SOCIAL SCIENCE RESEARCH 5, 349-383 (1976)

Background Characteristics in the U.S. Adult Population 1952-1973: A Survey-Metric Model

JAMES A. DAVIS
National Opinion Research Center and Dartmouth College

Data from 30 surveys (46,356 cases from 10 Michigan Election Surveys and 20 Gallup polls) are analyzed to study the interrelations of cohort, sex, race, region, religion, and education from 1952 to 1973.

A d-system model is developed that pools results from the various surveys into a set of parameters forming a flowgraph model. This technique allows one to see whether the surveys used are consistent and whether the results are consistent with outside sources such as census reports. (Generally speaking, the surveys were found to be consistent and the model showed generally good fit with census and NORC General Social Survey estimates, but significant dffferences in two or three parameters.) A computer model of the system is used to generate implications and extrapolations.

Substantively, the main results were: (1) a definite trend away from sex equality in education, especially for college attendance; (2) very little narrowing in race, region, and religious differences in education, until the most recent cohorts where Northern White Catholics and Southern White Protestants show virtual parity with "Yankees" (Northern White Protestant); and (3) a projection that shows discouragingly little narrowing of intergroup education differences in the general population for the next 5 years.

We here consider the cross-tabulations of age (cohort), sex, region, race, religion, and educational attainment in 30 national surveys from 1952 to 1973.

Why? (1) To study the "attainment process" as a dynamic system; that is, to see how social differences in educational attainment changed during the first two-thirds of the twentieth century. (2) To show how d-systems (Davis, 1976) can be used to construct models using categorical data that are analogous to econometric models. (3) To provide a tool for examining sample biases in national surveys.

Variables and Model

The first variable is cohort, or respondent's age recoded to year of birth (more exactly, reported age at interview subtracted from the year of the study). To use year and cohort is an arbitrary but nontrivial decision, because

Copyright (31976 by Academic Press, Inc. All rights of reproduction in any form reserved.

year, cohort, and age have a tricky interlock. If we know any two, the other is determined. (If the respondent was born in 1900 and was interviewed in 1960, he must have been age 60 at the time of the interview.) Consequently, we cannot generally introduce age, cohort, and year (often called period) as separate variables in the same analysis (see Mason, 1973). The choice of excluded variable is arbitrary. We chose to keep year because it seemed odd to study social change and exclude year. We chose to keep cohort to make sure that at least one variable in the model would change appreciably over 20 years. Of necessity, age will disappear, but we can keep our eyes open to trends associated with later years and earlier cohorts, and these can just as easily be described as age effects.

We grouped the cases into four cohorts, as shown in Table 1.

TABLE 1
Age Composition of Cohort Groups, 1947-1974

	Cohort			
Year	Vietnam	World War 11	Dep.	World War I
1974	18-34	35-50	51-67	68+
1973	18-33	34-49	50-66	67+
1972	18-32	33-48	49-65	66+
1971	18-31	32-47	48-64	65+
1970	18-30	31-46	47-63	64+
1969	18-29	30-45	46-62	63+
1968	18-28	29-44	45-61	62+
1967	18-27	28-43	44-60	61+
1966	18-26	27-42	43-59	60+
1965	18-25	26-41	42-58	59+
1964	18-24	25-40	41-57	58+
1963	18-23	24-39	40-56	57+
1962	18-22	23-38	39-55	56+
1961	18-21	22-37	38-54	55+
1960		21-36	37-53	54+
1959		21-35	36-52	53+
1958		21-34	35-51	52+
1957		21-33	34-50	51+
1956		21-32	33-49	50+
1955		21-31	32-48	49+
1954		21-30	31-47	48+
1953		21-29	30-46	47+
1952		21-28	29-45	46+
1951		21-27	28-44	45+
1950		21-26	27-43	44+
1949		21-25	26-42	43+
1948		21-24	25-41	42+
1947		21-23	24-40	41+

To form the grouping we simply worked with the data (prior to running any cross-tabulations) until we got reasonable case bases across the study period, 1952-1973. Although the cuts are arbitrary, we can give the four categories some verisimilitude as follows.

World War I. The oldest cohort consists of persons born in 1906 or earlier. At an extreme, the handful of respondents 80 years and older in the 1952 study were born in 1872 or before. All reached age 16 in 1922 or before 1

Interwar. The Interwar cohort was born between 1907 and 1923, and reached age 16 between 1923 and 1939, the boom and depression years between World War I and World War 11.

World War II and Korea. The third cohort was born between 1924 and 1939, and reached age 16 between 1940 and 1955. Their critical years for educational attainment span the period from World War 11 through the Korean War.

Vietnam. The youngest group, those born in 1940 and after, and reaching age 16 in 1956 and after, did not become eligible for sampling until the early 1960s. By the end of the decade, however, they were becoming a substantial portion of the adult population, as will be shown later.

For the Vietnam cohort, we dropped the lower age to 18 from 21 because national suneys moved their age limit to 18 with the advent of the "18-year-old vote." (The Constitutional Amendment was ratified in 1971.) In a biological sense the Vietnam cohort is artificially larger than its predecessors, but adulthood was actually increased sociologically (or at least politically) by 3 years during the late 1960s in the same way that cities have grown by annexing adjacent suburbs.

The second variable is sex, male and female.

The third variable combines race, region, and religion into a typology, or "social stereotypology." For the rationale, we begin with Table 2, the cross-tabulation of region (census South vs All Other or North), race (White vs Nonwhite) and religion (Protestant, Catholic, Jewish, Other, and None and No Answer), combining all cases from 10 Michigan Election studies, 1952 to 1972.

The variables are strongly associated: Catholics are predominantly (84.5%) Northern Whites; most Nonwhites (73.5% in the North, 92.8% in the South) are Protestants; Southern Whites are strongly (87.1%) Protestant.

To avoid small cell sizes, to finesse problems of causal ordering, and for simplicity, the 16 cells in Table 2 are collapsed into a sixfold typology in Table 3.

Yankees (39%), Northern Catholics (18%), Northern Blacks (5%), South

11n terms of the attainment process, it is useful to consider the year in which these groups reached age 16, near the time they decided whether or not to complete high school.

TABLE 2
Region, Race, and Religion, 1952-1972 Pooled

Region	Race	Protestant	Catholic	Jewish	Other, None, No Answer	Total
North	White	6,790	3,125	463	576	10,954
	Nonwhite	639	169	1	60	869
South	White	3,963	368	43	176	4,550
	Nonwhite	830	35	0	29	894
		12,222	3,697	507	841	17,267

TABLE 3
Data from Table 2 in Typology

Group	Region	Race	Religion	N	Proportion
Yankees	North	White	Protestant	6,790	.393
Northern Catholics	North	White	Catholic	3,125	.181
Northern Blacks	North	Black	Any	869	.050
Southern White Protestants	South	White	Protestant	3,963	.230
Southern Blacks	South	Black	Any	894	.052
Other			Any other	1,626	.094
			Total	17,267	1.000

ern White Protestants (23%), Southern Blacks (5%), and All Other (9%) will be treated as a single six-category nominal variable.

The fourth variable is educational attainment, trichotomized as less than high school graduate, high school graduate, and more than high school.

Figure 1 shows how these variables are specified as a d-system model. The model is specified within cohort (i.e., there are four submodels) because statistical analysis shows many of the coefficients vary significantly from cohort to cohort. This is one of the main empirical results and will be described in detail later.

Within each cohort group, we take sex and stereotype as prior and education as the dependent or sink variable. The diagram shows double-headed arrows between sex and stereotype since they have no plausible causal order. Statistical analysis, however, shows that sex is unrelated to stereotype in any cohort. We will, therefore, end up with a Simon-Blalock model with a zero coefficient that enables us to avoid choosing a causal direction for these two variables. The d-system approach requires that one category of each variable be chooser as a base or reference point (Davis, 1976). We chose male for sex, Yankee for stereotype, and less than high school for education.

to) Schematic
Wilhin cohor,
Sel
: ~~Educotion
~Stereolype

(b) Cotegories WBhin cohor!

<:emole~ sChopl

Northern Slock~
\ ~hern Blo~ > College
Other
Sose cotegories Sex = mole; SterePlype = 10nkee; Educolron= Less thon high schPPI

Eig.1. D-system model for hackground variables.

The system is thus set up to pose these questions: How did the educational attainments of Yankee males change from cohort to cohort? How did the educational attainments of women and members of various ascriptive social groups compare with those of Yankee men? How have these attainment differences changed from cohort to cohort? To what extent has cohort turnover increased or decreased social differences in educational attainment?

