MATCHACURVE-1 FOR ALGEBRAIC TRANSFORMS TO DESCRIBE SIGMOID- OR BELL-SHAPED CURVES

Chester E. Jensen and Jack W. Homeyer

INTERMOUNTAIN FOREST AND RANGE EXPERIMENT STATION
Ogden, Utah 84401
1970

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.

USDA Forest Service
Intermountain Forest and Range Experiment Station
1970

MATCHACURVE-1 FOR ALGEBRAIC TRANSFORMS TO DESCRIBE SIGMOID - OR BELL - SHAPED CURVES

Chester E. Jensen and Jack W. Homeyer

INTERMOUNTAIN FOREST AND RANGE EXPERIMENT STATION
Forest Service
U.S. Department of Agriculture
Ogden, Utah 84401
Joseph F. Pechanec, Director

CONTENTS

	Page
INTRODUCTION	1
AN EXAMPLE	1
ALTERNATIVE ARRANGEMENTS OF AN ANALYST'S CURVES IN TWO-DIMENSIONAL SPACE	6
MISCELLANEOUS NOTES ON MATCHACURVE	8
STANDARDS.	11

ABSTRACT

Graphed curves of the sigmoid- or bell-shaped classes can be scaled to and compared with the wide array of standard curves presented in this paper. If a standard can be found that suitably matches the graphed curve, the algebraic form specified for the standard can be fitted to any relevant data set by least squares. The algebraic forms utilized are all ratios of exponential functions.

INTRODUCTION

The analyst often graphically initiates a two-dimensional regression analysis, hand-fitting the expected form of an XY relation through a set of plotted data points.

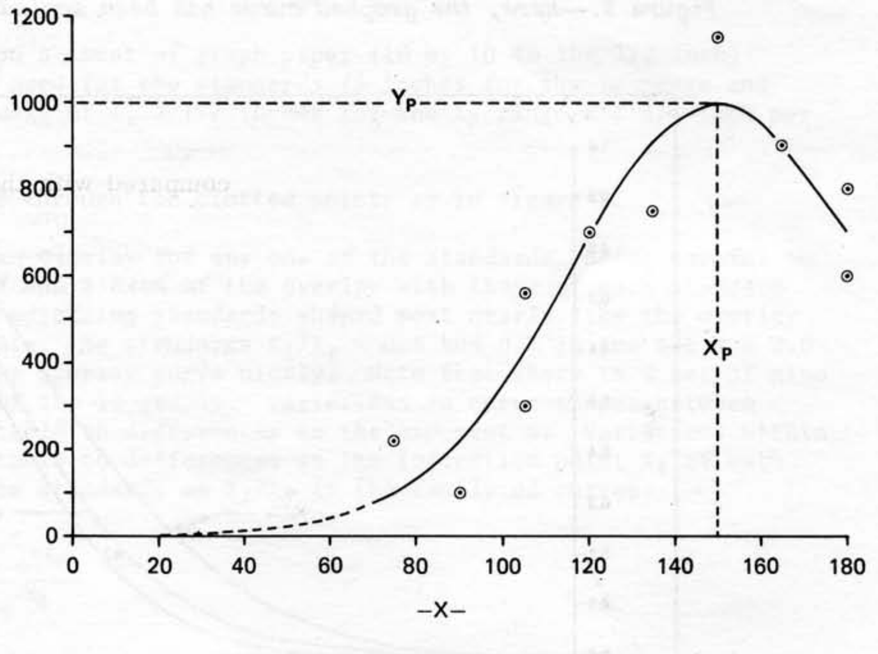
Assume he wants next to find an algebraic transform[1] of the independent variable X, such that transformed X values fitted to their corresponding Y values by least squares emulate the graphed form with acceptable accuracy. If the graphed form is linear, no search is required since the appropriate form is X itself. For curvilinear forms, however, this effort can be difficult and time consuming.

In this paper we have attempted to reduce the search effort required to find acceptable curves of the sigmoid- and bell-shaped classes. The analyst simply matches a scaled version of his graphed curve with graphed standards or curves (pages 11-22), and selects the two adjacent ones most like his own in shape. Interpolation between the transforms given for each standard results in identification of the best alternative offered by the system. It is this transform that finally is fitted to the data set by least squares.

AN EXAMPLE

Let's say that an analyst wants to generate a least squares-fit estimating 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 sigmoid upward with larger X values until a peak is reached, then sigmoid downward. The apparent data trend supports his expectation and he hand-fits such a curve through the data as shown in figure 1.

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



---
[1] Defined here as any nonlinear mathematical form of a variable.

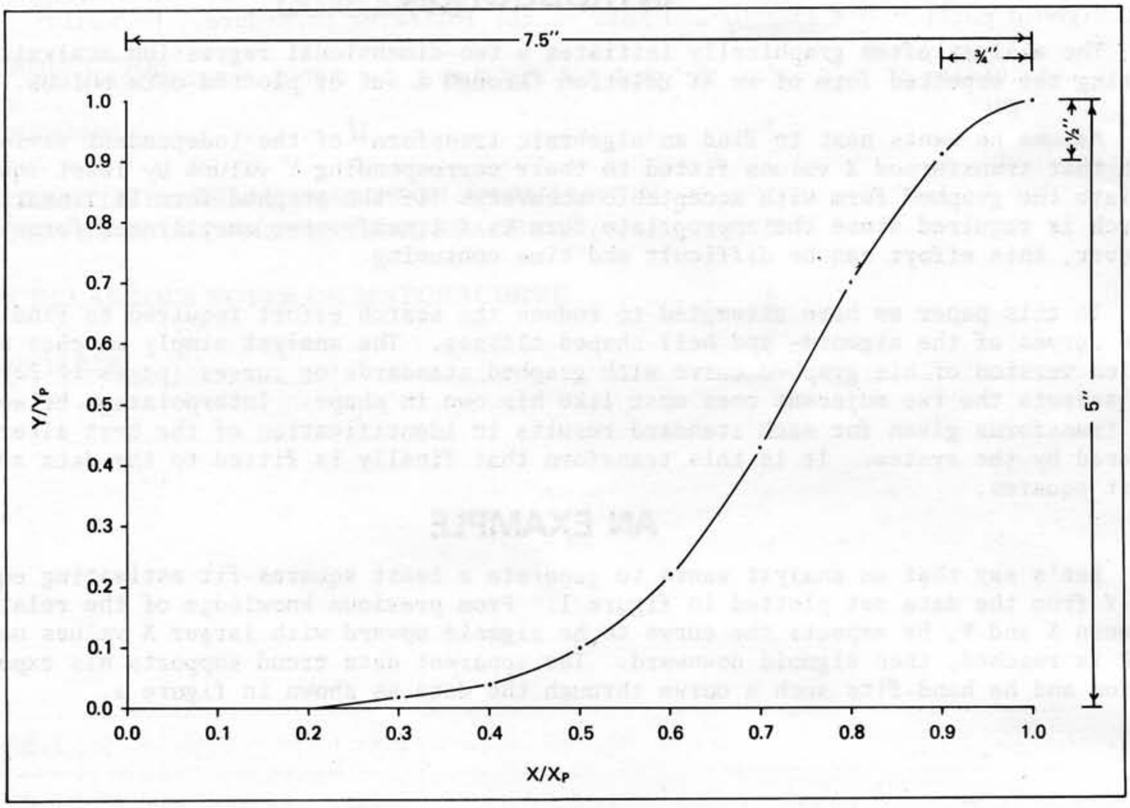


Figure 2.--Here, the graphed curve has been scaled to the standards.

