DOCUMENT RESUME

ED 336 048 HE 024 870

AUTHOR	Saupe, Joe L.
TITLE	A Technique for Producing a Double-Entry Expectancy Nomograph from Observed Proportions without Distributional Assumptions. AIR 1991 Annual Forum Paper.
PUB DATE	May 91
NOTE	20p.; Paper presented at the Annual Forum of the Association for Institutional Research (31st, San Francisco, CA, May 26-29, 1991).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Expectancy Tables; Higher Education; Institutional Research; Predictive Measurement; Predictor Variables; *Sample Size; *Statistical Analysis; Statistical Inference
IDENTIFIERS	*AIR Forum; *Nomographs; University of Missouri Columbia

ABSTRACT

This paper describes a procedure for smoothing the proportions of a double-entry expectancy table which might be used in higher education admissions advising or other functions. The product of the smoothing procedure is a nomograph, a medium for displaying expectancies which provides pairs of individual values for two predictor variables rather than pairs of ranges, as in the usual expectancy table. The paper demonstrates the procedure and the resultant nomograph, using the high school class percentile ranks, achievement test composite scores, and freshman year grade point averages of first-time freshmen of five consective entering classes at the University of Missouri, Columbia. The number of students available for developing the nomograph was 12,835, a large number. If it could be shown that the nomograph solution was satisfactory for that sample size, the question, "For what smaller sample sizes would it produce acceptable results?" could be raised. In a further step the study investigated the effects of varying sample size and minimum group size and found that samples smaller that 12,000 produce satisfactory expectancy nomographs but that the smallest sample possible may lie between 1,000 and 3,000. However, altering the minimum group size had little effect on the nomograph curves. The study concludes that valid expectancy nomographs can be produced from large samples using minimum group sizes of 50 and that these nomographs may be easier to use than the expectancy tables for which they are intended to substitute. Two figures, five tables and four references are included. (JB)

A TECHNIQUE FOR PRODUCING A DOUBLE-ENTRY EXPECTANCY NOMOGRAPH FROM OBSERVED PROPORTIONS WITHOUT DISTRIBUTIONAL ASSUMPTIONS

Joe L. Saupe
Professor, Higher and Adult Educations and Foundations
University of Missouri-Columbia

301 Hill Hall
University of Missouri-Columbia
Columbia, MO 65211

(314) 882-5123

Paper presented at the Thirty-First Annual Forum,
The Association for Institutional Research
The Westin St. Francis, San Francisco, California
May 26-29, 1991.

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-First Annual Forum of the Association for Institutional Research held at The Westin St. Francis, San Francisco, California, May 26-29, 1991. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum Papers.

Jean Endo
Chair and Editor
Forum Publications Editorial
Advisory Committee

Abstract

A TECHNIQUE FOR PRODUCING A DOUBLE-ENTRY EXPECTANCY NOMOGRAPH FROM OBSERVED PROPORTIONS WITHOUT DISTRIBUTIONAL ASSUMPTIONS

A procedure for smoothing the proportions of a double-entry expectancy table is described. The product of the procedure is a nomograph from which can be read expectancies from combinations of values of two predictor variables. The nomograph might be used in admissions advising or in establishing standards for the admission of freshman students. The procedure is used to construct nomographs for predicting proportions of freshman year grade point averages ≥ 2.0 and for proportions ≥ 3.0 from high school class percentile ranks and ACT Composite scores for a sample of first-time freshmen. Effects of sample size and of the minimum size of groups of students used in estimating nomograph curves on the stability of the curves are examined. Suggestions for additional work on deriving expectancy nomographs are given.

A TECHNIQUE FOR PRODUCING A DOUBLE-ENTRY EXPECTANCY NOMOGRAPH FROM OBSERVED PROPORTIONS WITHOUT DISTRIBUTIONAL ASSUMPTIONS[1]

An expectancy table is composed of proportions or probabilities which portray the predictive relationship between one or more, rarely more than two, predictor variables and a criterion variable (Schrader, 1965). Single-entry tables involve one predictor variable and double-entry tables are based upon two predictor variables. Expectancy tables may be constructed on the basis of distributional assumptions or on the basis of observed frequencies without distributional assumptions (Schrader, 1965; Morgan, 1988).

A common type of expectancy table displays the relationship between one or more predictors of success in college, e.g., high school class percentile rank and admissions test score, and a measure of success in college, e.g., first-term or first-year grade point average. Such tables may be used (a) in counseling prospective or newly admitted students, (b) in interpreting the relationship between the predictor and criterion variables and (c) in setting sliding-scale admission standards.

Expectancy tables constructed using distributional assumptions, e.g., bivariate or multivariate normality, have the advantage that the series of proportions are smoothed and can be extrapolated beyond regions in which appreciable numbers of observations fall (Morgan, 1988). Perrin and Whitney (1976) have shown that smoothing enhances the validity of expectancy table values. Of course, if the distributional assumptions are wrong, the expectancy values are likely to be invalid. Tables built directly from observed frequencies are likely to include irregularities (reversals) in the series of proportions. On the other hand, tables developed directly from observed frequencies are not dependent on distributional assumptions.

Normally, directly-derived expectancy tables cannot be smoothed without invoking distributional assumptions. Isotonic procedures can be used to remove reversals (Perrin and Whitney, 1976), but these procedures do not enable extrapolation. The purposes of this paper are (a) to introduce a procedure for smoothing the directly-derived proportions of a double-entry expectancy table without imposing distributional assumptions and (b) to investigate the stability of the smoothed expectancies under several conditions. The product of the smoothing procedure is a nomograph from which can be read expectancies. The nomograph is a "user-friendly" medium for displaying expectancies and it provides for reading expectancies from pairs of individual values of the two predictor variables rather than for pairs of ranges, as in the usual expectancy table.