Data

To estimate the parameters in the model we used 30 national surveys spread over the years 1952 to 1973. Ten are University of Michigan election year studies for the even years 1952 to 1972 (save for 1954 when only a token survey was fielded); the remaining 20 surveys, two a year in odd years, are American Institute of Public Opinion (Gallup) surveys.[2] (No 1953 AIPO file was available in the project files when analysis began.)

The Michigan studies are multistage area probability samples with predesignated respondents. The AIPO studies are multistage area probability samples with quotas applied at the final level.[3] All are designed to represent the adult U.S. population living in noninstitutional quarters. Although the designs differ in terms of textbook sampling theory, the two sets "should" give similar results for similar questions. Therefore, we will pool all 30 sets in

[2] Michigan studies were obtained through the courtesy of the Interuniversity Consortium for Political Research; the AIPO studies were provided by the Roper Public Opinion Research Center, Wflliamstown, Massachusetts.

[3] Eor a discussion of AIPO sampling designs, see Hastings, 1975, Appendix A-1.

TABLE 4
Studies and *N* Values

Year	Michigan Election Series	AIPO study number	*N*
1973		863, 868	2,987
1972	x		2,679
1971		825, 839	3,056
1970	x		1,621
1969		778, 784	3,083a
1968	x		1,608
1967		744, 747	2,992a
1966	x		1,278
1965		710, 713	3,214a
1964	x		1,631
1963		671, 675	3,168a
1962	x		1,284
1961		646, 649	3,008a
1960	x		837
1959		610, 614	3,069
1958	x		1,282
1957		586, 589	3,014
1956	x		1,730
1955		543, 546	3,052
1954			
1953			
1952	x		1,763
Total	10	20	46,356

aOriginal files are weighted in terms of "times at home" and education. For this analysis the files were adjusted to remove the weights (i.e., to count each case in the file equally).

the analysis, while keeping a weather eye open for systematic differences between odd and even years.

Table 4 lists the files, dates, and N.

Estimating the Parameters

All numbers (source values, arrow coefficients, intercept constants) in the model are proportions, p, or differences in proportions, *d.* In each survey the values are estimated from cross-tabulations, as explained in Davis (1976). In addition, we must decide whether, across the 20 years of studies, each value is always zero(for *d,* not of course, for p),is always some constant, *d* or p, shows a linear trend, or shows a nonlinear trend. The statistical rationale, drawing on results in Goodman (1963), is given in the Appendix. Assuming the reader to be familiar with the Appendix, we proceed to consider the estimates.

TABLE 5
Proportion in Each Cohort

Year	World War I	Interwar	World War II and Korea	Vietnam
1973	.121	.245	.290	.345
1972	.144	.250	.276	.329
1971	.161	.249	.299	.290
1970	.176	.278	.279	.266
1969	.175	.306	.337	.183a
1968	.211	.304	.319	.165
1967	.225	.318	.334	.124a
1966	.226	.333	.310	.131
1965	.267	.324	.312	.096
1964	.252	.342	.334	.073
1963	.272	.336	.342	.050
1962	.298	.354	.323	.024
1961	.333	.344	.298	025a
1960	.326	.378	.296	-
1959	.325	.361	.314	-
1958	.332	.397	.271	-
1957	.344	.399	.257	-
1956	.335	.412	.252	-
1955	.338	.418	.244	-
1954	-	-	-	-
1953	-	-	-	-
1952	.431	.410	.159	-

aOutlier in Table 4.

Cohort

Table 5 gives the cohort proportions, 1952 to 1973. The figures document the facts of life. Proportions in the World War I and Interwar cohorts decline steadily, while proportions in the two younger groups increase.

Table 6 reaches the same conclusions in a more rigorous fashion. To illustrate the techniques explained in the Appendix, we will walk through the numbers in some detail.

Row 1 gives the pooled proportions, p^, estimated as a weighted average of the yearly proportions, with weights inverse to the estimated sampling variance of each proportion.4

4Since the studies are multistage (cluster) samples, textbook formulas will tend to underestimate the sampling variances (Kish, 1957). Using the common empirical rule of thumb that clustered variances are twice as large, the computer was programmed to multiply all estimated variances by 2 (or, when relevant, all standard deviations by 1.414) here and throughout the analyses in this paper.

TABLE 6

Statistical Tests for Data in Table 3

Cohort

Vietnam

World War I WWII and(-19611967,

ResultTotal(-1955)InterwarKoreaTotaland 1967)1969)

(1)p.238.233.328.290.091 .117 .112

(2) Test B

Chi square920.7849.5328.2219.6 1786.5 1408.2 136.0

Probability<.001<.001<.001<.001 <.001 <.001 <.001

(3) Test C

Chi square36.223.417.6144.983.8 26.8 11.4

Probability.007.138.515<.001<.001 <.002 .182

(4) Test B - C

Chi square884.5826.1310.674.71702.71381.4 124.6

Probability<.001<.002<.001<.001 <.002 <.001 <.001

(5) r2957.969.943.321.961 .979 .989

(6) b -.0135 -.0142 -.0090 +.0041 +.0256 +.0288 +.0295

(7) Type IIIb 111 III IIIb IIIb IIIb III

Row 2 with its large chi squares and small probabilities shows that none of the series can be fitted with its constant, p^.

Row 3 tells us what happens when we try to fit the four series with linear trend lines, using weighted least squares, with weights for each year inverse to the sampling variance. Taking each cohort in turn: The data for World War I do not fit (p < .01). Inspection of the computer output, presenting the contribution of each year to the total chi square, however, showed that 1955 contributed 11.127 to the total chi square of 36.2. After removing this apparent outlier (second column from the left), we get a nice fit for the linear estimates, with the probability of .138 and an r2 of .969.5 The Interwar estimates give a "classic" type 111 or linear trend result, with Test C producing a probability of .515 and an an r2 of 943 The World War 11 and Korea data give a bad linear fit (probability for Test C < .001), although the B-C test shows significant improvement. Inspection of the raw data in Table 5 tells us why: The trend is curvilinear. Proportions in this cohort increase from 1952 to 1969, but decrease afterwards. The Vietnam cohort gives a bad fit on

Sr2 values were computed from unweighted regression, since we were in a hurry and we were not sure how to obtain them from our weighted regression computer program. Weighted and unweighted regressions give virtually identical results since the N values are spread out evenly across the study years.

the initial run. After removing 1961 and 1967 (with individual chi squares of 17.783 and 19.487) we get some improvement and finally obtain linear fit after removing 1969, which had an individual chi square of 13.589 on the second run. (Readers nervous about plucking outliers will note the consequences are merely to change the regression slopes in the third decimal.)

Three of our four series show quite reasonable linear trends (see r2 values in row 5 of Table 6). To avoid curve fitting, we can obtain the World War II and Korea proportions by subtraction. Thus, equations for the four cohort proportions, 1952 to 1973:

$P_{\text{World War I}} = 1.17 - .0142 * (\text{Year} - 1900)$ (1)
$P_{\text{Interwar}} = .91 - .0090 * (\text{Year} - 1900)$ (2)
$P_{\text{Vietnam}} = -1.81 + .0295 * (\text{Year} - 1900)$ (3)
$P_{\text{World War II and Korea}} = 1.00 - [(1) + (2) + (3)]$ (4)

(Equation 3, of course, should not be applied for years prior to 1962, since it will give negative proportions.)

Equations 1 through 4 have two uses. Technically, they enable one to produce estimated cohort distributions for any year one chooses, perhaps to check other data sets against the model, perhaps for simulations of the sort presented later in this paper. Substantively, they document perhaps the strongest social change in America in the last 20 years. While the coefficients look small, a steady decrease of 2.3% each year (.0142 + .0090 = .0232) in the older cohorts (and hence a 20-year decrease of 46%) represents a massive change, at least by comparison with other parameters in the model. Table 7 illustrates and gives one some "feel" for the fit of the estimates.

In 1952, 87.4% of the adult population came from cohorts born in

TABLE 7

Cohort Proportions in Selected Years

Cohort	Source	Year 1952	1962	1972
World War I	Raw data	.431	.298	.144
	Equations	.432	.290	.148
Interwar	Raw data	.410	.354	.250
	Equations	.442	.352	.262
World War 11 and Korea	Raw data	.159	.323	.276
	Equations	.128	.339	.276
Vietnam	Raw data	-	.024	.329
	Equations	-	.019	.314

1923 and before. By 1972, 59.0°,to were from cohorts born after 1923, and almost a third (31.4%) were from cohorts reaching age 16 in 1956 and after. The changing of the guard is not total (the groupings were chosen to prevent this), but it is substantial.

