

American Sociology Association

Sociocultural Behaviors Correlated with Census Undercount

Paper prepared for presentation in Special Session 2 15,
Saturday, August 22, to the American Sociological Association

by Dr. Leslie A. Brownrigg and Dr. Manuel de la Puente
U.S. Bureau of the Census

Dr. Manuel de la Puente:

(General introduction)

The next set of papers analyze the ethnographic sample. This data set compiles results of the match and resolution of independent alternative enumerations collected at 29 sites to the 1990 Decennial Census information keyed from census forms. Sites were in the continental United States and Puerto Rico. This was our final sample design:

SEE FIGURE II SAMPLE DESIGN
WITH % RACE/ETHNIC

The method of conducting and matching Alternative Enumerations to the census and resolving the Census Day status of all persons and housing units enumerated in either source is described in other papers. This sample was purposefully selected to find places where we could study behavioral correlates of undercount.

(The differential net undercount)

The differential net undercount of minorities in Decennial Censuses is one symptom of the larger more global problem of errors in the census. "Undercount" describes a net trend of errors. Net undercounts occur if there are more errors of omission (missing people and housing) than balancing errors of erroneously included population and housing or -- of course--if there are a large number of omissions even though most people counted are correctly enumerated. (The opposite of undercount is "overcount": the net trend of more erroneous inclusions than omissions.) The main method for estimating national net coverage until the 1990 Census was Demographic Analysis. In this method, unless erroneous inclusions result in overcount, they are not detected. Apparent trends in net coverage may be distorted by the overcount, or double count, of sub-populations with certain characteristics.

There has been a racial category for Black (under various terms) in historic censuses and most people identified in this race category in the Decennial Censuses are born and die in the United

States, therefore the Black population can be independently estimated from vital statistics and by carrying forward population found in past censuses. As a result, a separate demographic analysis is possible for the Black population. Demographic analysis has documented net undercount for the Black population and some age groups by sex and cohort patterns of coverage. Undercount was suspected for minority populations groups including those of Hispanic origin, Asian, American Indians among others. Undercounts were often alleged but difficult to document.

When we first set out to study the behavioral causes of undercount among minority populations, we envisioned the primary task as documenting and explaining omissions: why people and housing were missed by the census. We are able to do that, but omission is far from the whole story.

The independent ethnographers who conducted the research chose sample areas of minority populations where conditions suggested undercounts might occur. The ethnographic evaluation was coverage research deliberately set up -- positioned-- in places where our there were respondent behaviors believed to cause (or to be the context for) census undercount, where we predicted that undercounts would occur. The sample was deliberately biased to places that are considered difficult to enumerate by the Census Bureau. Barriers to enumeration included high incidence of residential mobility, irregular housing, motives for concealment, such as undocumented immigration status or illegal conversions of garages and back rooms into housing units, languages other than English, limited literacy, fears of outsiders.

Behaviors and evidence of these behaviors that constituted barriers to enumeration were systematically observed at three levels:

- for the site as a whole in its neighborhood context,
- for the physical housing unit
- for households -- and for individual persons. Systematic observations were made about reasons why people were predicted to be missed by the census, or had an ambiguous retaliation to their household, and to note who in the sample were recent immigrants or did not speak English well.

The behavioral observations were coded on check lists submitted by the participating ethnographic researchers. Observations from the neighborhood logs are treated as variables in the correlations below. The demographic characteristics collected for the primary purpose of matching Alternative Enumerations to Census, such as age, race, sex, marital status and relationship within the household, can be treated as a variable for individual cases of erroneous inclusion, omission, or correct enumeration; likewise, demographic patterns can be analyzed in variously defined sub-populations from the sample, or, as in this presentation, for each sample area.

In some sample areas where we fully expected that undercounts would be measured instead were found outliers of over count -- due to the compounding of census errors. There were omissions but a greater number of erroneous inclusions. In the sample areas where the higher net undercounts registered, erroneous inclusions in the census also “improved” the count.

Net coverage in the direction of undercount turned out to be a more rare phenomenon than errors of omission and erroneous inclusion which were found at all the sites.

Severe overcount and severe undercount are illustrated in Chart 2 (San Diego) and Chart 4 (Harlem) where the components of error and omission to the outcome can be observed. Chart 2 gives coverage results from San Diego where many residents are undocumented Mexicans; Chart 4 gives results from an inner-city Black neighborhood besieged by distribution of illegal drugs.

SEE CHARTS 2 & 4

People and housing units correctly enumerated by the 1990 Census in the ethnographic sample areas assumed a special importance: what characteristics distinguished the correctly enumerated from others who were omitted or were erroneously enumerated? Our colleagues here today will provide some answers (McKay 1992; Wobus, 1992; and see de la Puente 1992; Brownrigg 1991 and Brownrigg 1992a, and Bell 1992).

We will present patterns of association between measures of coverage error calculated for each site and site social characteristics. The site level correlations reported here do not permit inference about relationships between individual or household characteristics and coverage error. We will report demographic and behavioral correlates of rates of net coverage, error and omission at the site level.

(Dr. Leslie A. Brownrigg:)

At the level of the site, we explored whether net undercount was associated with the same or different sociocultural factors as net overcount, whether the same or different sociocultural factors produced observed rates of the error of gross omission and the error of gross erroneous inclusion. This took a new approach to what was predictive about our specialized sample: not that it was positioned to find undercounts (qua omissions) but rather to see what factors in minority, low income neighborhoods produced higher than usual incidence of errors in the census that sometimes netted to undercount.

(Variables to express aspects of coverage at the level of the site)

We defined a number of variables which express aspects of the net coverage at the level of the site.

The principal variable, TCOVER, computes the total site population (number of census records for people) keyed from all the census forms recovered from a given sample area (the "B" population) over the total population resolved to have been resident on Census Day applying Census Bureau rules (the "R" population"), less 1 .O.

TCOVER provides a negative expression for the site level rate of net undercount and positive expressions for net overcount. The formula is:

$$B/R-1 = \text{TCOVER (net undercount or net overcount)}$$

SEE SLIDE WITH DEFINITIONS

The TCOVER measure has a distribution close to a bell curve between net undercount and net overcount for the 29 sample areas. This distribution is illustrated in Histogram 1.

SEE HISTOGRAM 1
NET UNDERCOUNT OR OVERCOUNT IN
CENSUS ENUMERATION BY SITES OF THE ETHNOGRAPHIC SAMPLE

The ethnographic sample as a whole displays a slight rate of net undercount. The mean of all sites is 1.6 per cent undercount; the median is .6 per cent undercount, both less than the national undercount, both less than the differential undercount registered for separate minority groups by the Post Enumeration Survey.

Because values of rates of net undercount or net overcount are continuous in a range from -47 per cent (47% net UNDERcount) to a 53 per cent maximum (53% OVER count), we derived from TCOVER an absolute ranked variable for net coverage in the census.

The variable we call TERROR, for total error, collapses net undercounts (negative sign) and overcounts (positive) into an absolute measure of coverage error. TERROR expresses the "B" population (enumerated in the Census) over the "R" --resolved-- population minus the integer one.

SEE SLIDE HISTOGRAM 2

Histogram 2 shows TERROR's distribution. Note that the extremes are the 46% undercount and the 53% overcount but that 9 of the 29 sites registered between 0 and 2 undercount or overcount.

We also ranked net undercount or net overcount (TCOVER-X) dividing the cases almost evenly to rank into five ranges, sites with:

- 1= severe net undercount from 17 to 47 per cent,
- 2= moderate net undercount (2.9 per cent to 12 per cent net undercount),
- 3= "good censusing" with between or .6 percent undercount and .6 per cent overcount
- 4= moderate net overcounts (2 per cent - 6.9 per cent)
- 5= net overcounts between 8.9 per cent and 53 per cent.

(Male and female net coverage)

We calculated male and female net undercount or overcount in a similar fashion using only those records defined as for males or females. An intriguing observation is the inter-correlation of rates male and female net coverage with each other and with measures of the local rate of total net coverage.

SEE SLIDE OF MALE COVERAGE BY FEMALE COVERAGE

Since net undercounts have been consistently registered by Demographic Analysis among minority men, some census coverage improvement efforts have focused on males. At least in the ethnographic sample of difficult-to-enumerate, predominantly minority, low income neighborhoods, Census coverage of males is bundled with the coverage of females.

(Omissions and erroneous inclusions)

We created variables to express site level rates of omission and erroneous inclusion: the two components which interplay to produce undercount or overcount. The site level rate of omission (BROMIT) is expressed by the proportion of resolved Census Day population (the "R" population) which was omitted in the Census source. The site level rate of erroneous inclusion (BERRY) is the proportion of all records keyed from Census forms (the "B" population) which were resolved as erroneous enumerations.

Having calculated these various expressions of site level net coverage and its components, we tested all these coverage variables against others which expressed the demography and sociocultural aspects of the neighborhoods where the rates were registered.

(Demographic breakdowns : Tests of Age Groups and Sex by Age Groups)

We broke down the population keyed from census forms in the ethnographic sample -- resolved as the correctness of the enumeration--into age and sex groups to discover demographic characteristics of each site.

Since the sites were deliberately selected as likely places where the undercount occurs, the demographic profiles of the sites suggested that some key to the undercount might be found by comparing demographic profiles. Possible correlations between the proportion of an age group, or a sex/age group, or various combinations of age groups or sex/age groups observed in a site's raw census population and the six coverage measurements were explored.

(Age Group Tests)

We tested the percentage (proportion) of all records of persons keyed from Census forms in the ethnographic sample for people

- age 60 and older,
- age 60 and older, male;
- age 60 and older, female;
- age 35-59, male
- age 35-59, female;
- age 0-6 both sexes (KIDALL)
- age 0-17 both sexes,
- age 0-17 + 60 and older, that is, dependents;
- age 18-34 both sexes,
- age 18-34 male (MEN1 8)
- age 18-34 female,

to profile sites demographically. Then we compared the relative proportions to the coverage variables.

