How to Compute a Mean? The Chisini Approach and Its Applications

Rebecca Graziani and Piero Veronese

Scholars often consider the arithmetic mean as the only mean available. This gives rise to several mistakes. Thus, in a first course in statistics, it is necessary to introduce them to a more general concept of mean. In this work we present the notion of mean suggested by Oscar Chisini in 1929, which has a double advantage. It focuses students' minds on the substance of the problem for which a mean is required, thus discouraging any automatic procedure, and it does not require a preliminary list of the different mean formulas. Advantages and limits of the Chisini mean are discussed by means of examples.

KEY WORDS: Geometric mean; Harmonic mean; Power mean; Substitutive mean.

1. INTRODUCTION

In introductory textbooks in statistics, the mean is usually presented as a synthetic measure of a collection of observations providing information on the central tendency of the data. The mean considered and discussed in the present context is the *arithmetic mean* and its role of location index is emphasized by the usual comparison with other indexes, such as the median or the mode. A natural consequence of a similar approach is that students identify the notion of mean with that of arithmetic mean and use the latter in any context. The necessity of introducing them to a more general concept of mean is evident, but how can such a goal be reached? Recently, for example, Lann and Falk (2006) focused their attention on the class of the weighted (arithmetic) means. To make students understand that many kinds of weighted means exist and that the weights must be suitably chosen, they resort to several examples and show how the different weighting procedures may depend on the hypotheses underlying the problem, on sampling procedures, or on particular, namely geometric, requirements.

In this work we discuss another complementary way of achieving the previously mentioned aim. This approach, suggested by Chisini (1929), seems to us particularly suited to giving a good understanding of a notion of mean, which is more general than the notion of arithmetic mean. The approach has a double advantage. First, by discouraging any automatic procedure it makes students understand the substance of the problem for which a mean is required. Second, it does not require a preliminary (and necessarily incomplete) list of the different mean formulas. Dodd (1940) refers to the Chisini means as *representative* or *substitutive* means, to emphasize that they are useful in a *practical context* and to distinguish them from other synthetic standard location indexes. For example, if a car travels at speed v_1 over a distance s_1 and at speed v_2 over a distance s_2, different means of (v_1, v_2) can be formally computed. The Chisini idea, however, is that a suitable mean can be chosen only taking into account the specific problem to be solved (see the first example in Section 2). A nice discussion of the Chisini mean and an extension to random variables was given by de Finetti (1931). For an interesting review of the notions of utility and mean in the 1930s (see Muliere and Parmigiani 1993).

The structure of the article is as follows. In Section 2, we give a formal definition of the Chisini mean and in Section 3, we illustrate the theory by resorting to several examples. Finally, in Section 4, we present an extension of the concept of the Chisini mean that overcomes some limitations of the method, and offer some concluding observations.

2. THE CHISINI MEAN

In the middle of the last century, Dodd (1940) stated that there was no general agreement as to what constitutes a mean. The situation seems to be unchanged today. The only necessary condition, widely shared and referred to as the consistency property, is that the mean of a set of numbers all equal to a constant c should itself be equal to c.

Chisini (1929) pointed out, however, that in a practical context a mean should simplify the problem under investigation (by replacing several observations by a single value) so that the overall evaluation of the problem itself remains unchanged. Therefore, the main issue is the specification of the so-called *invariance requirement*, being a function of the observations, that we want to remain unchanged while replacing the observations by their mean. More formally, the Chisini mean of n homogeneous values $x_1, \ldots, x_n$, *with respect to* the invariance requirement f, is the number (if it exists) $\bar{x}$ such that:

$$f(\bar{x}, \bar{x}, \ldots, \bar{x}) = f(x_1, \ldots, x_n). \tag{1}$$

The preceding equation may have no solutions, so that the mean does not exist, or it may have several solutions of which each one may be used. If, for example, $n=2$ and $f(x_1, x_2) = x_1^2 + x_2^2$,

Rebecca Graziani is Assistant Professor, Department of Decision Sciences, Università L. Bocconi, Milan, Italy (E-mail: *rebecca.graziani@unibocconi.it*). Piero Veronese is Professor, Department of Decision Sciences, Università L. Bocconi, Milan, Italy (E-mail: *piero.veronese@unibocconi.it*). This work was partially supported by Università L. Bocconi. The authors are indebted to Professor D.M. Cifarelli for introducing them to the Chisini notion of mean. They also thank a Referee and the Editor whose criticisms and insightful comments improved the article.