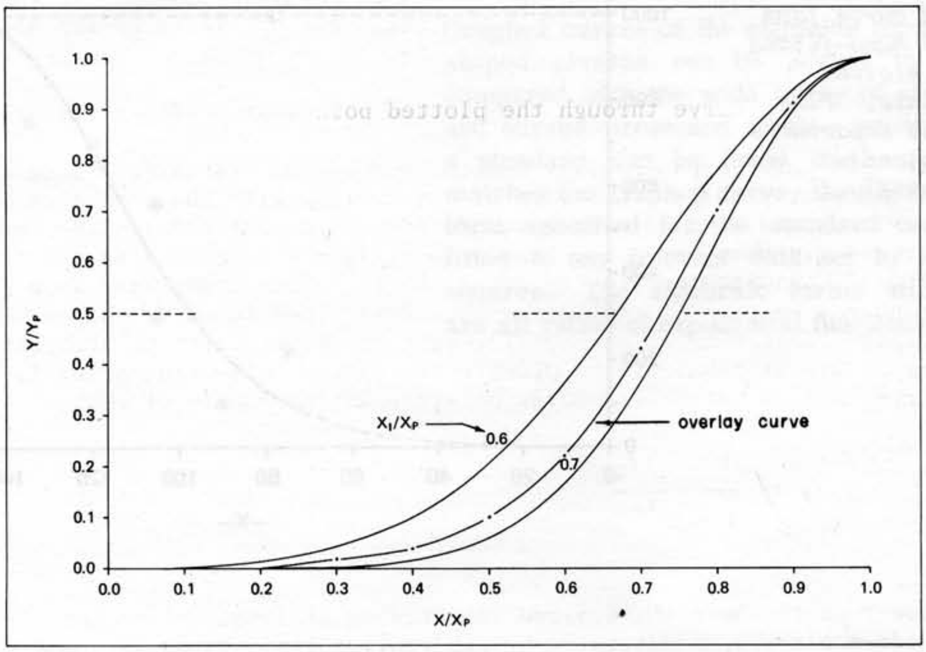


Figure 3.--The standards best matched to the original curve bracket the overlay curve.

Next, to find a suitable X transform to describe this curve, he would work with the left sigmoid portion ($0 \le X \le 150$) and hold to the following procedure. He would:

1.--Determine XY values at the peak of the curve, X_P and Y_P respectively in figure 1.

2.--Divide X_P into 10 equal parts[2] as shown below and determine from the curve the Y value at each of the 11 resulting X values.

Original data

X	0	15	30	45	60	75	90	105	120	135	150
Y	0	0	0	15	40	100	220	430	700	910	1,000

3.--Let the array of X values be called X_i and the array of Y values, Y_i. Then scale the X_i to X/X_P and the Y_i to Y/Y_P.

Scaled data

X/150	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
Y/1,000	0.00	0.00	0.00	0.02	0.04	0.10	0.22	0.43	0.70	0.91	1.00

4.--Plot these points on a sheet of graph paper (10 by 10 to the 1/2 inch) at the same scale[3] used for the standards (5 inches for the Y_P range and 1/2 inch per 1/10 unit of Y, 7-1/2 inches for the X_P range and 3/4 inch per 1/10 unit of X).

5.--Draw a smooth curve through the plotted points as in figure 2.

6.--Use this curve as an overlay for any one of the standards, being careful to match exactly the X and Y axes of the overlay with those of each standard examined, and find adjoining standards shaped most nearly like the overlay curve.[4] In this case, the standards $X_I/X_P = 0.6$ and 0.7 in the set $n = 2.0$ (fig. 3) bracket the overlay curve nicely. Note that there is a set of nine standards on each of the 10 graphs. Variations in curve shapes between graphs are attributable to differences in the exponent n. Variations within graphs are attributable to differences in the inflection point X_I of each curve, scaled to the standards as X_I/X_P in the family of curves:

$$Y/Y_P = \frac{e^{-T} - e^{-T_0}}{1 - e^{-T_0}}$$

[2] Use fewer parts where the desired sensitivity is less.
[3] This odd scale was necessitated by publication limitations in paper size.
[4] A light source behind the graphs to be compared will assist in the matching process.

where:

e = natural log base, 2.71828

$$T = \left| \frac{(X/X_P) - 1}{(X_I/X_P) - 1} \right|^n$$

T_0 = T, evaluated at X = 0; then $(X/X_P) = 0$

X/X_P = the scaled X-values, 0 to 1.0

X_I/X_P = the scaled inflection point when $0 < X_I/X_P < 1$

Having found an acceptable pair of adjoining standards that bracket the overlay curve, the analyst can interpolate between the scaled inflection points represented by the two and arrive at the final X transform. He can use proportional horizontal departure of the overlay curve from the left bracketing standard at $Y/Y_P = 0.5$ as the basis for interpolation. In figure 3, the overlay curve lies at about the 70 percent point between the scaled inflection points 0.6 and 0.7; so the interpolated $X_I/X_P = 0.67$.[5]

Then:

$$T = \left| \frac{(X/150) - 1}{0.67 - 1} \right|^{2.0}$$

and the X transform finally adopted will be

$$Y/Y_P = \frac{e^{-\left|\frac{(X/150) - 1}{0.33}\right|^{2.0}} - e^{-\left|\frac{0 - 1}{0.33}\right|^{2.0}}}{1 - e^{-\left|\frac{0 - 1}{0.33}\right|^{2.0}}}$$

This can be simplified to:

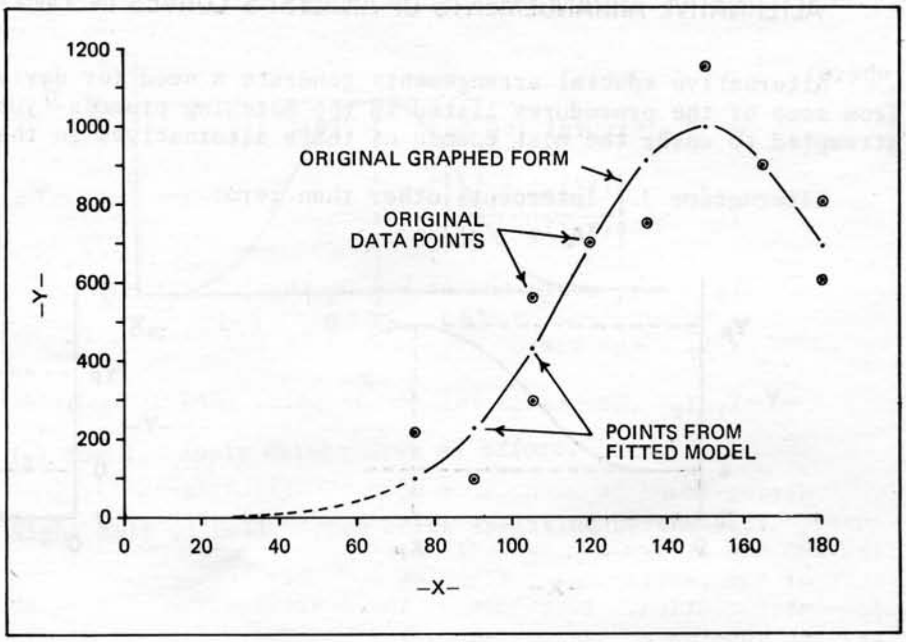
$$Y/Y_P = \frac{e^{-\frac{|(X/150) - 1|^{2.0}}{0.1089}} - e^{-9.18}}{1 - e^{-9.18}}$$

But, from standard tables of e^{-x}, $e^{-9.18} < 0.0002$ and, *in this case*, the expression can be further simplified to:

$$Y/Y_P = e^{-\frac{|(X/150) - 1|^{2.0}}{0.1089}}$$

[5] This same interpolating principle can be used *between* standards of different sets for refined n estimates, but this operation is not recommended for general application.

Figure 4.--Points from the fitted model have been plotted over the original graphed form.



Since Y/Y_P is only a scaling change for Y, we can substitute Y for Y/Y_P in the final sample linear model to be fitted by least squares as shown immediately below.

$$Y = \beta_0 + \beta_1 e^{-\frac{|(X/150) - 1|^{2.0}}{0.1089}}$$

Then the estimated β_0, β_1 ($\hat{\beta}_0$, $\hat{\beta}_1$), will adjust for the scalar differences.

With the aid of a desk calculator and e^{-x} or logarithm tables, the X transform required for each X value in the data set can be obtained and the least squares fitting process applied as usual.[e] It is preferable, however, if not necessary, to use a computer to arrive at the fitted model, in this case

$$\hat{Y} = -2.05 + 998.7\, e^{-\frac{|(X/150) - 1|^{2.0}}{0.1089}}$$

Points from the fitted model are compared to the original graphed form in figure 4.

It is clear that plotted points from the fitted model are an excellent match for the original graphed curve form within the data range of X.

