ESTIMATION OF YEAR-TO-YEAR COVARIANCES FOR A NATIONAL HEALTH SURVEY

Van L. Parsons, National Center For Health Statistics
3700 East-West Highway, Hyattsville, MD 20782

KEY WORDS: covariance, complex sample survey

I. Introduction

National demographic surveys conducted by U.S. government agencies are often designed to collect data on a periodic basis over a 10 year time period. One such survey is the National Health Interview Survey (NHIS), sponsored by the National Center for Health Statistics (NCHS) and produced with the cooperation of the U.S. Bureau of the Census. This survey is redesigned after each decennial census, with the current survey being implemented from 1985 to 1994. In this 10 year interval, data will be collected on a weekly basis, and annual estimates will be presented in the NCHS Current Estimates Series 10 Reports.

For any individual year, the survey has a classical hierarchal design structure, and thus, analyses are facilitated by the existence of textbook methodology. The analyses of combined years of data, however, are often complicated by the fact that the NHIS does not always have a classical design structure over consecutive years. This was the case for the NHIS during the data collection years 1985, 1986 and 1987. In 1987 a full design was implemented, but in 1985 and 1986, sample reductions of 25% and 50% were imposed which resulted in nonclassical design structures for any combined years.

In this paper we will discuss the dependencies of the NHIS over different years and discuss the estimation of the covariance between the annual NHIS estimators of total. Often, variance estimators for annual totals are available, and if an estimator of year-to-year covariance for totals can be obtained, then variances of linear combinations of yearly estimators can be computed. Appropriate linearization techniques may then be used for functions of totals.

II. NHIS Design Structure over Years

The NHIS is a multistage complex survey documented in Parsons and Casady (1986). We give a brief description of the full design.

The Primary Sampling Units (PSU) of the U.S. are metropolitan areas and counties. The largest metropolitan areas define self-representing (SR) strata while the smaller municipal divisions are clustered to form non-self-representing (NSR) strata. Within each NSR stratum two PSU's are selected. Next, each SR stratum or sample PSU is subdivided into at most three substrata. Within each substratum a systematic sample of geographical areas, called enumeration districts (EDs), is selected from a prior implicitly stratified list. Within each selected ED (and possibly its successor in the list) a secondary sampling unit (SSU) consisting of up to 20 clusters of housing units is formed. The housing units of a given cluster within a SSU are dispersed throughout the SSU, and the systematic nature of dispersion would suggest that for many variables, the total universe SSU household variation within a fixed SSU would be dominated by a within cluster component and not a between cluster component.

The PSUs and SSUs are selected only once at an initial phase of the survey. Each year these same PSUs and same SSUs are visited, but a different component cluster within the SSU will be interviewed. Thus, an annual sample cluster within a sample SSU may be thought of as one component from a collection of about 20. Additional sampling stages may occur within this annual sample cluster. Furthermore, new SSUs from new construction areas are added to the annual sample to keep coverage up-to-date.

In theory, the covariance between two estimators of annual total can be computed using the general equation

$$\mathrm{Cov}(A,B) = .5\,[\mathrm{Var}(A+B) - \mathrm{Var}(A) - \mathrm{Var}(B)] \quad (1)$$

where A and B represent the annual totals from two different years.

If a full NHIS design is used from year to year then the variance expressions of (1) for yearly or combined years will be of similar form, thus making expression (1) of practical use. If, however, the full design is not used for a given year, then a direct computation of a covariance may be less complicated than the use of (1). Such was the case for the NHIS for 1985, 1986 and 1987 having 75%, 50% and full designs. In the next sections we discuss the sample reduction designs and the direct evaluation of covariances.

PSU Reduction models for NHIS

It was decided that the NHIS would consider three scenarios for implementation: full, 75%, and 50% designs. Furthermore, these designs would be nested at the PSU level in order to achieve maximum cost savings. These full, 75% and 50% designs will denoted by the label d = f,q and h.

A stratification structure and first stage selection criteria were devised to facilitate the creation of the three NHIS designs.

i. SR and NSR strata were defined for the full design; 2 sample PSUs per NSR stratum would be selected.

ii. The "small" SR strata of the full design were paired to form additional NSR strata for use with the 50% subdesign; 1 sample PSU per NSR stratum would be selected.

iii. The original NSR strata of i) and the additional NSR strata discussed in ii) were paired for use with the 75% subdesign; 3 sample PSUs per NSR stratum would be selected.