A promising single age group indicator that correlated with total net coverage (TCOVER) and with the net coverage of the male population (MCOVER) was the proportion of children age 6 or under (KIDALL). And the proportion of children age 6 or under at the ethnographic sites had a significant negative correlation with the site level rates of omission.

SEE SLIDE OMISSION RATE BY CHILDREN AGE 0-6

(Sex Ratio)

Some of the sites with coverage problems had extremely high sex ratios, over 200. Variables were created to express sex ratio of the site population as a whole (SEXRATIO) and the sex ratio of the young adult population age 18-34 (SEXR). The sex ratio showing up in the raw census (B) population was calculated three ways:

- 1) data-defined as male divided by data-defined as female multiplied by 100;
- 2) data-defined as male of known age divided by data-defined as female of known age times 100 — eliminating persons defined only by sex but not age, and
- 3) SEXR18 for sex and age data-defined persons, the number of males between age 18-34 over females of the same age group multiplied by 100: the sex ratio of a particular age group at each site.

No correlation was found between any of these 3 ways of calculating sex ratio with any measures of the rate of net coverage, nor with the local rates of omission (BROMIT) nor erroneous inclusion (BERRY). Although some earlier studies of localized undercount have suggested sex ratios as a possible indicator of coverage, this preliminary test suggests that sex ratios could not be more randomized nor further removed from local rates of coverage. Although it seems counter-intuitive that male coverage is not necessarily worse where there are fewer or more males in the population relative to females, at least by this test in the ethnographic sample, sex ratio does not correlated with net coverage. Thus, localized sex ratio cannot be used to predict or spot undercount areas.

(Fertility ratios)

Other ratios driven by demographic theory describe a population's fertility. We defined several expressions of the fertility ratio:

- FERT1) younger children age 6 or less to their likely potential mothers at the site-- females age 18 to 34 : young children, young mothers;
- FERT2) all minor children, aged 0-17, to women 18-59 : all younger generation to all potential mothers and some grandmothers;

FERT3 younger children, aged 0-6 to older women age
60 years or older : a grandchildren to grandmothers
“fertility” ratio.

FERT1 (young children, young mothers) displayed the strongest correlations with all the Census coverage variables. FERT3, the “grandmother” fertility ratio, is also a promising indicator.

Plots 1-6 illustrate regressions with FERT1 (young children: women 18-34-- the likely pool of young mothers) and the six coverage variables, for all 29 sites. Points in these plots refer the letter of the cell in the sample design (See SAMPLE DESIGN FIGURE, where a= urban concentrations of Black population, b= rural concentrations of the Black population, etc.)

SEE PLOTS 1-6

(Additional plots 7- 12 demonstrate relations between fertility ratio and Census coverage variables after removing from the universe the four sites containing concentrations of Asians. Three of the Asian neighborhoods-- Chinatown in New York City, a Korean neighborhood in Queens, New York, and the Koreatown section of Los Angeles, California have demographic profiles featuring the higher proportions of people age 60 year and older than other sites in the ethnographic sample. This demography weakens the strength of correlations with FERT1 . Plots of 25 other than Asian sites display an improved slope for the strong correlations.)

SEE SLIDE OMISSION RATE BY FERTILITY

Further, fertility ratio expressed with young children and young women (FERT1) or expressed as all children and all potential mothers (FERT2) both significantly correlate negatively with with rates of omissions in the resolved population and correlate negatively with rates of erroneous enumeration in the census.

Therefore, fairly conclusively, at least in the sites of the ethnographic sample it appears that the higher the proportion of young children age zero to 6 years in a local population, or the higher the fertility ratio (using the measure of young children and younger women, potentially the children’s mothers). then the more accurate, complete and error-free is the Census enumeration of that population. Conversely, in places where there are very few young children or where the fertility ratio is low, errors in the census are high, especially errors of omissions with resulting net undercount

(Neighborhood variables)

We now turn to relationships between census coverage and neighborhood characteristics other than the local demography.

Each ethnographer who conducted an Alternative Enumeration submitted a three page check list to characterize the general neighborhood containing the contiguous housing units that made up the site. Traits selected to characterize neighborhoods were easily visible to observe or easily

verifiable “indices” of more complex social behaviors, such as gang or posse activity.

We examined correlations between neighborhood traits keyed from ethnographers’ on logs and the various coverage expressions. Since we had learned that the presence of young children and high fertility ratios were demographic hallmarks of areas of “good” Census enumeration we also examined what features of neighborhoods were present or absent in areas characterized by a high proportion of children or by high fertility ratios.

Safety is a concern. Census Bureau enumerators consider reputedly dangerous areas as difficult-to-enumerate and in dangerous neighborhoods, people might not wish to open their doors to enumerators. The ethnographers systematically reported on the presence or absence of recent crimes that indicated neighborhoods were not safe. They made five observations about neighborhood “safety” :

- 1 Did they observe a violent interpersonal crime (murder, rape, assault resulting in hospitalization) or did residents account that such a crime took place between January 1990 and when they concluded their Alternative Enumeration?
- 2 Did they see or hear about theft of personal property, a stolen car, or breaking and entering a house?
- 3 Were there open market in illegal drugs?
- 4 Were gangs or posses active?
- 5 Was there some other concern for safety?

It’s worth noting that in three of the ethnographic sites murders occurred during the specified time period and at two other sites, drugs were openly sold on the street, and these were places where net undercounts registered. However, statistically, only one of the Personal Safety indices displayed a significant correlation with any of the site level measurements of net coverage or gross omission or error. That correlation was a positive between evidence of gang or posse activity (SAFE-4) and higher net overcount! (per TCOVER and TCOVER_X, ranked).

In the neighborhoods with a local demography characterized by high fertility ratios, neighborhoods were dramatically more safe: none of the indicators of crime correlated with higher fertility

Other variables used to capsule neighborhoods’ social environment. Relative to surrounding jurisdiction, what was the cost and availability of housing of the type found in the neighborhood? What educational achievement was typical of adults there? What was the income source for the majority of households? What language is spoken in the neighborhood? In what language are its business signs? Is neighborhood literacy indicated by the presence of distribution points for reading materials like newspapers or facilities for adult education?

Some quality-of-life indicators were set up as dummy variables, for the presence (versus 0

absence). These included observed public drunkenness or homeless people in the street, graffiti, uncollected trash, at least one vacant building vandalized, vandalized vans or cars in the street, crowds out during day, and crowds out at night.

Looking for which neighborhood characteristics correlated with net undercount or overcount we found a negative correlation between the availability of housing of the type found at the ethnographic site and high undercount.

SEE SLIDE AVAILHUS

This negative correlation compared the rank of net coverage (from 1 severe undercount to 5 severe overcount) with availability of housing ranked from 1 (very scarce) to 4 (high vacancy). We interpret this as confirming an association between localized net undercounts and high vacancy rates (one of the hypothesis posited after the 1988 pilot of this method).

As far as the neighborhood language is concerned, we found that if the neighborhood language was other than English (for example, Spanish or Chinese) then the rate of net undercount of females (FCOVER) was higher. None of the other of net or gross coverage measurements correlated significantly with language or literacy variables at the level of the neighborhood or site. We believe that the single correlation pointing to language as a factor in female coverage may be related to the prevalence of women as respondents to the Census generally in the ethnographic sample. It is also important to note that since many sites were ethnically mixed and several languages other than English were spoken at several sites, language barriers to Census enumeration may function at the household level and rather than at the neighborhood level.

Neighborhoods characterized by lower levels of household income were correlated with lower educational achievement of adults, open drug sales and vandalized buildings-- however, we do not find that such neighborhoods produced any consistent pattern of net Census coverage. Certain traits reflecting deteriorated, high crime, and altogether less desirable neighborhoods characterized the Black urban sites in the ethnographic sample more than the other settings. These neighborhoods were hard-to-enumerate, but produced varied, inconsistent Census results: both undercounts and overcounts

Generally, children and young mothers were rare in the ethnographic sample sites. The places chosen to illustrate "hard-to-enumerate" where undercounts were thought likely to occur were dominated by young adults. The two sites where fertility ratios were highest and net coverage was good were both located in housing projects of single family dwellings where the similarly constructed houses were set on lots. The sites were on an American Indian reservation and in a community of Mexican American farm workers. In another study (Brownrigg 1991), one of us found that "regular" housing such as this --and that found in some larger apartment buildings -- is associated with more accurate coverage of housing and excellent population coverage due to lower rates of omitting whole households when occupied housing is not found by the Census.

Discovery of a relationship between low fertility ratios and poor coverage in the Census offers the potential that “hot spots” of undercount could be detected during the census-in- process. Our own sample is too small for generalization but the larger data sets from the Post Enumeration Survey could be explored to confirm or rule out this suggestion.

(Introduction to other papers in the session)

In this paper, we reviewed some results at the level of the site. In the ethnographic sample data, it is possible to define sub-populations for study. Peter Wobus collected the records of Asians from the entire sample to draw contrasts between individual Asians who were correctly enumerated in the Census versus those whom the Census were omitted. Ruth McKay analyzed patterns of omission within households in the population of Hispanic households from the predominantly Hispanic sites that Manuel de la Puente analyzed in a 1992 study. Jacqueline Hagen describes in ethnographic depth the results of the coverage evaluation she personally conducted.

