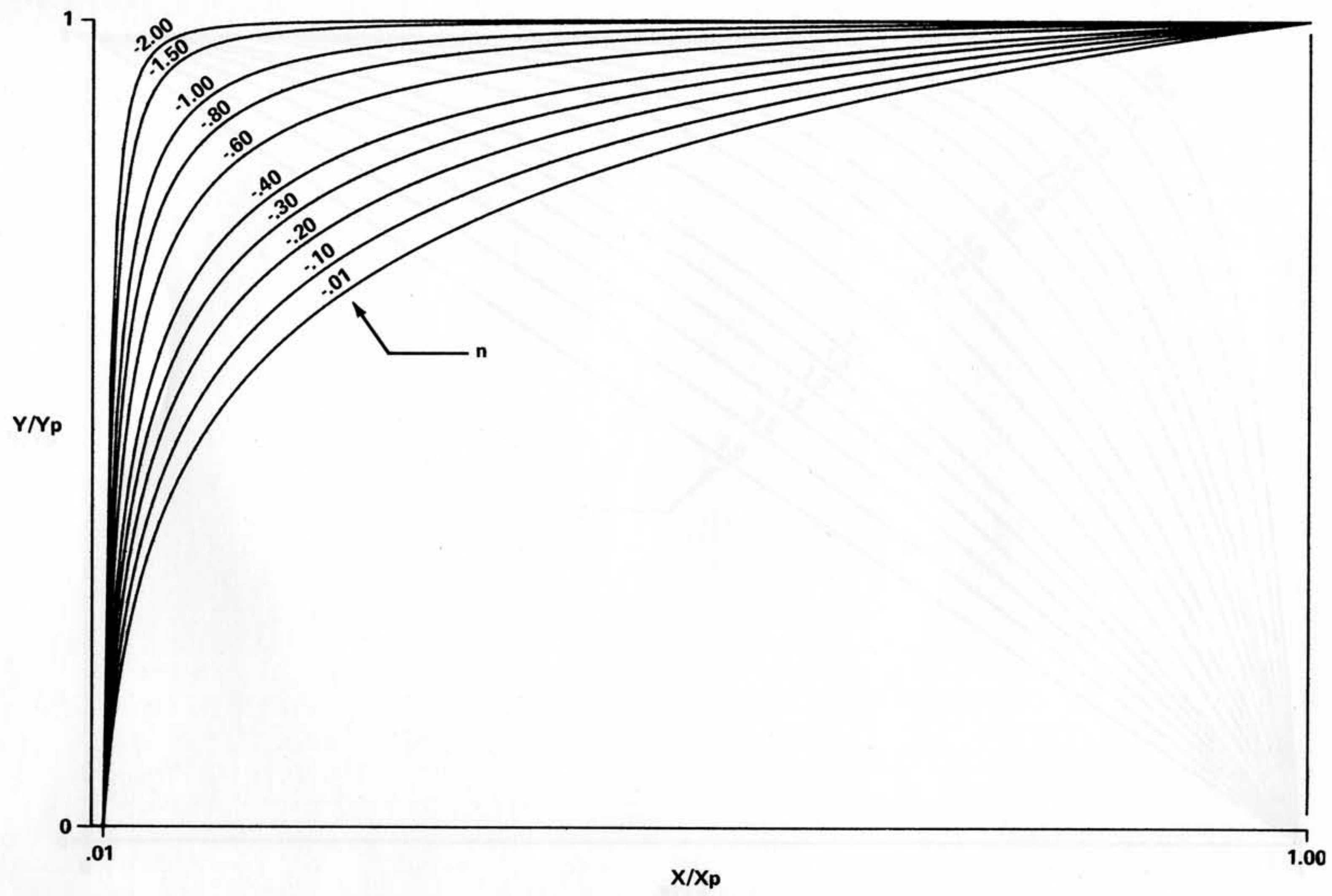
B-3.--Standards for position B, set 3--(X-transform to be fitted by least squares = $(X)^n$, $.01X_p \leq X \leq X_p$*)*