The collapsing in steps ii) and iii) was done using methodology similar to that done in i) (see Massey et al (1989)). There were some departures from this principle in practice, but for simplicity of discussion we assume no deviations.

The actual PSU sampling alluded to in i) - iii) above can be modeled as follows:

Consider a collapsed stratum consisting of the paired strata as defined in iii). Assume that the two component strata have N_1 and N_2 population PSUs, labeled $1,2,\ldots,N_1$ and $N_1+1,N_1+2,\ldots,N_1+N_2$, respectively.

For the full design, d=f, 2 PSUs are draw from each original NSR stratum using Durbin's (1967) procedure. Sample selection is done independently between the two components of the collapsed stratum. The indicator variables for PSU inclusion in the full design will be denoted:

$a_{fi} = 1$ (0) if PSU i is (not) selected, $i = 1,2,\ldots,N_1+N_2$.

The sample PSUs for the 50% and 75% designs are subsampled from those of the full design sample. The subsampling is modeled conditionally as follows.

Let i_1,i_2,i_3,i_4 be the four selected sample PSUs in a collapsed stratum for use with the full design; let $1 \le i_1 < i_2 < N_1+1 \le i_3 < i_4 \le N_1 + N_2$.

Next, generate X_0,X_1,X_2 independent symmetric Bernoulli random variables, and let $X_i' = 1-X_i$.

For the 50% and 75% designs, d=h and d=q, we define the indicator variables for PSU inclusion to have the conditional distributions specified below for a given set of full design sample PSUs.

$$a_{hi_1} = X_1,\ a_{hi_2} = X_1',\ a_{hi_3} = X_2,\ a_{hi_4} = X_2',$$

and

$$a_{qi_1} = X_0 + X_1X_0',\ a_{qi_2} = X_0 + X_1'X_0',$$

$$a_{qi_3} = X_0' + X_2X_0,\ a_{qi_4} = X_0' + X_2'X_0.$$

We note that the a_{hi}'s and a_{qi}'s are functions of the a_{fi}'s , and X_0,X_1,X_2, that

$a_{hi} \le a_{qi} \le a_{fi}$ (nesting of designs f,q,h for PSU inclusion), and

$$\sum_{i=1}^{N_j} a_{fi} = 2, \quad \sum_{i=1}^{N_j} a_{hi} = 1 \text{ for } j = 1,2,$$

$$\text{and} \quad \sum_{j=1}^{2} \sum_{i=1}^{N_j} a_{qi} = 3.$$

This procedure was used within each of the collapsed strata of iii).

Within PSU Sampling and Reduction

A fixed sample PSU is subject to second and higher levels of sample selection. Over two different years the surveys will share common SSUs, but the clusters of housing units sampled each year will be different. We assume the sample SSUs are selected at some initial time and then systematically labeled 1 to 4 for defining possible subdesigns. With this scheme, nested full, 75% and 50% sample SSUs can be constructed, e.g., dropping label 4 and dropping labels 1,4 will form 75% and 50% designs.

For the derivation of covariances, we shall assume that the full design SSUs are chosen by simple random sampling without replacement (SRSWOR), and any subsample of SSUs will also be chosen conditionally by SRSWOR.

The SSU reduction only occurs in the SR PSUs of the 50% and 75% designs; any PSU in a collapsed stratum will not experience a sample reduction at the second stage level. If an SR PSU in a reduced design had full design sampling fraction f then its reduced sampling fraction will be f/2 for the half design and (3/4)f for the 75% design.

Covariance within PSUs

Assuming the SRSWOR model for second stage sampling, the within PSU covariance between annual estimators of PSU totals can be determined. Consider the survey as having $m(d_0)$ sample SSUs selected at some initial stage by SRSWOR from a universe of M SSUs. Typically, if designs d_1 and d_2 are specified for respective years y_1 and y_2, and design d_1 is designated to have an equal or larger sample than design d_2, but no more than $m(d_0)$ units, then a sample of $m(d_1)$ SSUs for design d_1 will be chosen by SRSWOR from the $m(d_0)$ sample SSUs, and a sample of $m(d_2)$ SSUs will be chosen by SRSWOR from the $m(d_1)$ SSUs, thus imposing a nested structure. An unbiased estimator of a PSU total for year y is

$$\hat{X}(d,y) = \sum_{j=1}^{m(d)} \hat{X}_j(y)\ [M/m(d)],$$

where $\hat{X}_j(y)$ is an unbiased estimator of SSU j's total.

