Tools of the Trade:

STANDARD ERROR: THE STATISTICIAN'S HEDGE

As with any measurement of a population, mortality or incidence rates have a certain amount of chance variation associated with them and, therefore, we cannot be 100% certain that any difference in the observed rates represents a statistically significant difference in the actual or real rates of any two or more population groups. However, although 100% certainty of an observed rate can never be realized, a level of certainty can be computed and used when examining the difference between these rates. To accomplish this, a 90% or 95% "confidence interval" (CI) is usually determined for a rate. In other words, a range is computed for the observed rate which should include the actual or real rate 90 or 95 percent of the time. To determine this range or confidence interval, the standard error of the rate must first be computed. A standard error is a measure of the variability of the observed rate. With this value, a 95% confidence interval can be computed for an age-adjusted rate using the following formula:

95% Confidence Interval = Age-Adjusted Rate ± (1.96 x Standard Error)

A 90% confidence interval can also be determined by substituting 1.645 for the 1.96 in the formula listed above.

Let us look at some of the actual data from this report and see how this works. Please refer to the data table on the following page. The 1983 age-adjusted male cancer incidence rate for the south central area was 366.29 (per 100,000) with a S.E. of 6.21. By applying the 95% formula above, one can be 95% certain that the actual rate is between 354.12 and 378.46. The female rate is listed as 293.16 with a S.E. of 5.11. This means that we can be 95% certain that the actual rate for females is between 283.14 and 303.18. Since the two ranges, 354.12 to 378.46 and 283.14 to 303.18, do not overlap, we can say there is a statistical significance between the rates of cancer found in this area for men and for women.

If you scan over the data contained in the table, you will notice that the S.E.s for the rural counties are much higher. To maintain the level of 95% confidence in the rates for these less populated areas, the range associated with the actual rate must be expanded. A small population has this effect because the overall rate is greatly affected by adding or subtracting just one or more cases of cancer. A few cases added to a densely populated county's rate will not have nearly the affect on its rate as compared to adding a few cases to a sparsely populated county's rate. The consideration here is that the additional case(s) may have occurred by chance in one year rather than another, thus affecting the rate. This is why the S.E.s have such an importance when analyzing the data contained in the cancer incidence report, i.e., the data are for one year, and in many cases, for counties that have rather low numbers of actual cases.

Careful use and understanding of standard errors and confidence intervals can assist even those unfamiliar with "statistics" to interpret rates - those imprecise measurements of a population that should never be taken for granted.

NUMBER OF CANCER CASES AND AGE-ADJUSTED (1970 U.S. STANDARD) INCIDENCE RATES PER 100,000 POPULATION BY COUNTY and SEX, SOUTH CENTRAL PENNSYLVANIA, 1983

County	Number	Males Rate	S.E.	Number	Females Rate	S.E.
TOTAL AREA	3,226	366.29	6.21	3,356	293.16	5.11
Adams	126	366.83	31.49	116	279.77	26.39
Bedford	55	206.65	27.39	53	161.46	22.33
Blair	270	368.62	21.70	333	329.13	18.54
Cumberland	314	365.84	19.81	341	297.42	16.22
Dauphin	490	413.53	17.94	505	309.65	13.81
Franklin	177	295.43	21.70	191	249.28	18.21
Fulton	8	116.13	39.43	16	195.51	48.71
Huntingdon	77	355.61	38.90	76	281.87	32.50
Juniata	24	232.06	45.11	34	242.39	42.55
Lancaster	677	373.99	13.80	726	308.60	11.64
Lebanon	236	399.74	24.99	218	289.72	19.84
Mifflin	92	365.17	36.93	96	285.57	29.48
Perry	73	420.64	46.80	63	298.48	37.20
York	607	378.53	14.73	588	293.18	12.13