INDIANS AND SMUDGES ON THE CENSUS SCHEDULE

by Donald S. Akers and Elizabeth A. Larmon
U. S. Bureau of the Census

This is an explanation of how certain anomalies in the reported characteristics of American Indians in the 1960 Census are to be explained by systematic smudging of the Census schedules.

The Bureau of the Census has long recognized that the census counts are subject to error. Every effort is made to identify the errors, to correct them if possible, and to seek means of preventing their recurrence. One of the most useful means of identifying errors is to review the census results for reasonableness. To trace the source of the error, once its existence is discovered, may at times require an intensive exercise in deductive logic as well as an intimate knowledge of census procedures.

One such exercise in identification of a census error by deduction, this time by experts outside the Bureau, received some publicity a few years ago. Coale and Stephan in their article "The Case of the Indians and the Teen-Age Widows,"1/ relating to the 1950 Census, noted that the tabulated number of widowed males under age 20 was excessive and that the excess was greater the younger their age. They were able to demonstrate that the error occurred when the census information was transferred to punch cards. When the card was punched one column off, an excessive number of widows and Indians would be generated, all of whom would be in their teens or twenties. Coale and Stephan concluded by saying:

> "The Bureau of the Census changed over in 1960 to data sensing machinery to transcribe information onto magnetic tape, and the specific problem of a shift in columns is no longer relevant to census operations. The new set of processing operations poses new problems of error control for the Bureau, and may possibly cause misleading figures to show up in new and unsuspected ways in small cells. Users must continue to regard such data with special care."

As expected, a variety of irregularities and inconsistencies have appeared in the 1960 Census counts. Some relate to the counts of Indians once more. They are small compared to the total population but are large enough to distort the statistics for the Indian population.

Only the total number of Indians and their sex distribution were tabulated from the complete count. Their age distribution and all other characteristics were tabulated from the 25-percent sample and published in the report Nonwhite Population by Race, U.S. Census of Population: 1960, Final Report PC(2)-1C.

In these tabulations there is a marked excess of Indians at ages 55 to 59 (Table A); the number in this age group exceeds those 50 to 54

years old by one half. The only reasonable explanation for this excess is census error.

TABLE A.—AMERICAN INDIANS 35 YEARS OLD AND OVER, BY AGE: 1960
(Based on 25-percent sample)

Age	Number	Year of birth
Total, 35 and over	162,783	1920 and before
35-39	28,389	1920-24
40-44	22,929	1915-19
45-49	21,711	1910-14
50-54	20,767	1905-09
55-59	31,560	1900-04
60-64	11,830	1895-99
65-69	9,975	1890-94
70-74	6,857	1885-89
75 and over	8,765	1884 and before

At first, it was thought that the error in age might be related to an excess that had been observed in the count of the total population at age 59 (Table B). However, there are incongruities in the characteristics of Indians at these

TABLE B.—POPULATION 50 TO 64 YEARS OLD, BY AGE, ACCORDING TO SAMPLE AND COMPLETE COUNT: 1960

Age	Sample Count	Complete Count
50 to 54	9,696,502	9,605,954
50	2,035,449	-
51	1,979,215	-
52	1,951,935	-
53	1,856,757	-
54	1,873,146	-
55 to 59	8,595,947	8,429,865
55	1,801,394	-
56	1,702,574	-
57	1,697,037	-
58	1,536,568	-
59	1,858,374	-
60 to 64	7,111,897	7,142,452
60	1,504,160	-
61	1,441,041	-
62	1,386,263	-
63	1,369,684	-
64	1,410,749	-

ages which are not to be found in the total population. There are quite unreasonable excesses in the percent single and the percent counted as Other Relative at these ages (Table C). Half of the employed at these ages have neither occupation nor industry of worker reported.

TABLE C.--SELECTED CHARACTERISTICS OF AMERICAN INDIANS, 35 YEARS OLD AND OVER, BY AGE: 1960
(Based on 25-percent sample)

Characteristic	35 to 44	45 to 64	65 and over
Percent single:			
Male	13.7	25.9	7.3
Female	7.5	22.0	3.0
Percent other relative	12.9	27.2	18.4
Percent of employed:			
Occupation not reported	10.1	51.6	11.8
Industry not reported	9.3	33.8	11.2

These peculiarities in the age distribution and characteristics of the Indian population seem to be limited to a few areas, places where Indians represent a relatively small proportion of the total population. The excess at age 55 to 59 over other ages, and the anomalies in other characteristics are largely limited to the urban population and are most striking in a few States, such as New York, Illinois, Florida, Kansas and Texas.

