Developments in Business Simulation & Experiential Learning, Volume 26, 1999

THE NEED TO MEASURE VARIANCE IN EXPERIENTIAL LEARNING AND A NEW STATISTIC TO DO SO

John R. Dickinson, University of Windsor
James W. Gentry, University of Nebraska-Lincoln

ABSTRACT

The measurement of variation of qualitative or categorical variables serves a variety of essential purposes in simulation gaming: describing/summarizing participants' decisions for evaluation and feedback, describing/summarizing simulation results toward the same ends, and explaining variation for basic research purposes. Extant measures of dispersion are not well known. Too, simulation gaming situations are peculiar in that the total number of observations available--usually the number of decision periods or the number of participants--is often small. Each extant measure of dispersion has its own limitations. Common limitations are that, under ordinary circumstances and especially with small numbers of observations, a measure may not be defined and the maximum theoretical value may not be attainable. This paper reviews existing measures of dispersion and describes a new measure which is always defined and for which a maximum value is always attainable.

CONSIDERATION OF VARIANCE

Roles of Variance in Experiential Learning

Measurement in the domain of experiential learning does not have a rich history (Gentry *et al.* 1998). Recent Association for Business Simulation and Experiential Learning (ABSEL) conferences, for example, have seen considerable discussion about the measurement of learning, without much consensus being reached in terms of the types of criterion variables to be measured. To the extent that educators have even attempted to measure learning, analyses have generally been limited to the investigation of mean differences of outcome variables, despite frequent admonitions that greater concern should be given to the learning process rather than performance outcomes.

We contend that insight into process can be gained via examination of within-student and across-student variance in decision-making. For example, if the role of an experiential exercise is to foster trial-and-error learning, the variance in a student's decisions will be greatest when the trial-and-error stage is at its peak. To the extent that the exercise works as intended, greater variances would be expected early in the game play as opposed to later, barring the game world encountering a discontinuity.

On the other hand, it is possible that decisions may be more consistent early and then more variable later if the student is trying to recover from early missteps. Similar to what Ross (1991) found with salespeople who are in danger of not making quota, students desperate to finish in the black may make radical decisions (conceivably outside company policy). A measure of variance could be used in the game's diagnostic software to signal the instructor that the nature of decision-making has changed dramatically.

Another central issue in the administration of simulation games is comparability across worlds or industries. Due to a number of factors, the worlds/industries may not be directly comparable:

- Defining parameters may be varied by the instructor to make the simulated environment different across worlds/industries.

- Worlds/industries may comprise different numbers of competing companies.

- The dynamics of competition may vary, with some worlds/industries collectively employing more aggressive or conservative strategies or the mix of aggressive-conservative competitors varying across worlds/industries.

- Companies may “go missing” during the competition differentially across worlds/industries. Is it fair to reward more highly the second-place finisher who makes large profits in a world where several competitors have defaults than the highest performer in a more conservative, more competitive world, even though that highest performer has generated less profit?

Measures of variation across a number of input and output variables could help make company performances more comparable for evaluation purposes.

Types of Input and Output Variables

Strategic decisions in business simulation games are predominantly quantitative. Many comprise dollar amounts, either in the form of expenditures or prices. Other “amounts” include numbers of employees, numbers of distributors, units of products manufactured or ordered, and so on. Many output variables are also quantitative, including basic sales, profit, and other income statement results, inventory levels, perceptual map distances, indices of morale or product quality, and so on.

For purposes of monitoring participant performance, providing feedback to participants, and the conducting of basic research, it is often necessary to summarize the distributions of both input and output variables. For quantitative variables, appropriate descriptive statistics are well known and are usually classified into measures of central location (e.g., mean, median, mode) and measures of dispersion (e.g., range, percentiles, variance, standard deviation, coefficient of variation) (cf. Keller and Warrack 1997, Chapter 4).

Other variables related to simulation games are qualitative or categorical. That is, they are decisions or results of kind or type rather than degree (Dickinson 1995). Examples include decisions relating to advertising messages, sales promotion types, product mix, types of research and development, types of marketing research studies, menus of ethical alternatives, and so on. More specifically, for instance, *Micromatic* (Scott *et al.* 1992) makes available a list of nine types of competitive information participants may purchase. *The Marketing Game!* (Mason and Perreault 1987) features a menu of five advertising types.

For qualitative variables, too, it is often useful to summarize their distributions. Percentage distributions, either univariate or in some crosstabular form, are the most common type of descriptive statistic for qualitative variables. Otherwise, description of such variables most often takes the literal form of bar and pie charts (cf. Anderson, Sweeney, and Williams 1996, Chapter 2; Siegel and Morgan 1996, pp. 23-26).