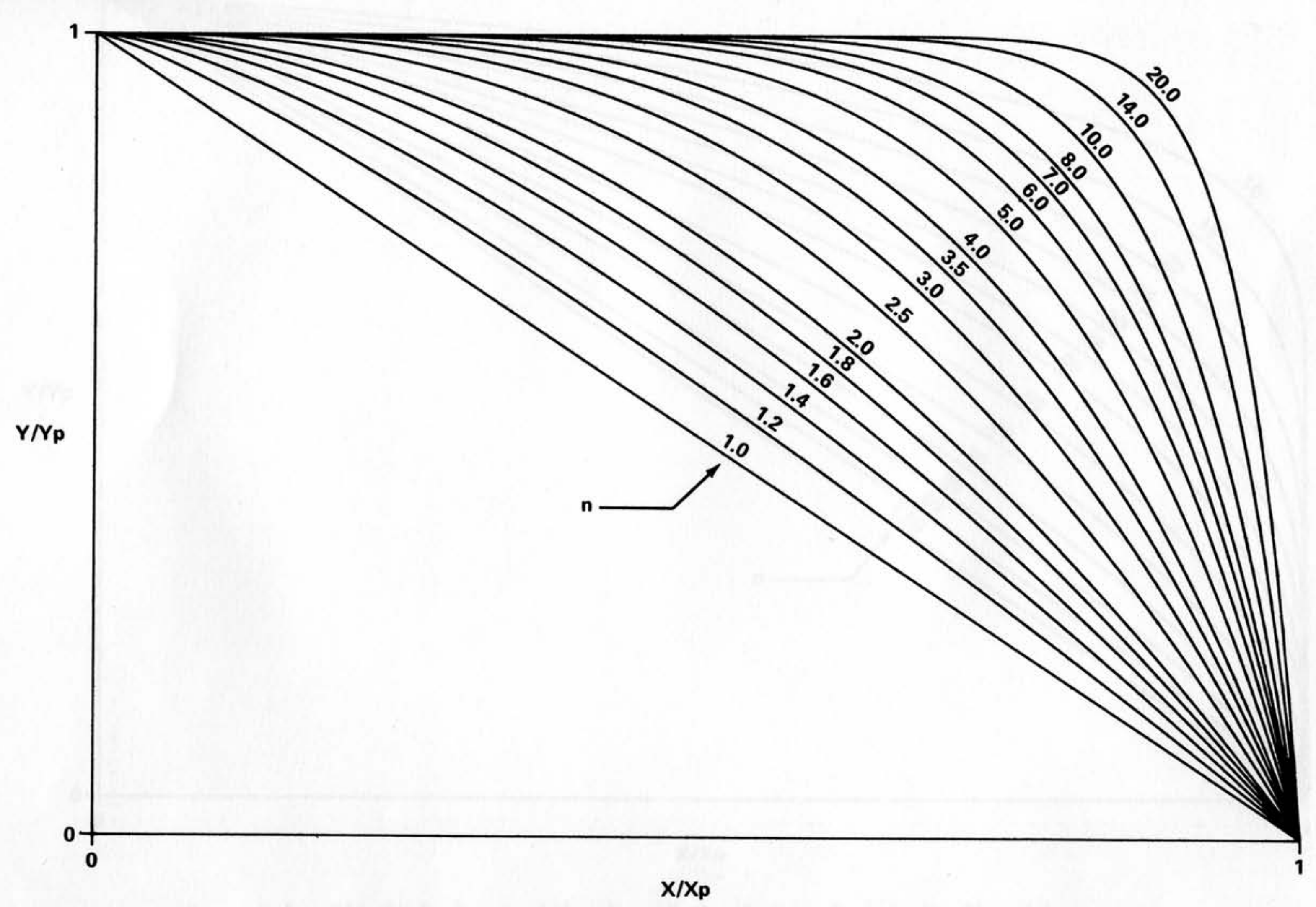
C-1.--Standards for position C, set 1--(X-transform to be fitted by least squares = $(X_p-X)^n$, $0 \leq X \leq X_p$)



