ESTIMATION OF S.E.'S FOR MEANS AND PROPORTIONS FROM THE NMIHS

Wayne E. Johnson
Centers for Disease Control/National Center for Health Statistics
Room 915, 6525 Belcrest Road, Hyattsville, MD 20782

KEY WORDS: SUDAAN, STANDARD ERRORS

"Quick and dirty" methods of approximating standard errors are compared with results from SUDAAN[1], using data from the livebirths sample of the 1988 National Maternal and Infant Health Survey.

In round numbers, ten thousand birth certificates were selected from among the four million recorded in the United States during 1988. The universe of records was sorted by race (black, nonblack) and within race by birthweight (<1500, 1500<= & <2500 and >=2500). The sampling fractions for these birthweight categories were: 1/14, 1/55 and 1/113 for blacks, 1/29, 1/160 and 1/720 for nonblacks. The basic sampling weight is the reciprocal of the sampling fraction. Additional information was used to form nonresponse adjustment cells within the strata. Thus, sampling weights vary within strata.

Means are estimated for the mother's age, height and weight; percentages are estimated for the conditions of being married, low age (< 19), and low education (< 12th grade).

The results of the five "quick and dirty" estimation methods are compared with SUDAAN on the full sample, the six race/birthweight sampling strata, the two collapsed race strata and the three collapsed birthweight categories.

The "quick and dirty" methods make no use of the survey design; all but one of them make some use of the sample weights. SAS's PROC UNIVARIATE was used to compute the estimates. UNIVARIATE calculates the sample mean as:

$$XBAR_w = SUM_{i-1,\ldots,N}(W_i * X_i) / SUM_{i-1,\ldots,N}(W_i)$$

and the variance as:

$$SUM_{i-1,\ldots,N}(W_i * (X_i - XBAR_w)^2) / d$$

d depends upon the VARDEF option:

$$VARDEF = \begin{cases} DF \\ WDF \end{cases} ==> d = \begin{cases} N-1 \\ SUM_{i-1,\ldots,N}(W_i) - 1 \end{cases}$$

The first method is to treat the data as though it were collected in a simple random sample; in this case, actual weights are ignored (replace the W_i above with 1) and VARDEF=DF so that d=N-1. As will be seen in the tabulations, especially those for marital status and education, ignoring the sample weights can lead to gross errors in the estimates of population parameters. Therefore, this method should not be used. The unweighted estimate of the population parameter and its standard error are shown in the third column, "UNWTD ESTMAT", and the fourth column, respectively.

Methods 2 - 5 will use the sample weight, W_i, or a constant multiple, KW_i. In the formula for UNIVARIATE's mean, let W_i be replaced by KW_i; then K is a factor common to both numerator and denominator: $XBAR_w$ is unchanged. Accordingly, weighted estimates for means or percentages will only be tabulated in the fifth column: "WTD ESTMAT". The argument above, applied to the formula for UNIVARIATE's variance, shows that multiplication of the weights by a constant also leaves the variance unchanged when VARDEF=WDF (ignoring the -1 in the denominator). When VARDEF=DF, the effect is to multiply the original weighted variance by the given constant (or to multiply the original weighted S.E. by the

square root of that constant). These effects will be seen in the tabulations. For these reasons, VARDEF=DF, rather than WDF, is used with the constant scale changes in the fourth and fifth methods explained below.

The second method is to allow UNIVARIATE to use the sample weights to estimate the mean and variance with VARDEF=DF. The parameter estimate appears in the fifth column: "WTD ESTMAT" and its standard error appears in the sixth column. This method greatly increases the variance over that in method one, by a factor approximately equal to the mean weight, which is nearly 379 in this survey. Thus the standard error is multiplied by about 20. The tabulations show that this grossly exceeds the estimate for standard error produced by SUDAAN. Therefore, this method ought not to be used.