The Data

The smoothing procedure is illustrated using the high school class percentile ranks (HSCPR), ACT Composite scores (ACT) and freshman year grade point averages (GPA) of the first-time freshmen of five consecutive entering classes for a large mid-western university. There were 12,835 students in the five classes who had complete data on the three variables. Table 1 is the directly-derived expectancy table showing proportions of these students, classified by HSCPR and ACT score ranges, who earned a GPA of at least 2.0.

Table 1. Unsmoothed Expectancy Table, Proportions of Students Whose GPA ≥ 2.0 (N = 12,835)

ACT SCORE RANGE	HIGH SCHOOL CLASS PERCENTILE RANK RANGE							TOTAL
	0-39	40-49	50-59	60-69	70-79	80-89	90-99	
29-36	0.58	1.00	0.90	0.61	0.86	0.90	0.97	0.94
27-28	0.55	0.74	0.79	0.68	0.82	0.88	0.96	0.89
25-26	0.57	0.62	0.65	0.69	0.83	0.87	0.95	0.85
23-24	0.38	0.59	0.63	0.72	0.84	0.86	0.94	0.80
21-22	0.46	0.61	0.53	0.63	0.73	0.83	0.93	0.73
19-20	0.42	0.50	0.48	0.69	0.72	0.83	0.90	0.69
16-18	0.33	0.45	0.44	0.59	0.70	0.79	0.83	0.61
1-15	0.29	0.32	0.45	0.49	0.50	0.61	0.76	0.48
TOTAL	0.40	0.52	0.53	0.64	0.74	0.84	0.94	0.75

While the pattern of the proportions, generally, is as expected, with the smallest values in the lower left corner and the highest in the upper right corner, there as reversals at several points in the table, mainly towards the upper .eft corner of the table where the numbers of students are small. The smoothing procedures described here are used to convert the proportions of Table 1 into a nomograph.

The Procedure

The steps of the procedure used to construct an expectancy nomograph are as follows:

1. Starting in the upper right corner of the bivariate, HSCPR and ACT score, distribution of students and moving down and to the left, successive groups of students are formed. In the initial application of the procedure the minimum group size was set at fifty.

   The bivariate distribution is first divided into regions defined by HSCPR ranges. The ranges are narrow, but each must include an acceptable number of students. The HSCPR ranges overlap in order to make maximum use of the available data. The ranges used with the present data are: 0-39, 35-44, 40-49, 45-54, 50-59, 55-64, 60-69, 65-74, 70-79, 75-84, 80-89, 85-94, 90-99 and 95-99.

   The students in the highest HSCPR range with the highest ACT score are counted. If the count is 50 or greater, a group is defined. If not, the students with the next lower ACT score are added and if 50 or more students are now included, the first group has been formed.

   When a group has been formed, the procedure moves to the next lower ACT score and the process is repeated until a second group is formed. This process is repeated until the lowest ACT score for the HSCPR range is reached. If, after that ACT score is included, at least 25 students have been counted, then those students define a group. If less than 25 students have been counted, these students are added to the last previously defined group.

This process is repeated with the next lower HSCPR range and continues through all HSCPR ranges until the lowest ACT score in the lowest HSCPR range is reached. At this point all of groups have been formed.

2. The mean HSCPR, the mean ACT score and the proportion of students who were successful (PS), e.g. had a GPA $\geq$ 2.0, are calculated for each group.

3. Pairs of groups are used to define points for which PS values are stated as even-tenth values, i.e., .90, .80, .70, ..., .10. The even-tenth value of the PS for a point is an estimate of the proportion of students with the HSCPR and ACT score of the point who are successful, e.g., earn a GPA of at least 2.0. Points are developed as follows:

   The groups are sorted from high to low on the basis of PS values and the PS values are divided into ranges which are separated by even-tenth values.

   The group with the highest PS is paired with a group in the next lower range of PS values. Assuming the first group is in the PS $\geq$ .90 range, this group is paired with each of the groups with .80 $\leq$ PS < .90 and the distance between the first group and each of the groups with which it is paired is calculated,

   $$D = SQRT[(mnR_1 - mnR_2)^2 + (mnA_1 - mnA_2)^2],$$

   where mnR_1 and mnR_2 are the mean HSCPRs and mnA_1 and mnA_2 are the mean ACT scores for the two groups. The pair of groups which has the smallest distance is selected.

   The group with the next lower PS is then paired with the remaining groups in the .80 $\leq$ PS < .90 range and a second pair is identified using the smallest distance criterion. This process is continued until all groups with PS $\geq$ .90 have been matched with groups in the .80 $\leq$ PS < .90 range. If there are more groups in the higher range, groups in the lower range cannot be reused until all of the groups in that range have been used; after all have been used, then all are candidates for the next match. No group in the lower range can be used more than twice until all have been used twice.

When all groups in the PS $\geq$ .90 range have been paired, the pairing process continues with groups in the .80 $\leq$ PS $<$ .90 range being paired with groups in the .70 $\leq$ PS $<$ .80 range. The process continues until groups in the .10 $\leq$ PS $<$ .20 range are paired with those in the PS $<$ .10 range or until there are no more groups to pair.

Each pair of groups then defines a point which has the following three values:

PS is the even-tenth value spanned by the PS values of the two groups.