Sex

The second parameter to be estimated is the proportion female within each cohort. Table 8 gives the raw figures.

Table 9 summarizes the statistical analysis of sex composition over time within the four cohorts.

The Interwar and Vietnam results are clearly of the type designated Case II in the Appendix. We can fit the data handily with pooled constants of .519 for Interwar and .542 for Vietnam.

For the World War I cohort, things are more complex. On the first run, the data fell in Case IV (i.e., significant fluctuations across the years but no linear trend). Inspection of the raw numbers in Table 8 suggests differences between the odd years (AIPO) and the even ones (Michigan). Analyzing the two groups separately confirms this impression. The AIPO results can be fitted

TABLE 8
Proportion Female by Cohort and Year

Year	World War I	Interwar	World Warll and Korea	Vietnam
1973	.460	.479	.538	.527
1972	.622	.581	.552	.542
1971	.460	.504	.544	.505
1970	.588	.550	.531	.599
1969	.423	.480	.543	.534
1968	.562	.541	.577	.568
1967	.408	.478	.551	.580
1966	.553	.548	.550	.577
1965	.480	.519	.541	.542
1964	.580	.509	.564	.537
1963	.471	.518	.539	.554
1962	.561	.510	.588	.432
1961	.474	.528	.545	.557
1960	.532	.533	.590	
1959	.449	.498	.579	
1958	.493	.534	.571	
1957	.450	.537	.590	
1956	.514	.552	.605	
1955	.431	.521	.571	
1954				
1953				
1952	.543	.527	.565	

TABLE 9
Statistical Tests for Data in Table 5

Result	Total	World War I SRC	AIPO	Interwar	WW 11 and Korea	Vietnam
(I)p	.486	.552	.451	.519	.558	.542
(2) Test B						
Chi square	74.3	10.5	7.7	20.3	10.7	8.8
Probability	.001	.310	.568	.501	.968	.990
(3) Test C						
Chi square	78.1	4.8	7.7	19.6	6.0	8.3
Probability	.001	.779	.532	.353	.996	.689
(4) Test B - C						
Chi square	-	5.7	0.0	0.7	4.7	0.5
Probability	-	<.02	N.S.	>.30	<.05	>.30
(s) r2	.023	.611	.001	.023	.427	.029
(6) b	+.0018	+.0039	.0000	-.0009	-.0024	-.0018
(7)Type	IV	11 ? 111 ?	11	11	111-111	11

by a constant, p = .451. The Michigan data hover between a constant, p = .552, and a linear trend. Parsimony argues for the constant, but demography (sex differentials in mortality in the later ages) suggests the linear trend is more apt, and we chose the latter.

Why the odd-even difference? We do not know for sure, but suspect that the quota element in the AIPO samples may force a constant sex composition in all age groups, whereas the full-probability methods in the Michigan series allow the changing sex composition in the oldest cohorts to come through.

The results for the World War II and Korea cohort are also a bit ambiguous. A constant p of .558 fits very well, but there is a significant (p < .05) linear trend, a decrease in females over time. Sticking with demography against parsimony, we opt for the constant.

We are now ready to state the next four parameters in the model: the proportion female in each of the cohorts.

World War I cohort = .31 + .0039 * (Year - 1900) (5)
Interwar cohort = .5 19 (6)
World War II and Korea cohort = .558 (7)
Vietnam cohort= .542 (8)

Stereotype

To add the third variable, the region-race-religion typology, we must examine: (a) correlations between sex and stereotype categories in each cohort; and (b) the intercept values (i.e., the stereotype category proportions for males in each cohort).

Luckily there is no association between sex and stereotype in any of the cohort groups. When we ran the percentage differences for female vs male by Northern Catholic, Northern Black, Southern White Protestant, and Other we found that d = .000 fits the data in each cohort-by-year data set. Substantively, this confirms one's expectation that sex ratios are identical in subcultural groups. Technically, it absolves these survey houses from the occasional allegation that they underrepresent black males. Methodologically, it allows us to use sex and stereotype in the same model without worrying about causal direction.

The results for the intercepts are rather interesting but need some explanation for readers unfamiliar with d-systems. Consider, for example, Fig. 2, the results for sex and stereotype in the Interwar cohort. For the moment, we take the numbers at face value without worrying about how they were estimated. We arbitrarily treat sex as the prior variable with a value of .519 (Eq. 6). In theory there should be arrows running from female into each of the five nonbase categories, but we know their values to be zero, so we leave them out. The technique requires each category of the second variable to have an intercept value, namely the proportion in the category for those in the base category of the source (i.e., men). Thus, the .211 intercept value for Catholic means that .211 of the males in the Interwar cohort are Northern White Catholics.

To find the total proportion Catholic following flow graph principles, we multiply the source value times its arrow and add the residual: (.519 * .000) + (.211) = .211. That is, the intercept values are estimates of the

211

035

~ 1 zo~
Cr'tho | /
Nor~hern 310ck /

S1S Fennole: Sou~hern Wh'le Protesiont
~ Southern 310ck

Olher ~

t \043

108

Fig. 2. Sex and stereotype group within interwar cohort.

TABLE 10
Intercept Values for Stereotypology

	Cohort							
	World War I		Interwar		World War 11 and Korea		Vietnam	
Category	p^	crp	p	crp^	p	op^	p	,,p
Catholic	.153	.0065	.203	.0066	.214	.0075	.211	.0089
Northern Black	.026	.0029	.035	.0030	.048	.0039	.038	.0055
Southern White Protestant	.224	.0075	.207	.0067	.195	.0072	.191	.0112
Southern Black	.049	.0039	.045	.0034	.038	.0035	.042	.0058
Other	.101	.0054	.110	.0051	.113	.0058	.190	.0112
Yankee (by subtraction)	.447		.400		.392		.328	

marginal proportions for the stereotype groups in the special case where the second variable is independent of the first.

When the intercept proportions were run across years within cohorts, all 20 runs (fne nonbase categories of stereotype by four cohort groups) fell in Case II. Thus, the data show constant stereotype proportions over time within cohorts. Table 10 summarizes.

In the absence of interactions we can estimate the standard deviations of the sampling distributions of p values (as explained in Davis, 1976; see also footnote 4 of this paper). These are shown below and to the right of the pooled proportions. If we now think of each row in Table 10 as a set of k proportions with known standard deviations, we can use the techniques explained in the Appendix to test whether the proportions are homogeneous; that is, whether the stereotype categories are independent of cohort.

The four Catholic proportions are not homogeneous (chi square for Test B = 75.7 for 3 d.f.). However, when the .153 value for World War I is removed, the other three do not differ significantly from p= .211 (chi square = 1.6, p = .667). Presumably, these results reflect the direct and indirect contributions of turn-of-the-century immigration, such that the Northem Catholic proportion increased .058 in cohorts born after 1906.

Proportions for Northem Black vary significantly among cohorts (chi square = 21.0, p < .001), showing a general but not perfect pattem of increase from earlier to later cohorts. The increasing proportion Northern Black in the population presumably reflects the epochal migration of Southern Blacks to northem cities under the impetus of industrialization and war-induced economic opportunities.

The proportions for Southern White Protestant (.224, .207, .195, .191) are significantly heterogenous (chi square = 9.9, p = .019) and show a pattem of steady, albeit small, declines.

For Southern Blacks, the four proportions are homogeneous. The pooled value, p = .043, fits the data (chi square = 4.7, p = .192).

The residual category, Other, is not homogenous (chi square = 52.5, p < .001). When, however, the Vietnam cohort value, p = .190, is removed, the other three are easily fitted with the constant, p = .108 (chi square = 2.6, p = .536). The proportion Other was a steady .108 until the Vietnam cohort, where it jumped to .190. Since Other is a melange of Jews, Southern White Catholics, Other Religions, and No Religion, this trend is difficult to interpret.

We can now add the following parameters for stereotype intercept values to the model:

Catholic (if cohort = WW 1) = .153 (9a)

	All 0ther	= .211	(9b)
Northern Black	(World War 1)	= .026	(lOa)
	(Interwar)	= .035	(lOb)
	(World War 11)	= .048	(10c)
	(Vietnam)	=.038	(lOd)

Southem White Protestant (World War 1)	= .224	(lla)
(Interwar)	= .207	(11 b)
(World War II)	= .195	(llc)
(Vietnam)	= .191	(11 d)
Southem Black	= .043	(12)

Other	All 0ther	= .108	(13a)
	(Vietnam)	=.190	(13b)

To find a theme in these 13 numbers, let us calculate the proportion Yankee in the four cohorts by summing the category intercepts and subtracting their total from 1.000 (thus, Table 11).