FIGURE II: SAMPLE DESIGN WITH % RACE/ETHNIC

RACE/ETHNICITY	Concentrated		Heterogeneous
	Urban	Rural	Urban/Suburban
Black	CELL A	CELL B	CELL C
	98 % 89 % 95 % 98 %	92 % 98 %	53 % 31 %
Indian		CELL D 95 % 62 % 86 %	
Asian	CELL F 97 % 73 % 84 %		CELL G 58 % 39 % 18 % 44 %
Hispanic	CELL H 91 %	CELL I 96 % 53 %	CELL J 24 % 61 %
	CELL K 94 % 97 % 61 %		CELL L 52 % 70 % 83 %

TOTAL N . 29

% = Percent of race/ethnic group indicated per the Alternative Enumeration

CENSUS COVERAGE EXPRESSIONS

$$\text{TOTAL CENSUS ERROR} = \text{ABS}(\text{TCOVER})$$

$$\text{CENSUS ERROR (MALES)} = \text{ABS}(\text{MCOVER})$$

$$\text{CENSUS ERROR (FEMALES)} = \text{ABS}(\text{FCOVER})$$

CENSUS COVERAGE EXPRESSIONS

TOTAL
COVERAGE
(TCOVER) = (RAW CENSUS COUNT) / (RESOLVED POP) - 1

MALE
COVERAGE
(MCOVER) = (RAW CENSUS COUNT: MALES) / (RESOLVED POP) - 1

FEMALE
COVERAGE
(FCOVER) = (RAW CENSUS COUNT:FEMALES) / (RESOLVED POP) - 1

C H A R T I

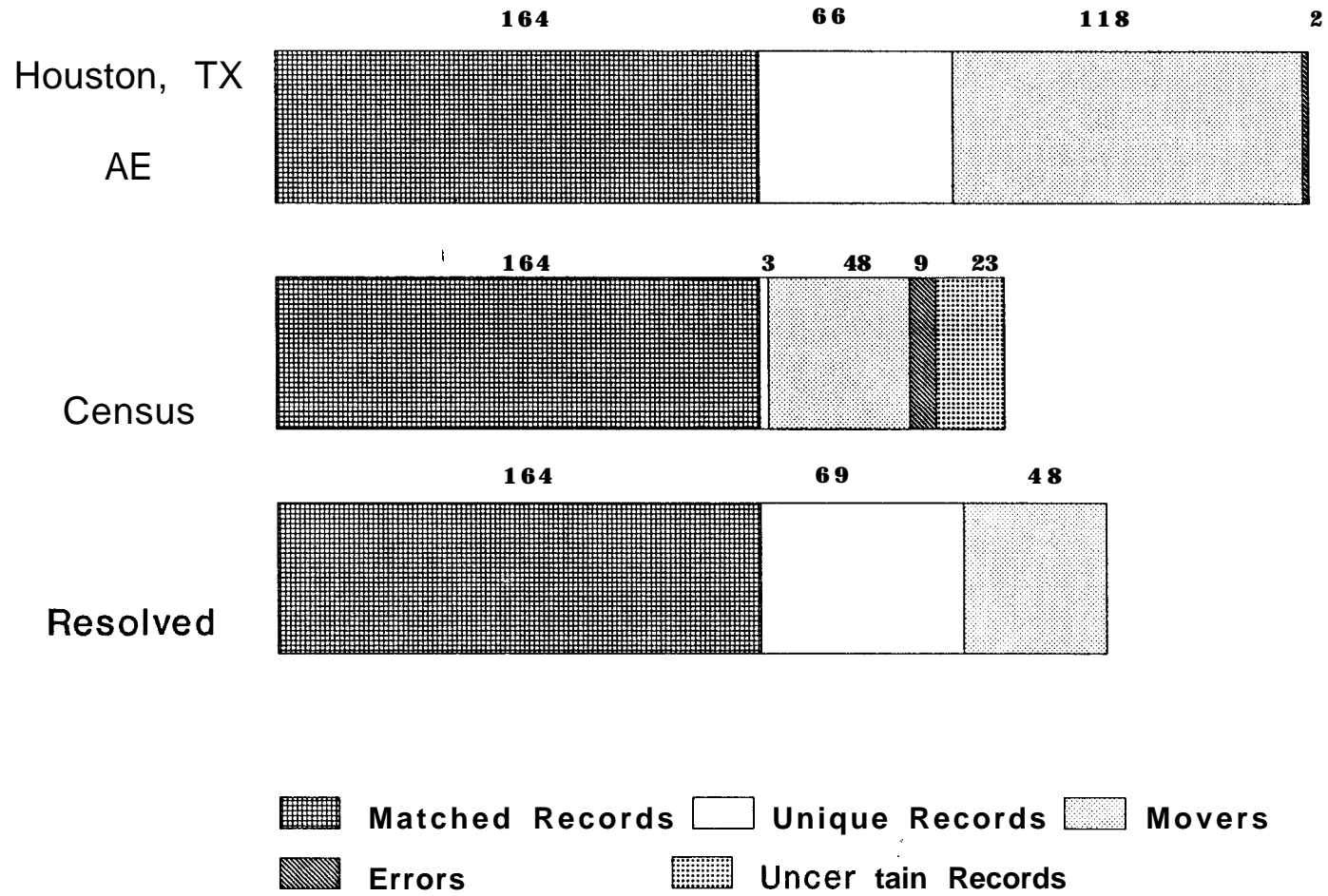
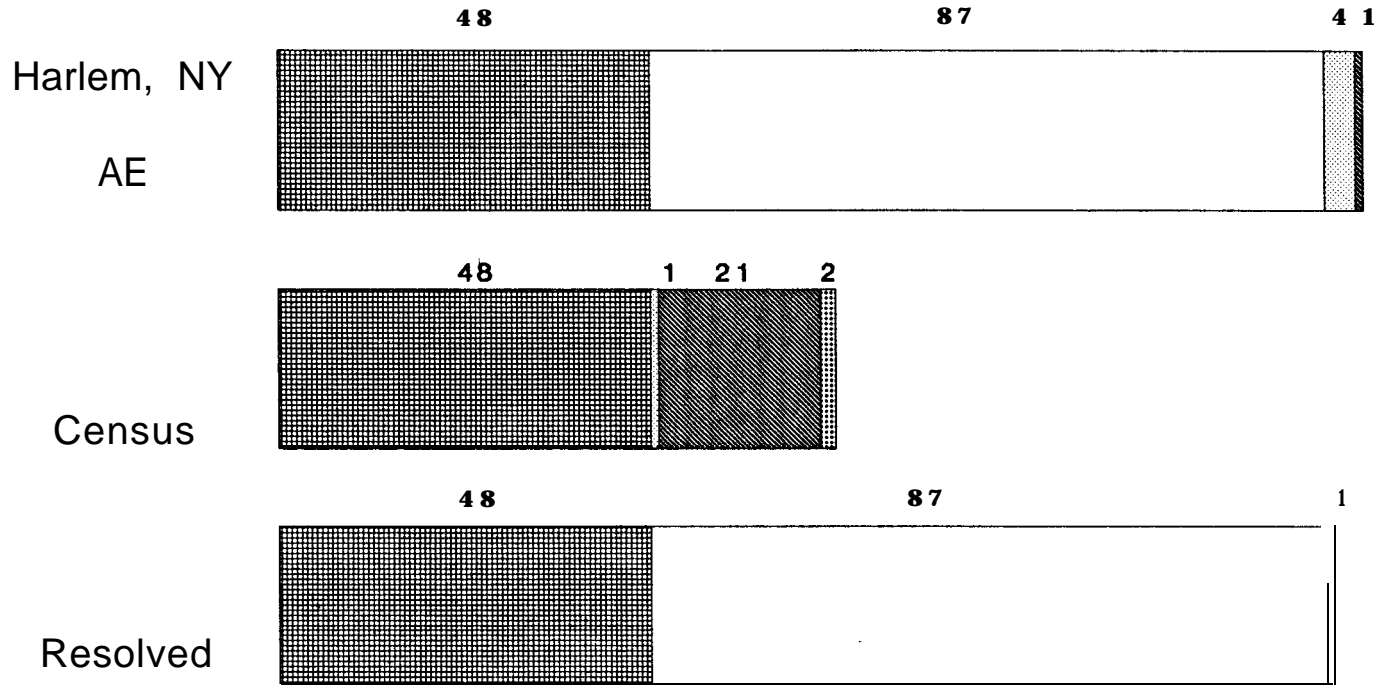


CHART IV



Matched Records
 Unique Records
 Movers

Errors
 Uncertain Records

CENSUS COVERAGE EXPRESSIONS

TOTAL
COVERAGE
(TCOVER) = (RAW CENSUS COUNT) / (RESOLVED POP) - 1

MALE
COVERAGE
(MCOVER) = (RAW CENSUS COUNT: MALES) / (RESOLVED POP) - 1

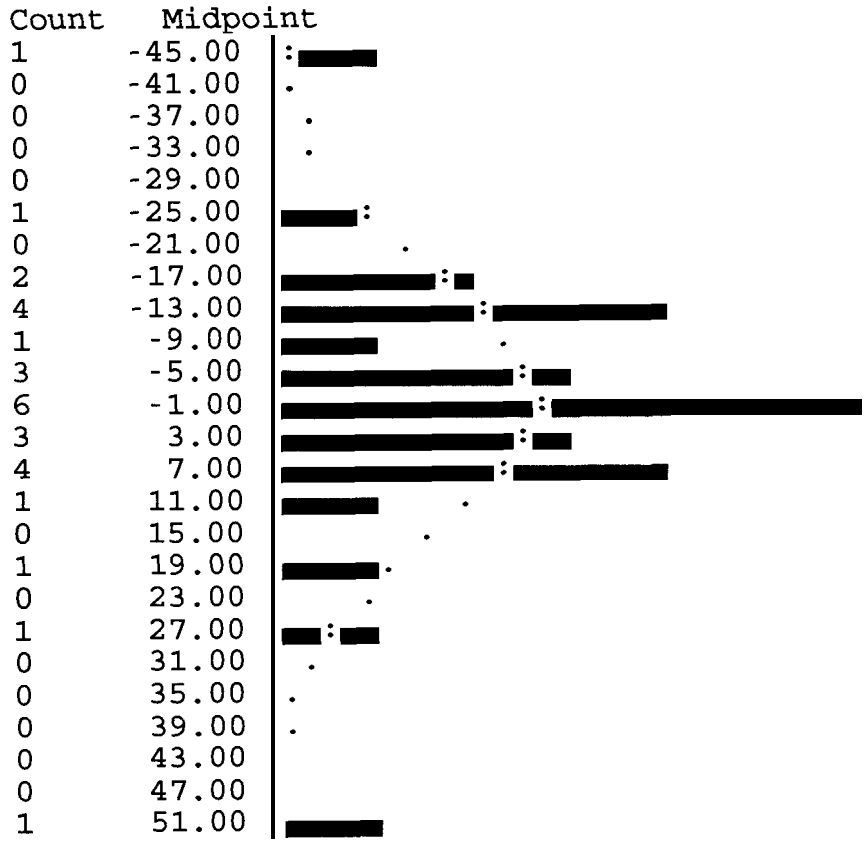
FEMALE
COVERAGE
(FCOVER) = (RAW CENSUS COUNT:FEMALES) / (RESOLVED POP) - 1

Histogram 1

Site Level Rates of Net Undercount or
in the Census Enumeration by Sites of the
Ethnographic Sample