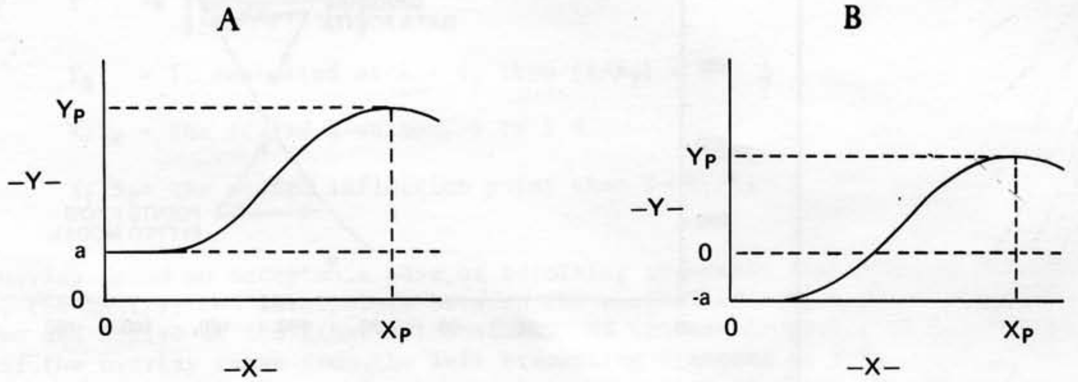
---

[e] The hand calculations for this fitting process, although not recommended, are shown on pages 9 and 10.

ALTERNATIVE ARRANGEMENTS OF ANALYST'S CURVES IN TWO - DIMENSIONAL SPACE

Alternative spacial arrangements generate a need for deviations--generally minor--from some of the procedures listed in the matching process[7] just presented. We have attempted to cover the most common of these alternatives in the discussion that follows.

Alternative 1. Intercepts other than zero:



Here, the sigmoid curvature never drops below the intercept *a*, regardless of whether *a* is positive or negative. To apply Matchacurve, substitute (Y-*a*) for Y.

Using the data from figure 1 with an intercept = +20, the curve would be as pictured in A of Alternative 1, and the actual Y values would be as tabulated below.

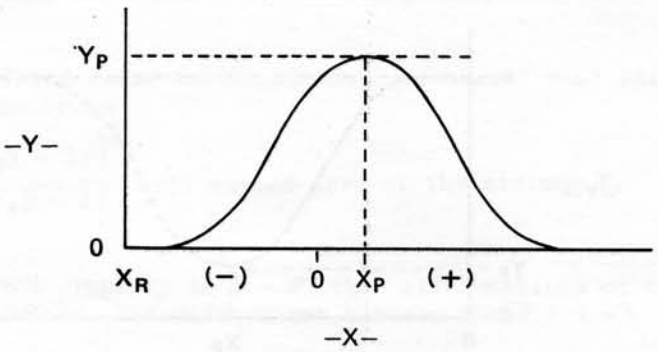
X	0	15	30	45	60	75	90	105	120	135	150
Y	20	20	20	35	60	120	240	450	720	930	1,020
Y-*a*	0	0	0	15	40	100	220	430	700	910	1,000

The X_i and associated (Y-*a*) values would be those used to identify the mathematical form in Matchacurve using the procedures already described. Having identified an appropriate X transform, the analyst fits it through the original Y values by least squares as before, then the *estimated* intercept, $\hat{\beta}_0$, should be close to 20.

Given an intercept = -20, *add* 20 to each Y value and proceed with Matchacurve as before (see B of Alternative 1).

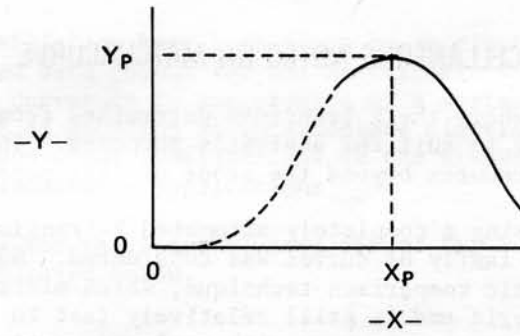
[7] This process will be referred to henceforth as "Matchacurve."

Alternative 2. X = 0 included in the range of X when Y >0:



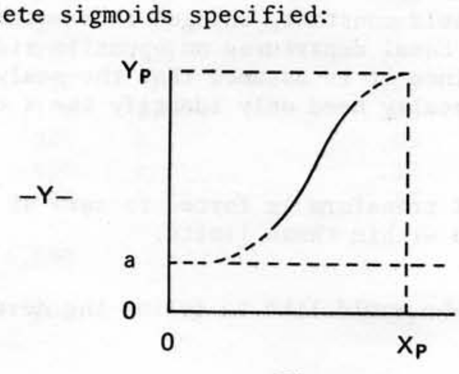
Substitute $(X - X_R)$ for X. Apply Matchacurve as before.

Alternative 3. Right half of bell-shaped curve specified by analyst:



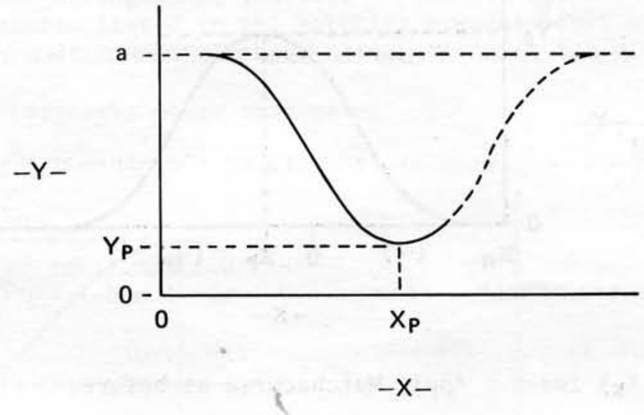
Plot the mirror image of the right side on the left as shown above and apply Matchacurve as before.

Alternative 4. Incomplete sigmoids specified:



If X_P, Y_P, and the intercept a are not present in the curve specified by the analyst, they must be estimated. Then, Matchacurve can be applied as before.

Alternative 5. Inverted bell-shaped curve, or portion thereof:



Plot the negative departures of the curve from the intercept as positive departures on the intercept, i.e., rotate the curve upward about *a*. Apply Matchacurve as before, fitting the X transform selected to the original Y values. $\hat{\beta}_0$ should be close to *a* and $\hat{\beta}_1$ should be negative.

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.

- 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.

- The portion of the curve to the *right* of X_P is a mirror image of the left in the Matchacurve transforms. Once *n* and the scaled inflection point X_I/X_P have been selected and held constant, changes in Y values depend only on departure of X/X_P from 1.0. Equal departures on opposite sides of 1.0 will result in equal Y values. Since it is assumed that the analyst's curve is bell-shaped (symmetrical), we really need only identify the X transform for one side (left), as in figure 1.

- Every Matchacurve X transform is forced to zero at $X = 0$ and at $X = 2X_P$ and should only be used within these limits.

- For those readers who would like to follow the development of the family of curves presented:

  a. At $X = 0$, e^{-x} has the value 1 and drops sigmoidally with increasing X values, eventually approaching zero and becoming asymptotic to the X axis.

b. For conceptual convenience, the right-hand, mirror image of e^{-x}, $1 - e^{-x}$, is next adopted. This function then is zero at X = 0 and increases sigmoidally approaching the value 1 and becoming asymptotic to 1 at larger values of X.

c. But with X_P fixed close to the minimum X (where $Y = a$) and $0 < X < X_P$, some $1 - e^{-T}$ values where

$$T = \left| \frac{(X/X_P) - 1}{(X_I/X_P) - 1} \right|^n$$ will exceed zero at the minimum X.

d. To maintain the property in $1 - e^{-T}$ that all functions of this family pass through the origin, all differences between 1 and $1 - e^{-T}$ are expanded to

$$\frac{1}{1 - e^{-T_0}}$$ and finally are subtracted from 1, or

$$1 - \left\{ \frac{1}{1 - e^{-T_0}} \right\} (1 - e^{-T}) = \frac{e^{-T} - e^{-T_0}}{1 - e^{-T_0}}$$

- The standards that follow have been drawn by an electronic plotter operating on computer-generated data points for the Matchacurve X transforms. In a few instances, sharp curvature is represented by a series of short, straight lines and is the result of adopting an intermediate plotting interval, 0.075 inch, for the X scale. These minor imperfections in the standards are judged to be unimportant in Matchacurve applications.

- Hand calculations for the least squares fit of the X transform, e^{-T}, shown on page 4 are tabulated below.