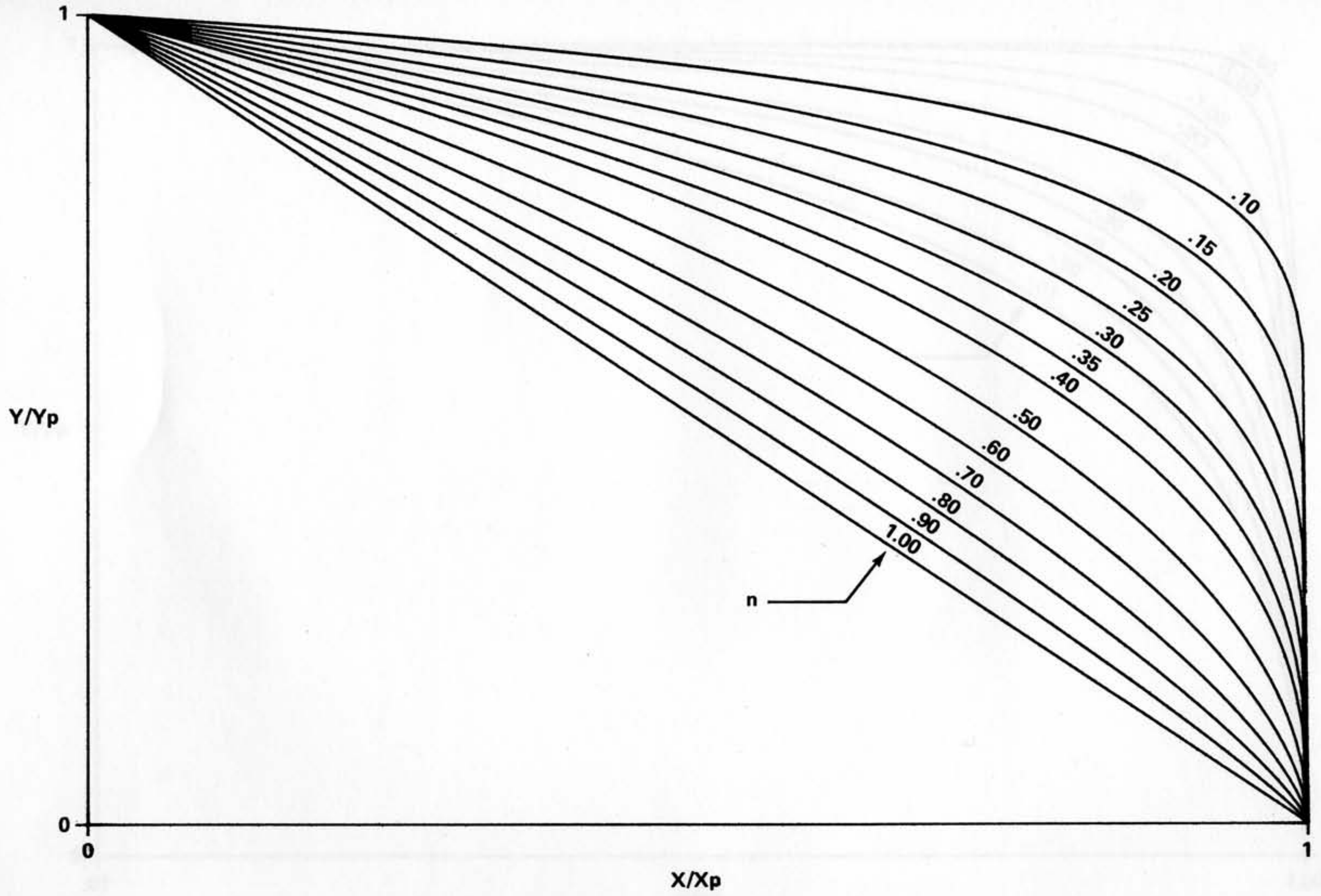
C-2.--Standards for position C, set 2--(X-transform to be fitted by least squares = $(X)^n$*,* $0 \leq X \leq X_p$*)*