$R = mnR_1 - \{(mnR_1 - mnR_2) \times [(PS_1 \quad PS)/(PS_1 - PS_2)]\}$.

$A = mnA_1 - \{(mnA_1 - mnA_2) \times [(PS_1 - PS)/(PS_1 - PS_2)]\}$.

R and A are estimates of the mean HSCPR and mean ACT score for the combined group of students with the even even-tenth PS.

4. A curve is then fitted to the scatter diagram of points (R,A) for each even-tenth value of PS. The curve specifies those pairs of HSCPR and ACT score values which predict the given proportion successful.

Observation of a number of scatter diagrams of points (R,A) generated by the process described here suggested that a curve which decreases at an increasing rate would, in most cases, better fit the points than a straight line. Consequently, the curve that is fitted for each even-tenth value of PS is

$A' = a + bR^2$.

Also, because the curve should pass through the geometric center of the points of the scatter diagram rather than through the means of either the vertical or horizontal arrays, the curve which minimizes the perpendicular deviations of points from the line is used (Ehrenberg, 1984). In order to remove the influence of the standard deviations of A and R^2, the values of R^2 are converted to values which have the same standard deviation as A before the perpendicular deviation fitting is carried out. It turns out that the parameters of the line fitted in this manner are

$b = -(s_A/s_{R2})$ and

$a = [(s_A/s_{R2}) \times mnR^2] + mnA$.

In Calculating mnA, mnR^2, s_A (sd of A) and s_{R^2} (sd of R^2), each value of A and R^2 is weighted by the sum of the numbers of students in the pair of groups which defined the point (R,A).

5. The resulting curve for each even-tenth proportion is drawn on a bivariate diagram which has HSCPR as the horizontal scale and ACT Composite score as the vertical scale. The result is the desired expectancy nomograph.

These five steps were carried out as follows in developing the expectancy nomographs discussed in this paper. Steps 1, 2 and 3 were accomplished by means of a PL/I program which runs on an IBM mainframe computer. The coordinates -- PS, R and A -- of the points produced in step 3 are the output of the PL/I routine. These coordinates were then input into a SAS program which produces the means and sums of squares required to calculate the parameters, a and b, of the curves which are fitted to the several sets of points. The means and sums of squares were entered to a pc spreadsheet which includes formulas for calculating parameters a and b for each value of PS and which also calculates points of the fitted curves.

This set of procedures May be more cumbersome than necessary. It was developed, before the perpendicular deviation, curve fitting approach was adopted. In retrospect, it might have been more efficient to extend the PL/I program to carry out step 4 calculations, rather than using SAS and the spreadsheet. On the other hand, the SAS routine also produces, for each PS, r^2 for R^2 and A and scatterdiagrams which are useful in interpreting the goodness of fit of the resulting curve.

The GPA ≥ 2.00 Nomograph and Solution Parameters

The expectancy nomograph for freshman year GPA ≥ 2.0 produced for the 12,835 students using the five steps just described is shown in Figure 1. The numbers on a nomograph curve are the "chances in ten" of earning a freshman year GPA of at least 2.0 for the values of HSCPR and ACT score which lie on the curve. The nomograph indicates that the student who has a high school class percentile rank of 50 and an ACT score of 16 has 4 chances in 10 or a probability of .40 of earning a freshman year GPA of at least 2.0. For the student with a HSCPR of 50 and an ACT score of 20 the probability of earning at least a 2.0 is .50.