Observation number	Original data X	Original data Y	X/150	$T = \frac{[(X/150) - 1]^2}{0.1089}$ *	$X' = e^{-T}$ **
1	75	220	0.5	2.30	0.100
2	90	100	.6	1.47	.230
3	105	300	.7	.83	.436
4	105	560	.7	.83	.436
5	120	700	.8	.37	.691
6	135	750	.9	.09	.914
7	150	1,150	1.0	.00	1.000
8	165	900	1.1	.09	.914
9	180	800	1.2	.37	.691
10	180	600	1.2	.37	.691
Σ		6,080			6.103
MEAN		608			0.6103

*Rounded to hundredths.
**Values from tables of e^{-x}.

Continuing the computations:

$\Sigma(X')^2 = 4.546327 \qquad \Sigma X'Y = 4529.160$

$\Sigma(x')^2 = \Sigma(X')^2 - ((\Sigma X')^2/n) = 4.546327 - ((6.103)^2/10) = 0.8216661$

$\Sigma x'y = \Sigma X'Y - (\Sigma X' \; \Sigma Y/n) = 4529.160 - ((6080)(6.103)/10) = 818.536$

$\hat{\beta}_1 = 818.536/0.8216661 = 996.19$

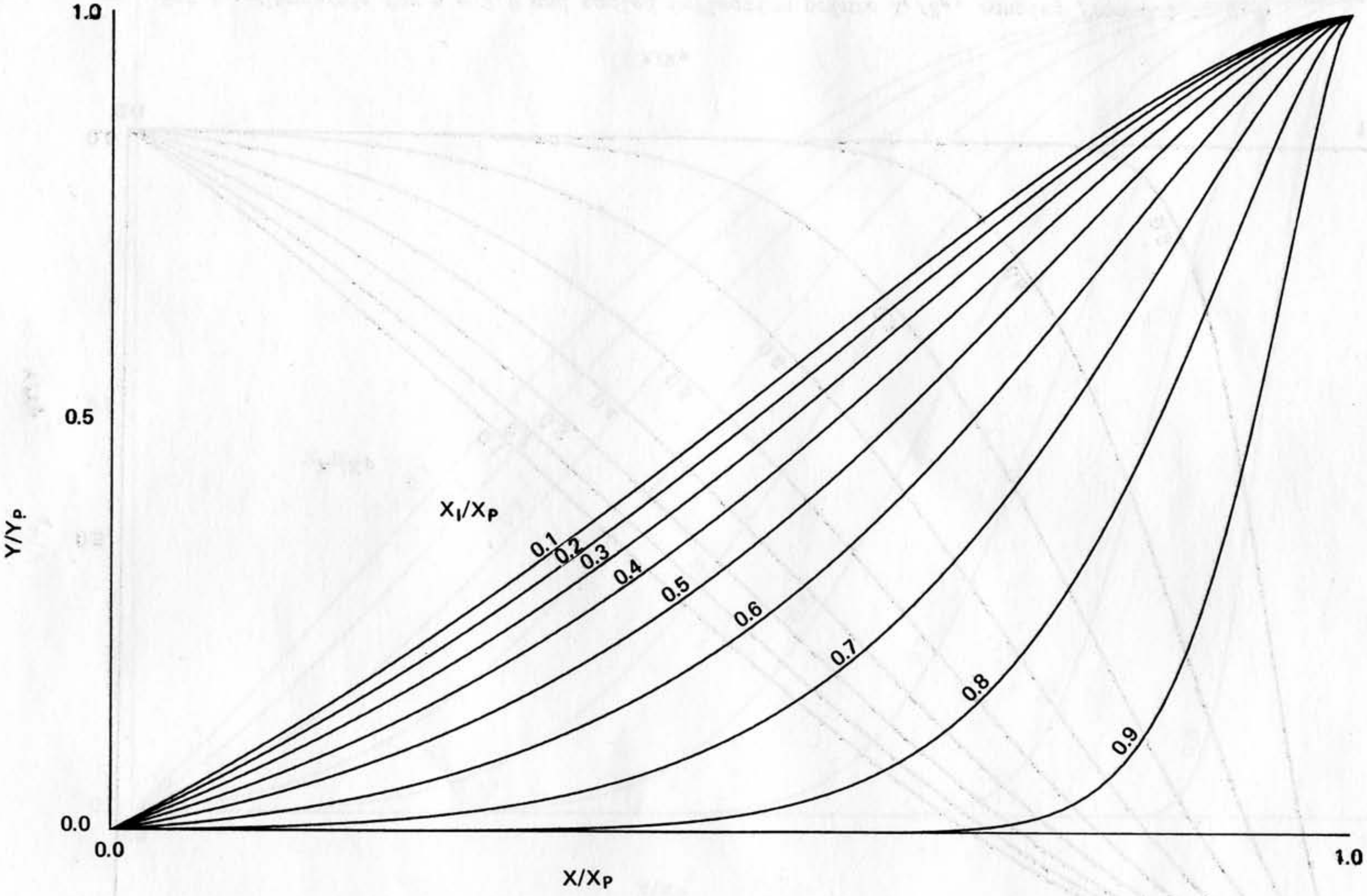
$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}' = 608 - 996.19(0.6103) = 0.0252$

and

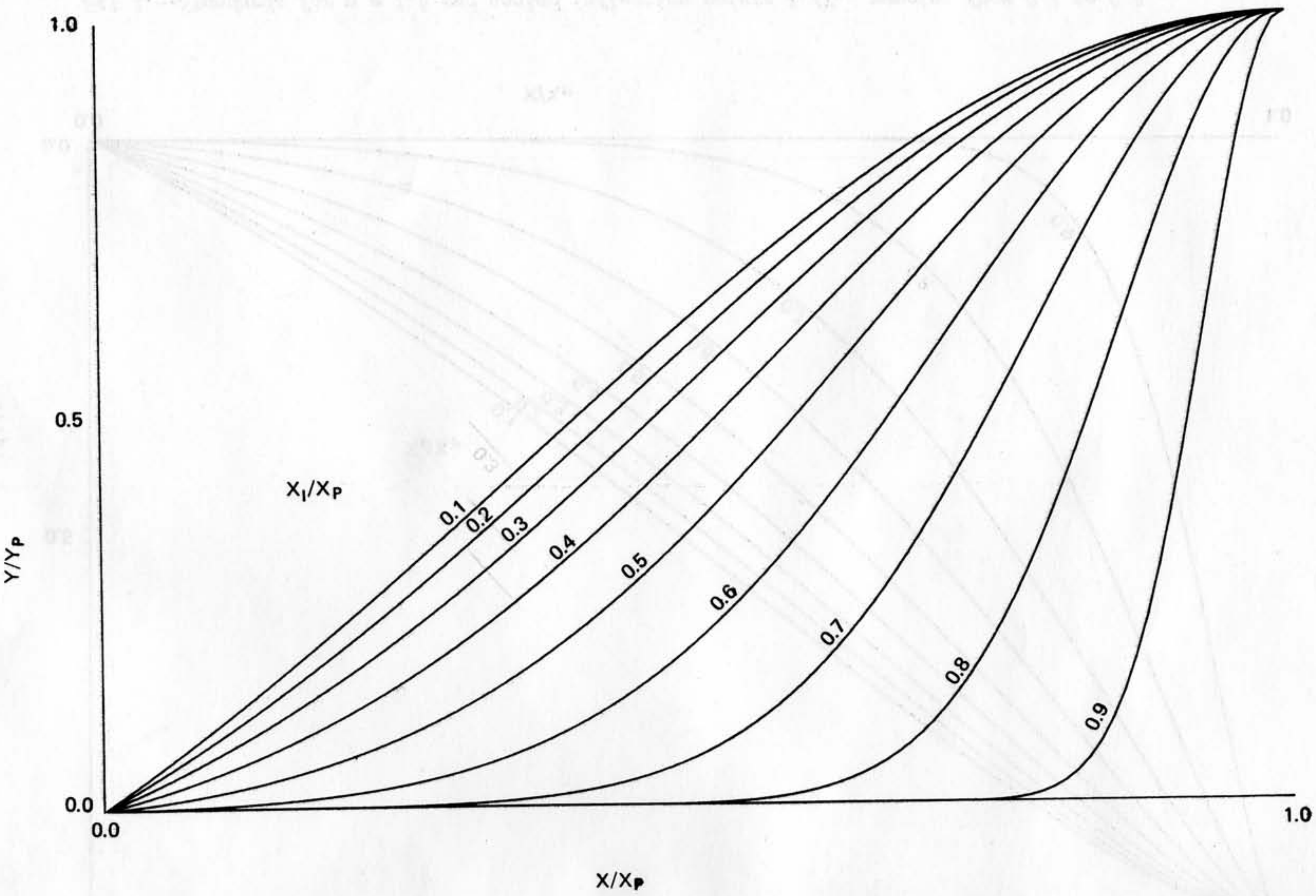
$\hat{Y} = 0.03 + 996.2\, X'$

Note that the desk calculator solution, wherein fewer significant digits were carried, resulted in slightly different constants than the more exacting computer solution given earlier where $\hat{\beta}_0 = -2.05$ and $\hat{\beta}_1 = 998.7$.