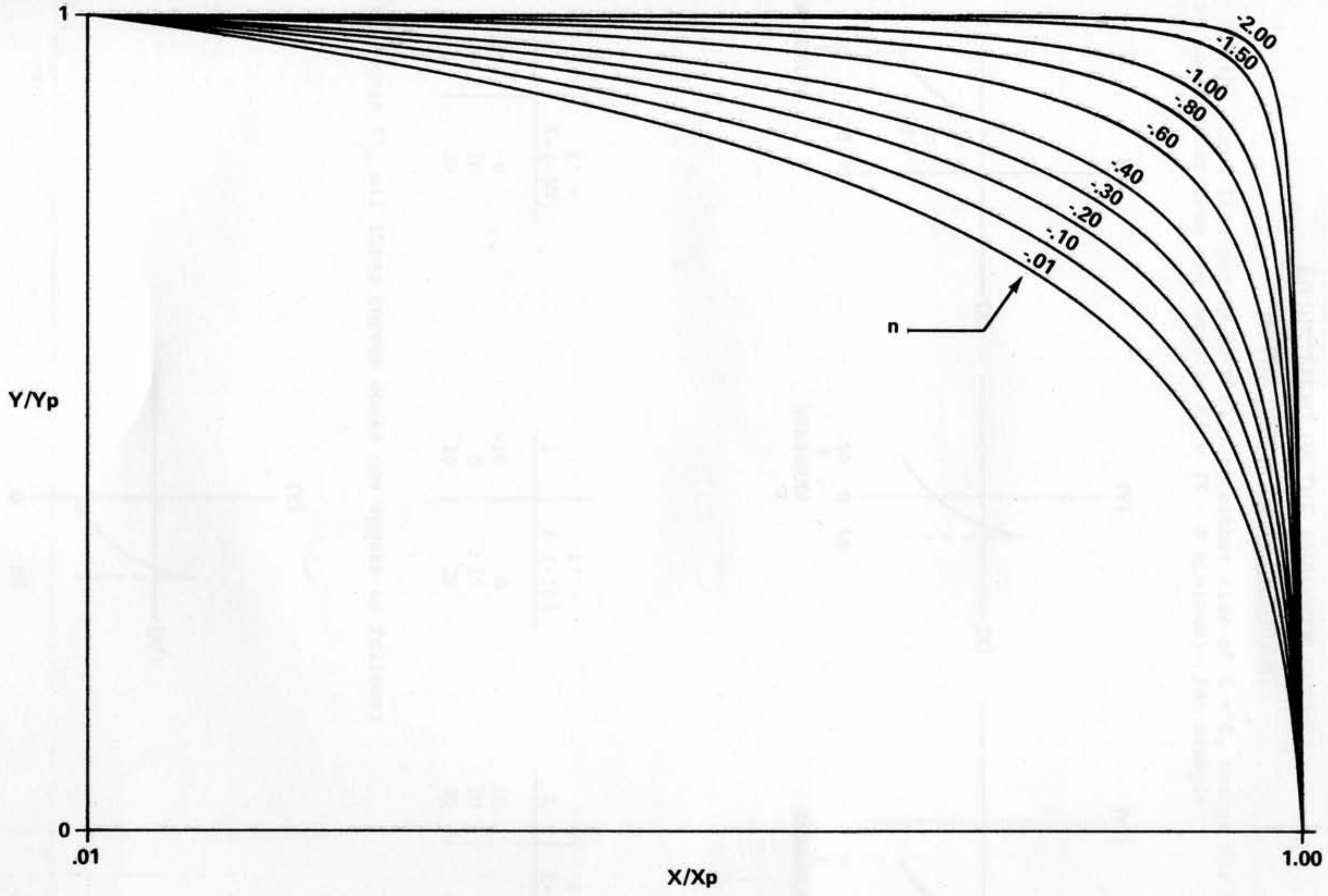
C-3.--Standards for position C, set 3--(X-transform to be fitted by least squares = $(X)^n$, $.01X_p \leq X \leq X_p$*)*



D-1.--Standards for position D, set 1--(X-transform to be fitted by least squares = $(X)^n$*,* $0 \leq X \leq X_p$*)*



D-2.--Standards for position D, set 2--(X-transform to be fitted by least squares = $(X_p-X)^n$, $0 \le X \le X_p$*)*

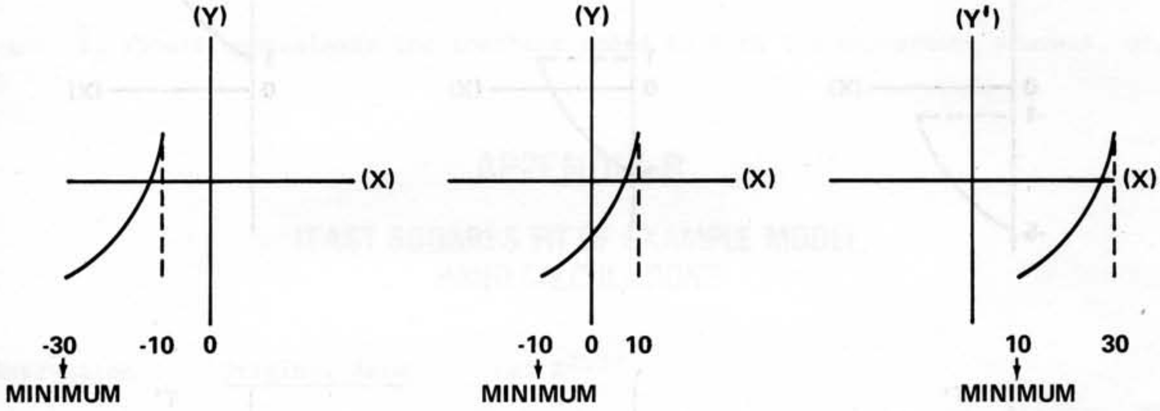


D-3.--Standards for position D, set 3--(X-transform to be fitted by least squares = $(X_p - X)^n$*,* $.01X_p \leq X \leq X_p$*)*

