Tools of the Trade:

CONFIDENCE INTERVALS for a CRUDE RATE

Whenever the Bureau of Health Statistics computes and releases a birth, infant death or any other commonly used crude (non-adjusted) rate, we usually attempt to stress to the user that, where either the number of events or the size of the population is very small, rates are subject to fluctuation due to chance alone. Where this is the case, inferences cannot be drawn with any validity about the causes or effects of the vital events under study, e.g., Forest County with no infant deaths has the lowest infant death rate one year but, the next year, one infant death gives it the highest county infant death rate in the state. We find that inferences drawn from crude birth, death, infant, neonatal and fetal death rates, based on too few cases or too small a population, especially for small counties and municipalities, are one of the most common errors made by users of vital statistics. We know of one state which no longer computes and releases annual infant death rates below the state level, for this reason.

The variability of a rate can be estimated by computing a confidence interval, a very handy statistical tool that is not very difficult to understand and use. A confidence interval (CI) is composed of two figures or a range of numbers - an upper and lower limit - computed specifically for a given rate. That range then has a 95 percent chance of containing the "true" rate or a rate unaffected by chance events. For instance, the infant death rate for Butler county residents in 1984 was 8.5 per 1,000 live births or

(17 Infant deaths / 1,989 births) x 1,000 = 8.5

If a 95% confidence interval was computed for that rate, an upper limit rate of 12.6 and a lower limit rate of 4.5 would be the result. It can then be said that the changes are 19 in 20 that the 1984 infant death rate for Butler county (given the prevailing conditions and with chance variations removed) lies between 4.5 and 12.6 per 1,000 live births.

The larger the number of events and the population being studied are, the more likely is the chance that the computed rate is closer to a true rate. For example, Philadelphia's resident infant death rate in 1984 was 15.5 (388 deaths, 24,979 births). The 95% confidence interval for that rate is 14.0 to 17.1 deaths per 1,000 live births. The range of the confidence interval declines as the actual numbers being studied increase.

To construct a 95% confidence interval for a rate, the following formulae can be used:

Upper Limit = (1000 / n) (d + (1.96 x square root of d))

Lower Limit = (1000 / n) (d - (1.96 x square root of d))

Where:
d = number of events upon which the rate is based
n = denominator of the rate (area population for crude birth and death rates, live births for infant death rates)

The step-by-step procedures involved in calculating a 95% confidence interval for a rate are listed below. Here are some other important procedures to keep in mind when computing CIs. To compute a 99% confidence interval (less commonly used than the 95% CI), replace 1.96 with 2.58. As a rule, if the rate is greater than 100 events per 1,000, the square root of *d* is replaced with the square root of *(d(1 - (d / n)))*. Also, if the rate is expressed as of a population other than 1,000 (such as 10,000 for a maternal death rate or 100,000 for a cause-specific death rate), this population figure is then used in place of the 1,000 as listed above.

Step-by-Step Calculation of a 95% Confidence Interval for a Rate

In Clearfield County, in 1984, there were 21 resident deaths due to motor vehicle accidents. The 1984 estimated population of the county is 84,497. The 1984 death rate due to motor vehicle accidents is:

(21 / 84,497) x 100,000 = 24.9 deaths per 100,000 population

Below are the steps involved in calculating the 95% confidence interval for this rate. Remember, however, that this is a rate per 100,000 population.

1. The square root of 21 = 4.583

2. 4.583 x 1.96 = 8.983

3a. 21 + 8.983 = 29.983

3b. 21 - 8.983 = 12.017

4. 100,000/84,497 = 1.183

5a. 1.183 x 29.983 = 35.470 (upper limit)

5b. 1.183 x 12.017 = 14.216 (lower limit)

Confidence Interval = 14.2 to 35.5 at 95%

The confidence interval calculated on the previous page demonstrates that the "true" rate for Clearfield county resident deaths due to motor vehicle accidents in 1984 lies between 14.2 and 35.5 deaths per 100,000 population. This somewhat large range of 21.3 is directly due to the relatively small number of deaths upon which this rate is based. The smaller the number of events, the more unreliable will be a rate based on that number - the larger will be the confidence interval, also. When dealing with rates based on a small numbers of events (c. 20 or less), one should consider using multiple-year rates or enlarging the geographic area being reviewed to obtain a higher number and, subsequently, a more reliable rate (with a shorter confidence interval).

To compute a multiple-year rate, one simply sums the numbers of annual events for the years being studied (five and ten-year rates are the most common). Then, sum the populations for those years or, if only one year of population is available, multiply that figure by the number of years being studied. For example, suppose an area had 20 deaths in 1979, 35 in 1980 and 28 deaths in 1981. If the 1980 population was 10,000, the three-year summary crude death rate can be computed as shown below:

$$((20 + 35 + 28) / (3 \times 10{,}000)) \times 1000 = (83 / 30{,}000) \times 1{,}000 = 2.77$$

By the way, computation of confidence intervals for multiple-year rates is also recommended.

Using confidence intervals when analyzing and presenting rates demonstrates not only a perceptive knowledge of your subject matter but also aptly qualifies and guides the results of any study you may conduct.

IMPORTANT NOTE: Confidence intervals should not be calculated and used for crude rates based on less than 10 events. Rates based on such very small numbers are definitely not reliable.