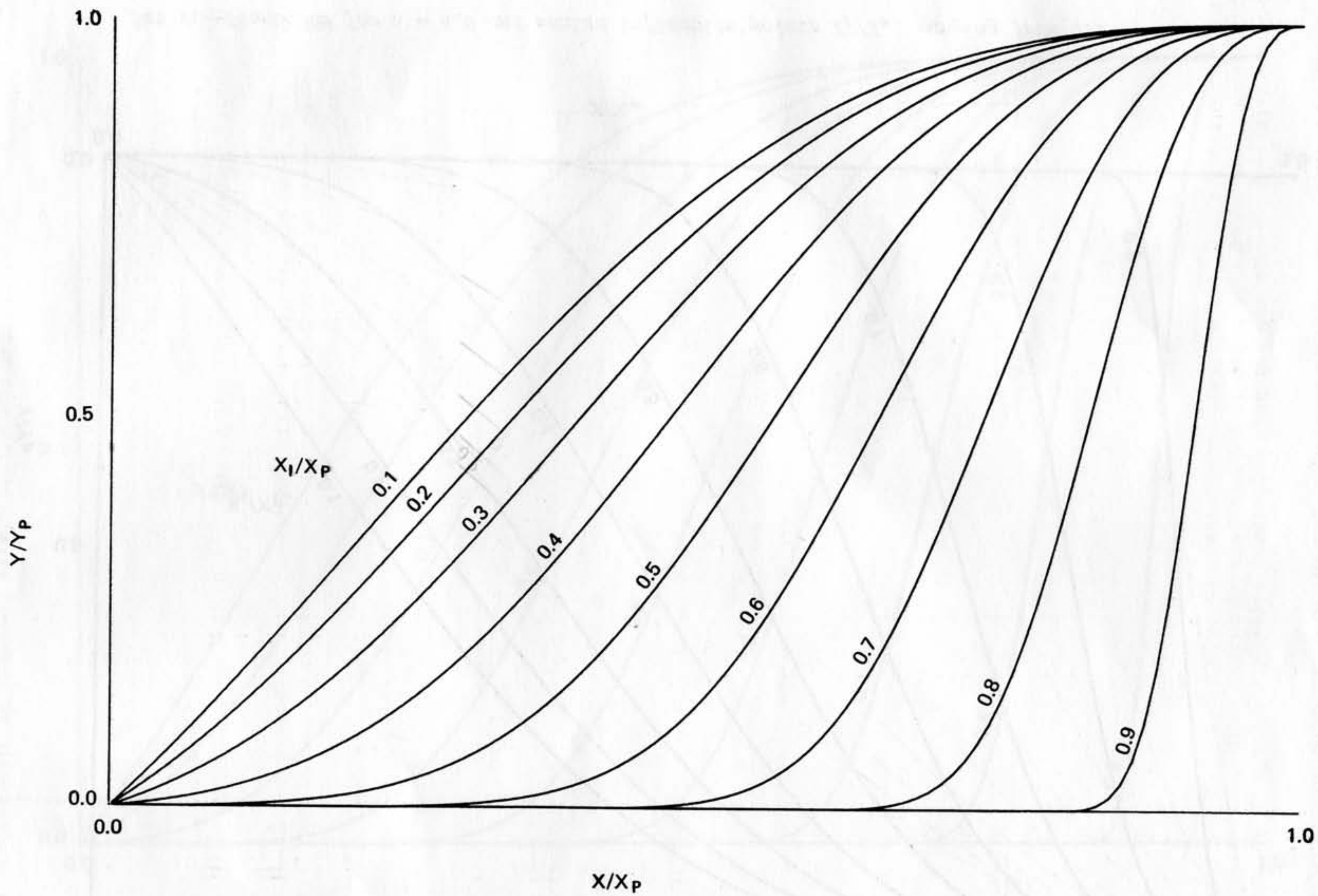
STANDARDS



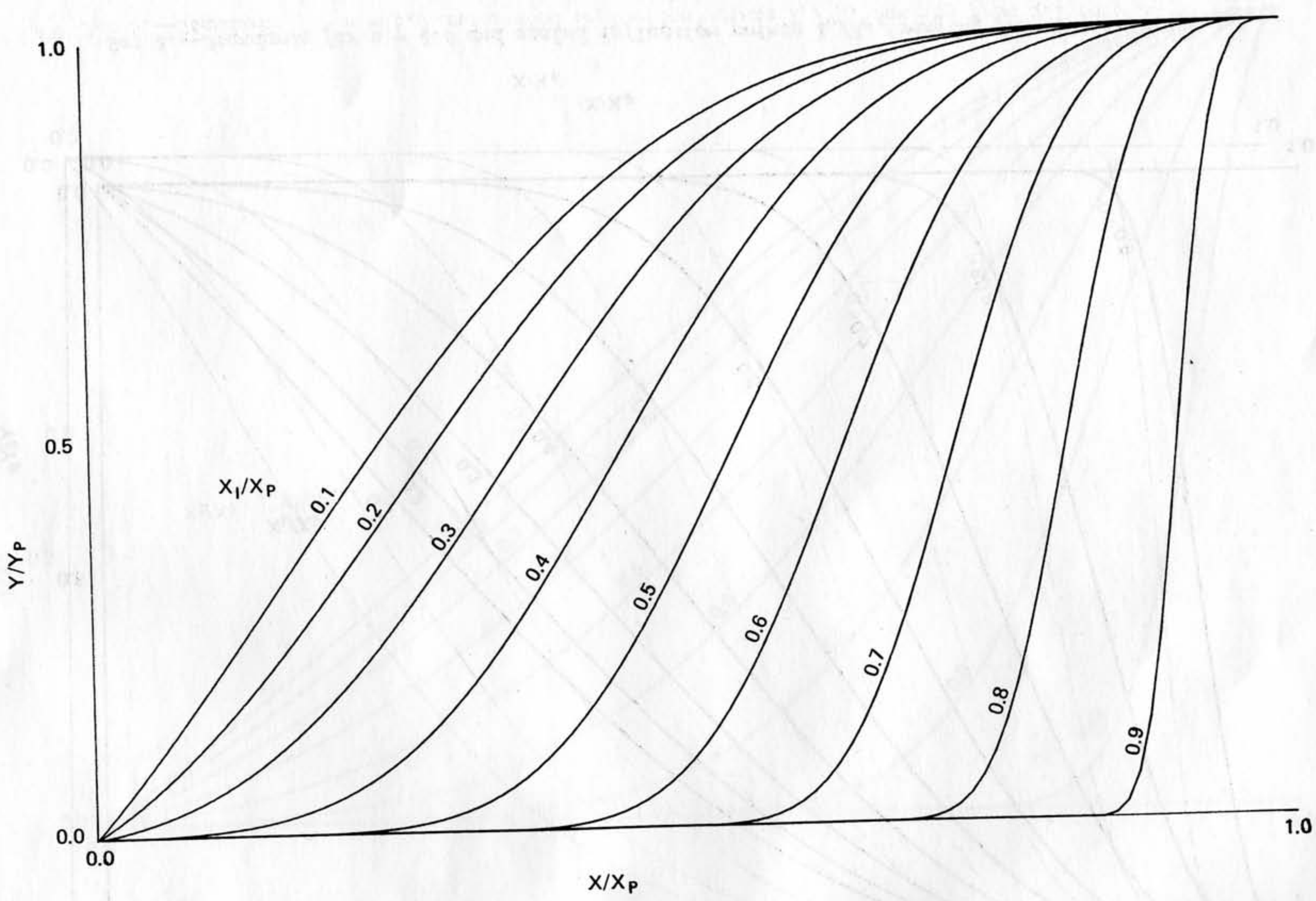
Set 1.--Standards for n = *1.5 and scaled inflection points* X_I/X_P*, ranging from 0.1 to 0.9.*