APPENDIX A

ADJUSTMENT OF THE ANALYST'S CURVE TO THE UPPER RIGHT QUADRANT

If the curve lies partly or wholly to either side of X = 0, change the X values of point coordinates from the curve to X' = (X - X minimum)--for example:

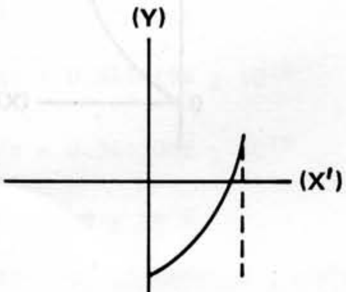


X	X' = X-(-30)
-30	0
-20	10
-10	20

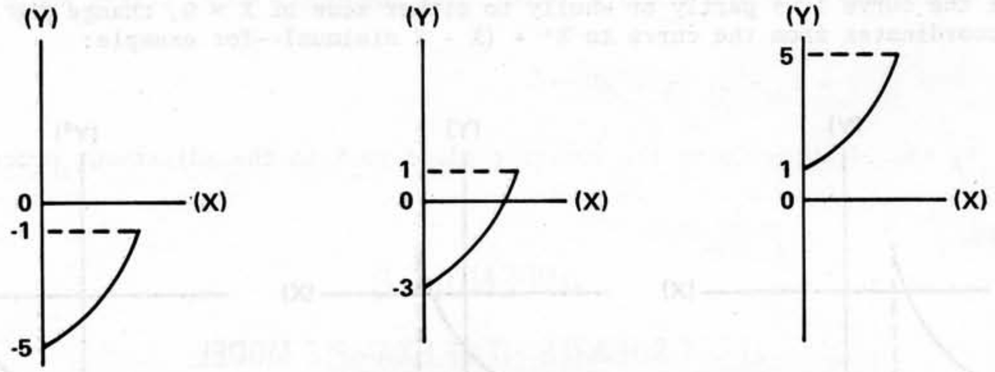
X	X' = X-(-10)
-10	0
0	10
10	20

X	X' = X-(10)
10	0
20	10
30	20

Plotted over X', all three curves would now appear as follows:



If the curve lies either partly or wholly above or below Y = 0, change the Y values of point coordinates from the curve to Y' = (Y - Y minimum)--for example:

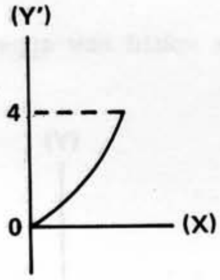


Y	Y' =Y-(-5)
-5	0
-3	2
-1	4

Y	Y' =Y-(-3)
-3	0
-1	2
1	4

Y	Y' =Y-(1)
1	0
3	2
5	4

Plotting Y' instead of Y, all three curves would appear as:



And, last, X' and Y' would be substituted for X and Y, respectively, when looking for a suitable form in the Standards.

In fitting an adopted form to a relevant data set by least squares, always *revert* to the *original* Y_i values--but *retain* any X-transforms in the fitted model. Assume, for instance, that the use of $X' = (X + 30)$ and $Y' = (Y + 3)$ had been necessary to orient the analyst's smoothed curve in figure 1 to the upper right quadrant as shown. The Y-axis would then be labeled Y' and the X axis, X'. All procedures would be identical to those shown for the example except that the final model fitted would be:

$$Y = B_0 + B_1 (X + 30)^{2.7}$$

--and, $\hat{B}_0$ should approximate the constant added to Y in the adjustment process, or, $\hat{B}_0 \doteq 3$.