Overcount

TCOVER



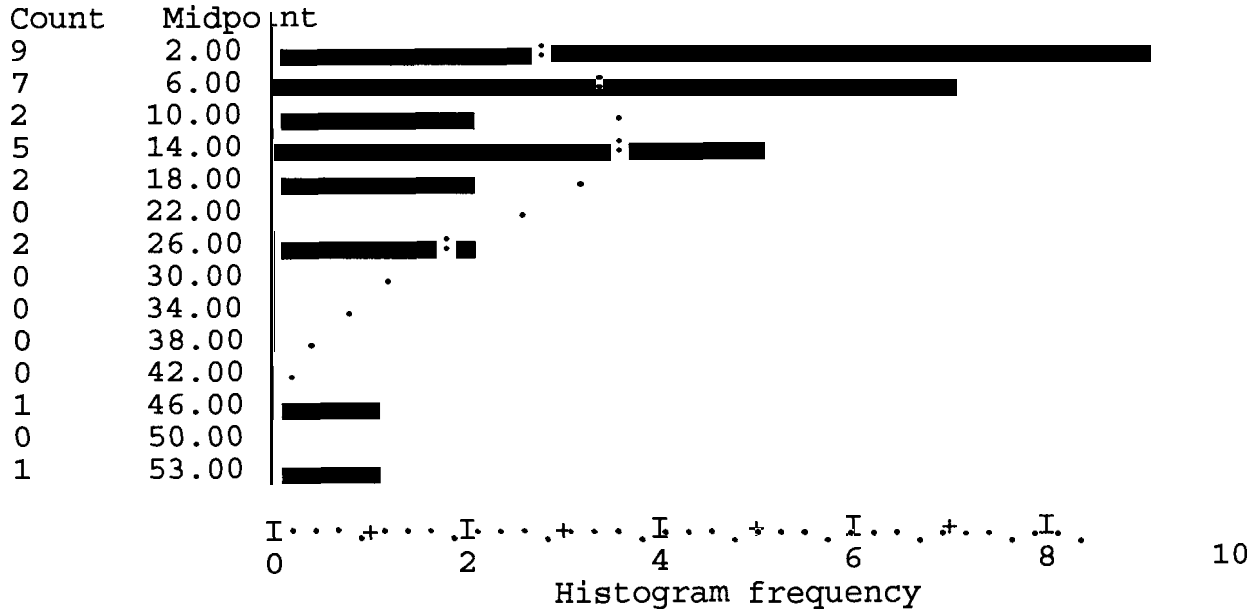
I . . . + . . . I . . . + . . . I . . . + . . . I . . . + . . . I . . .
0 2 4 6 8 10

Valid cases 29 Missing cases 0
Includes records with missing data on age or sex or both

Histogram 2

Site Level Absolute Net Coverage Errors
in the Census Enumeration by Sites
of the Ethnographic Sample

TERROR

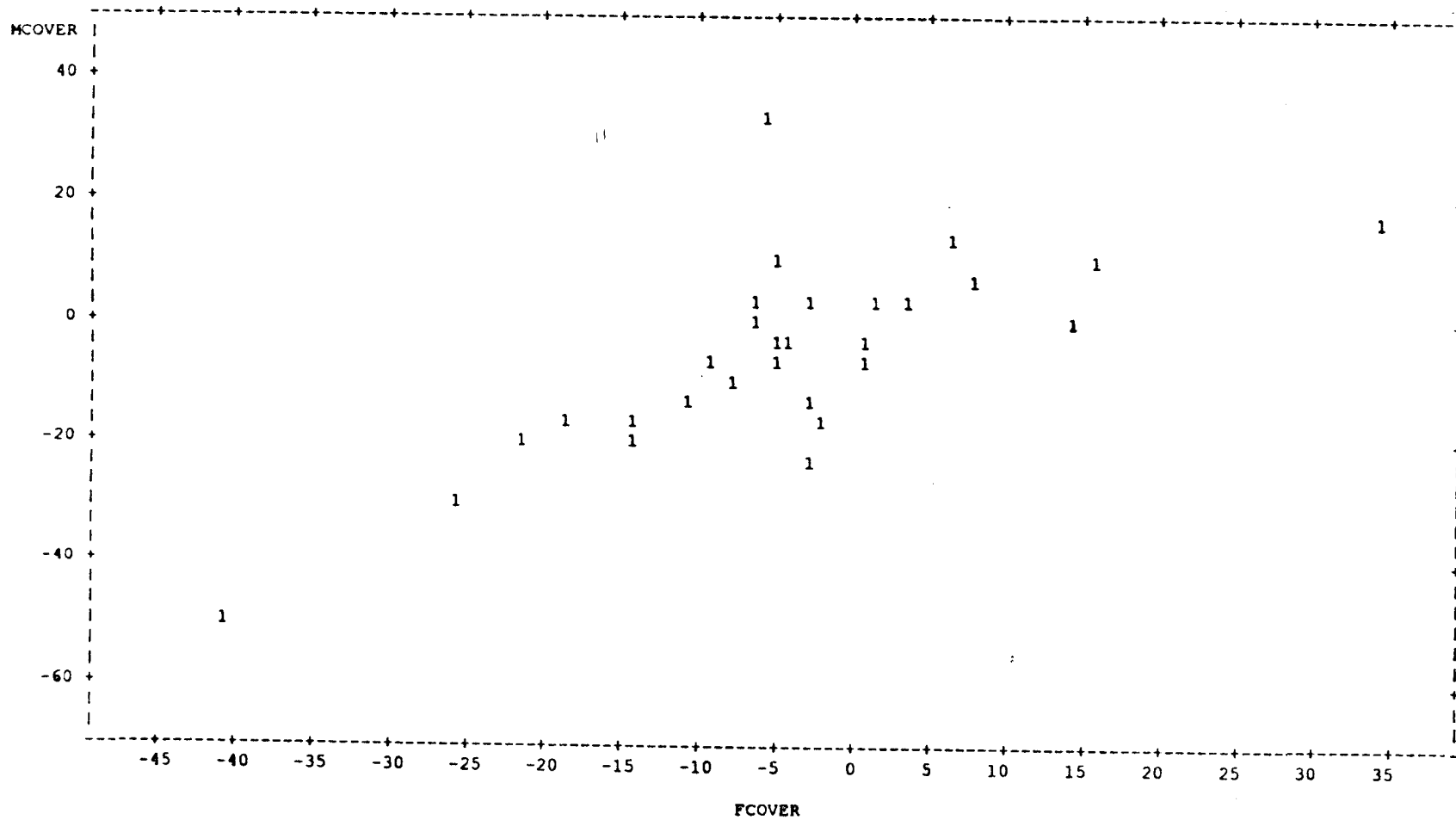


Valid cases 29 Missing cases 0

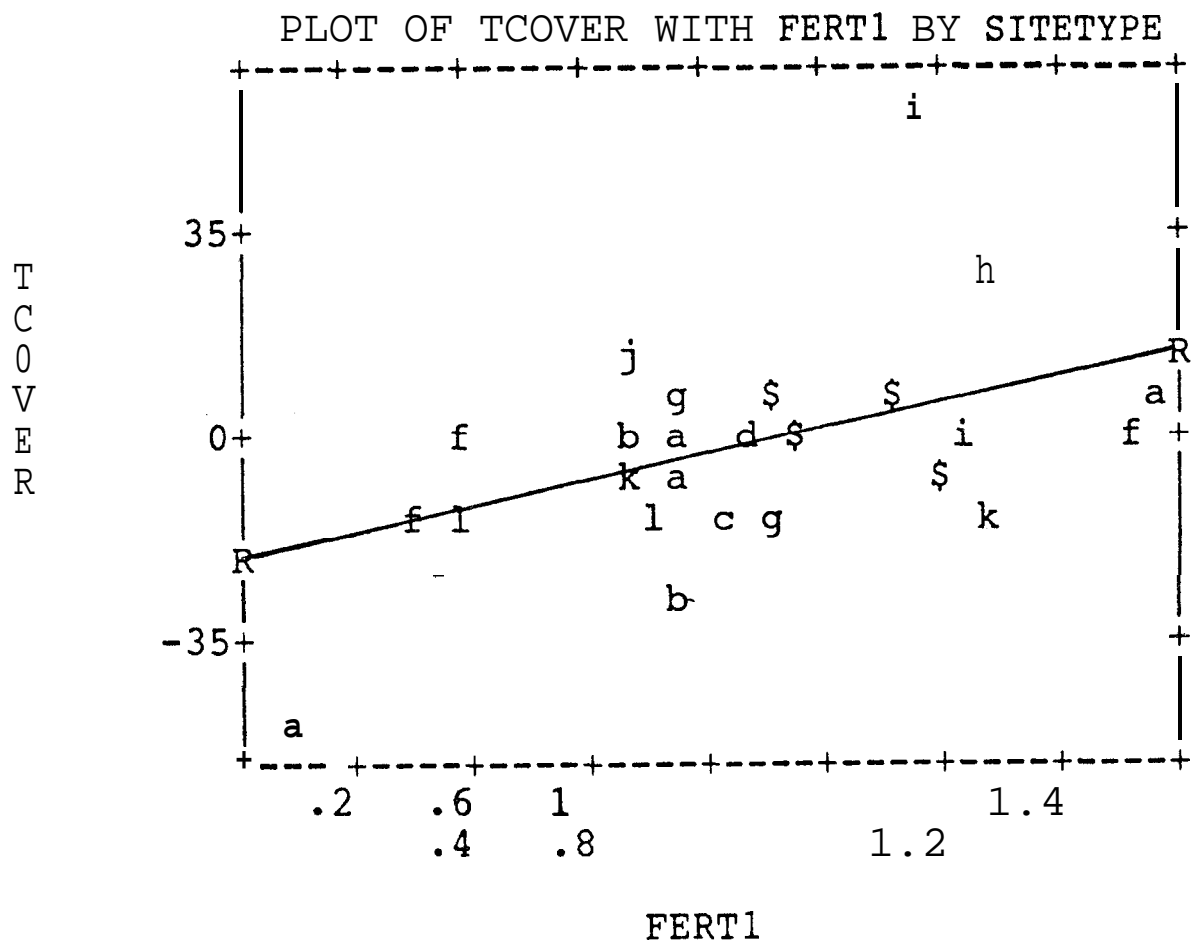
Includes records with missing data on age and sex in source files.