The within PSU covariance between yearly totals is

$$\mathrm{Cov}(\ \hat{X}(d_1,y_1),\hat{X}(d_2,y_2)\ ) =$$

$$M^2/m(d_1)\ [\ (1-f_1)\ S(y_1,y_2) + C_{ws}\ ] \qquad (2)$$

where

f_1 is the sampling fraction for design d_1,

$S(y_1,y_2)$ = the covariance between SSU totals of years y_1 and y_2:

$$S(y_1,y_2) = \sum_{j=1}^{M} \frac{(\ X_j(y_1)-\bar{X}(y_1)\ )\ (\ X_j(y_2)-\bar{X}(y_2))}{M-1}$$

with

$X_j(y)$ = total for SSU j in year y

$\bar{X}(y)$ = average SSU total for year y

and

$$C_{ws} = \sum_{j=1}^{M} \frac{C_j(y_1,y_2)}{M}$$

$C_j(y_1,y_2)$ = within SSU covariation due to all stages of subsampling within the SSU for years y_1 and y_2.

An estimator of $\mathrm{Cov}(\hat{X}(d_1,y_1),\hat{X}(d_2,y_2))$ is

$\widehat{\mathrm{Cov}}(\hat{X}(d_1,y_1),\hat{X}(d_2,y_2))$ =

$$M^2/m(d_1)\ \hat{S}(y_1,y_2:m(d_2)) \qquad (3)$$

where

$\hat{S}(y_1,y_2:m(d_2))$ =

$$\sum_{j=1}^{m(d_2)} \frac{(\hat{X}_j(y_1)-\hat{X}(y_1))(\hat{X}_j(y_2)-\hat{X}(y_2))}{m(d_2)-1},$$

the sample covariance of two estimated annual totals over the SSUs common to both samples.

It can be seen that the relative bias of this estimator is:

$$[\ (1/f_1 -1) + C_{ws} / S(y_1,y_2)\ ]^{-1}.$$

Thus if f_1 is small, the proposed estimator is approximately unbiased. This estimator can be used as the basis of covariance computation in the SR PSUs.

It should be noted that in 1986 the 50% sample SSUs in the SR PSUs did not subset the 1985 75% sample SSUs, but were selected by SRSWOR from the initial stage sample SSUs. The formulas (2) and (3) can be modified to reflect this by replacing $m(d_1)$ with $m(d_0)$ and f_1 with f_0, the sampling fraction for the initial SSU sample.

Horvitz-Thompson Estimators for NSR Strata and their Covariances

Consider a collapsed NSR stratum as discussed in the section on reduction models. For design d in year y, a Horvitz-Thompson estimator of a characteristic total X(y) in this collapsed stratum will be considered. We define:

$X_i(y)$ to be PSU i's total for year y

$\hat{X}_i(y)$ to be an unbiased estimator of PSU i's total for year y

a_{di} = 1 (0) if PSU i is (not) in sample for design d

π_{di} = Prob(PSU i in sample for design d)

π_{dij} = Prob(PSUs i and j in sample for design d).

The estimator of collapsed stratum total is

$$\hat{X}(d,y) = \sum_{i=1}^{N_1+N_2} \frac{\hat{X}_i(y)}{\pi_{di}}\ a_{di}.$$

Using the reduction models discussed earlier, the covariance for annual estimators of totals may be derived.

It will be convenient to partition the index set of all possible pairs of universe PSUs within a collapsed stratum,

$S_0 = \{(i,j) : 1 \leq i < j \leq N_1+N_2\}$ into sets:

$S_1 = \{(i,j) : 1 \leq i < j \leq N_1$ or $N_1+1 \leq i < j \leq N_1+N_2\}$ and

$S_2 = \{(i,j) : 1 \leq i \leq N_1 < j\}$.

