

## 5.5 Estimates of Variance

Two types of errors—sampling and nonsampling—affect sample survey statistics. The standard errors calculated for the NLSY97 reflect, for the most part, the magnitude of the sampling error. They also take into account some of the effects of nonsampling error related to response and enumeration, but they do not account for any systematic biases in the data.

Nonsampling error in surveys can be attributed to a variety of sources, including inability to obtain information about all persons in the sample, differences in question interpretation, inability or unwillingness of respondents to provide accurate information, inability to recall information, processing errors, imputation errors, and undercoverage. Several studies were undertaken to investigate the existence and prevalence of nonsampling error. Undercoverage was discussed extensively throughout this chapter, and several special studies were discussed in section 5.3 and are described in detail in Appendices D, E, L, M, N, O, and P.

Sampling error is the name given to the between-sample variation in sample-based estimates. These differences occur by chance, and the variability is measured by *standard errors* of the estimates. Sample estimates from a given survey design are unbiased when an average of the estimates from all possible samples would yield the true population value. In this case, the sample estimate and its estimated standard error can be used to construct approximate confidence intervals, or ranges of values, that include the true population value with known probabilities. Specifically, approximately 95 percent of the intervals from two standard deviations below the estimate to two standard deviations above would include the true population parameter.

The impact of departures from simple random sampling on the precision of sample estimates is often measured by the *design effect* (DEFF). The design effect is defined as the ratio of the variance corrected for the sampling design to the variance that would be obtained given a hypothetical simple random sample. Most complex, multi-stage designs result in a design effect greater than one; in other words, the variance of an estimate is greater than the variance that would be obtained had the data been based on a simple random sample.

To estimate the variance using information about the sample design, it is necessary to use statistical procedures such as Taylor series approximations, Balanced Half Samples, or Jackknife. For the NLSY97, NORC used the Taylor series procedure to calculate standard errors. Finally, the square root of the design effect, referred to as DEFT, is another useful measure provided for multiplication by simple random sample standard errors to obtain design-corrected standard errors.

Standard errors, design effects, and root design effects were calculated for six domains in the NLSY97 sample, including the full sample, males, females, Hispanics, non-Hispanic blacks, and other races/ethnicities, and for a total of 25 binary and continuous survey variables. They can be found in Tables 5.37 through 5.48. For the binary variables (odd-numbered tables), the sample size estimate (in percent), design and simple random sample standard errors (in percent), and design effect and root design effect are displayed. For the continuous variables (even-numbered tables), the total, mean, design and simple random sample standard errors of the mean, and design effect and root design effect are given. Summaries of the design effect and root design effect are also given for each set of variables in each domain, including the mean, standard deviation, minimum, median, and maximum.

A final table, Table 5.49, summarizes the root design effects among the six domains and two types of variables. As a rough rule of thumb, we recommend that NLSY97 analysts divide the statistical test statistics, derived under simple random sampling assumptions, by the median DEFT (or by its square in the case of  $\chi^2$  tests) prior to consulting the reference distributions of the tests and to determining the *p*-

values. This practice will approximately correct the test statistics for the fact that the NLSY97 is not based on a simple random sample, but instead upon correlated observations.<sup>16</sup>

---

<sup>16</sup> For more details, see Skinner, C.J., Holt, D., and Smith, T.M.F. (1989), *Analysis of Complex Surveys*, John Wiley & Sons, New York.  
or  
Lehtonen, R. and Pahkinen, E.J. (1994), *Practical Methods for Design and Analysis of Complex Surveys*, John Wiley & Sons, New York.