Measures of dispersion of qualitative variables, although they do exist, are not at all well known. None of eight contemporary business statistics textbooks (list available on request from the senior author) presents any such measures.[1] None of the SPSS DESCRIPTIVES, FREQUENCIES, or NPAR TESTS procedures includes any such measures (SPSS 1997, pp. 256, 382, 569).

[1] An exception is the variance of a binomial random variable.

Following is a brief review of existing measures of dispersion for qualitative variables. A new statistic which rectifies two common limitations of extant measures, and is particularly suited for application to simulation games, is then described.

THE CONCEPT OF VARIANCE OF A QUALITATIVE VARIABLE

The intuitive notion of variance or dispersion of a qualitative variable is easily illustrated. A population comprising all women and no men would have zero gender dispersion. A population comprising 80% women would be more concentrated in gender than a population comprising 60% women. A population comprising 50% women and 50% men would exhibit maximal dispersion. A population of 80% women would be of equal gender dispersion as a population comprising 80% men.

A simulation game participant selecting an informative advertising message in each of, say, 10 decision periods, would exhibit less variability than a participant selecting an informative message in each of five periods, a persuasive message in each of three periods, and a reminder message in each of two periods.

Finally, a simulation game providing a menu of, say, eight types of sales promotions has a greater *potential* variability than does a game providing a menu of five types of sales promotions. This parameter of measures of dispersion reflects the total number of categories comprising the variable, not the number of categories containing nonzero observed frequencies. "...the variation in a nominal variable depends in part on the number of its categories: the greater the number of classes, the greater the variation, other things being equal." (Reynolds 1984, p. 11)

Measures of dispersion for qualitative variables aim to summarize and quantify in a single number the above notions of variability.

EXISTING MEASURES OF DISPERSION

At least 10 measures of dispersion of qualitative variables have been developed, beginning as early as 1912. The most familiar of these are presented below, accompanied by a brief description of their characteristics. Let J indicate the number of categories of the variable, n indicate the total number of observations, n_j indicate the number of observations in category j (j=1 to J), and p_j indicate the proportion of observations in category j ($p_j=n_j/n$, j=1 to J).

Using this notation coupled with the concept of dispersion as described above, minimum dispersion exists when $n_j=n$ for some single j and $n_i=0$ for all $i \neq j$. Maximum dispersion exists when $n_1=n_2=...=n_J=n/J$ (i.e., $p_1=p_2=...=p_J$).

Index of Diversity (Gini's Concentration Measure) and Index of Qualitative Variation

Apparently the first measure of qualitative dispersion was a measure of concentration proposed by Corrado Gini in 1912 (Agresti and Agresti 1977, p. 206).

$$D = C = 1 - \sum_{j}^{J} \left(n_j / n\right)^2 = 1 - \sum_{j}^{J} p_j^2 \qquad (1)$$

Gini's measure currently is incorporated into the Logit Loglinear Analysis procedure of SPSS (Norusis 1994, p. 216).

The interpretation of D or C is straightforward: "The index of diversity gives the probability that a pair of randomly selected observations will be in different categories." (Reynolds 1984, p. 61) D or C ranges between 0 (when all observations fall within a single category) and (J-1)/J, i.e., $0 \leq D,C \leq (J-1)/J$. The maximum value of D or C is, obviously, variable, depending on the number of categories. Too, D or C cannot attain its theoretical maximum value when the number of observations, n, is not an integer multiple of the number of categories, J. That is, when n is not an integer multiple of J, it is not possible for all n_j to be equal.

Under some circumstances, namely, when comparing variation across variables comprising different numbers of categories, it is desirable for the theoretical range of the measure of dispersion to be fixed, specifically with a minimum value of zero and a maximum value of one. The Index of Qualitative Variation (IQV) transforms the Index of Diversity to this range.

$$IQV = \frac{J}{(J-1)} D = \frac{J}{(J-1)} \left(1 - \sum_{j}^{J} p_j^2 \right) \qquad (2)$$

As Reynolds (1984, p. 62) states, "The index of qualitative variation, which is a 'standardized' version of D because it takes into account the number of categories, has the same interpretation (it is the probability that a randomly selected pair of observations will be in different categories) except that its maximum possible value is 1.0."