MALE COVERAGE BY FEMALE COVERAGE

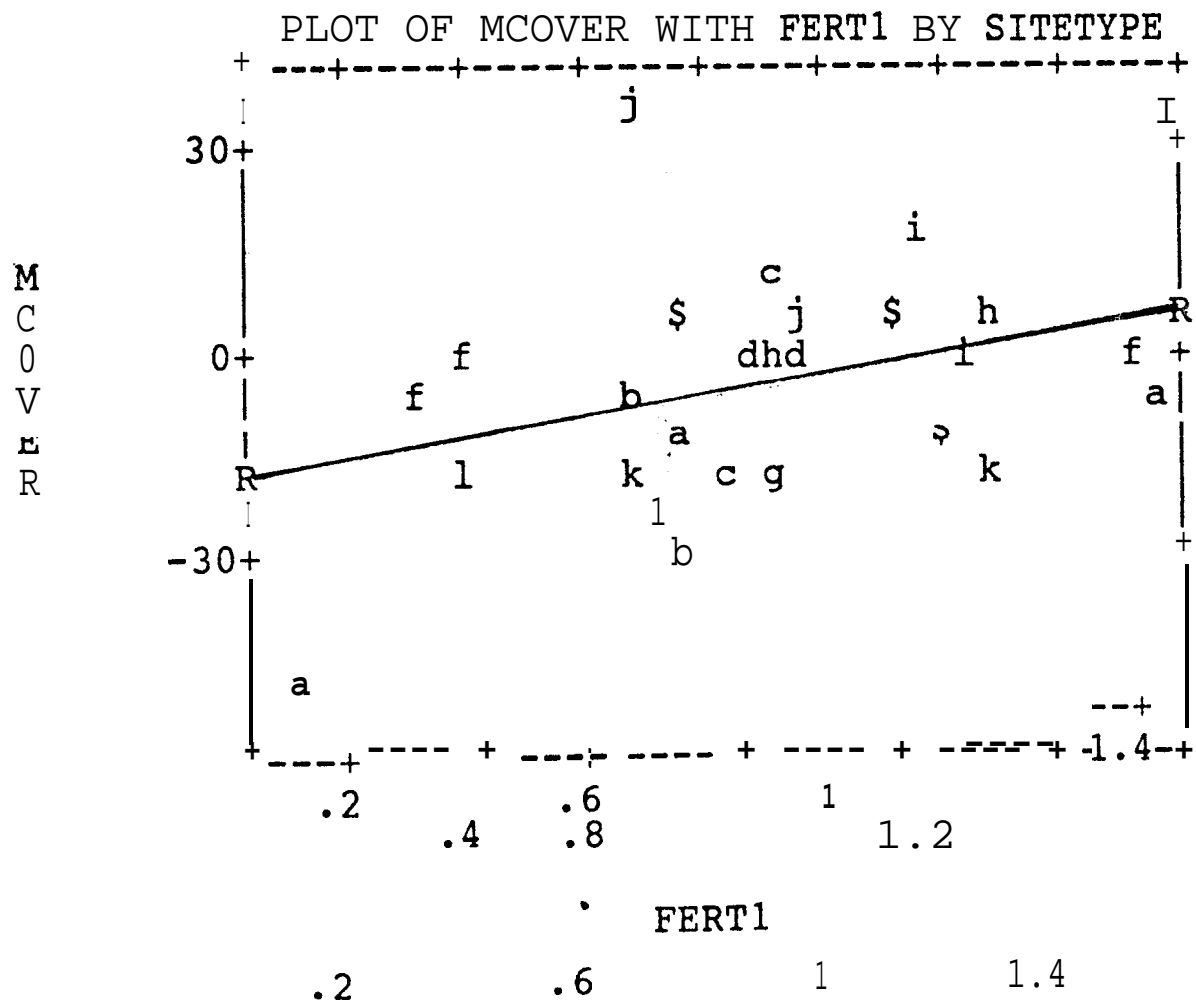
R = .7259 < .05



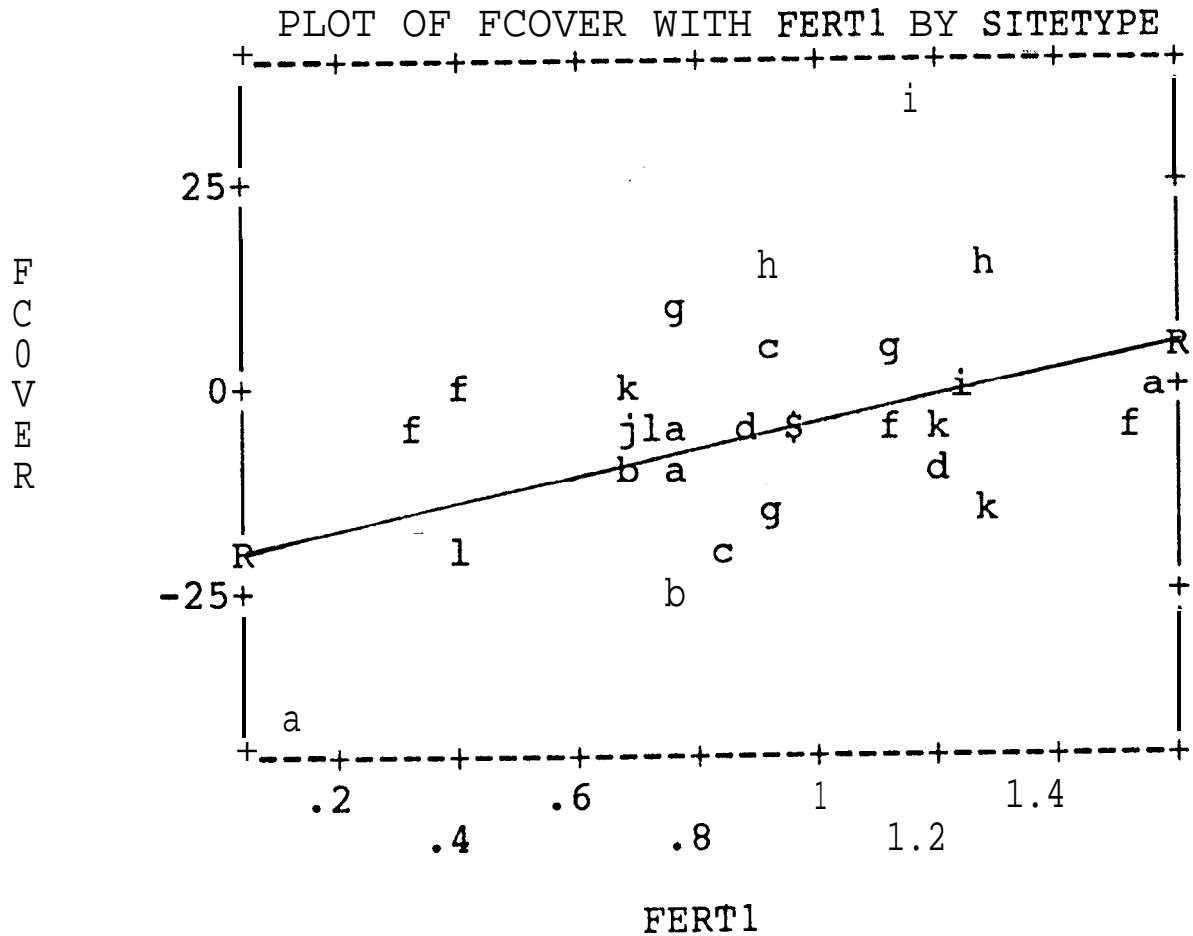
Plot 1
 Total Net Undercount or Overcount
 in the Census Enumeration
 with Fertility Ratio 1 (Children 0-6 to Females 18-34)
 Shown by Type of Site; All Ethnographic Sites