**Table 5.37. Variance Estimates for Overall NLSY97 Sample: Proportions**

PROPORTIONS (BINARY VARIABLES)	SAMPLE SIZE	PERCENT	DESIGN SE	SRS SE	DEFF	DEFT
R enrolled?	8983	97.45	0.21	0.17	1.52	1.24
R in labor force?	8984	17.49	0.57	0.40	2.03	1.42
R employed?	8984	17.17	0.55	0.40	1.89	1.38
R working >= 10 hrs/week? (alternative 1) <sup>17</sup>	8837	9.15	0.38	0.31	1.50	1.22
R working >= 10 hrs/week? (alternative 2) <sup>18</sup>	8837	11.99	0.45	0.34	1.75	1.32
R employed during school year?	8984	10.58	0.47	0.32	2.16	1.47
R ever worked during school year?	8984	19.77	0.58	0.42	1.91	1.38
R employed during summer?	8984	12.20	0.46	0.34	1.83	1.35
R ever worked during summer?	8984	15.99	0.53	0.39	1.85	1.36
R ever retained a grade?	7708	15.01	0.70	0.41	2.91	1.71
R ever skipped a grade?	7718	1.84	0.19	0.15	1.60	1.27
R ever in remedial classes?	3872	38.93	1.15	0.78	2.17	1.47
R have a computer at home?	5410	58.13	1.19	0.67	3.15	1.78
R have a quiet place to study?	5408	90.14	0.48	0.40	1.44	1.20
R's parent born in US?	7936	88.63	0.91	0.36	6.39	2.53
R's parent has worked at least 3 months since R born?	7927	95.00	0.53	0.24	4.88	2.21
<b>DEFF/DEFT SUMMARY (PROPORTIONS)</b>						
Mean					2.44	1.52
Standard Deviation					1.36	0.37
Minimum					1.44	1.20
Median					1.90	1.38
Maximum					6.39	2.53

<sup>17</sup> For this variable, all youths who reported working at job(s) more than 10 hours per week were coded as 1; youths reporting working less than 10 hours per week were coded as 0; and youths reporting "hours vary" were coded as 0. An alternative to this coding scheme is given in the following variable and described in the footnote below.

<sup>18</sup> For this variable, all youths who reported working at job(s) more than 10 hours per week were coded as 1; youths reporting working less than 10 hours per week were coded as 0; and youths reporting "hours vary" were coded as 1.

**Table 5.38. Variance Estimates for Overall NLSY97 Sample: Means**

MEANS (CONTINUOUS VARIABLES)	SAMPLE SIZE	MEAN	DESIGN SE	SRS SE	DEFF	DEFT
Number of weeks worked in 1996?	5398	8.11	0.30	0.20	2.25	1.50
Days/wk have dinner with family?	5356	5.19	0.04	0.03	1.78	1.33
Days/wk have fun with family?	5356	2.67	0.03	0.03	1.00	1.00
Hours/week on homework?	4723	5.99	0.10	0.07	2.04	1.43
Hours/week watching TV?	5382	17.26	0.25	0.17	2.16	1.47
Hours/week reading for pleasure?	5382	2.92	0.08	0.06	1.78	1.33
Parent's income	5211	22474.78	443.48	261.88	2.87	1.69
Spouse's income	3797	37698.98	991.44	599.69	2.73	1.65
Percent chance of R getting 4-year degree by age 30	3004	69.73	0.75	0.59	1.62	1.27
DEFF/DEFT SUMMARY (MEANS)						
Mean					2.03	1.41
Standard Deviation					0.57	0.21
Minimum					1.00	1.00
Median					2.04	1.43
Maximum					2.87	1.69

**Table 5.39. Variance Estimates for NLSY97 Males: Proportions**

PROPORTIONS (BINARY VARIABLES)	SAMPLE SIZE	PERCENT	DESIGN SE	SRS SE	DEFF	DEFT
R enrolled?	4598	97.40	0.28	0.23	1.48	1.22
R in labor force?	4599	18.59	0.70	0.57	1.51	1.23
R employed?	4599	18.22	0.68	0.57	1.42	1.19
R working >= 10 hrs/week? (a)	4517	9.37	0.50	0.43	1.35	1.16
R working >= 10 hrs/week? (b)	4517	12.49	0.60	0.49	1.50	1.22
R employed during school year?	4599	11.47	0.58	0.47	1.52	1.23
R ever worked during school year?	4599	21.80	0.76	0.61	1.55	1.24
R employed during summer?	4599	13.40	0.59	0.50	1.39	1.18
R ever worked during summer?	4599	17.76	0.65	0.56	1.35	1.16
R ever retained a grade?	3950	18.11	0.92	0.61	2.27	1.51
R ever skipped a grade?	3957	1.58	0.23	0.20	1.32	1.15
R ever in remedial classes?	1907	39.75	1.41	1.12	1.58	1.26
R have a computer at home?	2791	57.97	1.39	0.93	2.23	1.49
R have a quiet place to study?	2791	90.68	0.61	0.55	1.23	1.11
R's parent born in US?	4080	88.34	1.06	0.50	4.49	2.12
R's parent has worked at least 3 months since R born?	4074	94.97	0.58	0.34	2.91	1.70
<b>DEFF/DEFT SUMMARY (PROPORTIONS)</b>						
Mean					1.82	1.32
Standard Deviation					0.84	0.26
Minimum					1.23	1.11
Median					1.50	1.22
Maximum					4.49	2.12