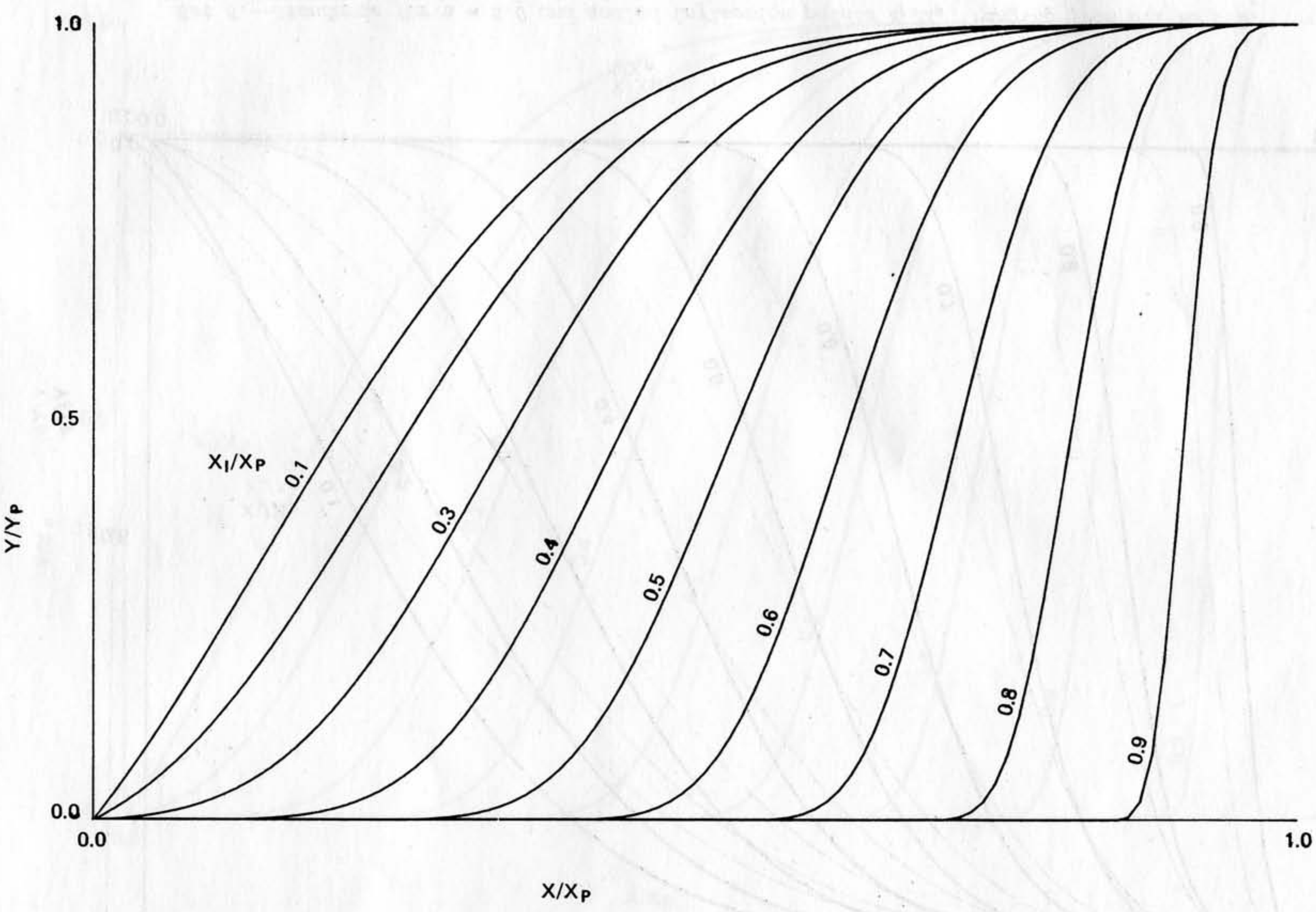
Set 2.--Standards for n = *2.0 and scaled inflection points* X_I/X_P*, ranging from 0.1 to 0.9.*



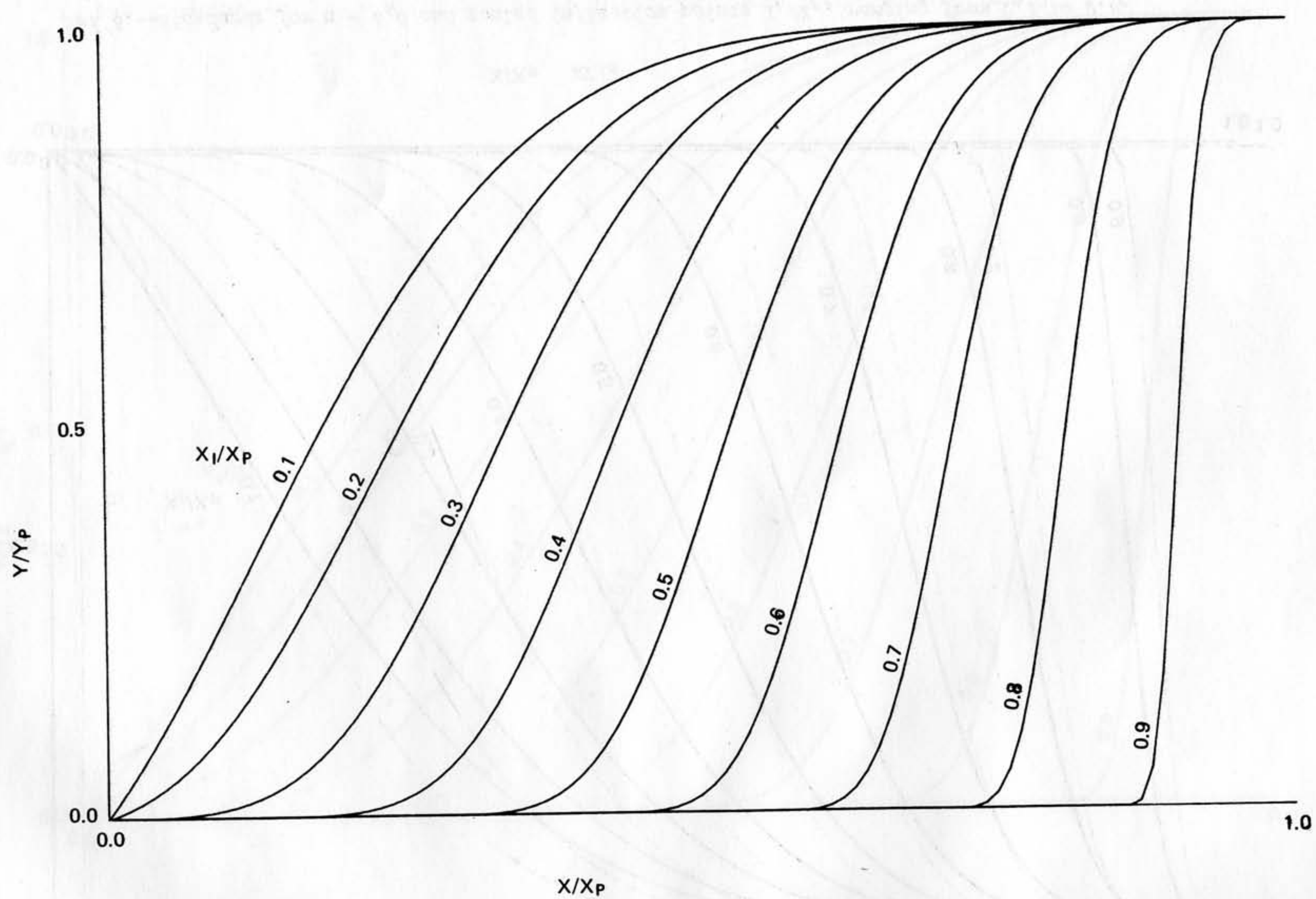
Set 3.--Standards for n = *3.0 and scaled inflection points* X_I/X_P*, ranging from 0.1 to 0.9.*



