

## GYTS Dataset Help File

The GYTS uses a two-stage sample cluster design. To reflect the complex sample design, there are two sample design variables on your data set named **STRATUM** and **PSU** (acronym for Primary Sampling Unit).

The variable **STRATUM** usually consists of two schools that are paired so that both schools have similar enrollment sizes.

However, sometimes a **STRATUM** may have only one school. For example, if a school has 100% chance of being in the selected school list (due to large enrollment) it will be the only school in that stratum, and we call this type of school a Certainty School.

In most cases, the Primary Sampling Unit represents a school. If the school is a Certainty School then the PSUs are the classes within the school.

The sampling weight variable is named **FINALWGT**.

Each student in the data set is assigned a sampling weight, which accounts for the following:

- Selection probability of the school
- Selection probability of the class
- Distribution of the population by grade and sex
- Non-responding schools
- Non-responding students
- Non-responding classes

Point estimates and 95% confidence intervals can be calculated using several software packages for statistical analysis of correlated data. Below are sample codes for EPIINFO, SUDAAN and STATA.

## EPIINFO Sample Code:

FREQ CR3 STRATAVAR = Stratum WEIGHTVAR=FinalWgt PSUVAR=PSU

ILL	Freq	%
+	20	35%
-	37	65%
Total	57	100%

Weight

P S U

Output to Table

Frequency of

Stratify by

☐ All (\*) Except

Clear Save Only Help OK

## SUDAAN Sample Code:

```
proc sort data= sasdata.dataset;  
by stratum psu;  
run;  
  
proc crosstab data = sasdata.dataset design = wr;  
nest stratum psu/missunit;  
weight finalwgt;  
tables cr3;  
run;
```