**Table 5.40. Variance Estimates for NLSY97 Males: Means**

MEANS (CONTINUOUS VARIABLES)	SAMPLE SIZE	MEAN	DESIGN SE	SRS SE	DEFF	DEFT
Number of weeks worked in 1996?	2750	9.07	0.38	0.30	1.60	1.27
Days/wk have dinner with family?	2768	5.35	0.05	0.04	1.56	1.25
Days/wk have fun with family?	2769	2.71	0.05	0.04	1.56	1.25
Hours/week on homework?	2407	5.36	0.12	0.09	1.78	1.33
Hours/week watching TV?	2777	18.31	0.31	0.24	1.67	1.29
Hours/week reading for pleasure?	2780	2.28	0.10	0.08	1.56	1.25
Parent's income	2659	22633.18	481.41	365.71	1.73	1.32
Spouse's income	2003	37146.39	1011.58	743.51	1.85	1.36
Percent chance of R getting 4-year degree by age 30	1519	66.23	1.02	0.87	1.37	1.17
<b>DEFF/DEFT SUMMARY (MEANS)</b>						
Mean					1.63	1.28
Standard Deviation					0.14	0.06
Minimum					1.37	1.17
Median					1.60	1.27
Maximum					1.85	1.36

**Table 5.41. Variance Estimates for NLSY97 Females: Proportions**

PROPORTIONS (BINARY VARIABLES)	SAMPLE SIZE	PERCENT	DESIGN SE	SRS SE	DEFF	DEFT
R enrolled?	4385	97.50	0.28	0.24	1.36	1.17
R in labor force?	4385	16.33	0.76	0.56	1.84	1.36
R employed?	4385	16.07	0.75	0.55	1.86	1.36
R working >= 10 hrs/week? (a)	4320	8.92	0.53	0.43	1.52	1.23
R working >= 10 hrs/week? (b)	4320	11.47	0.61	0.48	1.62	1.27
R employed during school year?	4385	9.64	0.59	0.44	1.80	1.34
R ever worked during school year?	4385	17.63	0.76	0.58	1.72	1.31
R employed during summer?	4385	10.95	0.57	0.47	1.47	1.21
R ever worked during summer?	4385	14.12	0.65	0.52	1.56	1.25
R ever retained a grade?	3758	11.75	0.73	0.52	1.97	1.40
R ever skipped a grade?	3761	2.11	0.27	0.23	1.38	1.17
R ever in remedial classes?	1965	38.11	1.40	1.10	1.62	1.27
R have a computer at home?	2619	58.31	1.37	0.96	2.04	1.43
R have a quiet place to study?	2617	89.56	0.70	0.60	1.36	1.17
R's parent born in US?	3856	88.93	0.92	0.50	3.38	1.84
R's parent has worked at least 3 months since R born?	3853	95.02	0.57	0.35	2.65	1.63
DEFF/DEFT SUMMARY (PROPORTIONS)						
Mean					1.82	1.34
Standard Deviation					0.53	0.18
Minimum					1.36	1.17
Median					1.67	1.29
Maximum					3.39	1.84

