PSYCHOMETRIKA—VOL. 13, NO. 2
JUNE, 1948

AN ABAC FOR THE SAMPLE RANGE

MARCO P. SCHÜTZENBERGER
PARIS, FRANCE

An abac is computed which gives the probability that at least $q\%$ of the whole population will be included in the range of a random sample of given size. Applications are suggested for testing homogeneity of a sampling from a given population.

Let a sample of N values be randomly drawn from an infinite continuous distribution. The present chart gives the probability p that at least q per cent of the whole population lie between the extreme values of the sample.

For instance, if the universe is that of the speeds in a given performance test, and the sample is constituted of N subjects passing in a given day, q will be the proportion to all subjects of those who will complete the task slower than the best of the sample and faster than the poorest.

It is easy to prove that the probability p, the percentage q, and the sample size N are related by

$$p = 1 - \left(\frac{m-1}{m}\right)^{N-1} \left(\frac{N+m-1}{m}\right)^{*}, \tag{1}$$

where

$$m = \frac{1}{1-q}. \tag{2}$$

For large enough N and m, (1) may be closely approximated by

$$p = 1 - (r+2)e^{-r-1}, \tag{3}$$

where

$$r = \frac{N}{m} - 1. \tag{4}$$

The important fact is that (1) holds for every *continuous infinite* distribution, even leptokurtic or platykurtic, skew or multimodal.

* Wilks, S. S. *Mathematical statistics*. Princeton, New Jersey: Princeton University Press, 1943. P. 93.

Use of the Abac

In ordinates are plotted, on a logarithmic scale, values of q (proportion of the whole population included in the sample range) from 50 per cent up to 998 per thousand.

In abscissas, on a logarithmic scale also, are values of the sample size N from 5 up to 1000.

Equi-probability curves are drawn for $p = 999/1000$, $99/100$, $90/100$, $75/100$, $50/100$, $25/100$, $10/100$, $1/100$, $1/1000$, so that interpolation can be made easily, and the corresponding values are written in the left and upper margins of the abac.

For example:

for $N = 20$ and $p = 99/100$, q is 75 per cent.
for $N = 200$ and $p = 999/1000$, q is 955 per thousand.

Applications

Two main applications may be suggested in the field of current psychology:

In forecasting the proportion of subjects included in the range of a given sample. For example, in a preliminary trial on 20 subjects of the speed test referred to above, extreme values of 45 and 135 seconds have been observed. Taking the equi-probability curve for $p = 99/100$, we may guess with 99 odds against one, that at least 75 per cent of all future subjects will score between 45 and 135 seconds. With 9 odds against one (that is to say, with $p = 90/100$) and 100 subjects instead of 20, this proportion would rise up to 960 per thousand.

An alternative use of the chart may be in testing a bias in sampling when there is a hint of range discrepancies and when classical tests are too difficult or lengthy to compute. For example:

Let us suppose we got from a large population the decilage of a test and that in a sample of seven, we found a subject in the first percentile and one in the last. Here q is at least 98 per cent, $N = 7$, and there is just a single chance out of a hundred ($p = 1/100$) that a random sample from the known population would be so scattered. Conversely, if in another sample of thirty subjects, we found none in the first nor in the 10th decile ($q \leqslant 80$ per cent), the probability is 99/100 that this sample would be a biased one.

Other applications of this chart could readily be found in every case where the distribution function departs from usual forms or is unknown.

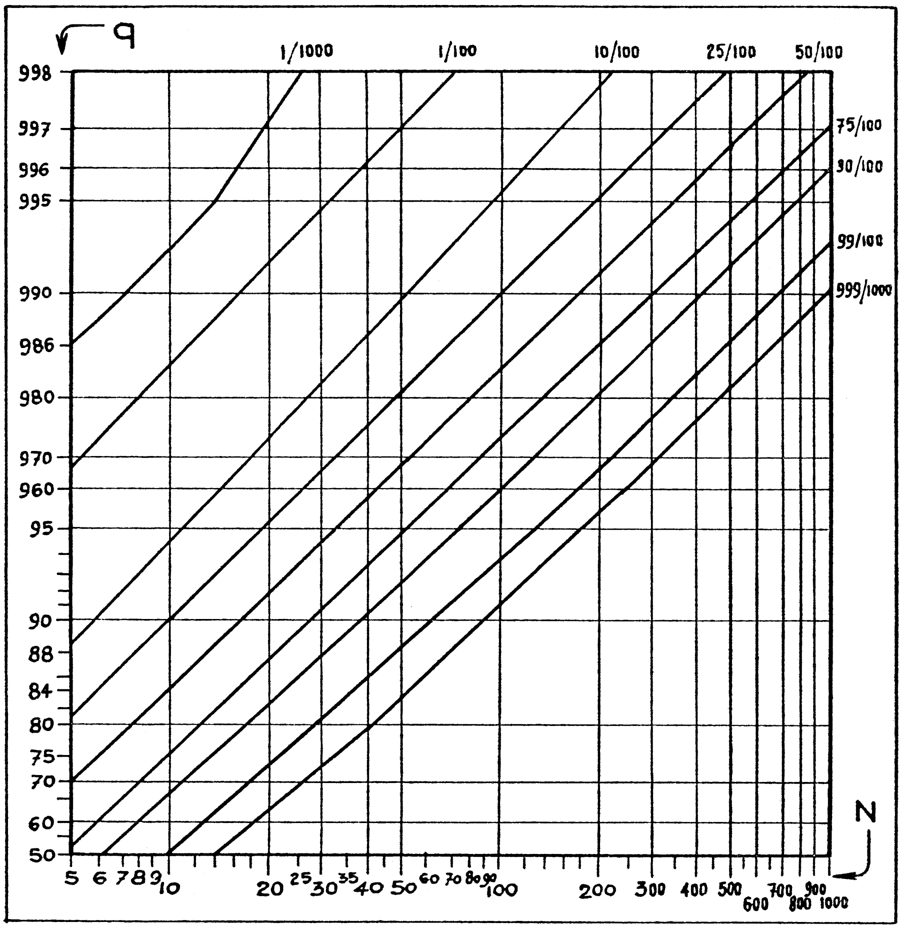
q
1/1000
1/100
10/100
25/100
50/100
75/100
90/100
99/100
999/1000
998
997
996
995
990
986
980
970
960
95
90
88
84
80
75
70
60
50
N
5 6 7 8 9 10 20 25 30 35 40 50 60 70 80 90 100 200 300 400 500 600 700 800 900 1000

FIGURE 1