Because other categories increase in newer cohorts, the proportion Yankee (non-Southern White Protestant) must decline. In the cohorts bom in

TABLE 11
Implied Proportions in 9a-1 3b

Cohort	Yankee	Southern White Protestant	Total White Protestant
World War 1	.446	.224	.670
Interwar	.396	.207	.603
World War 11	.395	.195	.590
Vietnam	.327	.191	.518

1906 or earlier, .446 were Yankees and two^thirds (.670) were White Protestants. By the time we reach persons bom in 1940 and after, Yankees had dropped to .327 and White Protestants to .518.

Equations 9a-13b and Table 11 underline a rather dramatic aspect of American social history in this century. At the tum of the century something like two^thirds of the population was White Protestant and 45% were in the Yankee subculture that has dominated American life from the Pilgrim Fathers to Gerald Ford. Under the onslaughts of Catholic immigration, black fertility, and in the case of other possibly rapid secularization, the figures among the youngest adults are rather different: one^third Yankee and a scant half (.518) White Protestant.

Education

To analyze educational attainment, its association with cohort, and its relationship with sex and stereotype, we begin with the intercept values–the educational attainment of Northern White Protestant Males. Table 12 gives the raw data.

The proportions are "constant" across time (as one would expect, since relatively few persons add substantial schooling during the adult years) and quite different among cohorts (each pair of adjacent proportions at the bottom of Table 12 is significantly different). Table 13 helps us see the pattern.

In the World War I cohort (persons reaching age 16 in 1922 or before), two-thirds (.687) failed to complete high school; of those who did complete high school, about half went on to a year or more of college.

For the Interwar cohort (persons reaching age 16 between 1923 and 1939), high school completion increased .270 (.687 - .417 = .270). However, only 41% of these high school graduates went on to college (.239/.239 + .344 = .410), possibly because of the Great Depression, though we have no hard evidence on this. Consequently, the college proportion increased .084.

For the World War II cohort (persons reaching age 16 between 1940 and 1955), there was another jump in high school completion (.417 - .296 = .121) and a retum to the "half college" proportion among high school graduates. This gives a distribution that is essentially even across the three categories of education and a .101 increase in college.

For the Vietnam cohort (persons reaching age 16 in 1956 and after), high school completion shows further growth (.296 - .150 = .146) and the college-high school ratio is again close to 50-50.

This extraordinary improvement in educational attainment is, of course, familiar, but one is always impressed with the figures. For example, the proportion of high school "drop outs" in the Vietnam cohort is about the same as the proportion of college attenders in the World War I category!

TABLE 12
Intercept Values for Education (Proportions among Yankee Males)

	High school				College			
Year	World War I	Inter- war	World War 11	Vietnam	World War I	Inter- war	World War 11	Vietnam
1973	.143	.242	.414	.382	.167	.341	.345	.455
1972	.110	.296	.307	.368	.096	.278	.474	.488
1971	.280	.365	.466	.504	.160	.189	.317	.307
1970	.170	.481	.354	.400	.132	.182	.378	.422
1969	.163	.324	.421	.481	.114	.258	.273	.405
1968	.105	.337	.434	.409	.140	.297	.368	.500
1967	.142	.408	.404	.400	.213	.279	.335	.475
1966	.109	.205	.230	.414	.152	.288	.328	.552
1965	.185	.314	.407	.548	.180	.276	.313	.310
1964	.145	.325	.352	.353	.203	.265	.375	.294
1963	.226	.444	.366	.346	.164	.213	.354	.324
1962	.175	.300	.340	.455	.105	.243	.358	.455
1961	.171	.371	.399	.385	.163	.254	.286	.462
1960	.113	.473	.238	-	.094	.236	.381	-
1959	.175	.350	.390	-	.187	.230	.349	-
1958	.146	.416	.414	-	.180	.303	.343	-
1957	.137	.353	.393	-	.149	.188	.246	-
1956	.184	.365	.417	-	.149	.222	.264	-
1955	.155	.356	.412	-	.151	.189	.247	-
1952	.119	.263	.236	-	.155	.234	.327	-
(l)p	.158	.344	.381	.425	.155	.239	.324	.415
(2) 0	.010	.013	.014	.026	.010	.011	.014	.025
(3) Test B								
Chi square	12.4	23.4	16.3	5.6	8.3	13.6	13.7	9.4
Probability	.928	.322	.751	.999	.993	.885	.882	.986
(4) Type	I	I	I	I	I	I		

TABLE 13
Education and Cohort for White, Northern, Protestant Males (Modeled Data)

	Cohort			
Education	World War I	Interwar	World War 11 and Korea	Vietnam
College	.155	.239	.324	.415
High school	.158	.344	.381	.425
Grade schoola	.687	.417	.296	.150
	1.000	1.000	1.000	1.000

aBy subtraction

We can now add the following parameters (education intercept values) to the model:

High school =	(World War I	.158	(14a)
	(Interwar)	.344	(14b)
	(World War 11 and Korea)	.381	(14c)
	(Vietnam)	.425	(14d)
College =	(World War 1)	.155	(15a)
	(Interwar)	.239	(15b)
	(World War II and Korea)	.324	(15c)
	(Vietnam)	.415	(15 d)

Now let us see how the numbers work out for Americans who are female, instead of male, or in stereotype groups other than Yankee. Table 14 gives the estimates for sex differences.

The first row in Table 14 gives the pooled sex difference in high school within cohort and across years. For example, the +.055 in the left-hand column means that females were 5.5% more likely to report high school than

TABLE 14
d Values for Sex and Education

	World War I	Interwar	World War 11 and Korea	Vietnam
High school				
(l)d	*+.055*	+.048	+.113	+.085
(2) *·7d*	.009	.010	.012	.018
(3) Test A				
Chi square	42.9	31.7	112.4	24.5
Probability	.005	.083	.001	.322
(4) Test *B*				
Chi square	7.3	9.3	17.2	4.2
Probability	.997	.987	.698	.999
(5) Type	11	11	11	1
College				
(l)d	+.011	-.037	_ 097	-.137
(2) *·7d*	.009	009	.010	.018
(3) Test A				
Chi square	13.4	26.1	102.4	62.0
Probability	.921	.248	.001	.001
(4) Test *B*				
Chi square	11.8	7.5	11.1	5.7
Probability	.945	.997	.960	.999
(5) Type	I	I	11	11

males, averaged across typology groups and years, within the World War I cohort.[6]

For high school, all four *d* values are positive. Women are consistently more likely to report high school graduation. The *d* value for Vietnam is not strictly significant ~(probability for Test A = .322), but we will keep it on the basis of comparison with census data (to be discussed later). The *d* values are not homogeneous (chi square = 21.9, $p < .001$ for Test B), but fall into two clumps. We can fit the data nicely with *d* values of +.052 for World War I and Interwar and +.105 for World War II and Korea. Thus,

dFemale-high school = (Interwar and Wor,}d War I) +.052 (16a)
(World War 11 and Korea
and Vietnam) +.105 (16b)

For college, the *d* valuos form a progression from virtually zero to negative; that is, men and women showed no difference in proportion college in the World War I group, but in later cohorts women were less likely to report college (chi square = 93.7, $p < .001$ for Test B on the *d* values). We decided to take the Interwar *d* at face value, although it is not significant, giving the following:

dFemale-college = (World War I)	·0OO	(17a)
(Interwar)	- ·037	(17b)
(World War II and Korea)	-.097	(17c)
(Vietnam)	- .137	(17d)

To interpret these findings, we can add Yankee females to the modeled data in Table 13, obtaining the results for females: by adding or subtracting the appropriate values from Eq. 14a through 15d. The results appear in Table 15.

The third row, grade school, is a good place to begin. We see that in the World War I cohort females were more likely to be graduated from high school *(d* = .052), but their advantage declines with later cohorts and reverses for the Vietnam group. Thus, a traditional educational advantage of females over males has eroded and possibly reversed.

Turning to the proportion of high school graduates going on for one or more years of college (estimated by the proportion college divided by the sum of college and high school), we see negative signs (female high school graduates are less likely to go on to college) and increasing magnitudes (the feminine deficit is stronger in the newer cohorts). However, for reasons discussed below, we should not take the difference between -.097 and -.137 too seriously.

