Sample Statistics

This page intentionally left blank.

Sample Statistics

The results from surveys that use samples such as the Youth Risk Behavior Survey (YRBS) are affected by two types of errors: non-sampling error and sampling error. Nonsampling error usually is caused by procedural mistakes or random or purposeful student errors. For example, nonsampling errors would include surveying the wrong school or class, data entry errors, and poor question wording. While numerous measures were taken to minimize nonsampling error for the Youth Risk Behavior Survey (YRBS), nonsampling errors are impossible to avoid entirely and difficult to evaluate or measure statistically.

In contrast, sampling error can be described and evaluated. The sample of students selected for your YRBS is only one of many samples of the same size that could have been drawn from the population using the same design. Each sample would have yielded slightly different results had it actually been selected. This variation in results is called sampling error and it can be estimated using data from the sample that was chosen for your YRBS.

A measure of sampling error that is often reported is the "standard error" of a particular statistic (mean, percentage, etc.), which is the square root of the variance of the statistic across all possible samples of equal size and design. The standard error is used to calculate confidence intervals within which one can be reasonably sure the true value of the variable for the whole population falls. For example, for any given statistic, the value of that statistic as measured in 95 percent of all independently drawn samples of equal size and design would fall within a range calculated from the standard error of that statistic.

If simple random sampling had been used to select students for your YRBS, it would have been possible to use straightforward formulas for calculating sampling errors. However, your YRBS sample design uses two-stage cluster sampling; thus it is necessary to use more complex formulas. Software packages like SUDAAN can be used to compute sampling errors (Research Triangle Institute, 2001).

SUDAAN computes rates, means, or totals and their standard errors from the data collected in a complex multistage sample survey. The statistical approach used for computing the standard errors is a first-order Taylor Series linear approximation of the deviations of estimates from their expected values. For more details on the Taylor method, see Woodruff (1971). For more details on the formulas used in SUDAAN, see Research Triangle Institute, 2001.

In addition to the standard errors, SUDAAN estimates the design effect for each statistic, which is the standard error using the given sample design, divided by the standard error that would result if a simple random sample of the same size had been used. A design effect value of one indicates that the sample design is as efficient as a simple random sample; a value greater than one indicates a tendency for greater sampling error due to the use of a more complex and less statistically efficient design. Since the YRBS uses clusters of students in schools and classes to lower survey costs and reduce survey burden, it is not surprising that the design effect for estimates generated from YRBS data is greater than one.

The standard error and design effect for all the questions in your YRBS are presented in the following table. The table also includes the 95% confidence interval (CI) limits. The CI may be interpreted by using the following example: if the estimate of the percentage of students who rode with someone who had been drinking alcohol is 36.6% and the confidence interval is (34.4 – 38.7) then there is a 95% probability that the true value of the "percentage of students who rode with someone who was or had been drinking alcohol" lies between 34.4% and 38.7%.

REFERENCES:

1. Research Triangle Institute. SUDAAN,® version 9.0.1 [software and documentation]. Triangle Park, NC: Research Triangle Institute; 2005.

2. Woodruff, R.S. (1971). "Simple method for approximating variance of a complicated estimate." Journal of the American Statistical Association 66: 411-414.