with x_1 and x_2 real, then $\bar{x} = \pm\sqrt{(x_1^2 + x_2^2)/2}$ are two possible means. Furthermore, note that if $x_1 = -3$ and $x_2 = 1$, then $\bar{x} = \pm\sqrt{5}$ and thus one of the mean values (namely $+\sqrt{5}$) is greater than $\max(x_1, x_2)$, showing that the Chisini mean might also be *external* to the interval of the observations (for a more relevant example, see de Finetti 1931). Consequently, it is reasonable to impose requirements on function f so as to obtain a unique mean value. Note first that when $x_1, x_2, \ldots, x_n$ are all equal, f turns out to be a function of one variable, ϕ say, i.e., $\phi(x) = f(x, x, \ldots, x)$. Now, if $y = \phi(x)$ is a continuous and strictly increasing function, then the Chisini mean is obtained from (1) as:

$$\bar{x} = \phi^{-1}(f(x_1, \ldots, x_n)). \tag{2}$$

For instance, if $f(x_1, \ldots, x_n) = w_1x_1 + w_2x_2 + \ldots + w_nx_n$, where w'_is are nonnegative constants not all equal to zero, then $\phi(x) = \sum_{i=1}^n w_i x$ and, from (2), we obtain:

$$\bar{x} = \phi^{-1}\left(\sum_{i=1}^n w_i x_i\right) = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}.$$

Clearly, there are situations in which no particular requirement can be identified so that the Chisini approach cannot be used. Moreover, it cannot lead to identification of some important location indexes such as the median or the mode of a statistical distribution (see Section 4 for a discussion of this issue).

It must be stressed that each invariance requirement identifies a particular mean and that the converse is also true (see Table 1). Very often, however, in a practical situation it is difficult to choose *directly* a particular mean, whereas an invariance requirement can be easily specified. Thus Table 1 can be seen as a useful tool for the interpretation of the classical means, but in the practical setting of Chisini approach it is not necessary.

3. EXAMPLES

We illustrate the aspects of the Chisini mean previously discussed by several examples. The first three are very simple and can be used in every introductory course, whereas the remaining ones show how this approach can also be fruitful in more complex situations.

Mean Traveling Speed (Chisini 1929). Assume that a car travels at speed v_i over a distance s_i for $i = 1, \ldots, n$. A natural requirement to be satisfied is that the total time $T = \sum_{i=1}^n s_i/v_i$ spent in travel remains unchanged when replacing the observed speeds by their mean. Therefore, Equation (1) gives $(s_1/\bar{v}) + \cdots + (s_n/\bar{v}) = (s_1/v_1) + \cdots + (s_n/v_n)$ whose solution is $\bar{v} = \left(\sum_{i=1}^n s_i\right)/\left(\sum_{i=1}^n s_i/v_i\right)$. Thus, we can recognize in the last expression a harmonic mean of the v_i's (see row 3, Table1) but the preliminary knowledge of such a mean is not necessary to solve the problem.

If, however, we were interested in another value depending on the various speeds such as the total consumption of oil c, the speed mean would change. Assume that $c = a + s(v - v_0)^2$, where a and v_0 are constants depending on the car used, so in this case the invariance requirement turns out to be

$$f(v_1, \ldots, v_n) = na + \sum_{i=1}^n s_i(v_i - v_0)^2, \tag{3}$$

leading to

$$\bar{v} = v_0 + \left[\left(\sum_{i=1}^n s_i(v_i - v_0)^2\right)\Big/\left(\sum_{i=1}^n s_i\right)\right]^{1/2}.$$

Note that such a mean does not directly appear in any row of Table1, being indeed a quadratic mean of translated v_i's.