6With three variables, sex-stereotype-education, the question of interaction effects arises, but they turned out to be negligible in the data sets we ran. In the dozens of conditional d values run for this section of the analysis, only six significant interactions occurred (applying Test B across conditions rather than times) and they showed no consistency across years or effects.

TABLE 15
Education, Sex, and Cohort for White Northern Protestants

		Cohort			
Education	Sex	World War I	Interwar	World War 11 and Korea	Vietnam
College	Male	.155	.239	.324	.415
	Female	.155	.202	.227	.278
Female	*d*	.000	-.037	-.097	-.137
High school	Male	.158	.344	.381	.425
	Female	.210	.396	.486	.530
Female	*d*	+*.05* 2	+.052	+.105	+.105
Grade school	Male	.687	.417	.296	.150
	Female	.635	.402	.287	.192
Female	*d*	-.052	-.015	-.009	+.042
College, along high school,	Male	.495	.410	.460	.494
and college	Female	.425	.338	.318	.344
Female	*d*	-.070	-.072	-.142	-.150

Since the female advantage in high school has eroded and the female disadvantage in college among high school graduates has increased, the result is a sharp, progressive growth of the sex difference in college.

In absolute terms, female educational attainments have improved steadily from cohort to cohort, but not as fast as male attainments. Consequently, the relationship between sex and education has shifted from curvilinear (males more likely to drop out of high school and more likely to enter college) to a clear-cut male advantage.

Next, let us apply the same analysis to the five stereotypology groups: Northern Catholics, Northern Blacks, Southern White Protestants, Southern Blacks, and Others. The detailed statistical results will not be presented to save space; instead, Table 16 gives the final estimated coefficients after applying the tests in the Appendix to the time series data (finding no time trends within cohorts) and then testing the intracohort effects for homogeneity.

To interpret the figures, let us look at high school completion, summing the coefficients for college and high school to get differences in high school completion between Yankee males and those in various other stereotype categories (Table 17). Beginning with the proportions and reading across the rows of Table 17, we see a steady increase in high school for all groups. In particular, the two black groups move from less than 10% in World War I to

TABLE 16
Coefficients for Typology and Education, within Cohort

Coefficient	Cohort: World War I	Interwar	World War 11 and Korea	Vietnam
College				
Southern Black	-.144	-.144	-.144	-.144
Northern Black	-.121	-.121	-.121	-.121
Southern White Protestant	-.030	-.030	-.030	-.030
Northern Catholic	-.071	-.071	-.071	.000
Other	.000	+.038	+.148	+.148
High school				
Southern Black	-.145	-.291	-.263	-.101
Northem Black	-.104	-.195	-.136	.000
Southern White Protestant	-.065	-.122	-.067	.000
Northern Catholic	.000	.000	.000	.000
Other	-.028	-.028	-.150	-.150

TABLE 17
Coefficients and Proportions for High School Diploma among Males
(Summed from Table 11)

Category	Cohort: World War I	Interwar	World War 11 and Korea	Vietnam
p				
Southern Black	.024	.148	.298	.595
Northern Black	.088	.267	.448	.719
Southern White Protestant	.218	.431	.608	.810
Northern Catholic	.242	.512	.634	.840
Other	.285	.593	.703	.838
Yankee	.313	.583	.705	.840
d				
Southern Baclk	-.289	-.435	-.407	-.245
Northern Black	-.225	-.316	-.257	-.121
Southern White Protestant	-.095	-.152	-.097	-.030
Northern Catholic	-.071	-.071	-.071	.000
Other	-.028	+.010	-.002	-.002
Mean	-.142	-.193	-.167	-.080

strong majorities in the Vietnam cohort. Southern Blacks, by 1940 (World War II and Korea) had high school completion rates as high or higher than Yankees in the World War I groups! But, since Yankee proportions were also increasing, it is equally useful to look at the columns, especially for d values. From left to right:

In the World War I cohort, the stereotype groups were strung out in a clear-cut order: Yankee-Other-Northern Catholics-Southern White ProtestantsNorthem Black-Southern Black.

In the Interwar cohort, Yankee high school completion jumped .270, Others and Catholics increased about the same amount, but Blacks and Southern Whites improved less, so their d values became more negative. As a result, the subgroup differences in high school around 1940 were as strong as or stronger than they had been at the turn of the century.

In the World War 11 and Korea cohort, the gap began to close for Blacks and Southern Whites, while Catholics kept the same distances from Yankees and others dropped trivially.

Finally, in the Vietnam Cohort, the Catholic disadvantage disappears, the Southern White coefficient nears zero, and the Black coefficients drop substantially. But the Black coefficients are still far from zero. The gap between Northern Blacks and Yankees is as large as the Catholic gap for World War 1. The gap for Southern Blacks in the Vietnam group is about as large as their World War I baseline.

In sum, all social groups in the United States have shown a phenomonal increase in high school completion in this century, but the gap between Yankees and other major groups remained substantial until after World War 11. In the Vietnam cohort, differences among the three White categories are trivial, but Blacks, especially Southern Blacks, are far from equality in high school completion.

Table 18 gives the relevant figures for college. We can use the figures in Table 18 to summarize the pattems for each group.

TABLE 18

Estimates of Proportion of High School Graduates Going on to College (Among Males)

	Cohort			
Category	World War I	Interwar	World War 11 and Korea	Vietnam
Proportions				
Southern Black	.458	.642	.604	.455
Northern Black	.386	.442	.453	.409
Southern White Protestant	.573	.485	.484	.475
Northern Catholic	.347	.323	.399	.494
Other	.544	.467	.671	.672
d (vs Yankee)a				
Southern Black	-.037	+.232	+.144	-.039
Northern Black	-.109	+.032	-.007	-.085
Southern White Protestant	+.078	+.075	+.024	-.019
Northern Catholic	-.148	-.087	-.061	-.001
Other	+.049	+.057	+.211	+.178

aSee Table 12 for Yankee proportions.

Southem Blacks are consistently disadvantaged in high school completion, especially the Interwar and World War II and Korea cohorts. However, among those with high school diplomas, the college rate is not consistently below Yankees. The key problem for Southern Blacks is high school completion.

Northem Blacks show a declining disadvantage in high school completion and a persistent, albeit small and not entirely consistent, disadvantage in college.

Southern White Protestants, despite lower high school completion, show higher college proportions for high school graduates in the early cohorts, but the differences disappear over time. In the Vietnam cohort their educational pattern is essentially the same as Yankees.

Northem Catholics, prior to Vietnam, were lower in both high school and in college attendance among high school graduates. In the Vietnam cohort these differences vanish and their educational pattern is virtually the same as for Yankees.

Others show little difference in high school completion, but an increasing advantage in college. Since the group is so heterogeneous, we cannot interpret the result.

In a nutshell, subgroup differences in educational attainment declined very little prior to World War II. Differences among White groups vanished in the Vietnam cohort, leaving a declining but nontrivial race difference as the remaining obstacle to equality in education among major social groups in contemporary America.

Validity

The data have a certain amount of *prima facie* credibility because they include 46,356 cases from 30 national surveys by two outstanding survey organizations. Furthermore, the figures show a pleasing amount of internal consistency: the linear trends are smooth and the fixed parameters show only insignificant wobbles from year to year. Save for sex composition in the World War I cohort, the Michigan Election and AIPO Studies give virtually identical results, despite the differences in their sampling techniques. Nevertheless, it is useful to compare our parameters with other data sets. To illustrate, we shall use the U.S. Census 1970 Public Use Sample and the NORC General Social Surveys for 1972, 1973, and 1974.[7]

To proceed, we draw on the following useful property: with d-systems (like log linear models) one can always construct a full cross-tabulation table

[7]The General Social Survey (GSS) is an annual sampling of continental U.S. adults (18 years of age and older) living in noninstitutional quarters. The survey employs a multistage area probability sample with quotas at the block level and is supported by a grant from the National Science Foundation.

TABLE 19
Cohort Proportions, Modeled and Census, 1970

Cohort	Model Proportion	Sampling standard deviation	Census	Differencea	Difference/ sigma
Vietnam	.255	.005	.299	(+ 044)b	8.800
World War 11 and Korea	.289	.006	.276	(-.013)	2.167
Interwar	.280	.006	.261	(-.019)	3.167
World War 1	.176	·004	.164	(-.012)	3 000

aMean absolute difference = 1.021.

bDifferences in parentheses are significant at the .05 level.

from the system parameters and an appropriate N8 That is, one can choose an arbitrary year and *N* and proceed to construct the full 144 cell (4 X 2 X 6 X 3) cross-tabulation for cohort-sex-stereotype implied by the model or any appropriate table collapsing variables or categories. To illustrate, let us consider the 1970 census sample, beginning with cohort. Table 19 gives the necessary figures.