Shannon's Entropy Measure (H) and Index of Relative Uncertainty (H')

Shannon's Entropy Measure (H), which is also called the Information Index (Agresti and Agresti 1977, p. 208), is defined as follows:

$$H = -\sum_{j}^{J} p_j \log(p_j) \qquad (3)$$

As with Gini's Concentration Measure, Shannon's Entropy Measure is incorporated into the Logit Loglinear Analysis procedure of SPSS (Norusis 1994, p. 215). And, as with the Index of Diversity/Gini's Concentration Measure, there exists a standardized version of the measure with a range of 0 to 1. The Index of Relative Uncertainty (H') is defined as:

$$H^1 = \frac{-\sum_{j}^{J} p_j \log(p_j)}{\log J} \qquad (4)$$

H and H' have the rather serious limitation that they are not defined where n_j=0 for any j. That is, where one or more categories of the variable contains zero observations, p_j=0 for those categories, log (p_j) is undefined, and H and H' may not be used as measures of dispersion. While the occurrence of n_j=0 may be unusual in studies of populations and ecology where the measures are commonly used (Agresti and Agresti 1977, p. 208), they are likely to occur in the context of simulation games where the total number of observations is often small.

Deviation from the Mode (DM)

The mode, of course, is the most common (and perhaps the only valid) measure of central location of a categorical variable. The more pronounced or extreme the modal frequency compared with other category frequencies, the less is the dispersion of the total distribution. The Deviation from the Mode is defined as follows (Reynolds 1977, p. 30):

$$DM = 1 - \frac{\sum_{j}^{J} (n_{mode} - n_j)}{n(J-1)} \qquad (5)$$

where n_{mode} is the number of observations in the modal category. While the theoretical range of DM is from 0 to 1, as with most other measures of dispersion, the theoretical maximum of 1 is only attainable when n is an integer multiple of J.

Poisson Index of Dispersion (χ^2)

Finally, Plackett (1974, p. 9) describes the Poisson Index of Dispersion (PID):

$$\chi^2 = \frac{\sum_{j}^{J} (n_j - M)^2}{M} = \frac{J \sum_{j}^{J} n_j^2}{n} - n \qquad (6)$$

where M is the mean frequency count across the J categories. χ^2 is a special case of the Pearson chi-squared statistic and, in addition to serving as a descriptive measure of dispersion, may be used for inference to test a hypothesis of homogeneity (Plackett 1977, pp. 8-10):

H_0: $p_1=p_2=...=p_J=1/J$.

The Poisson Index is an inverse measure of dispersion; it theoretically equals 0 when H_0 is true. However, as with the other measures of dispersion, this value is not attainable when n is not an integer multiple of J.

Note that the formula for the PID is not the same as the formula for the common variance of the J frequency counts. The latter would have as its denominator either J or J-1.

Other Measures of Dispersion

There exist additional measures of dispersion. These include the variance or standard deviation of the J frequency counts, Weiler's Index of Discrepancy, and the ratio of actual to maximum variance of the J frequencies. The last two are directly related to the Index of Diversity/Gini's Concentration Measure and the Index of Qualitative Variation (Agresti and Agresti 1977, pp. 209,210). The common variance and standard deviation equal zero when all n_j are equal, but both have a variable upper limit depending on both n and J.

LIMITATIONS OF EXISTING MEASURES

Theoretical Maximum Dispersion Limit Not Generally Attainable

As discussed above, a conceptual maximum dispersion distribution is where an equal number of observations fall into each category of a variable. Under this condition, several extant measures of dispersion have defined theoretical values:

Index of Diversity	(J-1)/J
Index of Qualitative Variation	1
Index of Relative Uncertainty	1
Deviation from the Mode	1
Poisson Index of Dispersion	0

However, the above conceptual definition cannot be realized where n is not an integer multiple of J. Though obviously n may some times be an integer multiple of J, the more general situation is where it is not. Under this more general condition, it is not possible for there to be an equal number of observations in each category and *none of the values listed above is attainable*.

To assess the extent of difference between the theoretical and attainable values, measures were calculated for a variety of combinations of numbers of variable categories (2 through 10 inclusive) and sample sizes. Actual sample sizes varied between (J+1) and 10, 20, 50, 100, and 500, with observations being distributed as nearly evenly as possible among the J categories.

(Sample sizes of n=J were not investigated since, in that case, n is an integer multiple of J and sample sizes of n<J were not investigated for the reason described below.)

From Table 1, for example, for a dichotomous variable, when n=3, the Index of Qualitative Variation can only attain a value of .89, compared with its maximum theoretical value of 1. The Deviation from the Mode can only attain a value of .67, compared with its maximum theoretical value of 1. The Poisson Index of Dispersion can only attain a value of .11, compared with its theoretical maximum dispersion value of 0. For J=6 and n=9, the respective deviations between the attainable and theoretical values are .022, .067, and 1.0 (not in Table 1).

For those measures which vary directly with dispersion, the deviation between the theoretical and attainable values is greatest when the total sample size, n, is only marginally greater than the number of variable categories, J. The Poisson Index of Dispersion varies inversely with dispersion and its deviation is greatest for large numbers of observations.

