ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 225, 2009

*Wiesław Wagner**

REMARKS ON CLASSICAL MEANS FOR ONE AND MANY SAMPLES

Abstract

In the paper there were given basic designations for one and many samples in reference to classical means. In the next there were introduced the optimization criteria defining classical means for one sample, also for the distributive series of size, and for many samples and gives several examples, after which comes recapitulation of the paper.

Key words: descriptive statistics, classical means, optimization criteria.

1. Introduction

In descriptive statistics, to the basic activities of elaborating numerical data from a sample there is classified determination of numerical characteristics. There are differentiated two groups of them: classical and positional. In the first one, there are most often applied characteristics of position, and, among them, classical means (see Zając, 1988; Jóźwiak and Podgórski, 1992; Luszniewicz and Słaby, 1996; Wagner, 2002). They play an important role in interpretation of data, and also they are often used to determine other numerical characteristics. Basically, there are differentiated three means: arithmetic, harmonic and geometric. Each of them is determined on different principles, namely on linear, inverse and power transformations. The question arises if a common principle of derivation can be found for them, and if, along with this, they can be treated as special cases of a wider class of means, both in the case of one sample, as well as of many samples. In the paper such a principle will be derived.

* Professor, University of Information Technology and Menagement in Rzeszów.

The further outline of the paper is the following. In the second chapter there were given basic designations for one and many samples in reference to classical means. In the next chapter there were introduced the optimization criteria defining classical means for one sample, also for the distributive series of size, and for many samples. The fourth chapter gives several examples, after which comes recapitulation of the paper.

2. Designations

Designations for one sample:

X – examined characteristic;
n – size (quantity) of the sample;
$P_n = (x_1, x_2, ..., x_n)$ – statistical sample of observation, but we assume that all values of the sample are positive;

$\overline{x}, \overline{h}, \overline{g}, \overline{x}_2, \overline{x}_p$ – means: arithmetic, harmonic, geometric, quadratic, power;

$\{(\tilde{y}_j, f_j): j = 1, 2, ..., k\}$ – point distributive series of size, where: $\tilde{y}_j$ center of j-th class interval (in particular the differentiated value of the characteristic X), f_j size of j-th class and k number of class intervals, where $n = \sum_{j=1}^{k} f_j$.

Designations for many samples:

q – number of samples;
$n_1, n_2, ..., n_q$ – size of samples;
$x_{k1}, x_{k2}, ..., x_{kn_k}$ – observations of k-th sample, $k = 1, 2, ..., q$,
$n = \sum_{k=1}^{q} n_k$ – total size of samples;
$\overline{x}_k, \overline{h}_k, \overline{g}_k, \overline{x}_{2,k}, \overline{x}_{p,k}$ – means for k-th test: arithmetic, harmonic, geometric, quadratic and power.

Beside the given designations, we introduce one more, the function $f(z)$, monotonic and continuous of positive values, transforming observation for the sample, and its inverse function is designated as $f^{-1}(.)$.

3. Optimization criteria

3.1. Criterion for one sample

Let a designate classical mean from statistical sample P_n. Its form is determined by optimization criterion:

$$Q(a) = \min_a \sum_{i=1}^{n} [f(x_i) - f(a)]^2 \tag{1}$$

i.e. mean a is the lowest value of the function $Q(a)$, for the set transforming function f. Extremum of the function $Q(a)$ is determined from the necessary condition $Q'(a) = 0$, neutralization of the derivative

$$Q'(a) = -2\sum_{i=1}^{n} [f(x_i) - f(a)] f'(a) = -2f'(a) \left\{ \sum_{i=1}^{n} f(x_i) - nf(a) \right\}$$

Since the function $f(a)$ in its assumption is non-zero, then after equating to zero of the last equation and after reductions, we receive the equation $\sum_{i=1}^{n} f(x_i) = nf(a)$, and hence the expression for the mean from the sample a

$$a = f^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i)\right] \tag{2}$$

In order to show that the function $Q(a)$ takes the lowest value for the determined a, we will make use of the final condition. The second derivative of the function $Q(a)$ takes the form

$$Q''(a) = -2f''(a)\left\{ \sum_{i=1}^{n} f(x_i) - nf(a) \right\} + 2n\left[f'(a)\right]^2.$$

It reduces itself for the set extreme point (2) to $Q''(a) = 2n\left[f'(a)\right]^2$, and it means that it always takes positive value. By this it was shown that the solution (2) guarantees the minimum of the function (1).