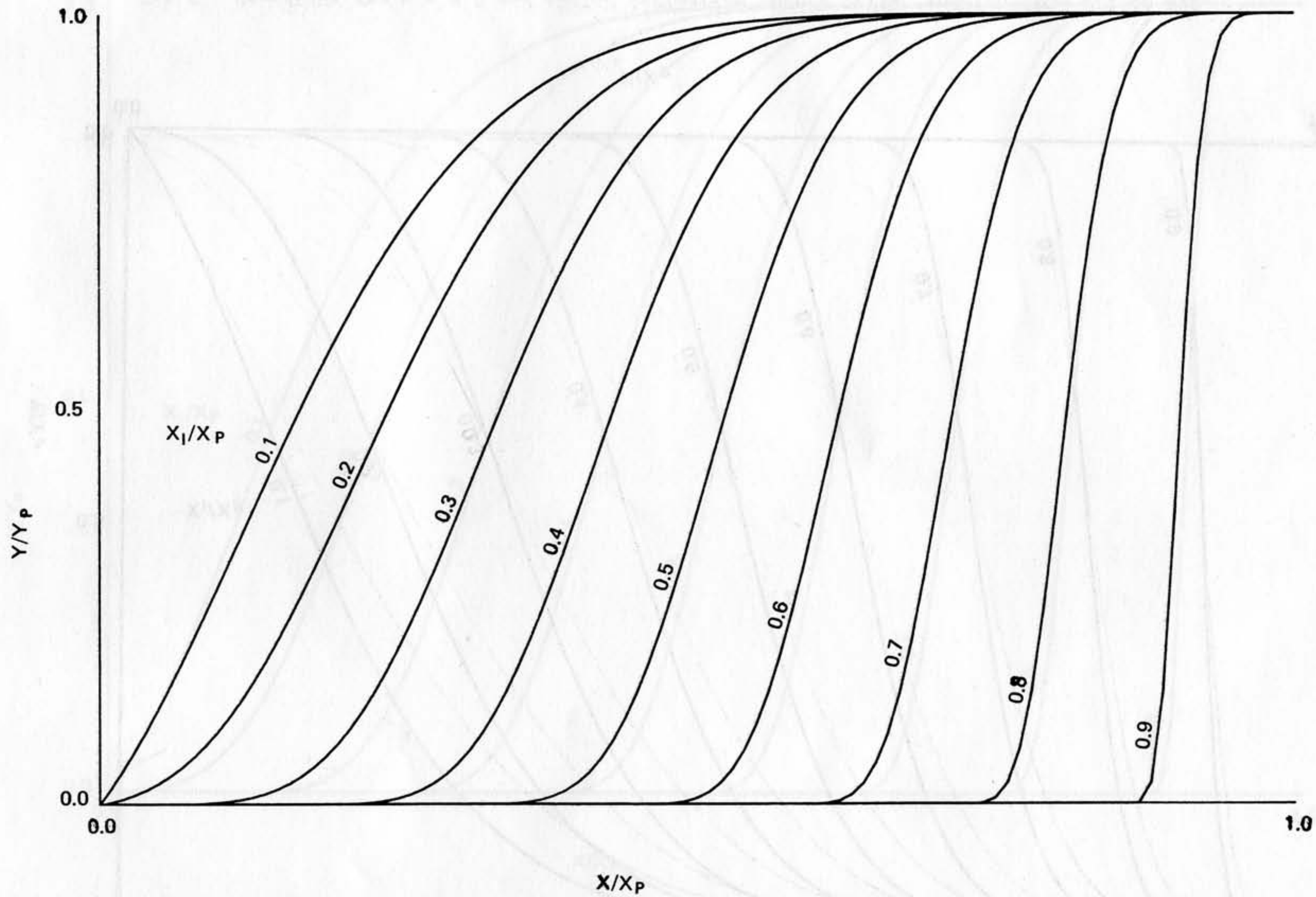
Set 4.--Standards for n = *4.0 and scaled inflection points* X_I/X_P, *ranging from 0.1 to 0.9.*



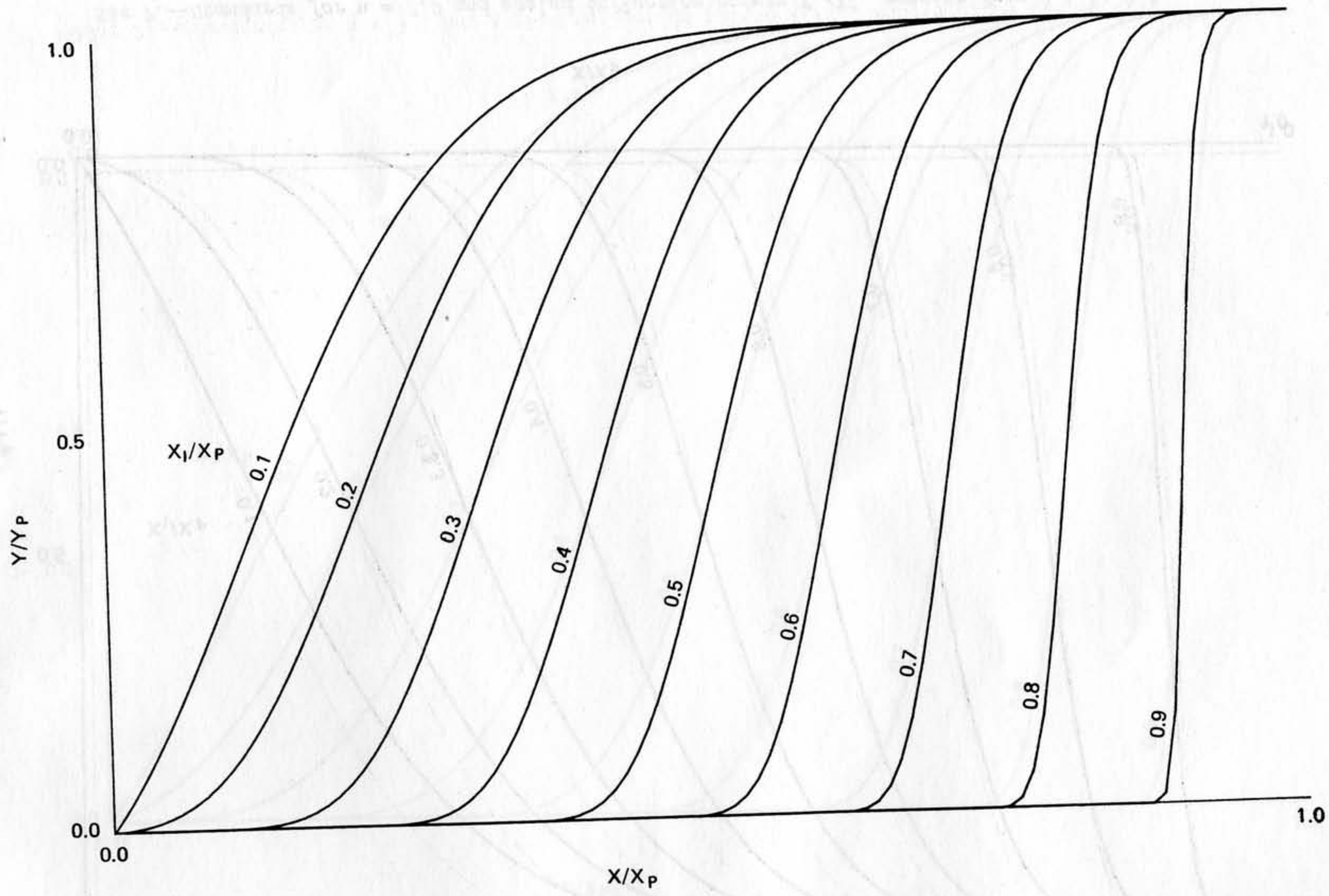
Set 5.--Standards for n = 5.0 and scaled inflection points X_I/X_P, ranging from 0.1 to 0.9.