**Table 5.42. Variance Estimates for NLSY97 Females: Means**

MEANS (CONTINUOUS VARIABLES)	SAMPLE SIZE	MEAN	DESIGN SE	SRS SE	DEFF	DEFT
Number of weeks worked in 1996?	2648	7.09	0.36	0.28	1.65	1.28
Days/wk have dinner with family?	2588	5.03	0.05	0.04	1.56	1.25
Days/wk have fun with family?	2587	2.63	0.04	0.04	1.00	1.00
Hours/week on homework?	2316	6.63	0.14	0.11	1.62	1.27
Hours/week watching TV?	2605	16.17	0.30	0.24	1.56	1.25
Hours/week reading for pleasure?	2602	3.58	0.11	0.10	1.21	1.10
Parent's income	2552	22309.21	574.34	375.22	2.34	1.53
Spouse's income	1794	38307.95	1403.80	959.07	2.14	1.46
Percent chance of R getting 4-year degree by age 30	1485	73.42	0.96	0.79	1.48	1.22
DEFF/DEFT SUMMARY (MEANS)						
Mean					1.62	1.26
Standard Deviation					0.41	0.16
Minimum					1.00	1.00
Median					1.56	1.25
Maximum					2.34	1.53

**Table 5.43. Variance Estimates for NLSY97 Hispanics: Proportions**

PROPORTIONS (BINARY VARIABLES)	SAMPLE SIZE	PERCENT	DESIGN SE	SRS SE	DEFF	DEFT
R enrolled?	1898	95.68	0.68	0.47	2.09	1.45
R in labor force?	1898	12.33	1.00	0.75	1.78	1.33
R employed?	1898	11.96	1.01	0.74	1.86	1.36
R working >= 10 hrs/week? (a)	1882	6.89	0.75	0.58	1.67	1.29
R working >= 10 hrs/week? (b)	1882	8.91	0.86	0.66	1.70	1.30
R employed during school year?	1898	6.87	0.73	0.58	1.58	1.26
R ever worked during school year?	1898	14.46	1.05	0.81	1.68	1.30
R employed during summer?	1898	7.95	0.87	0.62	1.97	1.40
R ever worked during summer?	1898	11.01	1.04	0.72	2.09	1.44
R ever retained a grade?	1509	17.72	1.33	0.98	1.84	1.36
R ever skipped a grade?	1512	1.77	0.41	0.34	1.45	1.20
R ever in remedial classes?	741	36.00	2.33	1.76	1.75	1.32
R have a computer at home?	1152	37.76	2.10	1.43	2.16	1.47
R have a quiet place to study?	1152	87.76	1.26	0.96	1.72	1.31
R's parent born in US?	1591	51.66	3.29	1.25	6.93	2.63
R's parent has worked at least 3 months since R born?	1586	86.82	1.50	0.85	3.11	1.76
<b>DEFF/DEFT SUMMARY (PROPORTIONS)</b>						
Mean					2.21	1.45
Standard Deviation					1.31	0.34
Minimum					1.45	1.20
Median					1.81	1.34
Maximum					6.93	2.63

**Table 5.44. Variance Estimates for NLSY97 Hispanics: Means**

MEANS (CONTINUOUS VARIABLES)	SAMPLE SIZE	MEAN	DESIGN SE	SRS SE	DEFF	DEFT
Number of weeks worked in 1996?	1140	4.79	0.49	0.34	2.08	1.44
Days/wk have dinner with family?	1149	5.01	0.10	0.07	2.04	1.43
Days/wk have fun with family?	1146	2.64	0.08	0.06	1.78	1.33
Hours/week on homework?	965	6.63	0.21	0.17	1.52	1.24
Hours/week watching TV?	1143	18.55	0.57	0.40	2.03	1.42
Hours/week reading for pleasure?	1143	2.83	0.18	0.14	1.65	1.28
Parent's income	936	18629.97	920.59	511.86	3.23	1.80
Spouse's income	693	25313.38	1093.97	750.26	2.13	1.46
Percent chance of R getting 4-year degree by age 30	573	66.01	1.54	1.34	1.32	1.15
DEFF/DEFT SUMMARY (MEANS)						
Mean					1.98	1.39
Standard Deviation					0.55	0.18
Minimum					1.32	1.15
Median					2.03	1.42
Maximum					3.23	1.80