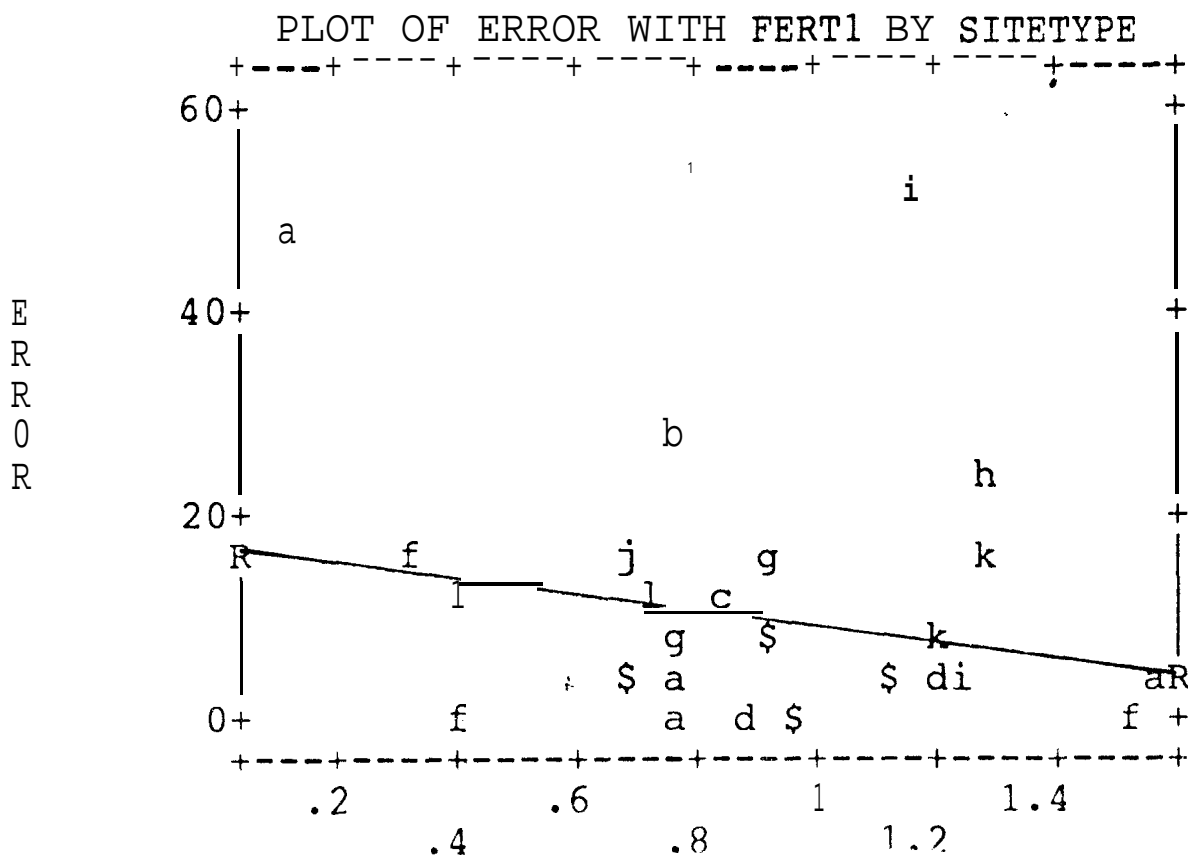
Plot 2
 Male Net Undercount or Overcount
 in the Census Enumeration
 with Fertility Ratio 1 (Children 0-6 to Females 18-34)
 Shown by Type of Site; All Ethnographic Sites



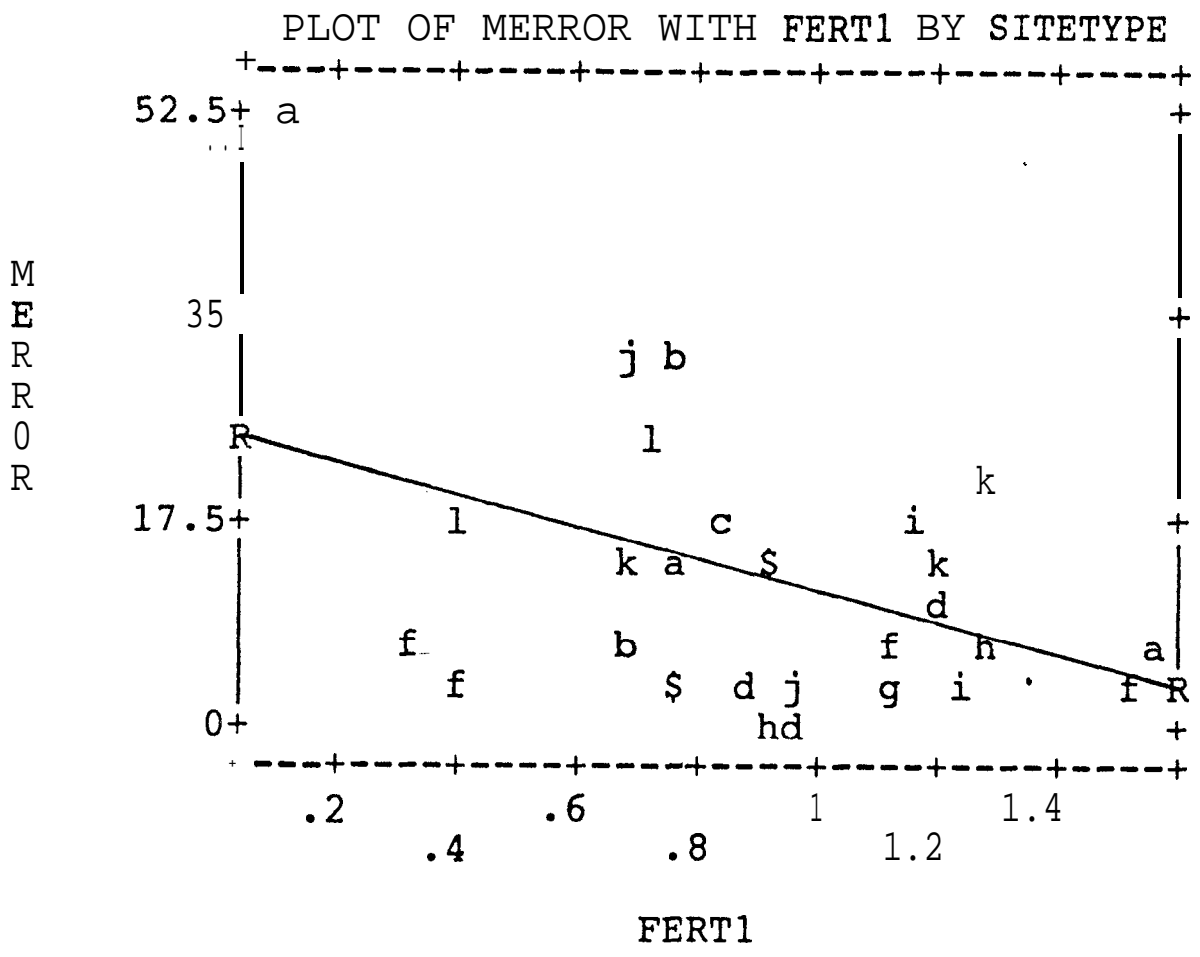
Plot 3
 Female Net Undercount or Overcount
 in the Census Enumeration
 with Fertility Ratio 1 (Children 0-6 to Females 18-34)
 Shown by Type of Site; All Ethnographic Sites



Plot 4
 Total Net Coverage Error
 in the Census Enumeration
 with Fertility Ratio 1 (Children 0-6 to Females 18-34)
 Shown by Type of Site; All Ethnographic Sites

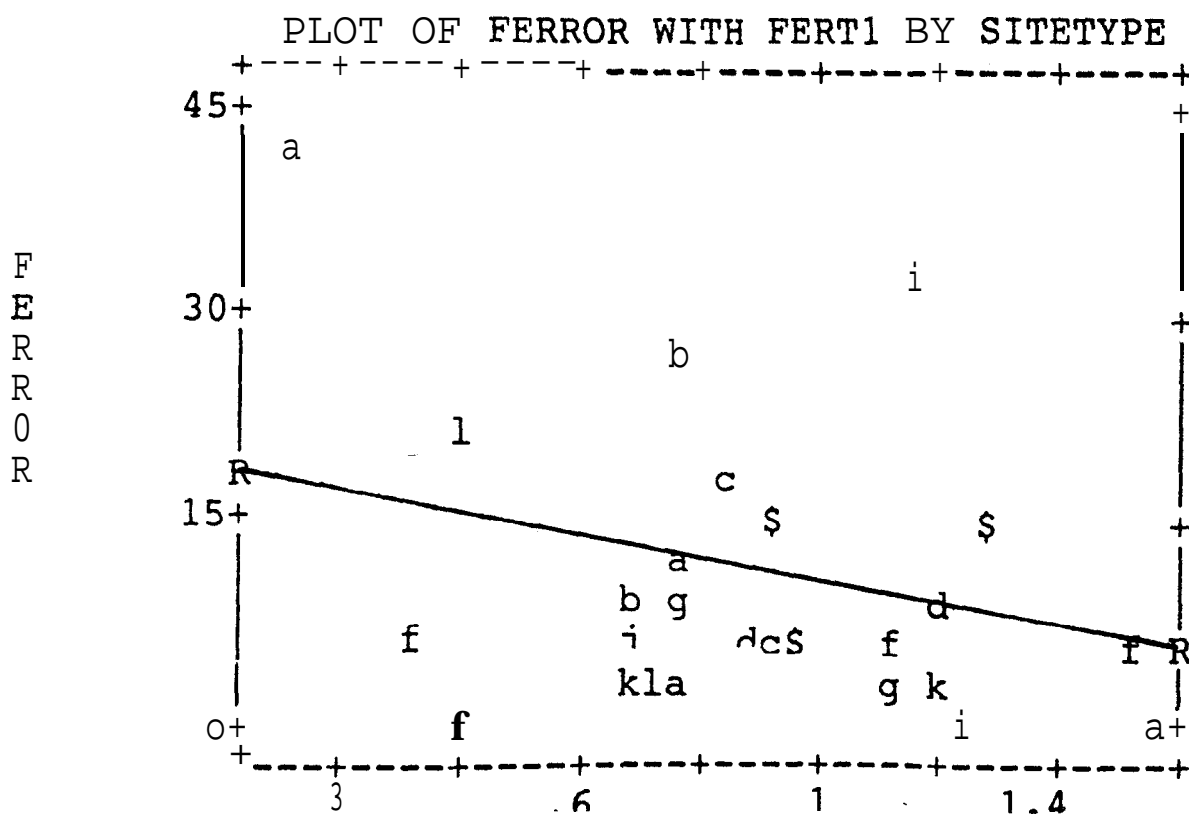


Plot 5
 Male Net Coverage Error
 in the Census Enumeration
 with Fertility Ratio 1 (Children 0-6 to Females 18-34)
 Shown by Type of Site; All Ethnographic Sites