Mean Interest Rate. Assume that P dollars are invested for n periods of time, with i_k being the interest rate for period k ($k = 1, \ldots, n$), and that the interest is compounded. If we want an interest rate that leaves the compound amount unchanged after the n periods of time, then the function f must be set equal to $f(i_1, i_2, \ldots, i_n) = P\prod_{k=1}^n (1 + i_k)$. Consequently $\phi(\bar{i}) = f(\bar{i}, \ldots, \bar{i}) = P(1 + \bar{i})^n$ and the Chisini mean is

$$\bar{i} = \phi^{-1}(f(i_1, i_2, \ldots, i_n)) = \sqrt[n]{\prod_{k=1}^n (1 + i_k)} - 1. \tag{4}$$

Implicitly we have shown that $(1 + \bar{i})$ is the geometric mean of $(1 + i_k)$, (see row 4, Table1).

Table 1. Examples of invariance requirements (f) and the corresponding Chisini means. The weights w_i's are assumed to be nonnegative and not all equal to 0.

	f	Conditions on x_i's	Mean	Name
1)	$\sum_{i=1}^n w_i x_i$	– –	$\frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$	weighted arithmetic mean
2)	$\sum_{i=1}^n w_i x_i^2$	$x_i \geq 0$	$\sqrt{\frac{\sum_{i=1}^n w_i x_i^2}{\sum_{i=1}^n w_i}}$	weighted quadratic mean
3)	$\sum_{i=1}^n w_i x_i^{-1}$	$x_i > 0$	$\frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n w_i x_i^{-1}}$	weighted harmonic mean
4)	$\prod_{i=1}^n x_i^{w_i}$	$x_i > 0$	$\left(\prod_{i=1}^n x_i^{w_i}\right)^{\frac{1}{\sum_{i=1}^n w_i}}$	weighted geometric mean
5)	$\sum_{i=1}^n w_i x_i^k$	$x_i > 0$, ($k \neq 0$, $k \in \mathbb{R}$)	$\sqrt[k]{\frac{\sum_{i=1}^n w_i x_i^k}{\sum_{i=1}^n w_i}}$	weighted power mean
6)	$\sum_{i=1}^n w_i e^{x_i}$	$x_i \geq 0$	$\log \frac{\sum_{i=1}^n w_i e^{x_i}}{\sum_{i=1}^n w_i}$	weighted exponential mean

Mean Exchange Rate. Let $r_1, \ldots, r_n$ be the exchange rates Eur/USD, for n days, so let $\bar{r}_a$ be the corresponding arithmetic mean. Clearly, the exchange rates USD/Eur are given by $w_i = 1/r_i$, $i = 1, \ldots, n$ and let $\bar{w}_a$ be their arithmetic mean. Now $\bar{w}_a = (1/n)\sum_{i=1}^{n} r_i^{-1} \neq 1/\bar{r}_a$ and thus the arithmetic mean rate is incoherent depending on the "country perspective". But the point is, why choose a simple arithmetic mean? Look at the problem from the Chisini approach. Suppose we change, in n different days, x_i€ at a rate r_i so as to obtain $y_i = r_i x_i$ \$. If we look for the mean rate Eur/USD, which leaves unchanged the total capital in USD, we have $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} r_i x_i = \sum_{i=1}^{n} \bar{r} x_i$, which trivially implies $\bar{r} = \sum_{i=1}^{n} r_i x_i / \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i / \sum_{i=1}^{n} x_i$. Now reverse the perspective. If we start with y_i\$, the Chisini mean rate USD/Eur, which leaves unchanged the total capital in Euro, is given by $\bar{w} = \sum_{i=1}^{n} w_i y_i / \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i / \sum_{i=1}^{n} y_i$. Thus, once the practical issue to be addressed has been specified, the Chisini approach identifies a weighted (not simple) arithmetic mean as the suitable one and no incoherency appears ($\bar{w} = 1/\bar{r}$).

Mean Geyser Eruptions Waiting Time. We reexamine here an interesting example discussed by Lann and Falk (2006). Suppose that the authorities of a park want to publicize the *mean time* that a visitor has to wait for an eruption of a geyser. In this case, the invariance requirement to consider is clearly the total waiting time of all visitors. Thus, if n_i visitors arrive in the i-th interval between two eruptions and wait t_i ($i = 1, \ldots, k$), then $\bar{t} = (\sum_{i=1}^{k} n_i t_i)/(\sum_{i=1}^{k} n_i)$, i.e., a weighted arithmetic mean of the t_i's, (see row 1, Table 1). Now, if we assume that the visitors arrive at random during the opening time of the park, n_i is proportional to ℓ_i, the time interval between eruptions i and $i+1$, i.e., $n_i = c\ell_i$, for a suitable constant c. Furthermore, because the arrivals are supposed to take place at random during *any* time interval, each visitor's waiting time can be assumed to be half the interval's length, i.e., $t_i = \ell_i/2$. Thus the previous mean becomes

$$\bar{t} = \frac{\sum_{i=1}^{k} \ell_i \ell_i/2}{\sum_{i=1}^{k} \ell_i} = \frac{1}{2}\frac{\sum_{i=1}^{k} \ell_i^2}{\sum_{i=1}^{k} \ell_i}.$$