To fill in Table 19, we:

(1) Obtained the cohort distribution (recoding age) in the 1970

Census sample, *N* = 13306.

8For example, given some appropriate N value and year we can find the sub-N values for each cohort from Eq. 1-4. Within each cohort we wish to construct a cross-tabulation for sex by stereotype. The marginal proportions for the four stereotype groups are given in Eq. 9a-13b, Yankee can be found by subtraction, and N values can be found by applying these proportions to cohort N values. Within each stereotype group, the proportion female is given in Eq. 5-8 and the proportion male by subtraction from 1.000 (the sex proportion is, of course, identical in each stereotype group within a cohort). Applying these proportions to the stereotype category N values, we have the frequencies for sex by stereotype within each cohort. We now wish to construct a table within each cohort where the rows are sex by stereotype and the columns are education. We begin with Yankee males, applying the proportions in Eq. 14a-lSd to get high school and college, and then obtaining grade school by subtraction. For males in other stereotype categories, we add or subtract the numbers in Table 16 for high school and college, and then find grade school by subtraction. We now have the educational proportions for males in all stereotype groups. For females, we add or subtract the coefficients in Eq. 14a-lSd with the results for males in the same stereotype category. When the educational proportions are applied to the *N* values in the sex by stereotype rows, we will have a frequency table for cohort-sex-stereotype-education that has the desired system parameters and adds up to the desired *N.* Richard Elmore, Dartmouth, 1977, has written a generalized computer program (in Dartmouth Basic) for producing tables from d-system parameters. For information, please write the author, Department of Sociology, Dartmouth College, Hanover, N.H. 03755.

(2) Found the model's cohort distribution for *N* = 13306, year = 1970, from Eq. 1-4;

(3) Calculated the standard deviations of the theoretical proportions for *N* = 13306 (as usual, correcting for multistage sampling, perhaps overcorrecting here);

(4) Found the differences between the observed (census) and expected (model) proportions;

(5) Divided the differences by the sigmas in step (3) recognizing as significant (at .05) any two-sigma discrepancies.

Table 19 indicates significant differences between the model and the 1970 Census. Census results have a greater proportion in the Vietnam cohort and smaller proportions in the older cohorts. In absolute terms, however, the results are not bad since the average absolute discrepancy is .022.

One's impulse is to infer that the model is biased since census results are generally considered the yardstick. However, more detailed study is called for, because a number of other possibilities for explaining this discrepancy come to mind:

The model data are spread out over 20 years, whereas the census data are collected for one specific month.

The model data are based on personal interviews with respondents, while the census data are reports by household informants.

Survey interviewers are generally more experienced than census enumerators.

The census universe probably includes "group quarters" cases (e.g., soldiers, students, prisoners) generally excluded from survey sampling frames.

Question wordings may not be identical.

Similar analyses can be made for other variables in the model. It is generally more instructive, however, to examine nonsource variables in terms of model parameters. Consider, for example, sex and education in the 1970 Census. John Fry has written an interactive computer program (MODDRIVE) that generates cross-tabulations for the full model or a model with variables collapsed out, given arbitrary *N* values and years. The program may be used as follows:

(1) The appropriate cross-tabulation is run off on the test data.

(2) The MODDRIVE program creates a cross-tabulation for the same *N* and year.

(3) The modeled data are run through CATFIT, a computer program for statistical analysis of d-systems. CATFIT gives the parameters and their standard errors (i.e., the sampling statistics for samples of exactly that size from a universe with the model's parameter values).

(4) The test cross-tabulation is run through CATFIT to obtain the

parameters.

(5) Observed and expected results are compared as in Table 19.

For sex by education within cohort, we have the following parameters: *Base proportior~s: (1)* proportion female (2) proportion high school among males (3) proportion college among males. *d values: (1)* female to high school, (2) female to college. Table 20 gives the results.

To interpret Table 20, consider, for example, the *d* value from female to high school in the Vietnam cohort. The figures tell us:

(a) The model *d* equals +.105.

(b) In samples of size 3977 (the number of Vietnam cases in the

1970 Census sample), this *d* has a standard deviation of .024.

TABLE 20

Sex by Education Within Cohort (Model vs 1970 Census)

		Among males			d values	
Cohort	Result	Female	High school	College	Female High	Female College
Vietnam	Model					
	Value	.542	.392	.427	+.105	-.137
	Sigma	.013	.017	.018	.024	.023
	Census	.516	.384	.395	+.067	-.074
	Difference	(-.026)a	-.008	-.032	-.038	(+.063)
	Difference/sigma	2.000	.471	1.778	1.583	2.739
World War 11	Model					
and Korea	Value	.558	.334	.307	+.105	-.097
	Sigma	.011	.016	.016	.023	.020
	Census	.525	.310	.323	+.120	-.102
	Difference	(-.033)	(-.034)	+.016	+.015	-.005
	Difference/sigma	3.000	2.125	1.000	.652	.250
Interwar	Model					
	Value	.519	.296	.211	.052	-.037
	Sigma	.011	.016	.014	.021	.018
	Census	.519	.271	.210	.057	-.036
	Difference	.000	-.025	-.001	+.005	+.001
	Difference/sigma	.000	1.563	.071	.238	.056
World War I	Model					
	Value	.583	.131	.128	.052	.000
	Sigma	.014	.016	.016	.021	.020
	Census	.575	.131	.139	.040	.004
	Difference	-.008	.000	+.011	-.012	+.004
	Difference/sigma	.571	.000	.688	.571	.016
	Mean	-.017	-.017	-.002	-.008	+.016

aValues in parentheses are significant at the .05 level.

(c) The actual census cross-tabulation gives a *d* of +.067.

(d) The census *d is* less positive, the discrepancy being -.038.

(e) The difference is not significant at the .05 level.

For Table 20, as a whole, we draw these conclusions:

(1) The model has more females than the census, but only in the World War II and Vietnam cohorts.

(2) The high school and college proportions for the base category male give a pretty fair fit.

(3) The *d* values for sex and education fit nicely, save for the college *d* in the Vietnam cohort, discussed' above. The model gives a strongish -.137 for the female deficit in college in the Vietnam cohort. The census estimate of -.074 is significantly less extreme, although clearly nonzero. The result confirms our decision to treat the value as nonzero, but it suggests that we should be cautious about inferring an accentuation in the *d* from World War II and Korea to Vietnam. (The 1972, 1973, and 1974 NORC General Social Suneys, to be discussed later, give an average of-.110 for this parameter.)

A similar analysis can be made for stereotype and education, although we have to collapse our stereotype groups to White, Northern Black, and Southern Black, respecifying the system with White as the base category (because the census does not ask religion). Table 21 gives the results, *a la* Table 20.

We interpret Table 21 as follows:

(1) The model has fewer Northern Blacks than the census. All four discrepancies are significant as is the average of +.017. No such differences appear for Southern Blacks.

(2) Educational proportions in the base category White are reasonably compatible with the census figures. The census gives about .025 fewer high school graduates and about +.010 more college graduates, but the discrepancies are not uniform or consistently significant.

(3) The *d* values (race differences in education) are generally compatible.

To summarize Tables 19, 20, and 21: the model generally matches the 1970 Census user sample, the important discrepancies being:

(1) The census estimates a higher proportion in the Vietnam cohort, possibly because of differences in sampling frames.

(2) The census shows fewer young females in the Vietnam and World War II and Korea categories. One suspects that the surveys underestimate the notoriously "never-at-home" younger adult males.