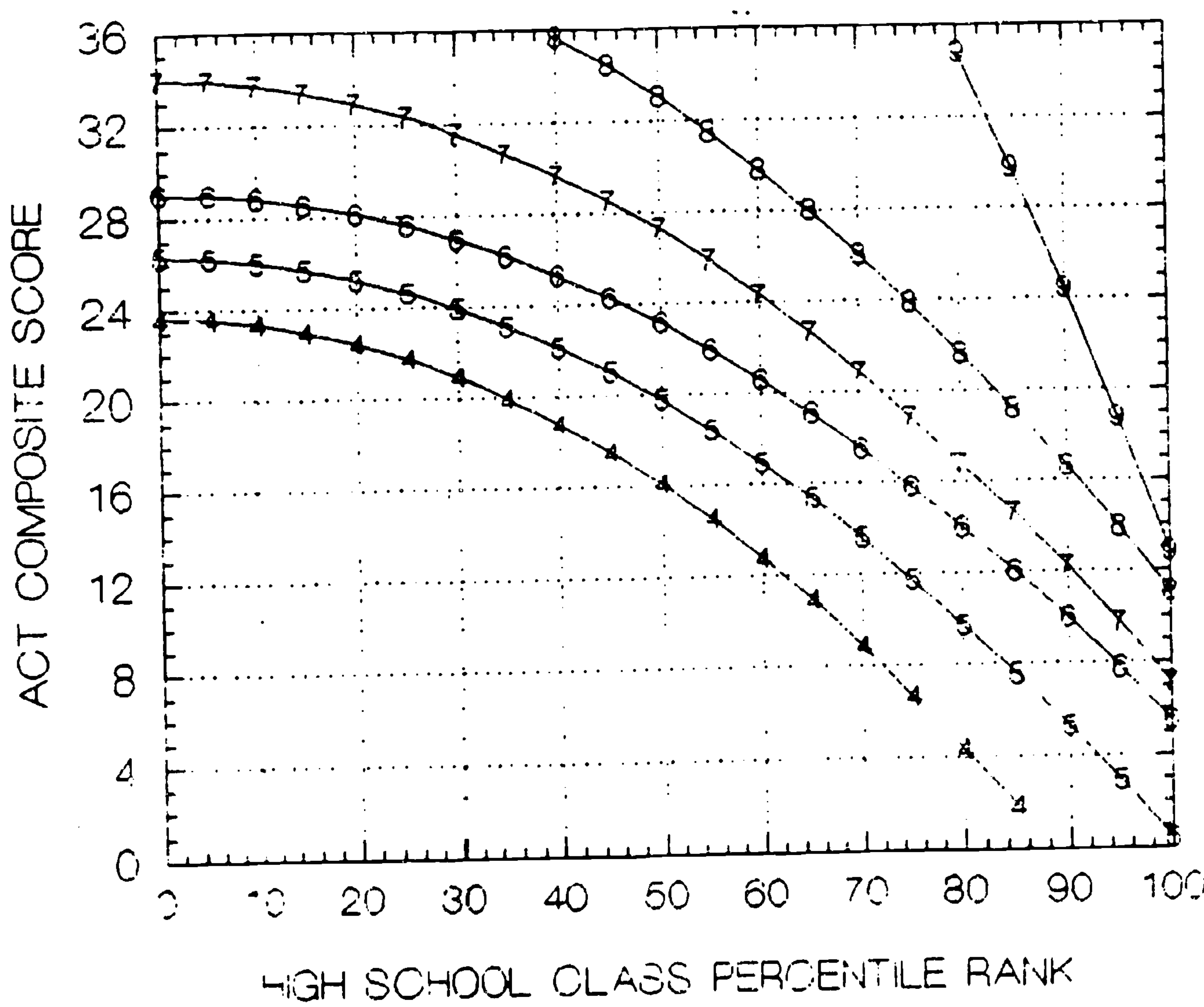

Figure 1. Expectancy Nomograph for GPA ≥ 2.0

Nomograph curves for PS = .30 and PS = .20 were calculated (see Table 3), but did not follow the pattern of the other nomograph curves, so are not plotted on the nomograph of Figure 1.

Parameters of the nomograph solution which are indicative of the validity of the solution are characteristics of the fitting of each nomograph curve. Table 2 displays these parameters. For each curve, the number of points generated in step 3 of the procedure, the number of students involved in defining these points and the average size of the groups (step 1 of the procedure) are included and reflect the amount of data involved in generating

Table 2. Parameters of Solution for Nomograph for GPA ≥ 2.0)

Prob.	Numb. Points	Total Students	Avg Grp Size	r	b	a
0.90	37	13,584	183.6	-0.70	-0.0061	73.94
0.80	36	10,684	148.4	-0.80	-0.0029	40.38
0.70	27	5,749	106.5	-0.83	-0.0027	34.00
0.60	33	5,810	88.0	-0.81	-0.0024	29.06
0.50	32	5,187	81.1	-0.80	-0.0026	26.27
0.40	24	3,272	68.2	-0.86	-0.0030	23.64
0.30	17	1,665	29.0	-0.75	-0.0021	19.41
0.20	4	279	34.9	-0.92	-0.0010	15.14
0.10	2	80	20.0	--	--	--

the curve. Clearly the 12,835 students are clustered toward the upper ranges of ACT scores and HSCPRs. Notice also that the average group size for PS = .30 and .20 fall below 50 indicating that groups involved in defining the points for these curves come from the lowest ACT score ranges. This may explain why the curves for these PS values were atypical.

Also shown in Table 2 for each nomograph curve is the correlation between A and R^2 and the parameters of the regression line, $A' = a + bR^2$. The typically high correlations are indicative of the validity of the resulting curves. The generally regular progression of the intercept parameter, a, for the several curves also suggests a solution that appropriately fits the data. The coefficients, b, of the several curves do not exhibit a regular pattern or a regular progression from one to the next. This absence of regularity in values of b is apparent in the differences between adjacent points at which the curves cross HSCPR = 100 in Figure 1. Even with these differences the overall symmetry of the nomograph curves suggests a satisfactory solution for the curves plotted.

A step which might be added to the procedure would smooth the values of a and, particularly, b as shown in Table 2 in order to remove irregularities from the nomograph. In the present case, the values of a may need very little adjustment, but some smoothing of the values of b might not only increase the regularity of the curves now plotted, but might also allow the curve for PS = .30 or even the one for PS = .20 to be become symmetrical with the others and plotted on the nomograph. The extension of the procedure to smoothing values of a and b was not pursued in the present project.

Effects of Varying Sample Size and Minimum Group Size

The number of students, 12,835, available for developing the expectancy nomograph shown in Figure 1 is quite large. If it is concluded the nomograph solution is satisfactory for that sample size, the question, For what smaller sample sizes will it produce acceptable results?, can be raised. Similarly the minimum group size of 50, used in step 1 of the procedure, was set arbitrarily. Will the procedure work for smaller minimum group sizes?, is another question that can be asked.

Sample Size. First, the procedure was applied to two random halves of the original sample.[a] Points of the curves which resulted for each random half, as well as the corresponding points calculated for the total sample are shown in Table 3. (A blank row in the table indicates that two or fewer points were generated by the procedure and that, consequently, a curve could not be fitted to the data.) The correspondence of the three curves seems to be quite good for PS = .40, .50, .70 and .80. The three curves for PS = .90 differ appreciably only at the extremes. For PS =.60 the total sample and second half sample curves are nearly the same, but the curve generated from the first half sample differs from the other two PS = .60 curves by up to 4 ACT score points. Smoothing of the parameters a and b might reduce or eliminate the discrepancy for PS = .60. The parameters of the curves fitted from the half samples were not remarkably different from those of the total sample.