The covariances can be expressed:

$\mathrm{Cov}(\hat{X}(d_1,y_1), \hat{X}(d_2,y_2))$ =

$$\sum_{(i,j)\in S_0} \left(\frac{X_i(y_1)}{\pi_{d_1 i}} - \frac{X_j(y_1)}{\pi_{d_1 j}} \right) \left(\frac{X_i(y_2)}{\pi_{d_2 i}} - \frac{X_j(y_2)}{\pi_{d_2 j}} \right) \times H(d_1,d_2)\, G_{ij}(d_1,d_2)$$

$$+ \sum_{i=1}^{N_1+N_2} \mathrm{Cov}_w(\hat{X}_i(y_1), \hat{X}_i(y_2)) \frac{\pi_{fi}}{\pi_{d_1 i}\,\pi_{d_2 i}} D(d_1,d_2) \quad (4)$$

where

$\mathrm{Cov}_w(\hat{X}_i(y_1), \hat{X}_i(y_2))$ is the within PSU covariance having the form expressed in (2), and scale factors $D(d_1,d_2)$, $H(d_1,d_2)$, and $G_{ij}(d_1,d_2)$ are defined:

Design	Scale Factors		
d_1,d_2	H	G	D
f, f	1	$(\pi_{fi}\pi_{fj} - \pi_{fij})\ I(S_1)$	1
f, q	3/4	$(\pi_{fi}\pi_{fj} - \pi_{fij})\ I(S_1)$	3/4
f, h	1/2	$(\pi_{fi}\pi_{fj} - \pi_{fij})\ I(S_1)$	1/2
q, q	1	$(\pi_{qi}\pi_{qj} - \pi_{qij})$	3/4
q, h	1	$(.75\pi_{fi}\pi_{fj} - .5\pi_{fij})\ I(S_1)$	1/2
h, h	1	$\pi_{hi}\pi_{hj}\ I(S_1)$	1/2

where

$I(S_1)$ = 1 (0) if (i,j) is (not) in the set S_1,

$\pi_{qi} = (3/4)\pi_{fi}$, $\pi_{hi} = (1/2)\pi_{fi}$,

$\pi_{qij} = (1/2)\pi_{fij}\ I(S_1) + (1/2)\ \pi_{fi}\ \pi_{fj}\ I(S_2)$.

Note that

1. If $y_1 = y_2$ and $d_1 = d_2$ the above formulas will reduce to the variance of the yearly estimator.

2. By an algebraic manipulation, it can be observed that the covariances for the (f,f), (f,q), and (f,h) designs are equivalent.

Estimators of Covariance Between Annual Estimators of Total

There are two cases to consider. If two consecutive years have the same design then the data from the two years can be combined at the PSU and SSU level and considered as one sample from the specified design. Variances for combined and individual years can be estimated, and an estimate of covariance can thus be computed from relation (1). It is usually the case that variance estimation software is available for a specific design, and thus, covariance estimation can be implemented using variance estimation techniques. We shall focus upon the case of having different designs in consecutive years.

When the annual surveys have different designs, the above mentioned technique would be difficult to implement because the combined surveys would not have a classical design structure. We consider direct estimators of (4).

Suppose that the designs (d_2,d_1) are nested so that d_2 is contained in d_1 at the first stage, i.e, $(d_2,d_1) = (h,q)$, (h,f) or (q,f). A natural approach to estimating (4) is to use the crossproduct deviations of sample annual totals. To do this we need at least two PSUs within each collapsed stratum with each containing two years of data; furthermore, these two PSUs must be in the same half of the collapsed stratum. We approach the problem as follows.

To simplify our expressions we let

$$\hat{X}'_{di} = \hat{X}_i(y)/\pi_{di}\ ,\ d = d_2,\ d_1$$

where i is a sample PSU of design d.

For some PSUs this variable will be defined for d_1 but not d_2 (or equivalently y_1 but not y_2). To define an estimator of crossproduct deviation over the two designs (or years) we introduce the variable:

$$\hat{X}''_{d_2i} = \hat{X}'_{d_2i}\, a_{d_2i} + \lambda_i\, \hat{X}'_{d_1i}\, (1-a_{d_2i}) - \lambda_i C\, \hat{X}'_{d_1i}$$

where C = 1/2 for (h,q) and (h,f), and C = 1/4 for (q,f) and λ_i is an arbitrary constant.

Note that $\hat{X}''_{d_2i}$ is defined for all sample PSUs in the d_1 design.