Determining the means with the pattern (2) depends directly on the selected transforming function $f(.)$. Depending on the selection of this function, the appropriate means are presented as follows:

Function $f(z)$	Kind of mean
z	arithmetic
$1/z$	harmonic
$\ln z$	geometric
z^2	quadratic
z^p	power

The pattern (2) indicates a wide class of possible means. Especially the property occurs here. If we replace the function $f(z)$ with its linear transformation $h(z) = \alpha f(z) + \beta$, where the constants are positive, then the type of mean will remain invariant to the mean determined by $f(z)$. In order to show that it is so, we take the pattern (2). Let us write it in the form $h(a) = \frac{1}{n}\sum_{i=1}^{n} h(x_i)$, and hence $h(a) = \frac{1}{n}\sum_{i=1}^{n}[\alpha f(x_i) + \beta] = \frac{\alpha}{n}\sum_{i=1}^{n} f(x_i) + \beta = \alpha f(a) + \beta$.

It is yet worth paying attention here to the proof of the final condition at the set mean. For example, let us take a transforming function of the type $f(z) = \frac{1}{z}$, i.e. the harmonic mean. Taking into consideration the earlier-given general form $Q^{"}(a)$, we determine for it: $f(a) = \frac{1}{a}, f^{'}(a) = \frac{-1}{a^2}, f^{"}(a) = \frac{2}{a^3}$ and $f(x_i) = \frac{1}{x_i}$. After substitutions we receive

$$Q^{"}(a) = -2 \cdot \frac{2}{a^3}\sum_{i=1}^{n}\frac{1}{x_i} + \frac{4n}{a^3}\cdot\frac{1}{a} + 2n\left(-\frac{1}{a^2}\right)^2 = -\frac{4n}{a^4} + \frac{6n}{a^4} = \frac{2n}{a^4} > 0$$

where we used the equation $\sum_{i=1}^{n}\frac{1}{x_i} = \frac{n}{a}$.

3.2. Criterion for the distributive series of size

The optimization criterion for the distributive series of size is formulated similarly to (1). It takes here the form

$$Q(a) = \min_a \sum_{j=1}^{k} f_j[f(\tilde{y}_j) - f(a)]^2 \tag{3}$$

Proceeding similarly to the point 3.1, we receive

$$f(a) = \frac{1}{n}\sum_{j=1}^{k} f_j f(\tilde{y}_j) \tag{4}$$

3.3. Criterion for many samples

Taking designations for many samples, we now express the optimization criterion for determining the mean in the form

$$Q(a)=\min_{a}\sum_{k=1}^{q}\sum_{j=1}^{n_k}[f(x_{kj})-f(a)]^2 \tag{5}$$

which provides the solution

$$f(a)=\frac{1}{n}\sum_{k=1}^{q}\sum_{j=1}^{n_k}f(x_{kj}) \tag{6}$$

Assuming that the means for samples are known, i.e. there is set a set of pairs $\{(\overline{a}_k,n_k):k=1,2,...,q\}$, then the general means for the total sample are determined according to the criterion (3) and the pattern (4), after making changes in: $\tilde{y}_j\to\overline{a}_k$ and $f_j\to n_k$, and summation should be conducted according to the index k within the scope 1, 2, ..., q, which finally leads to

$$f(a)=\frac{1}{n}\sum_{k=1}^{q}n_k f(\overline{a}_k) \tag{7}$$

Formulas (2), (4), (6) and (7) will be illustrated in the next chapter.

