NOTE ON QUARTILES AND ALLIED MEASURES*

BY DUNHAM JACKSON

If a number of values $a_1, a_2, \cdots, a_n$ of a quantity x have been observed, the lower quartile of this set of observations may be roughly described as a number x_1 such that one fourth of the a's are less than x_1 and three fourths of them are greater than x_1. Something more is needed for an exact definition, inasmuch as the condition stated either leads to an indeterminate value or is impossible of realization, according to circumstances. If x_1 is defined as a value of x which reduces to a minimum the expression

$$S_1 = \sum_{i=1}^{n} \varphi_1(x - a_i),$$

where $\varphi_1(x) = \frac{3}{4}x$ for $x \geqq 0$, $\varphi_1(x) = -\frac{1}{4}x$ for $x \leqq 0$, there will always be at least one value of x_1 satisfying the condition, and this will agree with the value of the quartile as ordinarily understood, but if $n = 4k$ and $a_k \neq a_{k+1}$, when the a's are arranged in order of increasing algebraic magnitude, any number between a_k and a_{k+1} will meet the requirement. It is the purpose of this note to show that a unique determination results in all cases from a definition analogous to one which the author recently suggested for the median.† As in the previous instance, the definition is admittedly of theoretical rather than practical interest. The discussion is put in such a form as to apply equally well to an arbitrary percentile or other measure of similar character, the ratio $1:3$ being replaced by any other positive ratio.

Let $a_1, \cdots, a_n$ be a set of real numbers (not necessarily all distinct) arranged in ascending order of magnitude algebraically, and let c be an arbitrary number of the interval $0 < c < 1$. For $p \geqq 1$, let a function $\varphi_p(x)$ be defined as follows: $\varphi_p(x) = (1 - c)x^p$ for $x \geqq 0$, $\varphi_p(x) = c(-x)^p$ for $x \leqq 0$.

* Presented to the Society, October 28, 1922.

† *Note on the median of a set of numbers*, this BULLETIN, vol. 27 (1920–21), pp. 160–164.

The value of c will be kept constant throughout, and need not be indicated in the notation for the function φ. Let

$$S_p = \sum_{i=1}^{n} \varphi_p(x - a_i),$$

and let x_p stand for the value of x, or any value of x, which reduces S_p to a minimum. It is to be shown that x_p is uniquely determined for each value of $p > 1$; that as p approaches 1, x_p approaches x_1, if x_1 has a determinate value; and that in the contrary case, x_p approaches a definite limit belonging to the interval within which x_1 is indeterminate.

Consider first the case that cn is not an integer. Let k be the integer such that $k - 1 < cn < k$. Then x_1 definitely has the value a_k. For as x changes from a_k to $a_k + \delta$, at least k terms of the sum S_1 are increased, each by the amount $(1 - c)\delta$, and not more than $n - k$ terms are diminished, the amount of decrease in each case being $c\delta$ at most, so that the net change in S_1 is at least

$$k(1 - c)\delta - (n - k)c\delta = (k - cn)\delta > 0;$$

and as x changes from a_k to $a_k - \delta$, at least $n - k + 1$ terms are increased, not more than $k - 1$ are diminished, and the net change is at least

$$(n - k + 1)c\delta - (k - 1)(1 - c)\delta = (cn - k + 1)\delta > 0.$$

It will be shown that x_p is uniquely determined for each value of $p > 1$, and that $\lim_{p=1} x_p = a_k$.

When $p > 1$, the function $\varphi_p(x)$ is continuous and has a continuous derivative for all values of x, including $x = 0$. Since $S_p(x)$ is continuous and becomes infinite as x becomes infinite in either direction, it must have at least one minimum. A necessary condition for a minimum is the vanishing of $S_p'(x)$. But it is readily seen, either by inspection or by writing down the explicit formulas, that $\varphi_p'(x)$ always increases when x increases, so that $S_p'(x)$ likewise increases when x increases, and can vanish only once. This proves the existence and uniqueness of x_p.

Let ϵ be an arbitrarily small positive quantity, and let r be an index such that $a_r < a_k + \epsilon \leqq a_{r+1}$; it is clear that $r \geqq k$.