The third method differs from the second in that VARDEF=WDF. The standard error is shown in the seventh column. Empirically, for a given variable, SUDAAN's S.E. in column 10 is a constant multiple (square root of the design effect in column 11) of that in column 7. For this survey, the tabulations show that the design effects are nearly the same across the variables selected, so this is a good option for "quick and dirty" work.

The fourth method is to multiply all weights by the reciprocal of the mean weight,

$$N \ / \ SUM_{i=1,\ldots,N}(W_i)$$

(in this survey, 1/379).

The heuristic explanation for this normalization is that the weights sum to N over the entire sample; with VARDEF=DF, the degrees of freedom will be the actual number of observations. The standard error is shown in the eighth column: "SCALE 1". It is evident that this column does not bear the same systematic relationship to SUDAAN's S.E. enjoyed by the S.E. for the third method. Therefore, this method is not recommended.

The fifth method is due to Ecob and Williams[2]: multiplying all weights by

$$SUM_{i=1,\ldots,N}(W_i) \ / \ SUM_{i=1,\ldots,N}(W_i^2)$$

(in this survey, about .00135) and using VARDEF=WDF may result in a satisfactory approximation to the standard error. However, as shown above, all scale changes yield the same standard error when using VARDEF=WDF as in method 3. The calculation was repeated using SPSS-X, and again yielded the result of method 3. In the spirit of experimentation, VARDEF=DF was tried, also. The resulting S.E.'s are tabulated in column 9: "SCALE 2". It is evident from the tables that this was no more successful than the fourth method. As mentioned above, when one set of weights is related to another by a constant multiplier and VARDEF=DF, then the S.E.'s of the one are related to those of the other by the square root of the constant multiplier. This can be seen by comparing the S.E.'s for "SCALE 1" to those of "SCALE 2".

1.Shah, Babubhai V.,Barnwell, Beth G., Hunt, P. Nileen, and LaVange, Lisa M. (1991). SUDAAN User's Manual, Release 5.50. Research Triangle Institute, Research Triangle Park, NC, 27709

2. Ecob, R. and Williams, R.(June 1991). Sampling Asian Minorities to Assess Health and Welfare. Journal of Epidemiology and Community Health,45(2),93-101.

1988 National Maternal and Infant Health Survey
Mothers' Mean Age at Time of Birth

Column 1	2 NMISS: NOBSRVD	3 UNWTD ESTMAT	4 S.E. DF (N-1)	5 WTD ESTMAT	6 S.E. DF (N-1)	7 S.E. WDF SUM(W'S)	8 SCALE 1 S.E. DF (N-1)	9 SCALE 2 S.E. DF (N-1)	10 SUDAAN S.E.	11 DESIGN EFFECT
ALL:	0:9953	25.61	0.0580	26.28	1.0897	0.0560	0.0560	0.0401	0.0775	1.91
RACE:										
BLACK	0:5226	24.47	0.0799	24.41	0.8759	0.0789	0.0450	0.0322	0.0851	1.16
~BLACK	0:4727	26.87	0.0804	26.67	2.0508	0.0797	0.1054	0.0754	0.0921	1.34
BIRTHWEIGHT:										
LOW	0:1551	25.64	0.1518	25.78	0.8199	0.1508	0.0421	0.0301	0.1561	1.07
MEDIUM	0:1517	25.37	0.1536	25.68	1.8154	0.1522	0.0933	0.0668	0.1666	1.20
HIGH	0:6885	25.66	0.0686	26.32	1.5119	0.0670	0.0777	0.0556	0.0826	1.52
RACE BY BIRTHWEIGHT:										
BLACK/LOW	0:841	24.65	0.2034	24.67	0.9300	0.2000	0.0478	0.0342	0.1974	0.97
BLACK/MEDIUM	0:803	24.35	0.2074	0.21	1.8560	0.2051	0.0953	0.0682	0.2048	1.00
BLACK/HIGH	0:3582	24.45	0.0957	24.39	1.1884	0.0951	0.0611	0.0437	0.0947	0.99
~BLACK/LOW	0:710	26.80	0.2200	26.50	1.3844	0.2220	0.0711	0.0509	0.2252	1.03
~BLACK/MEDIUM	0:714	26.51	0.2206	26.20	3.2051	0.2210	0.1647	0.1179	0.2234	1.02
~BLACK/HIGH	0:3303	26.97	0.0934	26.69	2.8361	0.0949	0.1457	0.1043	0.0970	1.05

....