(3) The census shows higher proportions of Northern Blacks, but

TABLE 21

Stereotype by Education within Cohort (Model vs 1970 Census)

d values

Proportions		Northern	Southem	Northem	Southern					
Cohort	Result	Northern Black	Southem Black	High schoola	Black Collegea	Black High	Black High	Black College	College	
Vietnam	Model									
	Value	.038	.043	.451	.365	+.033	-.070	-.146	-.168	
	Sigma	.004	.004	.013	.013	.064	.058	.054	.048	
	Census	.060	.041	.412	.375	+.104	+.007	-.169	-.184	
	Difference	(+.022)b	-.002	(-.039)	+.010	+.040	+.077	-.023	-.016	
	Difference/sigma	5.500	.500	3.000	.769	.625	1.328	.426	.333	
World War 11	Model									~
and Korea	Value	.048	.043	.407	.265	-.103	-.232	-.118	-.137	t:
	Sigma	.004	.004	.011	.010	.049	.044	.038	.038	
	Census	.072	.045	.389	287	-.069	-.236	-.136	-.171	~
	Difference	(+.024)	+.002	-.018	(+.022)	+.034	-.004	-.018	-.034	t
	Difference/sigma	6.000	.500	1.636	2.200	.694	.091	.474	.895	
Interwar	Model									~
	Value	.035	.043	.340	.201	-.165	-.259	-.102	-.126	C
	Sigma	.004	.004	.011	.010	.048	.033	.038	.031	~
	Census	.050	.037	.310	.197	-.031	-.216	-.075	-.072	t
	Difference	(+.015)	-.006	(-.030)	-.004	(-.134)	+.043	+.027	+.054	
	Difference/sigma	3.750	-1.500	2.727	.400	2.792	1.303	.711	1.742	
World War I	Model									
	Value	.026	.043	.170	.136	-.088	-.130	-.103	-.126	
	Sigma	.004	.006	.011	.010	.051	.030	.034	.017	
	Census	.034	.050	.159	.149	-.026	-.086	- 096	-.086	
	Difference	(+.008)	.007	-.011	+.013	+.062	+.044	+.007	(+.040)	
	Difference/sigma	2.000	1.167	1.000	1.300	1.216	1.467	.206	2.353	
	Mean	+.017	.000	-.025	+.010	+.001	+.023	-.002	+.012	

aWithin base category, Northern White. bValues in parentheses are significant at the .05 level.

no general race discrepancy. No good reason for the Northern Black discrepancy comes to mind.

Similar analyses were carried out for the 1972, 1973, and 1974 NORC General Social Suneys. No significant discrepancies were observed for d values, but two consistent differences in marginal distributions turned up. Table 22 summarizes.

The General Social Survey, like the census, shows more Northern Blacks than the model, the difference being significant in each year.

TABLE 22
Observed-Expected Proportions, Model vs NORC General Social Survey

General Social Survey

Variable	Category	1972	1973	1974	Average
Cohort	Vietnam	-.003	-.022	-.014	-.013
	(.188)a(1.294)		(.778)		
	World War 11	-.004	+.025	+.003	+.008
	(.250)(1.563)		(.188)		
	Interwar	+.018	+.013	+.002	+.011
	(1.200)(.813)		(.125)		
	World War I	-.011	-.016	+.009	-.006
	(.846)(1.231)		(.818)		
Sex	Female	_ 047 b	-.013	-.012	-.024
	(2.765)(.722)		(.667)		
Stereotype		Yankee-.088b			_ 067b +
065b			-.032		
		(5.176)	(3.722) (3.333)		
	Northem	+.015	+.001	-.007	+.003
	Catholic		(1.071)	(.071)	(.500)
	Northern		+.028b	+ 042b + 03sb	
	+.035				
	Black	(4.000) (6.000)			(5.000)
	Southern White	-.023	+ oo9	_ oggb	-.038
	Protestant	(1.643)	(.643)	(7.071)	
	Southern Black	+ 054b	+ 007	+.004	+.022
		(7.724)	(1.000)	(571)	
	Other	+.014	+.008	+.002	+.008
		(1.273)	(.615)	(154)	
Education		College+ 036b	+ ossb	+.061b	+.051
		(2.571) (3.438) (3.813)			
	High school	_ 043b		_ 040b _ 037b	
		_ 040			
		(2.688)	(2.353)	(2.056)	
	Less than	-.003	-.015	-.024	-.014
	high school	(.176)	(.833)	(1.333)	

aValues in parentheses equal O - E/sigma O. bSignificant at .05 level.

The General Social Survey, unlike the census, shows an excess of college and a deficit in high school, the differences again being significant in each year. Whether these differences come from coding conventions or sampling is unknown.

All other differences are either trivial or inconsistent from year to year.

Substantive Projections

The computer program used to check validity can also be used to draw substantive conclusions. In d-systems, inferences about associations and variable distributions can be drawn following the principles of linear flow graphs (Davis, 1976); but in a system as complex as ours, exactly the same results can be obtained more easily with the computer program. For example, if we wish to estimate the increase in college from 1955 to 1965, we can ask the program to run off the educational distributions for the years 1955 and 1965 with a convenient N, say 1000, and compare the proportions.

The enterprise resembles "computer simulation," but has a crucial difference. Each number in the result is based on a statistical estimate rather than an attractive a *priori* assumption. The closest analogy, perhaps, is the econometric model. In this sense, we use the phrase "survey-metric model" to stress the similarity, while calling attention to an important difference: these models use only the categorical metric characteristic of surveys.

In the light of the results in the previous section, three corrections were made in the parameters:

(1) The intercept values for Northern Black were shifted to the

census estimates (.060, .072, .050, and .034, for Vietnam, World War II, Interwar, and World War I, in Eqs. 10a-lOd), since both the census and the NORC results suggest such a correction. Of mathematical necessity, the Yankee proportions are reduced, a plan that is somewhat arbitrary, but reasonable, since the Northern Catholic results in the General Social Survey match the model.

(2) Proportions female in Vietnam and World War II were changed

to the census values (.516 for Eq. 8 and .525 for Eq. 7).

(3) The coefficient for female to college in the Vietnam cohort

was changed to the census estimate -.074 (Eq. 15d).

Table 23 shows results for various runs of the computer model.

The numbers are about what one would expect from the previous discussions since the adjustments are not drastic and the independence of sex and stereotype means that two-variable tables will not differ much from three-variable tables in their patterns.

The results for educational d values, however, do have something to say. While the figures show the patterns implied by the within-cohort parameters,

TABLE 23
Selected Model Results for Years 1948-1980
Data base

Result	1948	1952	1956	1960	1964	1968	1972	1976	1980
Cohort									
Vietnam	-	-	-	-	.08	.20	.31	.43	.55
World War 11	.03	.13	.22	.31	.33	.30	.28	.25	.23
Interwar	.48	.44	.41	.37	.33	.30	.26	.23	.19
World War 1	.49	.43	.37	.32	.26	.20	.15	.09	.03
Sex									
Female	.509	.517	.524	.529	.533	.537	.538	.538	.537
Stereotype									
Yankee	.408	.404	.400	.396	.388	.379	.370	.360	.351
Northern Catholic	.183	.186	.189	.193	.196	.199	.202	.206	.209
Northem Black	.043	.046	.049	.052	.052	.051	.050	.049	.048
Southern White	.215	.213	.211	.209	.206	.204	.201	.199	.196
Southern Black	.043	.043	.043	043	043	043	043	043	043
Other	.108	.108	.108	.108	.114	.124	.143	.143	.153
Education									
College	.162	.171	.180	.190	.210	.235	.260	.285	.310
High school	.243	.258	.274	.289	.310	.333	.355	.378	.401
Less than high school	.596	.571	.546	.521	.480	.432	.385	.337	.289
Educational d values									
Female (vs Male)									
College	-.019	-.027	-.037	-.046	-.051	-.057	-.062	-.066	-.069
College and high school	+.038	+.032	+.026	+.020	+.016	+.016	+.016	+.019	+.024
Stereotype (vs Yankee)									
College									
Southern Black	-.141	-.141	-.141	-.141	-.140	-.137	-.136	-.137	-.138
Northern Black	-.110	-.107	-.105	-.105	-.107	-.111	-.115	-.121	-.126
Southern White	-.029	-.029	-.029	-.030	-.029	-.027	-.026	-.026	-.027
Northern Catholic	-.063	-.063	-.062	-.062	-.055	-.043	-.035	-.029	-.024
Other	+.026	+.038	+.051	+.063	+.089	+.114	+.134	+.146	+.151
College and high school									
Southern Black	-.353	-.358	-.363	-.369	-.360	-.344	-.329	-.319	-.309
Northern Black	-.239	-.228	-.221	-.219	-.218	-.216	-.214	-.214	-.212
Southem White	-.117	-.116	-.115	-.114	-.106	-.095	-.084	-.076	-.067
Northem Catholic	-.041	-.039	-.039	-.041	-.033	-.023	-.017	-.016	-.017
Other	+.001	+.002	+.004	+.005	+.024	+.042	+.050	+.048	+.039

the absolute changes are small and discouraging to those who favor increasing equality. Although the Vietnam cohort shows a remarkable convergence in school completion rates for these social groups, the extrapolations for 1976 and 1980 are surprisingly close to the extrapolation back to 1948. Such sluggish response is inevitable in cohort processes. Even in 1980, about 45% of the adult population will be from pre-Vietnam cohorts, and the social inequalities of their educations in the 1920s through the 1950s will be preserved, as in amber, for many years to come. If, as the students of achievement argue, education is the central variable in group SES differences, this country will retain a discouraging amount of intergroup educational differentials long after the equalitarian shifts of the last 15 years have become old history.