The last expression is a scalar transformation of the self-weighted mean of the ℓ_i's and clearly coincides with the result obtained by Lann and Falk (2006).

Mean Family Size. Lann and Falk (2006) recall the distinction between the average size of families and the *average number of children in the family of the average child*, which is an arithmetic mean of values representing the family size of each child in the dataset. The authors state that "distinguishing between the two means is not always easy" and correctly underline that such means are used for different purposes (see also Jenkins and Tuten 1992). Thus the Chisini approach appears useful. For example, suppose that k families, each with n_i children ($i = 1, \ldots, k$), live in a specific region and assume that the government of the region wants to give a benefit of D dollars for each child in a family. In this situation, the mean number of children for family should leave unchanged the total cost that the government must meet, i.e., $\sum_{i=1}^{k} D\bar{n} = \sum_{i=1}^{k} Dn_i$. Consequently (see row 1, Table 1), the mean family size is $\bar{n} = \sum_{i=1}^{k} n_i/k$, which is the usual average size of families. Now, it is known that the number of siblings has a relevant negative impact on the total number of years of education completed by an individual (see e.g., Blake 1989). Therefore if the government is interested in the mean number of children *per* family, from an educational point of view, it may consider as invariance requirement the total number of years of education completed by all children of the region. If we denote by c_{ij} the number of years of school completed by child j ($j = 1, \ldots, n_i$) of the family i and, following Blake's assumption, we assume that c_{ij} is (roughly) proportional to the size n_i of the family of the child, (i.e., $c_{ij} = rn_i$ for a positive constant r), the invariance requirement is

$$f(n_1, \ldots, n_k) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} c_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} rn_i. \qquad (5)$$

The Chisini mean is then

$$\bar{n} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} rn_i}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} r} = \frac{\sum_{i=1}^{k} n_i \cdot n_i}{\sum_{i=1}^{k} n_i}.$$

The last expression is the average number of children in the family of the average child given by Jenkins and Tuten (1992). Note that, in (5), n_i appears also as the maximum value taken by the index j, but in this case it should clearly not be substituted by $\bar{n}$.

Geometric problems. To discuss different types of mean, several authors resort to geometric requirements. Indeed, this is equivalent to specifying an invariance requirement so that their results can be seen as Chisini means. Maor (1977) considers a rectangle with sides x_1 and x_2 and shows that it is possible to construct different equivalent squares of side $\bar{x}$, according to the definition of "equivalent". Thus, if the two figures are requested to share the same perimeter, we must have $2x_1 + 2x_2 = 4\bar{x}$, which is an equation of type (1), whose solution is the arithmetic mean. Similarly, if we consider the area or the diagonal, instead of the perimeter, then $\bar{x}$ corresponds, respectively, to the geometric or the quadratic mean.

Lann and Falk (2006) use the same idea to introduce the self-weighted mean as the one which leaves unchanged the ratio S/P, where $S = 2x_1^2 + 2x_2^2$ is the sum of the areas of the squares built on each side of the rectangle and P is its perimeter. Finally Mateјаš and Bahovec (2008), starting from a similar problem, namely *the average dimension of a cuboid with dimension* $x_1 \times x_2 \times x_3$, rediscover that Equation (1) represents a key point to give a general definition of mean that they call f-mean.