Our estimator of the covariance for annual totals is

$$\hat{\mathrm{Cov}}(\ \hat{X}(d_2,y_1),\hat{X}(d_1,y_2)\ ) =$$

$$A \sum_{(i,j)\epsilon S_1} G_{ij}/\pi_{ij}\ (\hat{X}'_{d_1i}-\hat{X}'_{d_1j})(\hat{X}''_{d_2i}-\hat{X}''_{d_2j})\ a_{d_1i}\ a_{d_1j}$$

$$+ \sum_{i=1}^{N_1+N_2} \hat{\mathrm{Cov}}_w(\hat{X}_i(y_1),\hat{X}_i(y_2))\ a_{d_2i}/\pi_{d_2i}$$

with A = 2 for $(d_2,d_1) = (h,q)$ and A = 1 otherwise.

The within PSU covariance terms are estimated using expression (3).

As an example, suppose that PSUs in the collapsed stratum for the f,q and h designs are [(1,2):(3,4)], [(1,2):(3)], and [(2):(3)] respectively. The estimators of design-to-design (year-to-year) covariance are:

$$\hat{\mathrm{Cov}}\ (\hat{X}_f,\hat{X}_q) =$$

$$(\hat{X}'_{f1}-\hat{X}'_{f2})(\hat{X}'_{q1}-\hat{X}'_{q2}-\frac{\lambda_1}{4}\hat{X}'_{f1}+\frac{\lambda_2}{4}\hat{X}'_{f2})G_{12}/\pi_{12}$$

$$+(\hat{X}'_{f3}-\hat{X}'_{f4})(\hat{X}'_{q3}-\lambda_4\hat{X}'_{q4}-\frac{\lambda_3}{4}\hat{X}'_{f3}+\frac{\lambda_4}{4}\hat{X}'_{f4})G_{34}/\pi_{34}$$

$$+ \sum_{i=1}^{3} \hat{\mathrm{Cov}}_w(\hat{X}_i(y_1),\hat{X}_i(y_2)/\pi_{qi}$$

$$\hat{\mathrm{Cov}}\ (\hat{X}_f,\hat{X}_h) =$$

$$(\hat{X}'_{f1}-\hat{X}'_{f2})(\lambda_1\hat{X}'_{f1}-\hat{X}'_{h2}-\frac{\lambda_1}{2}\hat{X}'_{f1}+\frac{\lambda_2}{2}\hat{X}'_{f2})G_{12}/\pi_{12}$$

$$+(\hat{X}'_{f3}-\hat{X}'_{f4})(\hat{X}'_{h3}-\lambda_4\hat{X}'_{f4}-\frac{\lambda_3}{2}\hat{X}'_{f3}+\frac{\lambda_4}{2}\hat{X}'_{f4})G_{34}/\pi_{34}$$

$$+ \sum_{i=2}^{3} \hat{\mathrm{Cov}}_w(\hat{X}_i(y_1),\hat{X}_i(y_2)/\pi_{hi}$$

$$\mathrm{Cov}\ (\hat{X}_q,\hat{X}_h) =$$

$$2\ (\hat{X}'_{q1}-\hat{X}'_{q2})(\lambda_1\hat{X}'_{q1}-\hat{X}'_{h2}-\frac{\lambda_1}{2}\hat{X}'_{q1}+\frac{\lambda_2}{2}\hat{X}'_{q2})G_{12}/\pi_{12}$$

$$+ \sum_{i=2}^{3} \mathrm{Cov}_w(\hat{X}_i(y_1),\hat{X}_i(y_2)/\pi_{hi}$$

If the second stage sampling fraction is small then these estimators will be approximately unbiased. The constant λ_i in the estimators should be chosen to reduce the variability of the covariance estimator. If the PSU totals are known, the λ_i should be selected to be $[X_i(y_2)/\pi_{d_2i}]\ /\ [\ X_i(y_1)/\pi_{d_1i}]$. In practice, the PSU totals will be unknown, and the λ_i must be estimated. If it is assumed that the PSU total changes little from year to year then the λ_i can be estimated by π_{d_1i}/π_{d_2i} which is 1.5, 2, and 4/3 for $(d_2,d_1) = (h,q)$, (h,f), and (q,f) respectively.

Examples of Covariance

At the time this paper was written only the 1985 and 1986 NHIS data tapes were available. Thus, only the year-to-year covariance for 75% and 50% designs could be evaluated. Table I provides estimated correlations for 1985 and 1986 totals as previously discussed. Table II provides estimates of the correlations of means and percents. These values were produced by Taylor linearization.