Early efforts to explain the error failed, but in 1967, the investigation was reopened, when the Census Bureau was asked to make projections of the Indian population by age. The first step in the investigation was to obtain a printout for the State of Kansas showing the detailed information for all Indians 45 to 69 years old on the computer tape for the 25-percent sample. A portion of this printout, presented in Table D, shows the number of Indians in Kansas aged 45 to 69 by single years of age. Instead of a concentration at age 59, which would be expected if year of birth was rounded to 1900, there were excesses at ages 57 and 58. There were 90 Indians aged 57 and 58 where about 12 might have been reasonably expected. Of the 90 cases, 85 were recorded as born in 1902. Clearly, some bias was operating in favor of one particular year of birth.

It also appeared that in most of these cases only race (Indian) and year of birth (1902) were reported. For most of the 85 cases, the other 3 complete-count items were blank. Sex had to be allocated for 76 of the cases, marital status for 75, and relationship for 84.

The cases were scattered throughout the State. Generally, each Indian born in 1902 was the only Indian recorded in this enumeration district. In only one instance were there as many as four in the district.

Each piece of evidence made the case more puzzling until it was suggested that the schedules might have been smudged in some systematic way during transcription to the sample questionnaires. This explanation proved to be the correct one.

The 1960 Census schedules were designed for use by FOSDIC (Film Optical Sensing Device for Input to Computer). Entries were made by filling in the appropriate circles on the schedule with pencil. It was recognized from the first that smudging could be an important source of error. The FOSDIC scanner has no way of telling an accidental smudge on the circle from a purposeful marking by the enumerator. There would be little chance of a smudge causing an error if it occurred for an item already filled by the enumerator, because the enumerator's entry would be darker than the smudge. The FOSDIC scanner selects only one circle for each item, the one with the darkest marking.

The FOSDIC schedules in the 1960 Census were printed on both sides and were bound in spiral notebooks. When the enumerator made an entry, his pencil pressed the reverse side of the sheet against the sheet below. The point where pressure was applied on the sheet below could be a FOSDIC circle if the sheets were aligned exactly, for the sheets were identical. If there was already a

TABLE D.--TALLY FROM 25-PERCENT SAMPLE FOR INDIANS IN KANSAS 45 TO 69 YEARS OLD, BY SINGLE YEARS OF AGE: 1960

Age	Cases
45	8
46	10
47	11
48	7
49	13
50	9
51	9
52	12
53	8
54	8
55	7
56	5
57	60
58	30
59	7
60	3
61	3
62	4
63	5
64	3
65	0
66	3
67	5
68	5
69	4

pencil mark in this circle, it could leave a
smudge on the reverse side of the sheet being
marked. If the smudge happened to fall on a
blank circle on the reverse side of the sheet,
the FOSDIC scanner might read it as an entry.

It so happened that the circles for certain
characteristics in the panel of questions for the
first and third individuals on the FOSDIC ques-
tionnaire (Form PH-3 and PH-4) were back to back.
The left-hand page of the FOSDIC book had the
panel of housing questions at the top and the
first population panel at the bottom. The right-
hand page contained two population panels. Thus,
the first population panel, where information for
the head of the household was normally entered,
was backed by a population panel for the preced-
ing household. The FOSDIC circles for most of
the 100-percent items in these two panels coin-
cided exactly. The alignment of the circles is
shown in Table E. An inspection of the table will
show that if the first individual on successive
sheets is coded as Head of household, white, born
in the decade 1890-1899, the reverse side of the
top sheet may show smudges on the circles for an
Indian, born in the 1900's, and in the specific
year, 2.

The three smudges were sufficient to make
the computer mistake a blank panel for a person.
The FOSDIC scanner cannot tell whether a name has
been entered in a panel, so the computer must
rely on the presence of a coded entry to tell
whether the enumerator had meant to enter someone
or had intended to leave the panel blank. It was
recognized that smudging on a blank panel might
lead to the inclusion of the panel in the count of
population. To minimize such spurious counting,
the computer was instructed to count a panel as a
person only if there were at least two entries
among the five characteristics reported in the
complete count (relationship to Head of household,
sex, race, age, and marital status) of which at
least one was required to be relationship, sex,
or race. By this rule, the smudges on Indian race
and 1902 year of birth were just sufficient to
generate a fictitious person.