Next, expectancy nomographs curves were generated for random samples which were created to be one-fourth, one-eighth and one-sixteenth the size of the original sample. These samples turned out to include 3,219, 1,601 and 818

Table 3. ACT Scores Which Determine Nomograph Points for the GPA ≥ 2.0 Nomograph, Total Sample and Random Halves of Total Sample

PROP. WITH GPA ≥ 2.0	SAMPLE	HIGH SCHOOL CLASS PERCENTILE RANK																				
		0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
.90	Total																	35	30	24	19	13
	1st Half																	33	29	25	20	15
	2nd Half																34	31	28	24	21	17
.80	Total									36	34	33	32	30	28	26	24	22	19	17	14	11
	1st Half									35	34	33	31	30	28	26	24	22	19	17	14	11
	2nd Half								35	34	33	32	31	29	27	26	24	22	20	17	15	12
.70	Total	34	34	34	33	33	32	32	31	30	29	27	26	24	23	21	19	17	15	12	10	7
	1st Half	33	33	33	33	32	31	31	30	29	28	27	25	24	22	20	19	16	14	12	10	7
	2nd Half	35	35	35	35	34	33	33	32	31	29	28	27	25	23	21	19	17	15	12	9	7
.60	Total	29	29	29	29	28	28	27	26	25	24	23	22	21	19	18	16	14	12	10	8	6
	1st Half	33	33	33	32	32	31	30	29	28	27	25	24	22	20	18	16	14	11	9	6	3
	2nd Half	29	29	29	28	28	27	27	26	25	24	23	22	20	19	17	16	14	12	10	8	5
.50	Total	26	26	26	26	25	25	24	23	22	21	20	18	17	15	14	12	10	8	5	3	0
	1st Half	26	26	26	26	25	25	24	23	22	21	20	18	17	15	13	12	10	7	5	3	
	2nd Half	27	27	27	26	26	25	24	23	22	21	20	18	16	15	13	10	8	6	3	0	
.40	Total	24	24	23	23	22	22	21	20	19	18	16	15	13	11	9	7	4	2			
	1st Half	23	23	23	23	22	21	20	20	18	17	16	14	12	10	8	6	4	1			
	2nd Half	24	23	23	23	22	22	21	20	19	18	16	15	13	12	10	8	5	3	1		
.30	Total	19	19	19	19	19	18	17	17	16	15	14	13	12	10	9	7	6	4	2	0	
	1st Half	19	18	18	17	17	17	16	15	14	13	12	11	10	8	7	5	4	2			
	2nd Half	17	17	16	16	16	16	15	15	14	[illegible]	13	13	12	11	10	9	8	7	6	5	3
.20	Total	15	15	15	15	15	15	14	14	14	13	13	12	11	11	10	9	9	8	7	6	5
	1st Half																					
	2nd Half																					

students respectively. The nomograph points produced for these three sample. as well as for the original sample, are shown in Table 4. For the sample of 3,219 the curves for PS = .50, .60 and .90 were quite similar to the total sample curves. For the sample of 1,601 the curves for PS = .60 .70 and .80 are quite similar to the corresponding curves for the total sample. The curves for PS = .40 were unsatisfactory for each of the three smaller samples and none of the curves for the sample of 818 were similar to the corr sponding total sample curves.

Table 4. ACT Scores Which Determine Nomograph Points for the GPA ≥ 2.0 Nomograph for Varying Sample Sizes, Minimum Group Size = 50

PROP. WITH	SAMPLE	HIGH SCHOOL CLASS PERCENTILE RANK																				
GPA ≥ 2.0	SIZE	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
.90	12,835																	35	30	24	19	13
	3,219																	34	30	25	21	16
	1,601																35	31	27	23	19	14
	818																	33	30	26	22	17
.80	12,835									36	34	33	32	30	28	26	24	2	19	17	14	11
	3,219												35	33	31	28	25	22	19	16	12	8
	1,601											35	33	31	29	26	24	21	18	15	11	8
	818															34	30	26	22	17	12	7
.70	12,835	34	34	34	33	33	32	32	31	30	29	27	26	24	23	21	19	17	15	12	10	7
	3,219									35	34	32	29	27	25	22	19	16	12	9	5	1
	1,601	33	33	33	33	32	32	31	30	29	28	27	26	24	23	21	19	17	15	13	11	8
	818								36	34	32	31	28	26	24	21	18	15	12	9	5	1
.60	12,835	29	29	29	29	28	28	27	26	25	24	23	22	21	19	18	16	14	12	10	8	6
	3,219	29	29	29	29	28	28	27	26	26	25	23	22	21	19	17	16	14	12	10	7	5
	1,601	30	30	29	29	28	28	27	26	25	24	23	21	20	18	16	14	12	10	7	5	2
	818			36	35	35	34	33	31	30	28	26	24	21	19	16	13	10	6	2		
.50	12,835	26	26	26	26	25	25	24	23	22	21	20	18	17	15	14	12	10	8	5	3	0
	3,219	27	27	27	27	26	25	24	23	22	21	19	18	16	14	12	9	7	4	1		
	1,601	22	22	22	22	21	21	21	20	19	18	18	17	16	15	13	12	11	9	8	6	4
	818	23	23	23	22	22	22	21	21	20	20	19	18	17	17	16	14	13	12	11	9	8
.40	12,835	24	24	23	23	22	22	21	20	19	16	16	15	13	11	9	7	4	2			
	3,219	20	20	20	20	19	19	19	18	17	17	16	15	14	13	12	10	9	8	6	5	3
	1,601	20	20	20	20	19	19	18	18	17	16	15	14	13	11	10	9	7	5	3	1	
	818	18	16	18	16	17	17	17	16	16	15	15	14	14	13	12	11	10	9	8	7	6
.30	12,835	19	19	19	19	19	18	17	17	16	15	14	13	12	10	9	7	6	4	2	0	
	3,219																					
	1,601																					
	818																					
.20	12,835	15	15	15	15	15	15	14	14	14	13	13	12	11	11	10	9	9	8	7	6	5
	3,219																					
	1,601																					
	818																					