A Note or' the Method

The results reported here are not novel. Save for the introduction of religion, the differences are familiar to those who work with census data (e.g. Duncan, 1968). And for religion, these findings were anticipated (and greatly stimulated) by Andrew Greeley's work (1974).

If there is something novel in the materials presented, it is in method. Here, we think we have presented some potentially useful tools. d-Systems seem to form a practical technique for studying social change through pooling surveys, for assessing the agreement of multivariate contingency tables, and for developing sociological analogs to econometric modeling. Although the statistics are nothing more than glorified percentage tables, their system character, in particular their analysis thouugh flow graph techniques, means there is no reason the survey analyst must adopt multiple regression to indulge in causal analysis, change studies, extrapolation, and all the other heady and dangerous feats pulled off by econometricians.

We also feel the technique has considerable possibilities for classroom use. Given the parameters in this system, or in any other (NORC is now working on a model incorporating Father's and respondent's occupation), a computer simulation can be constructed for most any computer. Students can be sent to the computer with questions (e.g., "Have race differences in education changed in the last 20 years?") and receive scientifically responsible answers from simple percentage tables. In addition, the student will be spared the apparatus of significance testing and the "iffy" character of results from a single sample. There is no reason why sociologists cannot develop and disseminate similar models on a variety of variables so that students can learn sociology as a laboratory science without having to suffer term-long courses in statistics before entering their "labs."

APPENDIX

Evaluating Trends in Proportions and Differences in Proportions from Multiple Surveys

James A. Davis and D. Garth Taylor

Assume we have a set of proportions (e.g., Table 1) or differences in proportions (e.g., Table 5) from sample surveys over some period of time. We wish to decide whether the values

(A) are generally zero (when examining differences but not, of course, proportions);

(B) have a constant nonzero value;

(C) show a linear trend for the time in question; or (D) show a more complex pattern of change.

Using Goodman's techniques for comparing contingency tables (Goodman, 1963) one can make three tests.

(A) Null Hypothesis for d Values

The most general null hypothesis asserts that all *d* values could be independent samples from a universe where *d* = .00. To test the hypothesis, one:

(1) squares each *d,* divides the squared value by its variance, and sums;

(2) evaluates the sum as chi square with *K* degrees of freedom, where *K* *is* the number of samples.

If chi square is not significant, one accepts the null hypothesis.

(B) Constant Value for d or p

The next null hypothesis asserts that all *d* values or p values could be independent samples from the same universe value. To test the hypothesis, one:

(1) finds the pooled value, weighting each estimate by the reciprocal of its variance, as explained by Goodman;

(2) subtracts each value from the pooled estimate, squares *K* results, divides each squared result by its variance, and sums;

(3) evaluates the sum as chi square with *K-l* degrees of freedom.

If chi square is not significant, one concludes that the values are not significantly different across time and/or studies.

(C) Linear Trend for d and p Values

To test for a linear trend in the data, one:

(1) runs a weighted regression for year and p or *d,* weighting each

proportion or *d* inversely to its variance;

(2) calculates the difference between each value and its predicted

level (~), squares the difference, divides by the variance of the value, and sums;

(3) evaluates the sum of chi square with K-2 degrees of freedom.

It is also possible to examine differences among the A, B, and C chi squares. A - B has one degree of freedom and tests the null hypothesis that fitting a constant does not improve the model. B - C also has one degree of freedom and tests the null hypothesis that fitting a linear trend does not improve fit in comparison with the model using a constant. B - C is thus analogous to testing whether a regression line differs significantly from zero.

A *Typology of Outcomes*

Table Al gives a classification of possible outcomes and suggests how to interpret the results.

If A is not significant, we conclude that we can fit the data by assuming *d is* zero at all times. It is possible (though rare) for A - B or B - C to be significant when A is not. If so, we can fit the data better with a constant or a trend, but these alternative models are more complicated and not necessary. Parsimony suggests the *d* = .00 interpretation.

If A is significant but B is not, we can fit the data with a constant nonzero value, but not with a constant zero. If, in addition, A - B is significant and C is not, we have a type-case constant model. If neither A - B nor B - C

TABLE A1
A Typology of Test Outcomesa

A	*B*	*C*	*A - B*	*B - C*	Type	Interpretation
-					I	d=.00
+	-		+	-	11	Constant
			-	-	lla	Borderline constant
+	+	-		+	111	Linear trend
				-	Illa	Borderline linear trend
+	+	+		+	Illb	Rough linear trend
				-	IV	Nonlinear trend

a+ = Significant;-= nonsignificant.

is significant, one may interpret the results as a constant with borderline significance. The failure of A - B weakens the case for the model, but the failure of C and significance of A imply that linear trends or constant zeros are not plausible alternatives. It is also possible that B - C is significant. This means a linear model can improve fit, but the nonsignificance of *B* says that this is not necessary. Thus, the data show some tendency toward linear trend, but a linear model is not required. Parsimony argues for choosing the constant model.

When A and *B* are significant but C is not, the data show a linear trend since a straight line model fits the data, while constants and *d* = .00 do not. The significance of *B - C* confirms the lineaS interpretation, and its nonsignificance puts it in the borderline linear trend category.

When A, *B,* and C are all significant, interpretation depends on the outcome for *B - C.* If it is significant, one concludes that a linear model improves fit, but the straight line cannot account for the data. In such a case, there may be a monotonic curve at work, imperfectly approximated by the straight line, or the data may contain clear outliers (see below). "Rough linear trend" describes both possibilities. If, however, *B - C is* not significant, the suggestion is one of complex or erratic fluctuation that cannot be described by either a constant or a straight line. Again, inspection of outliers is called for.

If the data are proportions rather than *d* values, the A test is omitted since the existence of at least one nonzero proportion in the set is *ipso facto* rejection of the hypothesis that the common universe value is zero. Otherwise, exactly the same scheme applies to proportions.

A *Note on Outliers*

When comparing surveys over time, outliers are a distinct possibility. "Blips" in the trends can come from slight differences in question wording, differences in sampling procedures from study to study or among survey organizations, or from mechanical errors in processing old studies with incomplete documentation.

There seems to be no unambiguous procedure for spotting such deviant cases, but it seems useful to have a consistent rule. We have used the following:

(1) Except for nonlinear trends and rough linear trends, there is

no necessity to spot outliers unless they stand out like sore thumbs.

(2) To define outliers, we use a chi square value (vs the modeled

estimate) that would be larger for I d.f. than the general alpha value.

Since we generally use the .05 level (after correcting for multistage

sampling), we use a cell chi square of 5.000, which is a bit above .02, as

a criterion.

ACKNOWLEDGMENTS

Karen Newman, John Fry, D. Garth Taylor, Gregory Gaertner, Tom W. Smith, and A. Wade Smith contributed almost two years of tedious statistical and computer work that made this report possible. I am greatly indebted to their care, hard work, and imagination. The research was supported by National Science Foundation Grant No. GS-385 34.

REFERENCES

Davis, James A. (1975), "Analyzing contingency tables with linear flow graphs: dsystems," *in* Sociological Methodology (D. Heise, Ed.), pp. 111-145. Jossey-Bass, San Francisco.

de Sola Pool, 1., Abelson, R. P., and Popkin, S. L. (1964), Candidates, Issues, and Strategies: A Computer Simulation of the 1960 Presidential Election. MIT Press, Cambridge.

Duncan, B. (1968), "Trends in output and distribution of schooling," *in* Indicators of Social Change (E. B. Sheldon and W. E. Moore, Eds.), pp. 601-672. Russell Sage Foundation, New York.

Goodman, L. A. (1963), "On methods for comparing contingency tables," Journal of the Royal Statistical Society, Series A, 126, 97.

Greeley, A. M. (1974), Ethnicity in the United States: A Preliminary Reconaissance. John Wiley & Sons, New York.

Hastings, P. K., ed. (1975), Survey Data for Trend Analysis. The Roper Public Opinion Research Center, in cooperation with The Social Science Research Council, Williamstown, Massachusetts.

Kish, L. (1957), "Confidence intervals for clustered samples," American Sociological Review, 22, 154-165.

Mason, K. O., Mason, W. M., Winsborough, H. H., and Poole, W. K. (1973), Some methodological issues in cohort analysis of archival data," American Sociological Review 38, 242-258.