Reprinted from IFPress (2006) 9(2)2,3

Making Ratio Statements about Product Improvements

Daniel M. Ennis

Background: When making product comparisons, it is sometimes useful to be able to substantiate ratio statements. The purpose of this report is to consider how this might be accomplished using standard rating and difference testing methods.

In everyday experience, such as finding out what the time or temperature is, we often use interval scales. When we say that it is 8:30am, we mean that an interval of eight hours and thirty minutes has elapsed since midnight, which is an arbitrarily agreed-upon point to begin recording time for the new day. This is our zero point on an interval scale on which the difference between each number has the same meaning. The time elapsed between 8am and 7am is the same as that between 7am and 6am. We use the same ideas when we use Centigrade (Celsius) and Fahrenheit scales to record temperature. If we take ratios of the numbers on these scales, the results are not meaningful. For instance, although 100°C/20°C is a ratio of 5, the corresponding ratio for the same temperatures in Fahrenheit is 212°F/68°F which is not 5, it is 3.1. The same difficulty arises with time as we usually measure it. We could measure temperature and time on absolute scales for which there is an absolute zero point, such as the Kelvin scale, but for time we would have to measure time from the beginning of the universe. If this is really the case, how can we make ratio statements such as: "It took me twice as long to get to work today as it did yesterday"? The answer is that, although interval scales do not have ratio properties, distances between points on them do. For interval scales linearly related to each other, the ratios are the same for equivalent points on the scales. Using the Centigrade scale |20 - 0 |/|100 - 0| = 1/5; and using the same temperatures in Fahrenheit |68 - 32|/|212 - 32| = 1/5. Similarly, the statement about getting to work is a ratio of distances on the interval time scale and we don't need to know how old the universe is to make it.

Scenario: You are interested in estimating the ratio of perceived malodor reduction achieved by your product to that of a competitor. Both products are instant action air fresheners and there is an industrial standard malodor available on which to make the comparisons. Malodor evaluations are made by a representative sample of air care users in test chambers designed for odor evaluations. Four equivalent groups of 100 consumers each make the evaluations. In one group, consumers evaluate a chamber containing malodor only on a 7-point rating scale and also a chamber containing malodor and your air freshener. The 7-point scale is a word-anchored scale with "1" labeled "no malodor present" and "7" labeled "extreme malodor present". A second group evaluates a chamber containing malodor only and a chamber containing malodor and your competitor's air freshener. In the remaining two groups, consumers compare the two chambers ('malodor + product' and 'malodor only') with the instruction to choose the room with the most malodor present. One of these groups evaluates your product. The other group evaluates the competitor's product. This method is called the 2-alternative forced choice method (2-AFC). Data from the four groups is presented in Tables 1 and 2. All four groups are considered to be equivalent in this scenario.

Interval Scale Estimates: Rating methods, such as the one just described, cannot be assumed to provide interval scale information. One of the basic problems is that we cannot assume that differences between successive rating categories are equal. This problem was described in earlier technical reports and a method for extracting interval scale information from ratings methods was described[1,2]. It is also possible to obtain interval scale information from 2-AFC data[3,4,5]. In both cases, using a Thurstonian model, we assume that there is an intensity continuum on an interval scale and that each product's perceived intensities can be represented on this scale as random values from a normal distribution. One product is assumed to be arbitrarily placed at zero and the other at a value called δ. The difference between the means of the two products is δ and the distance between them on the interval scale is $|\delta|$. An estimate of δ, called *d'*, and its variance can be obtained by fitting the models to data using the method of maximum likelihood. Model fitting can be done using *IFPrograms*™. Very much like the example about how much longer it took to get to work, we can construct ratio statements about relative malodor reduction because we have estimated two distances on an interval scale.

Making Ratio Statements: Table 3 contains the results of modeling the data in Tables 1 and 2 to obtain *d'* values. The rating and 2-AFC results provide similar estimates of

Group	Condition	1	2	3	4	5	6	7
1	Malodor	2	9	23	16	19	15	16
	Malodor + Air Freshener A	31	31	25	7	4	2	1
2	Malodor	2	7	25	16	21	18	12
	Malodor + Air Freshener B	7	15	33	15	16	10	4

Table 1. Results of rating experiments for two groups of consumers. A represents your product; B is your competitor's product.

Group	Condition	Choice
3	Malodor	85
	Malodor + Air Freshener A	15
4	Malodor	65
	Malodor + Air Freshener B	35

Table 2. Results of 2-AFC experiments for two groups of consumers

the difference between the malodor control and treatment conditions, although the variances of the estimates from ratings are lower than the 2-AFC estimates. The ratio of the relative malodor reduction for your product is 1.49/ 0.56 = 2.7 times that of your competitor based on the ratings. After rounding, a similar ratio of 2.7 is obtained based on the 2-AFC results. However, there is more confidence in the rating experiment because the variances of the *d*'s are smaller and this will affect the confidence in the ratios.

