Tools of the Trade:

COMPARING CRUDE RATES or RATIOS

Part II: Independent Rates

As was discussed in Part I of "Comparing Crude Rates or Ratios," there are different calculations for determining whether a significant difference exists between two crude (non-adjusted) rates or whether the difference between them is due to random effects. When the two crude rates are dependent (i.e., rates from overlapping time periods such as 1960 to 1970 and 1965 to 1975, or from geographical hierarchy such as rates for a county and the state), the formulas discussed in Part I of this series apply. However, when comparing two rates that are independent (i.e., they do not include any of the same data or events in their numerator, such as death rates for two different counties), formulas different from the ones presented in Part I apply and are discussed in this Part II of "Comparing Crude Rates or Ratios."

Before we begin, however, please remember that, as was pointed out in Part I, these statistical tools for analyzing rates are crude estimations and rather conservative approaches. A user may wish to utilize the more precise and sophisticated calculations performed by computer software such as SAS, SPSS or MINITAB. Also, a crude rate is easily computed and usually based on the number of vital events and the total population for a specific area or group, i.e., number of births or deaths among a specific population per 1,000 (or 100,000) of that specific population. A ratio is simply a proportion or percentage, usually a rate per 100.

As with dependent rates, the formula for determining significance between two independent crude rates or ratios will be different according to the number of events (deaths, etc.) upon which the rates are based - either a small number of events (10-99) or a larger number of events (100 or more). Each set of formulas is discussed separately in the following. To summarize their differences, these formulas for independent rates are constructed around their 95% confidence intervals and, depending upon the number of events, utilize either the ratio of the two rates or the difference between the two rates to determine significance.

Crude Rates or Ratios Based on 10-99 Events

As a general rule, if one of the independent rates to be compared is based on 10-99 events, you may use the following formulas to compute the 95% confidence interval directly for the difference between the two rates to determine whether a significant difference exists.

This measurement involves a three-step process. First, we must compute the difference (D) between the two rates. This is done with the following simple formula:

$$\mathbf{D = r_1 - r_2}$$

where:
$\mathbf{r_1}$ **= rate for County 1**
$\mathbf{r_2}$ **= rate for County 2**

The 95% confidence interval (CI) for the difference is then computed using the following formula:

CI = D ± square root of (CL_1^2 + CL_2^2)

where:
CL_1 = confidence limit for County 1 rate
CL_2 = confidence limit for County 2 rate

This computation becomes a three-step process due to the need to construct a confidence limit or CL (the numerical value that determines the range of the confidence interval, such as ±1.5) for each rate before you can use the above formula to determine the CI. The formula for a 95% confidence limit is:

CL = 1.96 x (r / square root of d)

where:
d = number of events

What you are actually doing with **(r / square root of d)** is computing an estimation of the standard error for the crude rate (see "Standard Error"). If this confuses you even more, forget we mentioned it or look into it later when you have the time.

At the end of this three-step process, if the confidence interval (CI) includes the value of 0, then it can be stated that the two rates are not significantly different, with 95% confidence. Of course, if the CI does not contain 0, then the difference between the rates is considered significant, with 95% confidence.

To demonstrate this computation, let's compare Dauphin County's 1987 infant death rate with Chester County's 1987 infant death rate and see if these two independent rates were significantly different. In 1987, Dauphin County's infant death rate was 13.1 per 1,000 live births, with 45 infant deaths (events) among 3,431 resident live births. Chester County's infant death rate that year was 6.3, with 34 infant deaths (events) among 5,394 resident five births. In our computations, Dauphin County will be County 1 and Chester County will be County 2. We would first compute the difference between the two rates with:

r_1 = 13.1
r_2 = 6.3

D = 13.1 - 6.3 = 6.8

The next step would be to compute the confidence limit (CL) for each of the two rates. The step-by-step process for each rate follows:

CL_1 = 1.96 x (r_1 / square root of d_1)
CL_2 = 1.96 x (r_2 / square root of d_2)

CL_1 = 1.96 x (13.1 / square root of 45) = 1.96 x 1.9528 = 3.83
CL_2 = 1.96 x (6.3 / square root of 34) = 1.96 x 1.0804 = 2.12

By inserting the two values for CL_1 and CL_2 into the formula for the confidence interval (CI), we would get the following results:

CI = 6.8 ± square root of ($3.83^2 + 2.12^2$) = 6.8 ± square root of 19.1633 = 6.8 ± 4.38 = 2.42 to 11.18

Since the confidence interval does not contain the value of 0, you can state, with 95% confidence, that the difference between Dauphin County's 1987 infant death rate and Chester County's rate is statistically significant.

Crude Rates or Ratios Based on 100 or More Events

If both independent rates to be compared are based on 100 or more events, a better and less complicated (only two-step) alternative for testing the difference between these two types of rates is to construct a 95% confidence interval for the ratio (instead of the difference) between the two rates.

The formula for calculating the ratio (R) between the two rates based on 100 or more events is:

$R = r_1 / r_2$

where:
r_1 = rate for County 1
r_2 = rate for County 2

(Please note that whenever only one of the two rates is based on 100 or more events, then that rate must be used as r_2.)

The formula for the 95% confidence interval for the ratio between the two independent rates is:

CI = R ± 1.96 x R x square root of ($(1 / d_1) + (1 / d_2)$)

where:
d_1 = # of events for County 1
d_2 = # of events for County 2

If the resulting confidence interval for the ratio contains the value of 1, then a significant difference does not exist at 95% confidence. If it does not contain the value of 1, then it can be said that the two county rates are significantly different with 95% confidence.

To demonstrate the computation of this tool, let's use five-year summary infant death rates for Dauphin and Chester counties to see if the relationship between these two rates differs from the previous single year comparison that was based on fewer events. During the period 1983-1987, Dauphin County had 189 resident infant deaths (events) among 16,829 live births for an infant death rate of 11.2, while Chester county experienced 232 infant deaths (events) among 24,741 five births for a rate of 9.4 per 1,000 live births. Again, in our computations, Dauphin County will be County 1 and Chester will be County 2. The ratio (R) between the two rates would be computed as follows:

r_1 = 11.2
r_2 = 9.4

R = 11.2 / 9.4 = 1.2

The next step would be to compute the 95% confidence interval for R:

d_1 = 189
d_2 = 232

CI = 1.2 ± [1.96 x 1.2 x square root of ((1 / 189) + (1 / 232))]
= 1.2 ± [1.96 x 1.2 x 0.0980]
= 1.2 ± 0.23
= 0.97 to 1.43

In this case, since the confidence interval contains the value of 1, the two rates are not considered significantly different, with 95% confidence. This is in contrast to the single year comparison of the infant death rates for these two counties. Chester County's during the same period.

Summary

The formulas and calculations presented in Parts I and II of "Comparing Crude Rates or Ratios" may seem complicated or elementary to you based on your experience with statistics. However, we only hope to present the thought to our readers that the rates they have been using all these years and making assumptions with are not as reliable as they may seem. A rate is only an estimation of a given population's specific experience and is not an absolute measurement. Therefore, rates, especially crude rates, should not be used to make important decisions without some understanding and test of their reliability.

For more detailed information on comparing rates, see *Epidemiology in Health Services Management* by G. E. Alan Dever, Aspen Systems Corporation, Rockville, Maryland, 1984.