APPENDIX B

LEAST SQUARES FIT OF EXAMPLE MODEL, HAND CALCULATIONS

Observation number	Original data X	Y	Let $X^{2.7}$ = X'	X'Y	$(X')^2$
1	200	1	0.163229×10^6	0.1632229×10^7	0.2664170×10^{11}
2	300	1	etc.	etc.	etc.
3	370	13			
4	400	24			
5	500	5			
6	530	30			
7	600	18			
8	650	30			
9	660	61			
10	730	60			
11	750	40			
Σ		283	0.2915176×10^9	0.1094696×10^{11}	0.1158486×10^{17}
Mean		25.727	0.2650160×10^8		

Continuing the computations:

$$\Sigma (x')^2 = \Sigma(X')^2 - ((\Sigma X')^2/n) = 0.3859179 \times 10^{16}$$

$$\Sigma (x'y) = \Sigma(X'Y) - (\Sigma X'\Sigma Y)/n = 0.3447002 \times 10^{10}$$

$$\hat{B}_1 = \Sigma(x'y) / \Sigma x^2 = 0.8931957 \times 10^{-6}$$

$$\hat{B}_0 = \overline{Y} - \hat{B}_1 \overline{X}' = 25.72 - \hat{B}_1 (0.2650160 \times 10^8) = 2.06$$

and

$$\hat{Y} = 2.06 + (0.8932 \times 10^{-6}) X^{2.7}$$

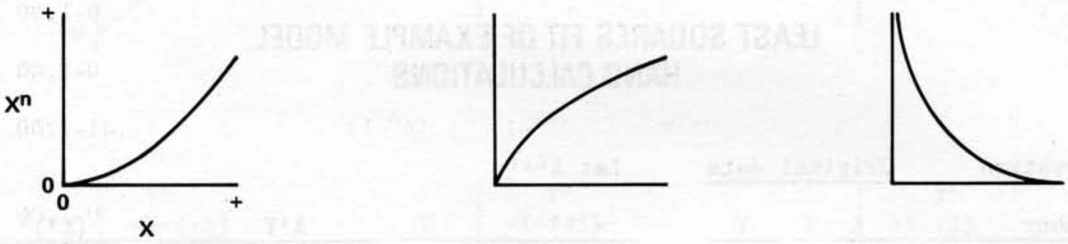
APPENDIX C

DOCUMENTATION OF STANDARDS

Curves of the three exponential sets are oriented in the upper right quadrant as shown below:

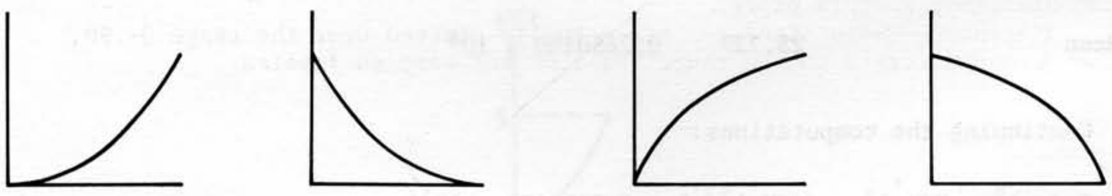
SET

1	2	3
$1.0 \leq n \leq 20.0$	$0.1 \leq n \leq 1.0$	$-2.00 \leq n \leq -0.01$



Each of these basic curve types above is presented in each of the four possible positions for the analyst's curve in the upper right quadrant.

A B C D



When the inherent exponential curve position differed from A, B, C, or D, it was necessary to reorient the exponential curve by reversing X-axes and/or rotating scaled X-transforms about Y_p.

The net result was the series of transforms scaled to 1.0 in both X and Y as indicated below:

Position	Set	Plotted form*	Plotting range in X
A	1	$Y = X^n$	0-1.00
	2	$Y = 1-(1-X)^n$	0-1.00
	3	$Y = [(.01^n-1)^{-1}]\ [(1-X)^n-1]$	**0- .99
B	1	$Y = (1-X)^n$	0-1.00
	2	$Y = 1-X^n$	0-1.00
	3	$Y - [(.01^n-1)^{-1}]\ (X^n-1)$	.01-1.00
C	1	$Y = 1-(1-X)^n$	0-1.00
	2	$Y = X^n$	0-1.00
	3	$Y = 1-[(.01^n-1)^{-1}]\ (X^n-1)$	.01-1.00
D	1	$Y = 1-X^n$	0-1.00
	2	$Y = (1-X)^n$	0-1.00
	3	$Y = 1-[(.01^n-1)^{-1}]\ [(1-X)^n-1]$	**0- .99

*These forms were submitted to a computer-controlled plotter to produce the Standards in this paper.

**Although these Standards are actually plotted over the range 0-.99, they are applicable to the range .01-1.00 and were so labeled.