Except, of course, for the parameters a and b, there was no identifiable relationship between parameters of the nomograph solutions for samples of 3,219 and 1,601 and whether or not the curve matched the curve for the original sample of 12,835. Satisfactory curves resulted when only 4 and 9 points were produced and unsatisfactory curves resulted from fitting curves to 14 and 27 points. Similarly, satisfactory curves were produced when the correlation between A and R^2 was as low as -.48 and unsatisfactory curves resulted when a correlation was as high as -.92.

It appears that samples smaller than 12,000 can produce satisfactory expectancy nomographs, but that the smallest sample for which a satisfactory nomograph might be produced, using the procedure described here, may lie somewhere between 1,000 or 1,500 and 3,000. With the smaller samples, it may be particularly important to adopt procedures to smooth the progressions of parameters a and b.

Group Size. To investigate effects of minimum group size, step 1 of the procedure for producing the expectancy nomograph was modified to make the minimum group size 30 and again to make the minimum size 10. Nomographs were produced by applying the modified procedures to the original sample. The nomographs curves produced by specifying minimum group sizes of 30 and 10 were, with very few exceptions, quite similar to the curves produced with the minimum group size set at 50. It turns out that by virtue of the large numbers of students in the original sample and the manner in which groups are formed, specifically the larger number of students with a given ACT score within a HSCPR range, the sizes of the groups created did not decline appreciably, even though minimum group size was lowered. The average sizes of the groups involved in fitting curves when the minimum group size was 50 for PS = .40 to .90 ranged from 68qq.2 to 183.6 (see Table 2); the average sizes decreased only to 46.8 to 146.9 when the minimum group size was set at 10. Thus, it is not surprising that altering the minimum group size had little effect on the nomograph curves produced.

Consequently, expectancy nomographs produced from samples of varying sizes when the minimum groups size was set at 10 were investigated. In addition to the original sample and the three smaller samples already described, a random sample of 404 students, approximately one-thirty second the size of the original sample was used. Points for all of the resulting nomograph curves, including the original sample curves with minimum group size = 50, are shown in Table 5.

Table 5. ACT Scores Which Determine Nomograph Points for the GPA ≥ 2.0 Nomograph for Varying Sample Sizes and Minimum Group Size = 10, Compared with Points for Sample Size = 12,835 and Minimum Group Size = 50