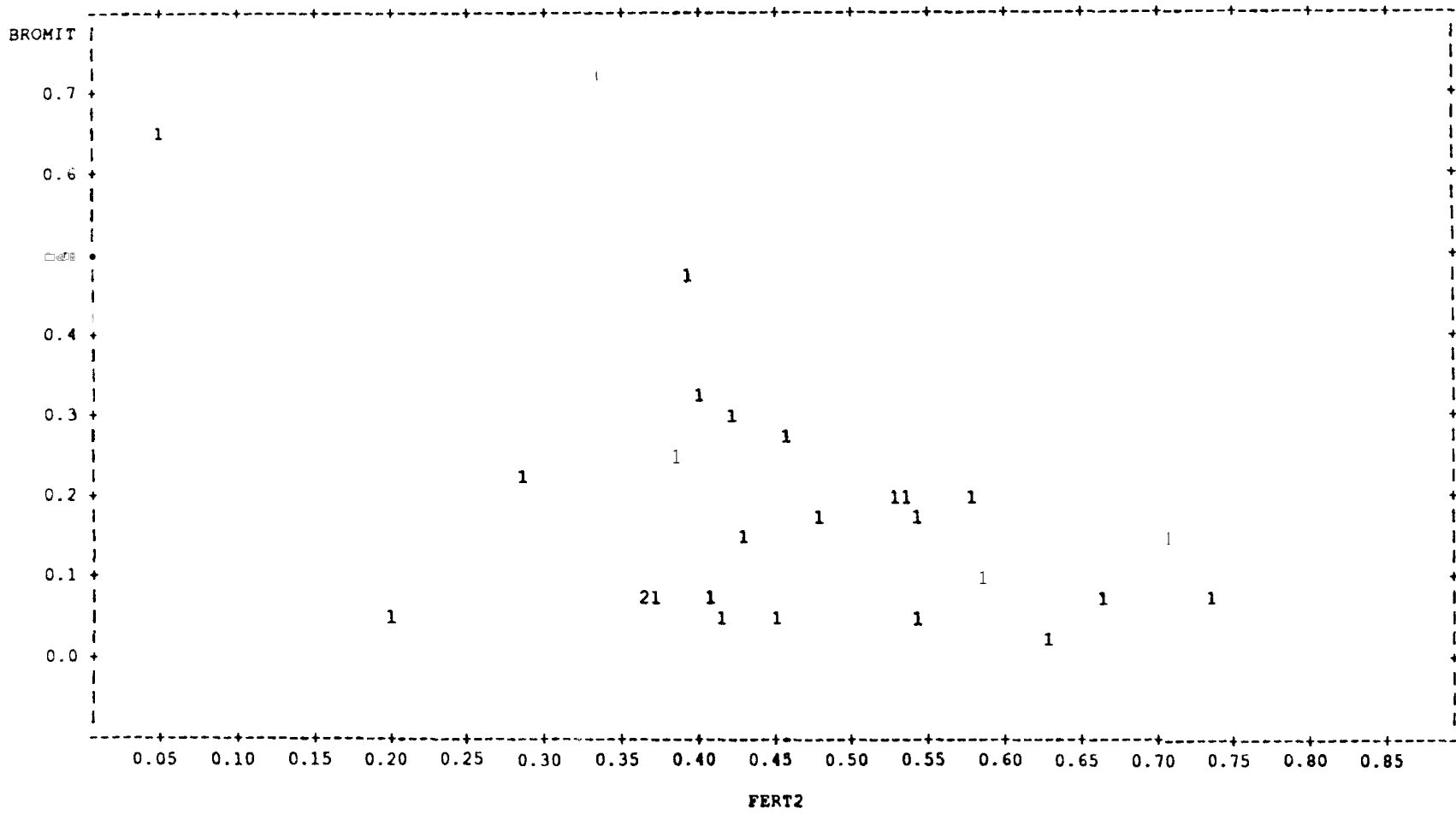
Plot 6
 Female Net Coverage Error
 in the Census Enumeration
 with Fertility Ratio 1 (Children 0-6 to Females 18-34)
 Shown by Type of Site: All Ethnographic Sites

Regression



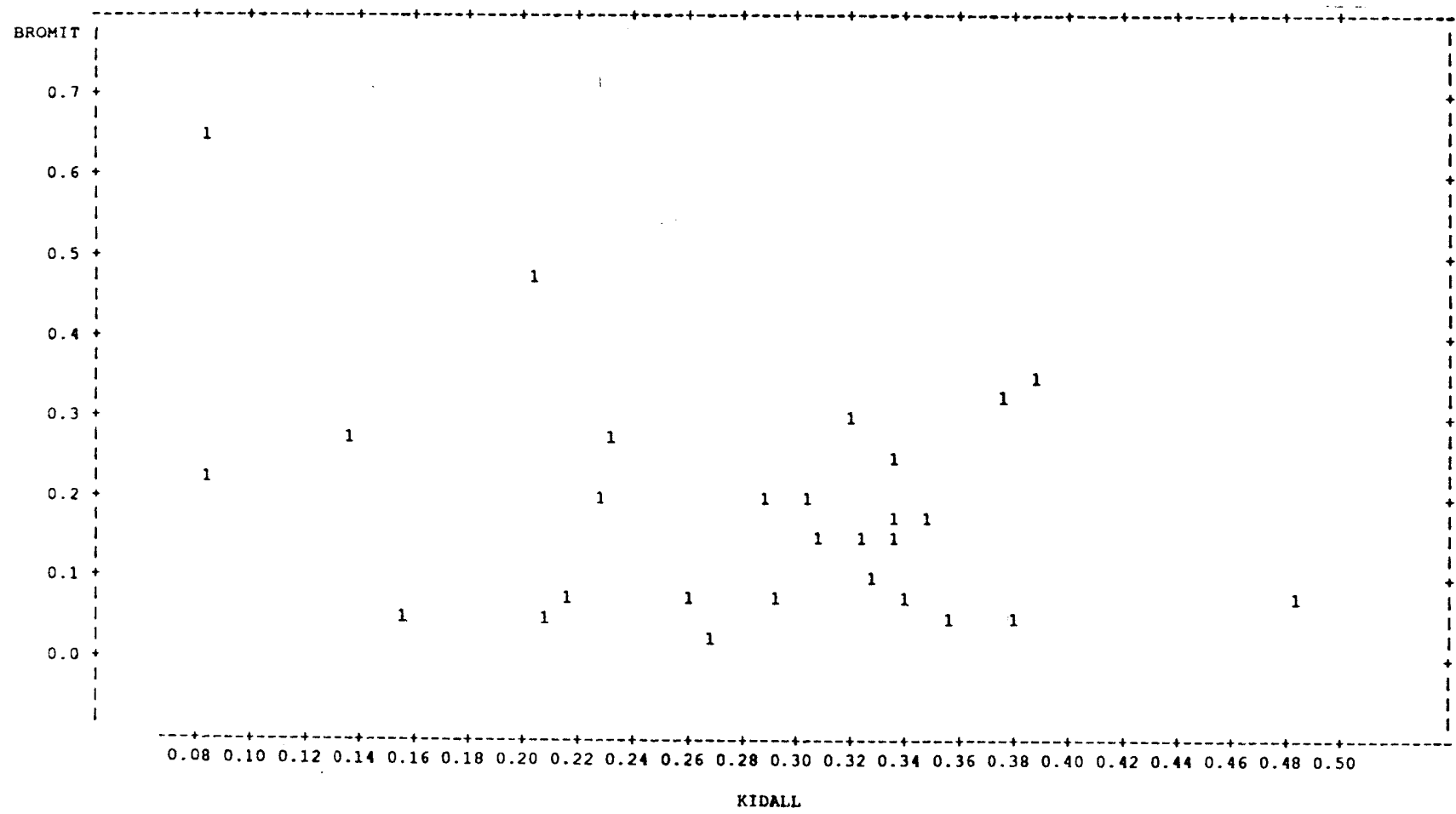
O M I S S I O N R A T E B Y F E R T I L I T Y