The Distribution of Ratios: The ratios for the rating and 2-AFC experiments are subject to error, so it would be useful to know the ratio corresponding to the lower 5% point of the distribution of estimated ratios. This distribution has been derived and Figures 1a and 1b display the relative likelihoods of possible ratios from the results shown in Table 3. It can be seen that based on the rating experiments, you may claim a ratio of 1.77 with 95% confidence. Based on the 2-AFC results you should claim a lower ratio of 1.61 with 95% confidence. It would not be appropriate to simply claim that your product reduces malodor 2.7 times better than your competitor, as estimated from the results in Tables 1 and 2. There is a reasonable possibility (based on 95% confidence) that the ratio may be as low as 1.77 based on the rating experiment or even as low as 1.61 based on the 2-AFC results. In order to obtain a more precise estimate of the ratio, it would be necessary to increase the sample size and recalculate the distribution of ratios.

Conclusion: Interval scales of time or temperature are commonly used to make day-to-day decisions about the fastest way to get to work or how warmly to dress. We commonly make ratio statements amount time intervals. Similarly, claims about ratios on interval scales of perceived intensity can be made provided that there is a reliable way of estimating these intensities. Thurstonian models provide a valuable first step in obtaining these estimates and their variances. In making ratio claims, it is important to realize that experimental estimates of ratios are subject to error. By determining the distribution of estimated ratios from an experiment, as shown in Figures 1a and 1b, it is possible to provide realistic estimates of the ratio of interest and avoid exaggerating the effect observed. These analyses will not only provide you with more confidence in your ratio statements, but provide a fair basis for comparing the performance of your product with that of your competitor's product.

	Ratings		2AFC	
Condition	*d'* *	variance	*d'* *	variance
Malodor + Air Freshener A	1.49	0.027	1.47	0.047
Malodor + Air Freshener B	0.56	0.022	0.54	0.033

Table 3. Results of modeling the data in Tables 1 and 2 to estimate *d'* values for air freshener products relative to a malodor control arbitrarily set at zero

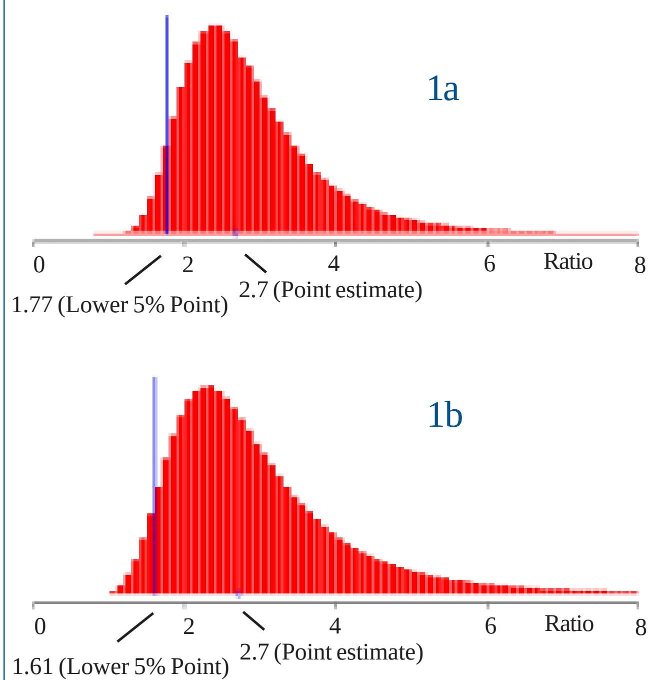

Figure 1. Distributions of ratios of malodor improvement for rating (1a) and 2-AFC experiments (1b). In both cases, a point estimate of the ratio is 2.7. For ratings the lower 5% point for the ratio is 1.77 and is 1.61 for the 2-AFC

References (see www.ifpress.com)

1. Ennis, D.M. (1999). Thurstonian models for intensity ratings. *IFPress*, **2**(3), 2-3.
2. Dorfman, D.D. and Alf, E., Jr. (1969). Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals - Rating method data. *Journal of Mathematical Psychology*, **6**, 487-496.
3. Ennis, D.M. (1998). Thurstonian scaling for difference tests. *IFPress*, **1**(3), 2-3.
4. Ennis, D.M. (1993). The power of sensory discrimination methods. *Journal of Sensory Studies*, **8**, 353-370.
5. Thurstone, L.L. (1927). A law of comparative judgement. *Psychological Review*, **34**, 368-389.