PROP. WITH GPA ≥ 2.0	SAMPLE SIZE	HIGH SCHOOL CLASS PERCENTILE RANK																				
		0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
.90	12.835*																	35	30	24	19	13
	12,835																34	31	28	24	20	16
	6,443																36	32	28	25	20	16
	3,219															35	33	30	27	24	21	17
	1,601												35	34	32	31	29	27	25	23	20	18
	818										36	35	33	32	31	29	28	26	24	22	20	18
	404										35	34	33	32	30	29	27	25	23	21	19	17
.80	12.835*									36	34	33	32	30	28	26	24	22	19	17	14	11
	12,835										35	33	32	30	28	26	24	21	19	16	13	10
	6,443									36	34	33	31	30	28	26	24	22	19	17	14	11
	3,219				36	35	35	34	34	33	32	31	29	28	27	25	23	22	20	18	16	13
	1,601	34	34	34	34	34	33	33	32	31	30	30	28	27	26	25	23	22	20	19	17	15
	818				36	35	35	34	34	33	32	31	30	29	28	26	25	23	22	20	18	16
	404								36	35	34	33	31	30	28	26	24	22	20	18	16	13
.70	12.835*	34	34	34	33	33	32	32	31	30	29	27	26	24	23	21	19	17	15	12	10	7
	12,835								35	34	32	31	29	26	24	22	19	16	13	10	6	3
	6,443						36	35	34	33	31	30	28	26	24	22	20	17	14	12	9	5
	3,219	35	35	34	34	34	33	32	31	30	29	28	26	25	23	21	19	17	14	12	9	7
	1,601	33	32	32	32	32	31	30	30	29	28	27	25	24	23	21	19	17	15	13	11	9
	818							35	34	33	32	31	29	27	26	24	21	19	17	14	11	9
	404						36	35	33	32	31	29	27	25	22	20	17	15	12	8	5	1
.60	12.835*	29	29	29	29	28	28	27	26	25	24	23	22	21	19	18	16	14	12	10	8	6
	12,835	31	31	31	31	30	29	29	28	27	25	24	23	21	19	17	15	13	10	8	5	2
	6,443	33	33	33	32	32	31	30	29	28	27	25	23	22	20	18	15	13	10	7	5	1
	3,219	32	31	31	31	30	30	29	28	27	26	24	23	21	19	17	15	13	11	8	6	3
	1,601	31	31	30	30	30	29	28	27	26	25	24	22	21	19	17	15	13	11	8	6	3
	818	30	30	30	30	29	29	28	27	26	26	24	23	22	21	19	17	16	14	12	10	8
	404																					
.50	12.835*	26	26	26	26	25	25	24	23	22	21	20	18	17	15	14	12	10	8	5	3	0
	12,835	25	25	25	25	24	24	23	23	22	21	20	19	17	16	14	13	11	9	7	5	3
	6,443	27	27	27	26	26	25	25	24	23	22	21	19	18	16	14	12	11	8	6	4	1
	3,219	28	27	27	27	26	26	25	24	23	22	20	19	17	15	14	12	9	7	4	2	
	1,601	26	26	26	26	26	25	24	24	23	22	21	19	18	16	15	13	11	9	7	5	3
	818	25	25	25	24	24	24	23	23	22	21	20	19	18	17	16	15	13	12	10	8	7
	404	23	23	23	23	23	22	22	21	21	20	19	18	17	16	15	14	12	11	10	8	6
.40	12.835*	24	24	23	23	22	22	21	20	19	18	16	15	13	11	9	7	4	2			
	12,835	22	22	21	21	21	20	20	19	18	17	16	15	14	12	11	9	8	6	4	2	
	6,443	22	22	22	22	22	21	21	20	19	18	17	16	15	14	12	11	9	8	6	4	2
	3,219	26	25	25	25	24	23	23	21	20	19	17	15	14	11	9	7	4	2			
	1,601	25	25	24	24	24	23	22	22	21	20	19	17	16	15	13	11	9	7	5	3	1
	818	24	24	23	23	23	22	22	21	20	19	18	17	16	15	13	12	10	8	6	4	2
	404	29	28	28	26	27	26	24	23	21	19	17	15	12	9	6	3	0	0	0	0	0
.30	12.835*	19	19	19	19	19	18	17	17	16	15	14	13	12	10	9	7	6	4	2	0	
	12,835	19	19	19	19	18	18	17	17	16	15	14	12	11	10	8	6	5	3	1		
	6,443	20	20	20	20	19	19	18	18	17	16	15	14	12	11	9	8	6	4	2		
	3,219	21	21	21	20	20	20	19	18	18	17	16	15	14	12	11	10	5	6	5	3	1
	1,601	19	19	19	19	19	18	18	18	17	17	16	16	15	14	13	12	12	11	9	8	7
	818																					
	404	40	40	39	38	36	34	32	29	26	22	18	14	9	3	0	0	0	0	0	0	0
.20	12.835*	15	15	15	15	15	15	14	14	14	13	13	12	11	11	10	9	9	8	7	6	5
	12,835	13	13	13	13	13	13	13	12	12	12	11	11	11	10	10	9	9	8	7	7	6
	6,443	17	17	17	17	16	16	16	15	15	14	13	12	11	11	9	8	7	6	5	3	2
	3,219	22	22	22	22	21	21	20	19	18	17	16	15	14	12	11	9	7	5	3	1	
	1,601																					
	818																					
	404																					

* Minimum Group Size = 50

For most PS values nomograph curves produced from the sample of 6,443 students with minimum groups size = 10 closely matched the original sample, minimum group size = 50, curves. In four cases (PS = .90, .70, .50 and .40) the curves developed from the sample of 3,218 seem satisfactory and in three cases (PS = .70, .60, and .50) satisfactory curves were produced from the sample of 1,601. Most of curves produced from even the smaller samples, 818 and 404, with minimum group size set at 10, did not depart greatly from the corresponding curves resulting from the larger samples and the larger minimum group size.

The parameters of the nomograph curves shown in Table 5 are of some interest. As the sample size decreases the numbers of points generated, the average group sizes and the correlations between A and R^2 decrease. Of course, minimum group sizes of 10 result in larger numbers of points than minimum groups sizes of 50. Apparently the larger numbers of groups and points offsets the effects of lower correlations between A and R^2 and result in generally accurate nomograph curves.

A Nomograph for GPA ≥ 3.0

An additional illustration of the expectancy nomograph is provided by Figure 2 which was produced from the original sample, minimum group size = 50, by changing the definition of success in step 2 of the procedure from GPS ≥ 2.0 to GPA ≥ 3.0. Curves for PS = .40, .60 and .80 are not plotted to avoid clutter in the nomograph (and also because the curve for PS =.40 lacked symmetry with the other curves). No points for PS =.90 were produced.

While this nomograph would seem to be satisfactory, it might be improved by smoothing parameters a and b in order to increase the regularity of the distances between nomograph curves when HSCPR = 100. Nomographs for GPA ≥ 3.0 were also produced from the two random halves of the original sample. The resulting two curves for each value of PS were essentially identical to each other and to the corresponding curve plotted in Figure 2, thus providing further support to the validity of the expectancy nomograph.