1988 National Maternal and Infant Health Survey
Percentage of Mothers Less Than 19 Years of Age

Column 1	2 NMISS: NOBSRVD	3 UNWTD ESTMAT	4 S.E. DF (N-1)	5 WTD ESTMAT	6 S.E. DF (N-1)	7 S.E. WDF SUM(W'S)	8 SCALE 1 S.E. DF (N-1)	9 SCALE 2 S.E. DF (N-1)	10 SUDAAN S.E.	11 DESIGN EFFECT
ALL:	0:9953	11.30	0.3174	7.90	5.2623	0.2703	0.2703	0.1935	0.3701	1.87
RACE:										
BLACK	0:5226	16.04	0.5076	15.59	5.5677	0.5018	0.2860	0.2047	0.5430	1.17
~BLACK	0:4727	6.07	0.3474	6.32	9.1016	0.3538	0.4676	0.3347	0.4332	1.50
BIRTHWEIGHT:										
LOW	0:1551	12.57	0.8421	11.32	4.3706	0.8046	0.2245	0.1607	0.8162	1.03
MEDIUM	0:1517	13.18	0.8689	11.07	9.6070	0.8055	0.4935	0.3533	0.8532	1.12
HIGH	0:6885	10.60	0.3711	7.66	7.2363	0.3205	0.3718	0.2661	0.3942	1.51
RACE BY BIRTHWEIGHT:										
BLACK/LOW	0:841	16.77	1.2889	15.79	5.8451	1.2574	0.3003	0.2149	1.2287	0.95
BLACK/MEDIUM	0:803	17.31	1.3359	15.74	11.6313	1.2853	0.5975	0.4272	1.2393	0.93
BLACK/HIGH	0:3582	15.58	0.6060	15.56	7.5708	0.6057	0.3889	0.2784	0.6059	1.00
~BLACK/LOW	0:710	7.61	0.9956	8.38	6.4874	1.0401	0.3333	0.2386	1.0990	1.10
~BLACK/MEDIUM	0:714	8.54	1.0468	9.02	15.5548	1.0720	0.7988	0.5717	1.1027	1.06
~BLACK/HIGH	0:3303	5.21	0.3866	6.16	12.5035	0.4183	0.6423	0.4598	0.4558	1.19

1988 National Maternal and Infant Health Survey
Mothers' Mean Height

Column 1	2 NMISS: NOBSRVD	3 UNWTD ESTMAT	4 S.E. DF (N-1)	5 WTD ESTMAT	6 S.E. DF (N-1)	7 S.E. WDF SUM(W'S)	8 SCALE 1 S.E. DF (N-1)	9 SCALE 2 S.E. DF (N-1)	10 SUDAAN S.E.	11 DESIGN EFFECT
ALL:	266:9687	64.47	0.0298	64.51	0.5693	0.0292	0.0292	0.0209	0.0403	1.91
RACE:										
BLACK	170:5056	64.55	0.0419	64.62	0.4651	0.0419	0.0239	0.0171	0.0455	1.18
~BLACK	96:4631	64.40	0.0423	64.49	1.0770	0.0418	0.0553	0.0396	0.0476	1.30
BIRTHWEIGHT:										
LOW	58:1493	64.26	0.0775	64.24	0.4209	0.0775	0.0216	0.0155	0.0813	1.1
MEDIUM	40:1477	64.05	0.0753	63.98	0.9101	0.0762	0.0468	0.0335	0.0856	1.26
HIGH	168:6717	64.61	0.0355	64.55	0.7899	0.0349	0.0406	0.0290	0.0430	1.52
RACE BY BIRTHWEIGHT:										
BLACK/LOW	34:807	64.32	0.1055	64.33	0.4889	0.1052	0.0251	0.0180	0.1054	1.00
BLACK/MEDIUM	25:778	64.14	0.1025	64.14	0.9277	0.1026	0.0477	0.0341	0.1033	1.01
BLACK/HIGH	111:3471	64.69	0.0507	64.69	0.6336	0.0507	0.0326	0.0233	0.0507	1.00
~BLACK/LOW	24:686	64.20	0.1141	64.18	0.7132	0.1144	0.0366	0.0262	0.1155	1.02
~BLACK/MEDIUM	15:699	63.95	0.1108	63.91	1.6217	0.1118	0.0833	0.0596	0.1141	1.04
~BLACK/HIGH	57:3246	64.53	0.0497	64.52	1.4872	0.0498	0.0764	0.0547	0.0501	1.01