TABLE 1
DEVIATIONS BETWEEN THEORETICAL AND ATTAINABLE MEASURE VALUES
(Numbers of observations)

	J=2	J=4	J=6	J=10
Index of Diversity	.056 (3)[a]	.030 (5)	.021 (8)	.021 (13)
Index of Qualitative Variation	.111 (3)[a]	.040 (5)	.025 (8)	.014 (13)
Deviation from the Mode	.333 (3)[a]	.200 (5)	.143 (7)	.091 (11)
Index of Relative Uncertainty	.082 (3)[a]	.041 (6)	.033 (8)	.026 (14)
Poisson Index of Dispersion	.111 (9)[b]	.333 (9)	.714 (7)	N.A.
	.053 (19)	.158 (19)	.263 (19)	.474 (19)
	.020 (49)	.061 (49)	.102 (49)	.184 (49)
	.010 (99)	.030 (99)	.052 (97)	.091 (99)
	.002 (499)	.006 (499)	.010 (499)	.018 (499)

a The greatest deviations for these measures occur with small numbers of observations and decrease as the number of observations increases. Figures in parentheses are the n for which the deviation is maximum.

b The Poisson Index of Dispersion is an inverse measure of dispersion and the greatest deviations occur with small numbers of observations and decrease as the number of observations decreases. Figures in parentheses are the n for which the deviation is maximum, starting with total observations of 10, 20, 50, 100, and 500, respectively.

Inability to Accommodate Frequencies of Zero

Shannon's Entropy Measure and its standardized version, the Index of Relative Uncertainty, require the taking of logarithms. However, the log of zero is undefined. Thus, where a n_j=0 for any j, a situation more likely to occur with the small total sample sizes typical of many simulation applications, these two measures cannot be calculated.

THE ANGSTA MEASURE OF DISPERSION

The angsta statistic rectifies the two limitations of several other measures of dispersion described above. Where n is not an integer multiple of J then it is not possible for n_j to be equal to n/J for each category. This is the condition in which several of the measures are not able to attain their theoretical maxima. Maximum dispersion in this case is realized when each of the "remainder" observations falls into a different category. Let "n\J" indicate the quotient of the division n/J truncated to an integer and "n MOD J" indicate the remainder of the division. Then maximum dispersion is realized when each of J-(n MOD J) categories contains n\J observations and each of (n MOD J) categories contains (n\J)+1 observations.

For example, for 18 observations on a variable comprising 5 categories, maximum dispersion is realized when two categories (=J-[n MOD J]) contain 3 observations (=n\J) and three categories (=n MOD J) contain 4 observations (=[n\J]+1). Minimum dispersion, of course, is realized when all n observations fall into a single category.

For a given set of data, under the condition of maximum dispersion of n observations among J categories, a maximum value is defined as the sum of the squared frequencies:

$$MAXDISP = (J-[n \text{ MOD } J]) (n\backslash J)^2 + (n \text{ MOD } J) ([n\backslash J]+1)^2$$

Under the condition of minimum dispersion, a minimum value is defined as the total number of observations squared:

$$\text{MINDISP} = n^2$$

Note that the actual numerical value of MAXDISP is *less than* the value of MINDISP.

For a given set of observations, then, a theoretical range equals MINDISP-MAXDISP.
Angsta is defined as:

$$Angsta = 1 - \frac{\sum_j^J n_j^2 - MAXDISP}{MINDISP - MAXDISP}$$

$$= 1 - \frac{\sum_j^J n_j^2 - MAXDISP}{n^2 - MAXDISP} \qquad (7)$$

When all n observations fall into a single category then:

$$n^2 = \sum_j^J n_j^2 = MINDISP \qquad (8)$$

and angsta=0.

When n is an even multiple of J and each category contains n\J= n/J observations, then:

$$\sum_j^J n_j^2 = MAXDISP \qquad (9)$$

and angsta=1.

When n is not an even multiple of J but the J categories contain as nearly equal n_j as possible, then angsta also equals 1.

Angsta is undefined when n=1.
In sum, two desirable properties of angsta are (1) $0 \le \text{angsta} \le 1$ for any J and n and (2) angsta is defined where $n_j=0$.

CONCLUSION

Though of obviously common usefulness, measures of dispersion for qualitative variables are not widely published. Several extant measures have been reviewed in this paper. All of these statistics are limited in their applicability by an inability to accurately measure maximum dispersion when n is not an integer multiple of J and, for some statistics, their being undefined when $n_j=0$ for any category. Not only are these two conditions not extraordinary, but they are more likely to be the norm, the latter especially in the context of the small observation bases often accompanying simulation applications and research.

The angsta measure introduced in this paper is not susceptible to either of these common limitations of existing measures of dispersion. It remains to examine additional properties of the angsta statistic. That research is presently underway.

(References available on request from the senior author.)