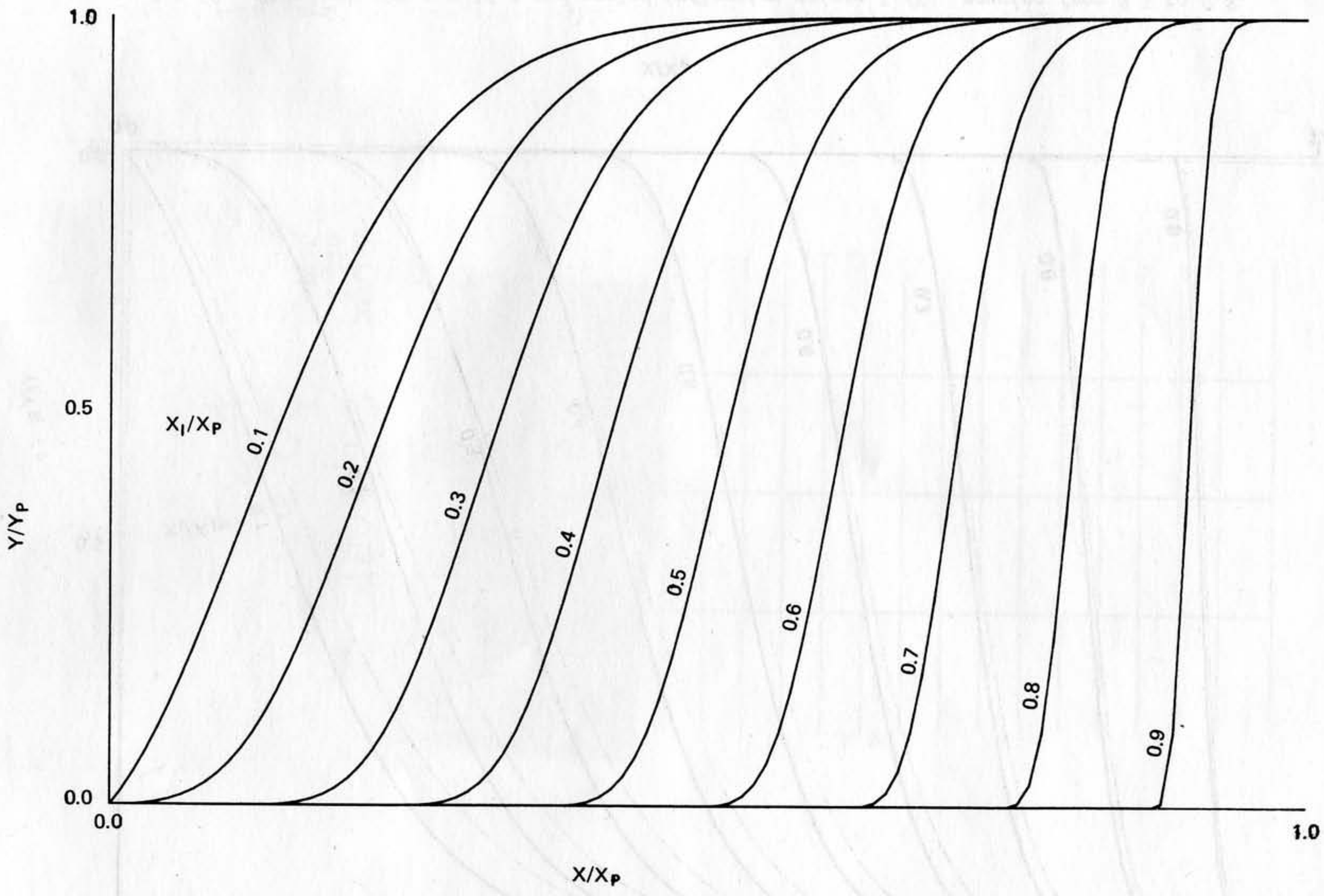
Set 6.--Standards for n = *6.0 and scaled inflection points* X_I/X_P, *ranging from 0.1 to 0.9.*



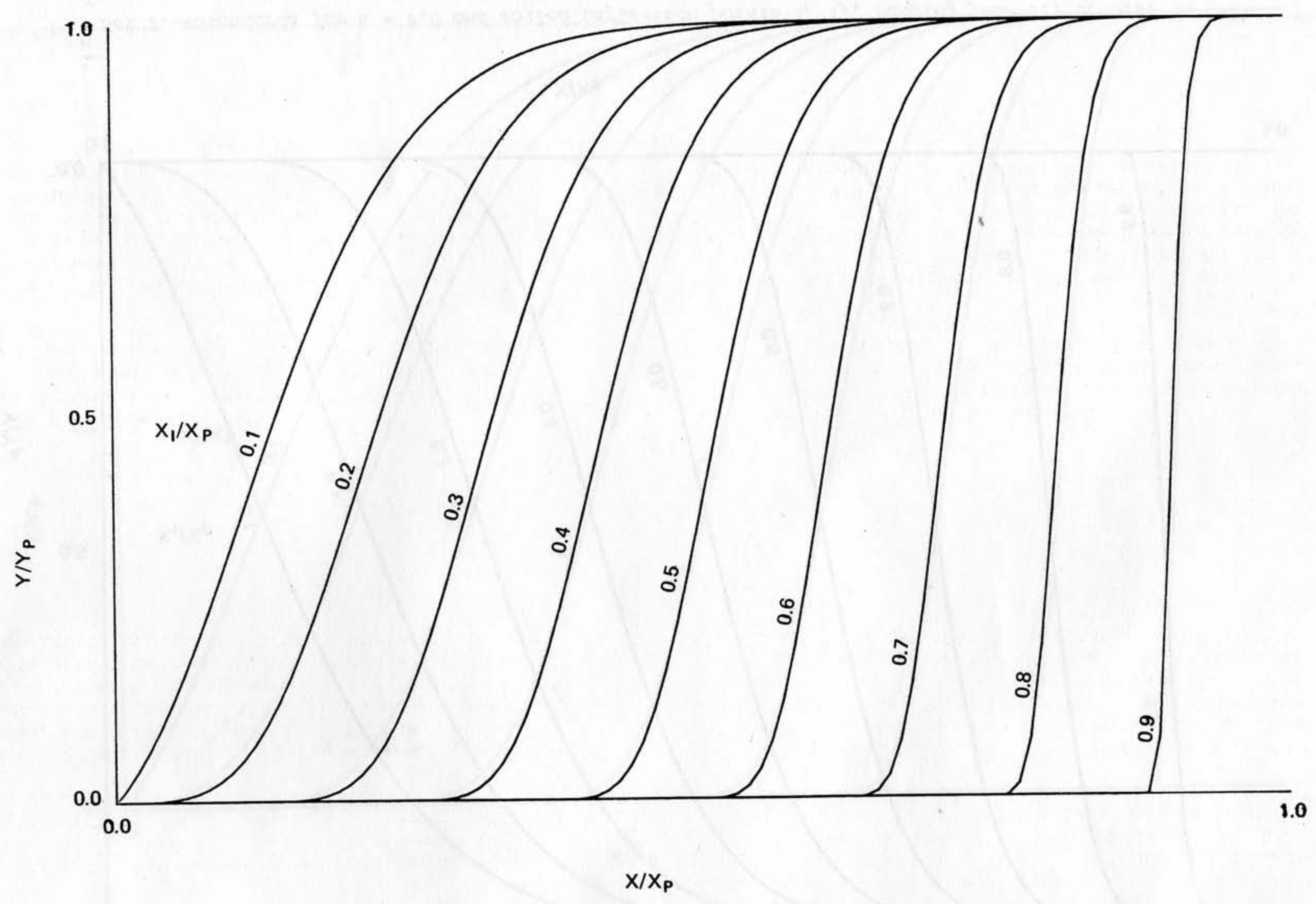
Set 7.--Standards for n *= 7.0 and scaled inflection points* X_I/X_P, *ranging from 0.1 to 0.9.*



Set 8.--Standards for n *= 8.0 and scaled inflection points* X_I/X_P*, ranging from 0.1 to 0.9.*



Set 9.--Standards for n = *9.0 and scaled inflection points* X_I/X_P, *ranging from 0.1 to 0.9.*



Set 10.--Standards for n = *10.0 and scaled inflection points* X_I/X_P, *ranging from 0.1 to 0.9.*

ABOUT THE FOREST SERVICE . . .

As our Nation grows, people expect and need more from their forests—more wood; more water, fish, and wildlife; more recreation and natural beauty; more special forest products and forage. The Forest Service of the U. S. Department of Agriculture helps to fulfill these expectations and needs through three major activities:

- Conducting forest and range research at over 75 locations ranging from Puerto Rico to Alaska to Hawaii.
- Participating with all State Forestry agencies in cooperative programs to protect, improve, and wisely use our Country's 395 million acres of State, local, and private forest lands.
- Managing and protecting the 187-million acre National Forest System.

The Forest Service does this by encouraging use of the new knowledge that research scientists develop; by setting an example in managing, under sustained yield, the National Forests and Grasslands for multiple use purposes; and by cooperating with all States and with private citizens in their efforts to achieve better management, protection, and use of forest resources.

Traditionally, Forest Service people have been active members of the communities and towns in which they live and work. They strive to secure for all, continuous benefits from the Country's forest resources.

For more than 60 years, the Forest Service has been serving the Nation as a leading natural resource conservation agency.