....

1988 National Maternal and Infant Health Survey
Mothers' Mean Weight Just Before Pregnancy

Column 1	2 NMISS: NOBSRVD	3 UNWTD ESTMAT	4 S.E. DF (N-1)	5 WTD ESTMAT	6 S.E. DF (N-1)	7 S.E. WDF SUM(W'S)	8 SCALE 1 S.E. DF (N-1)	9 SCALE 2 S.E. DF (N-1)	10 SUDAAN S.E.	11 DESIGN EFFECT
ALL:	389:9564	136.13	0.2990	134.77	5.4722	0.2802	0.2811	0.2012	0.3798	1.84
RACE:										
BLACK	244:4982	138.69	0.4353	139.53	4.8401	0.4357	0.2487	0.1780	0.4727	1.17
~BLACK	145:4579	133.35	0.4028	133.81	10.0987	0.3926	0.5188	0.3714	0.4465	1.29
BIRTHWEIGHT:										
LOW	75:1476	135.59	0.8065	135.09	4.3733	0.8035	0.2247	0.1608	0.8404	1.09
MEDIUM	55:1462	130.40	0.7234	129.01	8.5222	0.7130	0.4378	0.3134	0.7734	1.18
HIGH	259:6626	137.52	0.3566	135.12	7.5987	0.3357	0.3904	0.2794	0.4052	1.46
RACE BY BIRTHWEIGHT:										
BLACK/LOW	52:789	136.82	1.1172	136.74	5.1728	1.1133	0.2657	0.1902	1.1149	1.00
BLACK/MEDIUM	36:767	132.56	1.0207	132.67	9.3172	1.0299	0.4787	0.3426	1.0467	1.03
BLACK/HIGH	156:3426	140.49	0.5280	140.42	6.5940	0.5276	0.3388	0.2425	0.5279	1.00
~BLACK/LOW	23:687	134.18	1.1629	134.04	7.2725	1.1662	0.3736	0.2674	1.1783	1.02
~BLACK/MEDIUM	19:695	128.01	1.0163	127.44	14.6084	1.0070	0.7505	0.5372	1.0082	1.00
~BLACK/HIGH	103:3200	134.34	0.4686	134.13	14.0006	0.4683	0.7193	0.5148	0.4705	1.01