This explanation for the error in the Indian
data was confirmed by an examination of the cen-
sus schedules. The microfilms for selected enum-
eration districts where the printout recorded one
or more Indians born in 1902 were examined for an
entry of Indian. No actual entry of an Indian
was found, but there were eleven cases observed
where smudges had been left on Indian, and on the
decade of 1900-1909 and the specific year 2, for
date of birth. In every case, there were entries
of head, white, and the 1890 decade on both the
reverse side of the same sheet and on the face of
the preceding sheet. In one case, the smudges
fell on a panel where there were actual entries,
so that the FOSDIC scanner would have ignored the
smudges, but in the other cases they fell on blank
panels. On three of these, the smudges must have
been sufficiently faint for FOSDIC to pass over
them, because only seven cases were recorded in
the printout. Smudges were also found which
marked FOSDIC circles off-center or fell in spaces
where there were no circles. These, of course,

TABLE E.--THE ALIGNMENT OF CIRCLES
ON THE FOSDIC SAMPLE SCHEDULE

First Population Panel	Third Population Panel
Relationship:	Specific year of birth:
Head	2
Wife	1
Son/Daughter	0
Relative	-
Nonrelative	-
	Decade of birth:
Inmate	1960
Sex:	
Male	1930
Female	1920
Color or race:	
White	1900
Negro	--
Indian	1890
Japanese	1870
Chinese	1860
Filipino	1850
Other	--
Decade of birth:	Color or race:
1850	Filipino
1860	Chinese
1870	Japanese
1880	---
1890	Indian
1900	White
1910	---
	Sex:
1920	Female
1930	Male
1940	---
1950	---
	Relationship:
1960	Inmate
Specific year of birth:	
0	Son/Daughter
1	Wife
2	Head
3	---
4	---
5	---
6	---
7	---
8	---
9	---

had no effect on the count.

We estimate that there should have been successive entries of Head of household, white, born in 1890-1899, with an intervening blank panel, about 140,000 times in the sample schedules. We also estimate that spurious entries of Indian born in 1902 were recorded by the FOSDIC scanner about 3,500 times. So, the spurious entries were counted about 2.5 percent of the times that they could have occurred. (Multiplied by 4, the sampling ratio, the 3,500 recorded smudges would add 14,000 to the count of Indians).

Once the computer decided that it had found an Indian born in 1902, it ascribed many of the other characteristics, according to detailed specifications for assigning missing information. According to these specifications, persons with relationship missing were most often assigned to the category "Other Relative", and to single marital status. Occupation and industry characteristics were not assigned, however. This procedure explains the anomalous distribution of characteristics of Indians already noted and shown in Table C.

Having explained the excess of Indians at age 55 to 59 and their peculiar distribution of characteristics, we were still faced with the question of why the error was concentrated in certain areas. When one considers the nature of the error, it is reasonable to expect that it would occur with greatest frequency in areas with the largest population. The larger the total population, the larger the number of white heads of households born in 1890-99 and the greater the probability that such individuals would appear on successive schedules, creating the situation which could produce spurious Indians born in 1902. Indeed, most of the States with the largest excesses, in terms of absolute numbers, were the States with the largest population (Table F). The relative excess depended on the size of the Indian population in relation to the white population. Generally, the smaller the ratio of Indians to whites, the greater the distortion caused by the spurious additions. Thus, the estimated excess, in absolute numbers, is only about half as large in Illinois as in California, but the relative error in Illinois is three times as great as in California, because Illinois has a much smaller proportion of Indians.

TABLE F.—SELECTED STATISTICS RELATING TO THE EXCESS OF
INDIANS 55 TO 59 YEARS OLD IN THE SAMPLE, BY STATE
(States with 2,500 or more Indians)