It follows from the definition of φ that $\varphi_p'(x) = (1 - c)px^{p-1}$ or $- cp(-x)^{p-1}$, according to the sign of x, and hence

$$\frac{1}{p} S_p'(a_k + \epsilon) = (1 - c)(a_k + \epsilon - a_1)^{p-1} + \cdots + (1 - c)(a_k + \epsilon - a_r)^{p-1} - c(a_{r+1} - a_k - \epsilon)^{p-1} - \cdots - c(a_n - a_k - \epsilon)^{p-1}.$$

When p approaches 1, each of the first r terms on the right, apart from the factor $1 - c$, approaches the limit 1, and each of the remaining terms, apart from the factor c, but inclusive of the algebraic sign, approaches -1 or possibly 0. So

$$\lim_{p=1} S_p'(a_k + \epsilon) \geqq r(1 - c) - (n - r)c = r - cn \geqq k - cn > 0.$$

Similarly,

$$\lim_{p=1} S_p'(a_k - \epsilon) < 0.$$

For if the definition of r is changed so that $a_r \leqq a_k - \epsilon < a_{r+1}$, then $r \leqq k - 1$, and

$$\frac{1}{p} S_p'(a_k - \epsilon) = (1 - c)(a_k - \epsilon - a_1)^{p-1} + \cdots + (1 - c)(a_k - \epsilon - a_r)^{p-1} - c(a_{r+1} - a_k + \epsilon)^{p-1} - \cdots - c(a_n - a_k + \epsilon)^{p-1},$$

$$\lim_{p=1} S_p'(a_k - \epsilon) \leqq r(1 - c) - (n - r)c = r - cn \leqq k - 1 - cn < 0.$$

So the value x_p for which S_p' vanishes must be between $a_k - \epsilon$ and $a_k + \epsilon$ when p is sufficiently near 1, or, in other words, $\lim_{p=1} x_p = a_k$.

Suppose now that cn is an integer, and let $cn = k$. If it happens that $a_k = a_{k+1}$, reasoning similar to that presented above shows that $x_1 = a_k$ and $\lim_{p=1} x_p = a_k$, as before.

This special case being left aside, it is to be assumed that $a_k < a_{k+1}$. The definition of x_1 is satisfied by a_k or a_{k+1} or any intermediate value. For $p > 1$, on the other hand, x_p is seen to be uniquely determined, by the same argument as was used before. Furthermore, it is recognized that

$$\lim_{p=1} S_p'(a_k) = (k - 1)(1 - c) - (n - k)c = k - 1 + c - cn = c - 1 < 0,$$

$$\lim_{p=1} S_p'(a_{k+1}) = k(1 - c) - (n - k - 1)c = k - cn + c = c > 0,$$

so that $a_k < x_p < a_{k+1}$ when p is sufficiently near 1. It remains to be shown that x_p approaches a definite limit as p approaches 1.

Let x have a value between a_k and a_{k+1}. For $i = 1, 2, \cdots, k$, $x - a_i$ is positive, and

$$(x - a_i)^{p-1} = e^{(p-1)\log(x-a_i)}$$
$$= 1 + (p - 1) \log (x - a_i) + \tfrac{1}{2}(p - 1)^2 \log^2 (x - a_i) + \cdots$$
$$= 1 + (p - 1) \log (x - a_i) + (p - 1)^2\rho_i(x, p),$$

where $\rho_i(x, p)$ is a function which approaches $\frac{1}{2} \log^2 (x - a_i)$, and so remains finite, if x is held fast and p approaches 1. For $i > k$,

$$(a_i - x)^{p-1} = 1 + (p - 1) \log (a_i - x) + (p - 1)^2\rho_i(x, p),$$

where $\rho_i(x, p)$ again remains finite for fixed x as p approaches 1. If these values are substituted in the explicit expression for $(1/p)S_p'(x)$, there will be k terms each equal to $(1 - c)$ and $(n - k)$ terms each equal to $(- c)$, which will destroy each other, because of the relation $cn = k$, and each of the remaining terms will have a factor $p - 1$, so that we may write

$$\frac{1}{p(p-1)} S_p'(x) = (1 - c)[\log (x - a_1) + \cdots + \log (x - a_k)]$$
$$- c[\log (a_{k+1} - x) + \cdots + \log (a_n - x)] + (p - 1)\rho(x, p)$$
$$= \log \frac{[(x - a_1)\cdots(x - a_k)]^{1-c}}{[(a_{k+1} - x)\cdots(a_n - x)]^c} + (p - 1)\rho(x, p),$$

the function ρ remaining finite as p approaches 1.

As the exponents c and $1 - c$ are both positive, the fraction on the right increases steadily from 0 to $+ \infty$ as x goes from a_k to a_{k+1}, and the logarithm increases steadily from $- \infty$ to $+ \infty$, taking on the value 0 just once, say for $x = X$. For $x = X + \epsilon$, the logarithm is positive and independent of p, while the term $(p - 1)\rho(X + \epsilon, p)$ approaches zero as p approaches 1. So the value of the whole expression on the right is positive when $p - 1$ is sufficiently small. For a similar reason, the expression is negative for $x = X - \epsilon$, if p is sufficiently close to 1. This means that the root of $S_p'(x)$ is between $X - \epsilon$ and $X + \epsilon$ when $p - 1$ is sufficiently small, that is,

$$\lim_{p=1} x_p = X.$$

THE UNIVERSITY OF MINNESOTA