1988 National Maternal and Infant Health Survey
Percentage of Mothers Who Are Married

Column 1	2 NMISS: NOBSRVD	3 UNWTD ESTMAT	4 S.E. DF (N-1)	5 WTD ESTMAT	6 S.E. DF (N-1)	7 S.E. WDF SUM(W'S)	8 SCALE 1 S.E. DF (N-1)	9 SCALE 2 S.E. DF (N-1)	10 SUDAAN S.E.	11 DESIGN EFFECT
SAS VARDEF==> DENOMINATOR==>										
ALL:	0:9953	58.96	0.4931	74.42	8.5136	0.4374	0.4374	0.3131	0.5694	1.69
RACE:										
BLACK	0:5226	36.97	0.6678	36.76	7.4005	0.6670	0.3802	0.2721	0.7206	1.17
~BLACK	0:4727	83.27	0.5430	82.16	14.3247	0.5568	0.7359	0.5268	0.6818	1.50
BIRTHWEIGHT:										
LOW	0:1551	52.61	1.2683	55.24	6.8583	1.2626	0.3523	0.2522	1.2234	0.94
MEDIUM	0:1517	53.46	1.2811	59.72	15.0187	1.2592	0.7716	0.5523	1.2996	1.06
HIGH	0:6885	61.60	0.5862	75.57	11.6924	0.5178	0.6007	0.4300	0.6075	1.38
RACE BY BIRTHWEIGHT:										
BLACK/LOW	0:841	33.29	1.6260	30.88	7.4060	1.5932	0.3805	0.2723	1.5645	0.96
BLACK/MEDIUM	0:803	31.26	1.6368	28.88	14.4734	1.5994	0.7435	0.5322	1.5685	0.96
BLACK/HIGH	0:3582	39.11	0.8155	37.87	10.1306	0.8105	0.5204	0.3725	0.8063	0.99
~BLACK/LOW	0:710	75.49	1.6154	71.27	10.5916	1.6982	0.5441	0.3895	1.7883	1.11
~BLACK/MEDIUM	0:714	78.43	1.5403	73.23	24.0311	1.6570	1.2346	0.8837	1.7929	1.17
~BLACK/HIGH	0:3303	85.98	0.6042	82.72	19.6645	0.6579	1.0102	0.7231	0.7175	1.19

....

1988 National Maternal and Infant Health Survey
Percentage of Mothers with Less Than 12-th Grade Education

Column 1	2 NMISS: NOBSRVD	3 UNWTD ESTMAT	4 S.E. DF (N-1)	5 WTD ESTMAT	6 S.E. DF (N-1)	7 S.E. WDF SUM(W'S)	8 SCALE 1 S.E. DF (N-1)	9 SCALE 2 S.E. DF (N-1)	10 SUDAAN S.E.	11 DESIGN EFFECT
SAS VARDEF==> DENOMINATOR==>										
ALL:	23:9930	24.65	0.4325	20.20	7.8474	0.4029	0.4031	0.2886	0.5564	1.91
RACE:										
BLACK	16:5210	29.88	0.6342	29.17	6.9882	0.6297	0.3590	0.2570	0.6824	1.17
~BLACK	7:4720	18.88	0.5697	18.36	14.5018	0.5636	0.7450	0.5333	0.6569	1.36
BIRTHWEIGHT:										
LOW	7:1544	28.37	1.1476	27.71	6.1888	1.1391	0.3179	0.2276	1.1873	1.09
MEDIUM	2:1515	29.64	1.1736	28.97	13.8940	1.1654	0.7139	0.5110	1.3002	1.24
HIGH	14:6871	22.72	0.5055	19.57	10.8109	0.4786	0.5554	0.3975	0.59	1.53
RACE BY BIRTHWEIGHT:										
BLACK/LOW	5:836	32.78	1.6244	32.69	7.5410	1.6224	0.3874	0.2773	1.6270	1.00
BLACK/MEDIUM	0:803	33.87	1.6712	33.72	15.0979	1.6684	0.7756	0.5552	1.6764	1.01
BLACK/HIGH	11:3571	28.31	0.7540	28.52	9.4439	0.7556	0.4852	0.3473	0.7579	1.01
~BLACK/LOW	2:708	23.16	1.5866	24.45	10.0754	1.6152	0.5176	0.3705	1.6550	1.05
~BLACK/MEDIUM	2:712	24.86	1.6209	26.88	24.0991	1.6614	1.2381	0.8862	1.7242	1.08
~BLACK/HIGH	3:3300	16.67	0.6488	17.87	19.9359	0.6670	1.0242	0.7331	0.6909	1.07