It should be kept in mind that the correlations produced are subject to sampling variability. We do not have an estimate of the magnitude; our empirical results suggest that for the smaller subdomains the results are quite variable. Furthermore, the NHIS has been subject to "weighting" problems which adversely affect estimation. These problems are under study.

We observed that demographic and socio-economic variables usually showed a larger year-to-year correlation than health variables. This is due to the fact that both the first and second stage universe units are stratified by such variables. Correlations in the range 0.20 to 0.40 were common. Health variable correlation, however, seemed to range from 0.00 to 0.20.

References

Durbin, J. (1967), Design of multi-stage surveys for the estimation of sampling errors, App. Stat., 16, 152-164.

Massey, J.T., Moore, T.F., Tadros, W., Parsons, V.L. (1989) Sample Design and Estimation Procedures for the National Health Interview Survey, 1985-94, National Center for Health Statistics Series 2 Report, to appear.

Parsons, V.L., Casady R.J. (1986) Variance Estimation and the Redesigned National Health Interview Survey, Proceedings of the Section on Survey Research Methods of the American Statistical Association.

TABLE I

CORRELATION OF 1985 AND 1986 ESTIMATORS OF TOTALS

CHARACTERISTIC	TOTAL (1000s) (CV) 1985	TOTAL (1000s) (CV) 1986	CORR
US POPULATION	216.1 (0.87)	215.8 (1.30)	0.30
BLACK POPULATION	25.2 (3.51)	25.9 (5.11)	0.36
POVERTY POPULATION	26.1 (3.45)	26.4 (4.10)	0.23
DOCTOR VISITS	1143 (1.57)	1165 (1.82)	0.11
DOCTOR VISITS BY BLACKS	120 (4.60)	118 (6.34)	0.17
DOCTOR VISITS BY WOMEN	685 (1.82)	697 (2.18)	0.09
HOSPITAL DAYS	170 (2.68)	155.2 (2.69)	0.03
BED DAYS	1336 (2.48)	1416 (3.02)	0.12

TABLE II

CORRELATION BETWEEN 1985 AND 1986 ESTIMATORS OF MEANS AND PERCENTS

CHARACTERISTIC (subdomain)	PERCENT or MEAN 1985	(CV) 1985	PERCENT or MEAN 1986	(CV) 1986	CORR
% BLACK	11.68	(3.60)	11.99	(5.08)	0.38
% HISPANIC	8.17	(4.56)	8.48	(6.83)	0.27
% UNEMPLOYED	5.64	(2.62)	5.69	(3.36)	0.14
% 13+ YEARS EDUC.	28.41	(1.25)	29.09	(1.48)	0.38
% NO PHONE	7.27	(3.59)	6.95	(4.18)	0.15
% BORN IN JANUARY	8.17	(1.17)	8.21	(1.58)	0.01
% EXCELLENT HEALTH	39.22	(0.89)	39.15	(0.92)	0.12
(MALE)	42.53	(0.89)	42.59	(0.94)	0.11
(FEMALE)	36.19	(1.04)	36.01	(1.12)	0.11
(BLACK)	29.87	(2.62)	30.62	(3.04)	0.06
(NONBLACK)	40.45	(0.91)	40.31	(1.00)	0.05
DOCTOR VISITS	5.29	(1.42)	5.40	(1.70)	0.14
(MALE)	4.44	(2.06)	4.55	(2.16)	0.12
(FEMALE)	6.07	(1.69)	6.18	(1.96)	0.04
(BLACK)	4.74	(3.51)	4.55	(5.01)	0.05
(NONBLACK)	5.36	(1.52)	5.52	(1.79)	0.16
RESTRICTED ACTIVITY DAYS	14.85	(2.13)	15.26	(2.14)	0.14
(MALE)	12.82	(2.90)	13.13	(2.60)	0.17
(FEMALE)	16.70	(2.44)	17.21	(2.54)	0.14
(BLACK)	17.41	(4.58)	16.33	(4.78)	-0.03
(NONBLACK)	14.51	(2.17)	15.11	(2.32)	0.23
BED DAYS	6.18	(2.40)	6.56	(2.86)	0.17
(MALE)	5.19	(3.55)	5.37	(3.48)	0.20
(FEMALE)	7.08	(2.82)	7.65	(3.53)	0.16
(BLACK)	8.01	(5.12)	7.50	(6.65)	0.04
(NONBLACK)	5.94	(2.63)	6.43	(2.96)	0.18