**Table 5.45. Variance Estimates for NLSY97 Non-Hispanic Blacks: Proportions**

PROPORTIONS (BINARY VARIABLES)	SAMPLE SIZE	PERCENT	DESIGN SE	SRS SE	DEFF	DEFT
R enrolled?	2334	98.06	0.30	0.28	1.15	1.07
R in labor force?	2334	10.40	0.83	0.63	1.74	1.32
R employed?	2334	10.06	0.81	0.62	1.71	1.31
R working >= 10 hrs/week? (a)	2318	6.13	0.61	0.50	1.49	1.22
R working >= 10 hrs/week? (b)	2318	7.69	0.66	0.55	1.44	1.20
R employed during school year?	2334	5.05	0.61	0.45	1.84	1.36
R ever worked during school year?	2334	12.44	0.93	0.68	1.87	1.37
R employed during summer?	2334	7.04	0.61	0.53	1.32	1.15
R ever worked during summer?	2334	10.04	0.81	0.62	1.71	1.31
R ever retained a grade?	1997	27.22	1.90	1.00	3.61	1.90
R ever skipped a grade?	2005	2.85	0.49	0.37	1.75	1.32
R ever in remedial classes?	974	35.00	2.17	1.53	2.01	1.42
R have a computer at home?	1382	36.01	1.95	1.29	2.28	1.51
R have a quiet place to study?	1381	87.20	1.07	0.90	1.41	1.19
R's parent born in US?	2060	93.14	1.29	0.56	5.31	2.30
R's parent has worked at least 3 months since R born?	2057	91.13	1.27	0.63	4.06	2.02
<b>DEFF/DEFT SUMMARY (PROPORTIONS)</b>						
Mean					2.17	1.44
Standard Deviation					1.15	0.34
Minimum					1.15	1.07
Median					1.74	1.32
Maximum					5.31	2.30

**Table 5.46. Variance Estimates for NLSY97 Non-Hispanic Blacks: Means**

MEANS (CONTINUOUS VARIABLES)	SAMPLE SIZE	MEAN	DESIGN SE	SRS SE	DEFF	DEFT
Number of weeks worked in 1996?	1428	4.08	0.33	0.27	1.49	1.22
Days/wk have dinner with family?	1355	4.57	0.09	0.07	1.65	1.28
Days/wk have fun with family?	1361	2.79	0.08	0.06	1.78	1.33
Hours/week on homework?	1179	5.81	0.18	0.15	1.44	1.20
Hours/week watching TV?	1378	23.15	0.53	0.40	1.76	1.32
Hours/week reading for pleasure?	1379	2.92	0.14	0.13	1.16	1.08
Parent's income	1227	20082.21	868.83	438.59	3.92	1.98
Spouse's income	583	26160.00	1184.22	714.40	2.75	1.66
Percent chance of R getting 4-year degree by age 30	790	69.76	1.50	1.17	1.64	1.28
DEFF/DEFT SUMMARY (MEANS)						
Mean					1.96	1.37
Standard Deviation					0.86	0.28
Minimum					1.16	1.08
Median					1.65	1.28
Maximum					3.92	1.98

**Table 5.47. Variance Estimates for NLSY97 Non-Hispanic Nonblacks: Proportions**

PROPORTIONS (BINARY VARIABLES)	SAMPLE SIZE	PERCENT	DESIGN SE	SRS SE	DEFF	DEFT
R enrolled?	4731	97.64	0.26	0.22	1.40	1.18
R in labor force?	4732	19.90	0.66	0.58	1.29	1.14
R employed?	4732	19.60	0.65	0.58	1.26	1.12
R working >= 10 hrs/week? (a)	4617	10.22	0.48	0.44	1.19	1.09
R working >= 10 hrs/week? (b)	4617	13.46	0.54	0.50	1.17	1.08
R employed during school year?	4732	12.44	0.59	0.48	1.51	1.23
R ever worked during school year?	4732	22.34	0.72	0.60	1.44	1.20
R employed during summer?	4732	14.11	0.56	0.51	1.20	1.10
R ever worked during summer?	4732	18.20	0.64	0.56	1.31	1.14
R ever retained a grade?	4195	12.06	0.66	0.50	1.74	1.32
R ever skipped a grade?	4194	1.63	0.22	0.20	1.21	1.10
R ever in remedial classes?	2155	40.21	1.43	1.06	1.82	1.35
R have a computer at home?	2863	66.55	1.36	0.88	2.39	1.54
R have a quiet place to study?	2862	91.15	0.56	0.53	1.12	1.06
R's parent born in US?	4276	94.00	0.59	0.36	2.68	1.64
R's parent has worked at least 3 months since R born?	4275	97.19	0.46	0.25	3.38	1.84
DEFF/DEFT SUMMARY (PROPORTIONS)						
Mean					1.63	1.26
Standard Deviation					0.65	0.23
Minimum					1.12	1.06
Median					1.35	1.16
Maximum					3.39	1.84