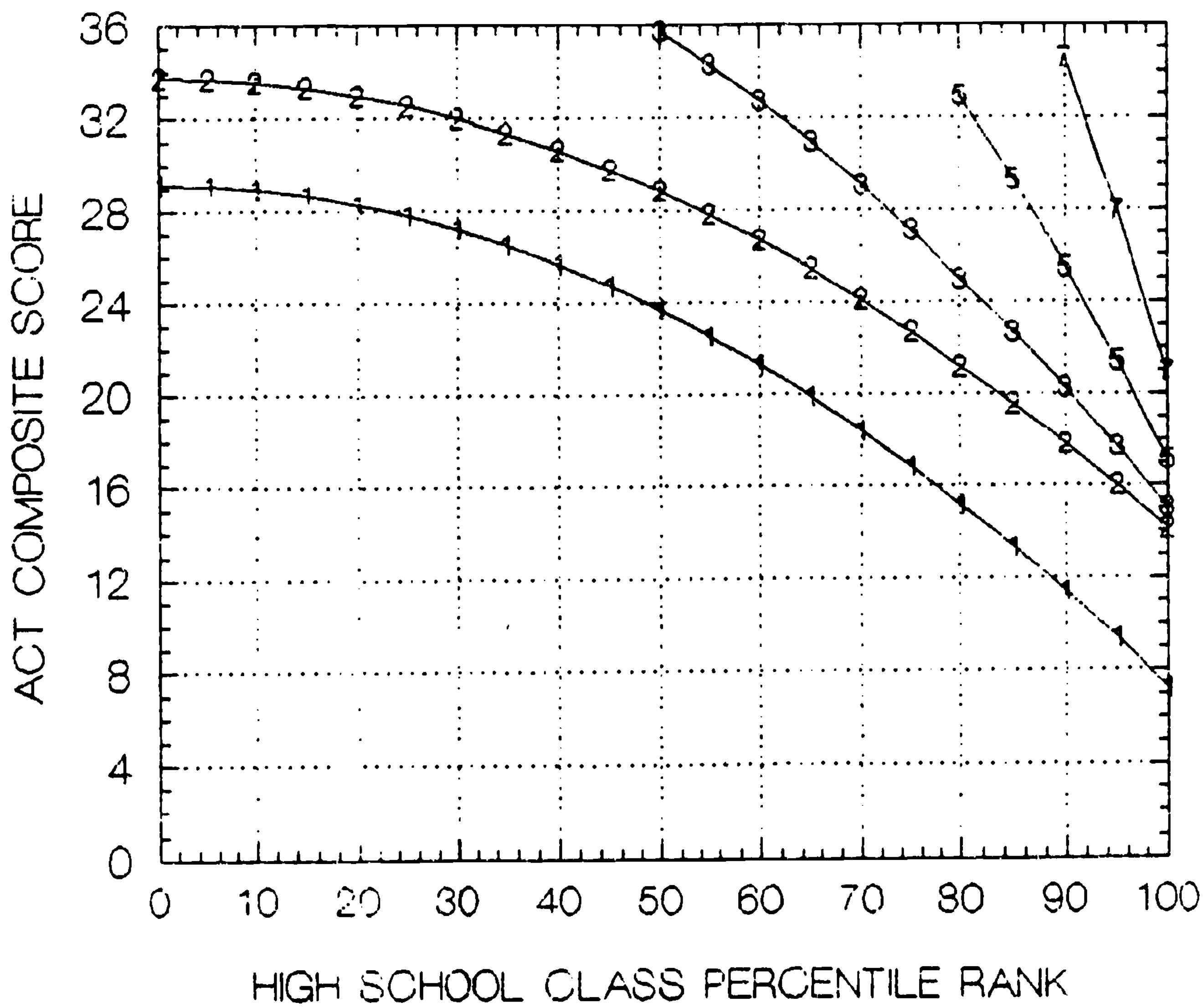

Figure 2. Expectancy Nomograph for GPA ≥ 3.0

Conclusions

Valid expectancy nomographs can be produced from large samples using minimum group size = 50 and these nomographs may be easier to use than the expectancy tables for which they are intended to substitute. However, additional work on the development of procedures for producing expectancy nomographs is needed. First, procedures for smoothing progressions of curve fitting parameters a and b over the several PS values are needed. It should be possible to accomplish the suggested smoothing without decreasing the accuracy of the curves, particularly at those points in the bivariate distribution where appreciable numbers of observations occur. Secondly, further exploration of the use, in the nomograph development procedure, of minimum group sizes of less than 50 and of the minimum numbers of observations needed to produce satisfactory nomographs is needed. Results reported here suggest that setting the minimum group size at 10 can produce accurate nomographs with considerably fewer than the 12,835 cases used in this study. Finally, a close examination of the nature of the individual groups formed in step 1 of the parocedure might lead to a reduction in the length of the HSCPR intervals. This variation in the procedure might be particularly beneficial with smaller samples and smaller minimum group sizes.

Notes

1. The assistance of Dr. Jon Maatta, who suggested the perpendicular deviation curve-fitting procedure, Mr. Michael Kurth, who wrote the PL/I program and integrated it with the SAS and spreadsheet processing, and Mr. Gary Moss, who ran the computer jobs which produced the data used in this paper, is gratefully acknowledged.

2. The SAS-produced "random halves" include numbers of cases, 6,443 and 6,399, which are not quite equal and which include a total number of students, 12,842, which exceeds that, 12,835, of the original sample. The reason for these discrepancies is not clear, but they are not believed to have harmed the data analysis reported in the paper.

References

Ehrenberg, A.S.C. (1984). Reply. American Statistician, 3(2), 162.

Morgan, G. (1988). Expectancy tables. In J. P. Keeves (Ed.) Educational research, methodology, and measurement: An international handbook. Elmsford, NY: Pergamon Press.

Perrin, D. W. and Whitney, D. R. (1976). Methods for smoothing expectancy tables applied to the prediction of success in college. Journal of Educational Measurement, 13(3), 223-231.

Schrader, W. B. (1965). A taxonomy of expectancy tables. Journal of Educational Measurement, 2(1), 29-35.