R = $-.4053$ $< .05$



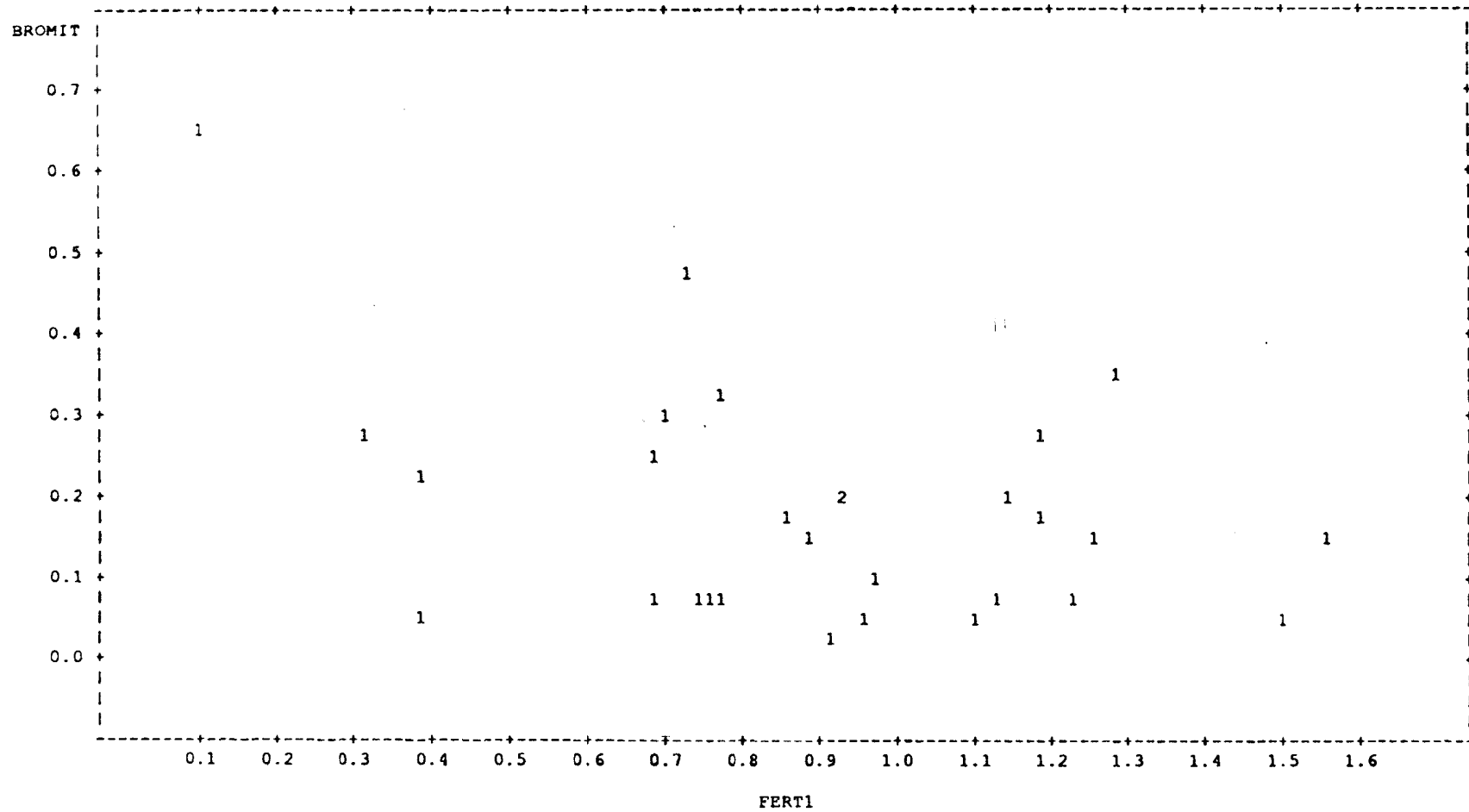
OMISSION RATE BY CHILDREN
AGE 0 to 6

$$R = 7.3624 < .05$$



MISSION RATE BY FERTILITY

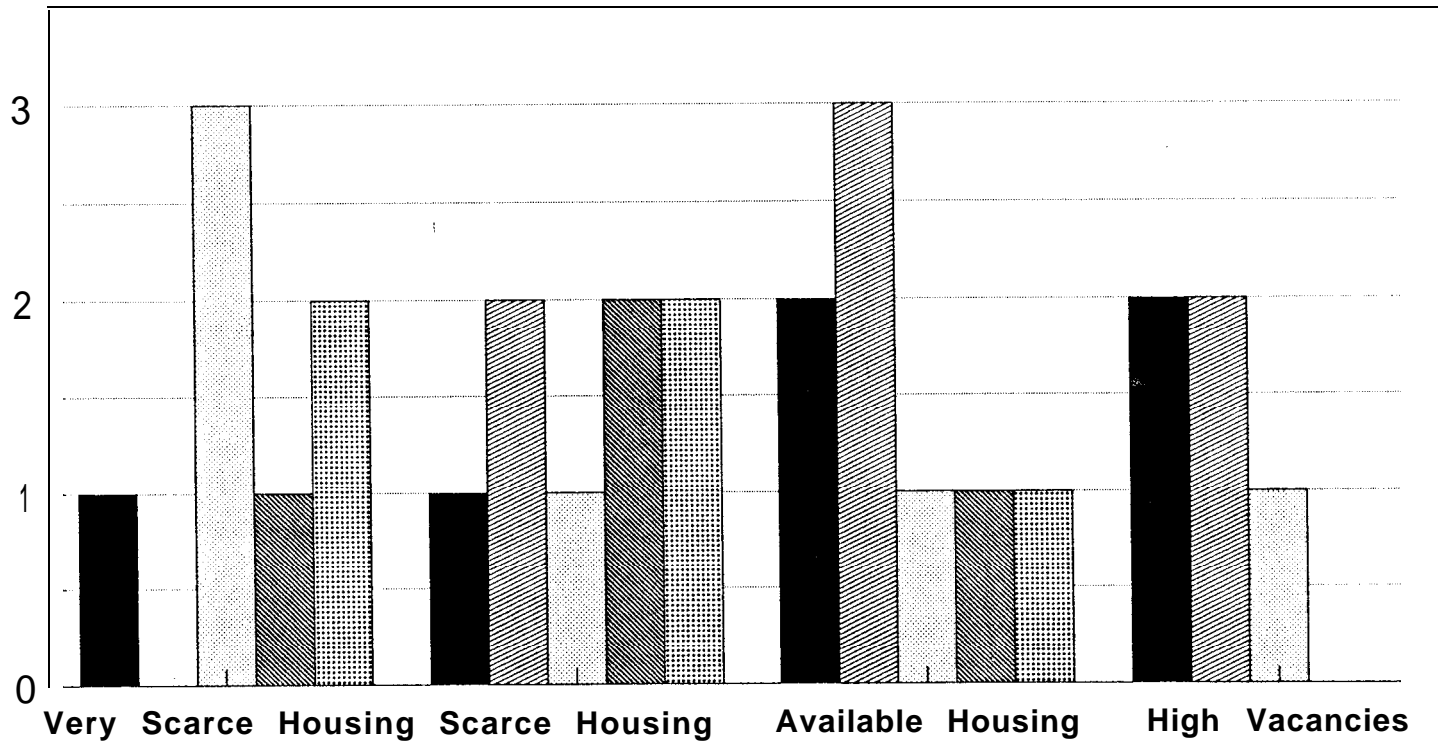
$$R = -.3991 < .05$$



Undercount/Overcount by Availability of Housing

SITES

$R = -.4181 < .05$

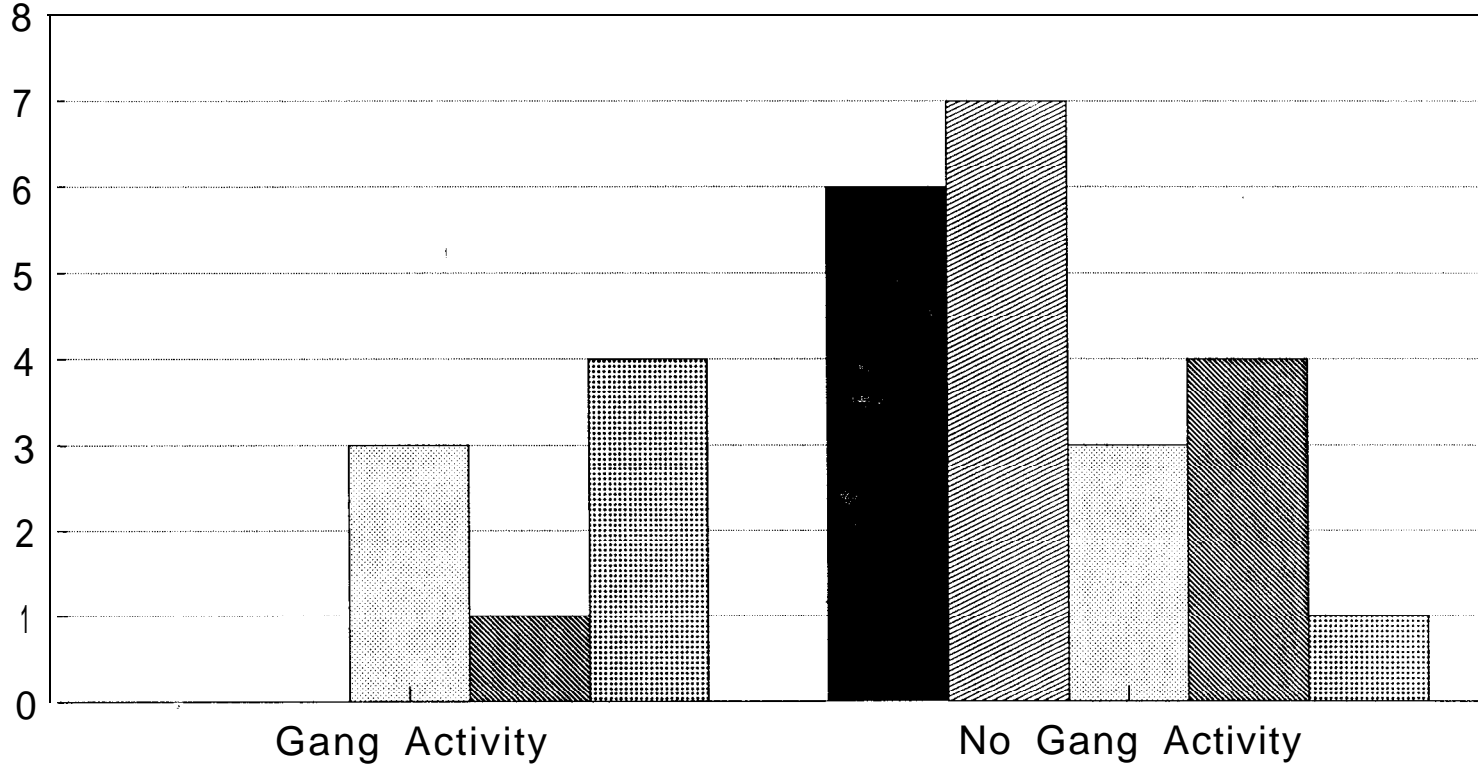


High Undercount
 Modest Undercount
 Good Censusing
 Modest Overcount
 High Overcount

Undercount/Overcount by Gang Activity

$R = .5637 < .05$

SITES



■ High Undercount ▨ Modest Undercount ▩ Good Censusing
▧ Modest Overcount ▤ High Overcount