4. Ilustrating examples

We are presenting various examples illustrating determining the means from the given criteria.

Example 1. Derivation of the geometric mean for one sample. We use the formula $\ln a=\frac{1}{n}\sum_{i=1}^{n}\ln x_i$. After transformations

$$\ln a=\sum_{i=1}^{n}\ln x_i^{1/n}=\ln\left(\prod_{i=1}^{n}x_i^{1/n}\right)=\ln\left(\prod_{i=1}^{n}x_i\right)^{1/n}$$

we have $a=\sqrt[n]{\prod_{i=1}^{n}x_i}=\overline{g}$.

Example 2. Derivation of the formula for the harmonic mean weighted from the distributive series. We use the formula (4) in the form $\frac{1}{a}=\frac{1}{n}\sum_{j=1}^{k}f_j\frac{1}{\tilde{y}_j}$, and hence we have $a=\left(\frac{1}{n}\sum_{j=1}^{k}\frac{f_j}{\tilde{y}_j}\right)^{-1}$, which leads to the weighted harmonic mean $a=\frac{n}{\sum_{j=1}^{k}\frac{f_j}{\tilde{y}_j}}=\bar{h}$.

Example 3. Determining the quadratic mean for many tests. We begin with the formula (6) in the form $a^2=\frac{1}{n}\sum_{k=1}^{q}\sum_{j=1}^{n_k}x_{kj}^2$, and hence we immediately have $a=\sqrt{\frac{1}{n}\sum_{k=1}^{q}\sum_{j=1}^{n_k}x_{kj}^2}$.

Example 4. Determining the power mean for many samples at the set power means of particular samples. We assume the data $\{(\bar{x}_{p,k}, n_k): k=1,2,...,q\}$ and apply the function $f(z)=z^p$. It leads to the formula (7) in the form $a^p=\frac{1}{n}\sum_{k=1}^{q}n_k\bar{x}_{p,k}^p$, which finally gives the power mean of the p-th degree, $a=\bar{x}_p=\left\{\frac{1}{n}\sum_{k=1}^{q}\bar{x}_{p,k}^p\right\}^{1/p}$.

5. Recapitulation

In the paper we showed usability of the transforming function in the optimization criterion of determining various classical means. Such a criterion allowed us to determine the means for one and many samples, including the distributive series.

It would be worth considering the presented problem of determining the means in the situation of incomplete samples.

References

Jóźwiak J., Podgórski J. (1992), *Statystyka od podstaw*, PWE, Warszawa.

Luszniewicz A., Słaby T. (1996), *Statystyka stosowana*, PWE, Warszawa.

Wagner W. (2002), *Podstawy metod statystycznych w turystyce i rekreacji*, t. I, Wyd. AWF, Poznań.

Zając K. (1988), *Zarys metod statystycznych*, PWE, Warszawa.

Wiesław Wagner

Uwagi o średnich klasycznych dla jednej i wielu prób

W statystyce opisowej do podstawowych czynności opracowywania danych liczbowych z próby zalicza się wyznaczanie charakterystyk liczbowych. Wyróżnia się dwie ich grupy: klasyczne i pozycyjne. W pierwszej, najczęściej stosuje się charakterystyki położenia, a wśród nich średnie klasyczne. Pełnią one ważną rolę w interpretacji danych, a także są często wykorzystywane do wyznaczania innych charakterystyk liczbowych. Zasadniczo wyróżnia się trzy średnie: arytmetyczną, harmoniczną i geometryczną. Każda z nich jest wyznaczana na innych zasadach, a mianowicie na przekształceniach liniowych, odwrotnościowych i potęgowych. Powstaje pytanie, czy można dla nich znaleźć wspólną regułę wyprowadzenia, a wraz z tym traktować je jako szczególne przypadki szerszej klasy średnich zarówno w przypadku jednej próby, jak i wielu prób. W pracy została taka reguła wyprowadzona.