**Table 5.48. Variance Estimates for NLSY97 Non-Hispanic Nonblacks: Means**

MEANS (CONTINUOUS VARIABLES)	SAMPLE SIZE	MEAN	DESIGN SE	SRS SE	DEFF	DEFT
Number of weeks worked in 1996?	2817	9.60	0.36	0.31	1.35	1.16
Days/wk have dinner with family?	2845	5.36	0.04	0.04	1.00	1.00
Days/wk have fun with family?	2842	2.65	0.04	0.04	1.00	1.00
Hours/week on homework?	2569	5.92	0.12	0.10	1.44	1.20
Hours/week watching TV?	2848	15.75	0.24	0.21	1.31	1.14
Hours/week reading for pleasure?	2847	2.93	0.11	0.09	1.49	1.22
Parent's income	3046	23463.63	518.66	358.46	2.09	1.45
Spouse's income	2516	40481.38	1153.45	788.36	2.14	1.46
Percent chance of R getting 4-year degree by age 30	1638	70.35	0.92	0.80	1.32	1.15
DEFF/DEFT SUMMARY (MEANS)						
Mean					1.46	1.20
Standard Deviation					0.41	0.16
Minimum					1.00	1.00
Median					1.35	1.16
Maximum					2.14	1.46

As shown in the tables above, the variables with the highest design effects throughout the domains are some of the parent variables, including parent's and spouse's income, whether the parent was born in the U.S., and whether the parent has worked for three or more months since the respondent was born. Furthermore, for a few variables, means and standard error estimates were based on small sample sizes. Variables which had sample sizes that were less than one-half of the total NLSY97 sample size (n=8984) were whether the respondent has taken remedial classes, spouse's income, and the percent chance, according to the parent, that the respondent will earn a four-year college degree by the age of 30.

Finally, Table 5.43 gives a summary of the mean, minimum, and maximum design effects and root design effects for the six NLSY97 domains. As is typical, the largest means belong to the overall category. Among the subgroups, the Hispanic and non-Hispanic black means are higher than those of the other races/ethnicities. Finally, the design effects for males and females are similar.

**Table 5.49. Comparison of Summary Statistics of NLSY97 DEFTs**

Group	Mean DEFF	Mean DEFT	Min DEFF	Min DEFT	Max DEFF	Max DEFT
Overall						
Binary	2.44	1.52	1.44	1.20	6.39	2.53
Continuous	2.03	1.41	1.00	1.00	2.87	1.69
Male						
Binary	1.82	1.32	1.23	1.11	4.49	2.12
Continuous	1.63	1.28	1.37	1.17	1.85	1.36
Female						
Binary	1.82	1.34	1.36	1.17	3.39	1.84
Continuous	1.62	1.26	1.00	1.00	2.34	1.53
Hispanic						
Binary	2.21	1.45	1.45	1.20	6.93	2.63
Continuous	1.98	1.39	1.32	1.15	3.23	1.80
Non-Hispanic Black						
Binary	2.17	1.44	1.15	1.07	5.31	2.30
Continuous	1.96	1.37	1.16	1.08	3.92	1.98
Other						
Binary	1.63	1.26	1.12	1.06	3.39	1.84
Continuous	1.46	1.20	1.00	1.00	2.14	1.46