4. POSSIBLE EXTENSIONS AND CONCLUSIONS

In the present article we suggest introducing students to the Chisini notion of mean, as a quite general approach to determine synthetic measures from a practical point of view. As shown in Table 1, several families of means well known in the literature can be obtained within the Chisini approach, through a suitable choice of the invariance requirement (namely the function f). This allows us to discuss, in a not purely introductory course, the specific properties and the characterizations of the different kinds of mean. In particular, the class of power means of order k can be considered (see row 5, Table 1). It is easily seen that, for $k=-1$, 1, 2, we obtain the harmonic, the arithmetic, and the quadratic mean respectively. Furthermore, this class also includes the geometric mean as a limit for k tending to 0. Thus it is interesting to prove results concerning the power means, the *associativity* for example, because they will automatically hold for many specific kinds of mean. We suggest, in particular, introducing the property of *monotonicity* in k: for a given set of observations, not all equal, the power mean of order r is always lower than the power mean of order s, if $r<s$. Thus an ordering among the more common means is easily established and consequently it is straightforward to recall, for example, that the harmonic mean is lower than the geometric mean (the equality holds, by *consistency*, only if all observations are equal).

Finally, as remarked in Section 2, the Chisini approach cannot lead to the identification of some important location indexes, such as the median or the mode of a statistical distribution. A generalization suggested by Herzel (1961) may be useful in overcoming this restriction. The main point is that there are situations in which it is unreasonable to require that the problem has to remain unchanged when the observations are replaced by a unique value. We can only ask that it varies as little as possible. Formally, for a given set of observations ($x_1, \ldots, x_n$), let $g(a; x_1, \ldots, x_n)$ be a real function measuring the change in the phenomenon determined by replacing the observations with a constant a and such that $g(a; a, \ldots, a) = 0$, for any a in the set of the x_is. Then the Herzel mean of ($x_1, \ldots, x_n$), with respect to g, is the value $\bar{x}$ such that $g(\bar{x}; x_1, \ldots, x_n) = \min_{a \in A} g(a; x_1, \ldots, x_n)$, for a suitable set A. The choice of A leads to means satisfying predetermined properties, such as positivity or coincidence with one of the observations themselves. Note that, if we set $g(a; x_1, \ldots, x_n) = |f(a, \ldots, a) - f(x_1, \ldots, x_n)|$, we obtain the Chisini mean as a solution. Viceversa, if $g(a; x_1, \ldots, x_n)$ allows us to write explicitly $a = h(x_1, \ldots, x_n)$, then h turns out to be the invariance requirement.

For example, suppose that $x_1, \ldots, x_n$ are earnings on a collection of fixed investments and that we look for a reasonable prediction p of a 'future' earning on a same type of investment. In this context, the Chisini approach cannot be used. More properly we might search for the unique value that minimizes the distance between our prediction p and the data. In particular, if we set g equal to $\sum_{i=1}^{n} |x_i - p|$, the mean earning p is the median of the x_i's. Alternatively, if we set $g(a; x_1, \ldots, x_n)$ equal to $\sum_1^n (1 - I_{\{a\}}(x_i))$, where $I_{\{C\}}(\cdot)$ is the indicator function of the set C, we obtain as mean earning p the mode of the x_i's.

[Received June 2007. Revised November 2008.]

REFERENCES

Blake, J. (1989), "Number of Siblings and Educational Attainment," *Science,* 245, 32–36.

Chisini, O. (1929), "Sul Concetto di Media," *Periodico di Matematiche,* 2, 106–116.

de Finetti, B. (1931), "Sul Concetto di Media," *Giornale dell'Istituto Italiano degli Attuari,* 2, 369–396.

Dodd, E. L. (1940), "The Substitutive Mean and Certain Subclasses of this General Mean," *Annals of Mathematical Statistics,* 11, 163–176.

Herzel, A. (1961), "Sulla Definizione Dei Concetti di Media e D'Interpolazione," *Metron*, 21, 126–138.

Jenkins, J. J., and Tuten, J. T. (1992), "Why Isn't the Average Child From the Average Family? and similar Puzzles," *The American Journal of Psychology,* 105, 517–526.

Lann, A., and Falk, R. (2006), "Tell Me the Method, I'll Give You the Mean," *The American Statistician,* 60 (4), 322–327.

Maor, E. (1977), "A Mathematician's Repertoire of Means," *Mathematics Teacher*, 70, 20–25.

Matejaš, J., and Bahovec, V. (2008), "Another Approach to Generalizing the Mean," *Teaching Statistics,* 30, 49–52.

Muliere, P., and Parmigiani, G. (1993), "Utility and Means in the 1930s," *Statistical Science,* 8 (4), 421–432.