State	Total Population	Ratio of Indian to White Population	Estimated Excess of Indians 55 to 59 Years Old 1/	
			Number	Percent
States				
Alaska	226,167	.083	-70	-17.5
Arizona	1,302,161	.071	435	24.8
California	15,717,204	.003	1,340	100.8
Colorado	1,753,947	.003	157	155.5
Florida	4,951,560	.001	653	315.5
Idaho	667,191	.008	39	25.2
Illinois	10,081,158	.001	744	322.1
Kansas	2,178,611	.002	314	251.2
Louisiana	3,257,022	.002	105	69.5
Michigan	7,823,194	.001	212	49.2
Minnesota	3,413,864	.005	441	95.9
Mississippi	2,178,141	.003	51	53.7
Montana	674,767	.033	152	29.3
Nebraska	1,411,330	.004	124	59.6
Nevada	285,278	.025	28	13.7
New Mexico	951,023	.064	390	32.8
New York	16,782,304	.001	2,225	232.5
North Carolina	4,556,155	.011	-8	-0.9
North Dakota	632,446	.019	-12	-3.7
Oklahoma	2,328,284	.031	474	19.4
Oregon	1,768,687	.005	239	101.7
South Dakota	680,514	.040	100	14.5
Texas	9,579,677	.001	428	203.8
Utah	890,627	.008	111	116.8
Washington	2,853,214	.008	509	87.2
Wisconsin	3,951,777	.004	307	68.4
Wyoming	330,066	.012	64	63.4

1/ The expected number was estimated by linear interpolation.

Tabulations of Indian statistics from the 25-percent sample were made only for States with 2,500 or more Indians. If sample statistics were available for States such as Pennsylvania and Ohio, with very large white populations but fewer than 2,500 Indians, the distortions in their data for Indians would, undoubtedly, be found to be very extreme.

The concentration of the error in urban areas has a similar explanation. Urban areas contained about 70 percent of the population in 1960 and, presumably, should have received about that proportion of the error in the Indian statistics. Moreover, since the proportion of the population which is Indian is only 0.1 percent in the urban areas but is 0.7 percent in rural areas the rate of error should be much greater in urban areas.

Although the distortions in the Indian data had been explained, the investigation was not complete until the extent of other errors due to smudging could be determined. The alignment of circles shown in Table E shows that a white head of household, if born in the decade of the 1900's would produce a smudge adding to the white population born in 1902 and, if born in the 1870's, could add to the Japanese population born in 1902. However, in the first case, the actual number of white persons born in 1902 is so large (about 1.5 million) that the estimated additions to this group as a result of smudging (about 23,000) would have a minor effect. In the second case, the number of whites born in the 1870's who were household heads in 1960 was relatively small and could have generated only a small number of smudges, if the rate of occurrence was the same as for the smudges that produced the Indians. Other potential sources of error due to smudging were considered, but none were found to meet the conditions necessary to produce noticeable distortions in the data.

Although this kind of systematic smudging has distorted the characteristics of the population, it has not added to the total count. The sample was adjusted to the complete count by a ratio estimation procedure. The Indian picked up by smudging had an equal chance with all other nonwhites in the same age-sex group of being retained in the sample count for this area when it was adjusted to the complete count.

The data on Indians from the complete count in 1960 must certainly be more accurate than from the sample count, for they were not affected by the smudging error. This is not to say that the complete count is entirely accurate. There are particular problems in enumerating Indians that would lead one to expect that even without processing errors the Indian data might be less accurate than the data for most other ethnic groups in the population.

Errors are inevitable in a project as involved as taking the census. The reduction of the number of errors depends in part on the ingenuity of the staff of the Bureau of the Census but also in part on the cost. Some types of errors could be virtually eliminated but the cost would be prohibitive. Each expenditure to reduce the likelihood of error must be balanced against the gain in accuracy to be expected.

Footnotes

1/ Coale, Ansley J., and Stephan, Frederick F., "The Case of the Indians and the Teen-Age Widows", Journal of the American Statistical Association, 57 (1962), 338-47.

Source Notes for Tables

Table A.--U.S. Bureau of the Census., U.S. Census of Population: 1960, Subject Reports. Nonwhite Population by Race. Final Report, PC(2)-1C.

Table B.--U.S. Bureau of the Census, U.S. Census of Population: 1960. Detailed Characteristics. United States Summary. Final Report PC(1)-1D.

Table C.--U.S. Bureau of the Census, U.S Census of Population: 1960, Subject Reports. Nonwhite Population by Race. Final Report, PC(2)-1C.

Table D.--U.S. Bureau of the Census. Unpublished Data.

Table F.-U.S. Bureau of the Census, U.S. Census of Population: 1960, General Population Characteristics, United States Summary, Final Report PC(1)-1B, and Subject Reports, Nonwhite Population by Race, Final Report PC(2)-1C.