StatNews #03

Spurious Results with Ratios
June 1996

Ratios are used in many fields to adjust or normalize one measure for another in order to make comparisons or rankings. In economics, national indices for wealth are formed by the ratio of wealth to population size, examples being per capita income and gross national product per capita. In nutrition, the weight of people relative to their frame size is captured as body mass index (weight in kg / height in meters squared).

Regression analysis is the standard way to adjust one measure for another. Any observations falling on the regression line are thought of as being equal relative to the covariate. Observations above the line are relatively large and ones below the line are relatively small. In regression analysis, a constant is included to estimate the value of the response variable when the covariate is at zero. A ratio is a special case of regression, equivalent to fitting without a constant or forcing the constant to be zero.

In human biology, physiology, and nutrition, there is increasing awareness that the use of such ratios can lead to spurious results. A workshop at the April 1996 Experimental Biology meetings was devoted to these concerns, which stem from the implicit assumptions that the relationship between the numerator and denominator of a ratio is a straight line with an intercept of zero. Recent studies have demonstrated that often these linear and zero-intercept assumptions are not met, with the consequence that proper adjustment for the denominator measure is not made.

A simple illustration of the consequences of violating the assumption of zero intercept comes from data collected on hand shape by a 7^{th} grade class this Spring. Hand length and width were measured on ten children, with the objective of studying length adjusted for width. A ratio index of shape was formed from hand length divided by width, with the idea that long, thin hands will have a high value of this ratio.

Alternatively, simple linear regression was used to find the best fitting line relating length to width. Then, a residual index of shape was formed as the distance of each observed length from the length predicted by the line.

When the shapes of the ten hands were ranked (from 1 to 10) from least to most thin using the residual index, the corresponding ranks using the ratio index were 1, 3, 5, 6, 4, 2, 9, 8, 10, and 7. This was not a very close correspondence, given, for example, that the 2nd thinnest hand on the residual index was the 6th thinnest on the ratio index. Further evidence of the lack of correspondence was that the correlation coefficient between the residual and ratio indices was only 0.8. The reason for the discrepancy was that, although length and width were linearly related, the estimated intercept was 10.175 cm, not zero.

A second concern about the use of ratios is that, even if the linear and zero-intercept assumptions are reasonably met for certain ratios, often analysts will statistically relate two ratios that share the same denominator. Recent studies have demonstrated that two ratios can appear to be related even when the numerator measures are clearly completely independent.

The use of ratios as response variables in regression should be avoided if possible in favor of adjusting for the denominator measure by including it as a covariate in the regression. If ratios are used, one simple way to mitigate these concerns and to ensure that complete adjustment has been made is to include the denominator of the ratios used as a covariate.

Two references are appended.

Allison-D-B. Paultre-F. Goran-M-I. Poehlman-E-T. Heymsfield-S-B. Statistical considerations regarding the use of ratios to adjust data. International Journal of Obesity 19(9). 1995. 644-652.

Goran-M-I. Allison-D-B. Poehlman-E-T. Issues relating to normalization of body fat content in men and women. International Journal of Obesity. 19(9). 1995. 638-643.

Author: Edward A. Frongillo, Jr.