DETERMINATION OF THE DISTRIBUTION OF SAMPLE SIZES REQUIRED IN TIME STUDY

by

JAGADEESAN V. POOLA

B. E., University of Madras, 1960

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1963

Approved by:

George F. Schrader
Major Professor

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Stop watch time study is the most commonly used method of measuring work in industry today. Time study is used to determine the time required by a qualified and well trained person working at a normal pace to do a specified task.[1] The time required to perform the elements of an operation may be expected to vary slightly from cycle to cycle. In work measurement, the determination of an acceptable time value to be assigned to an activity is often based on average task or cycle time obtained from timing several cycles of the activity with a stop watch.

In the early years of time study practice, 15 or 20 readings were thought to be sufficient, but without any substantiation. Time study is a sampling process; consequently the greater the number of cycles timed, the more nearly the results will be representative of the activity being measured. Idealistically, it would be desirable to time a very large number of cycles, but due to economic factors, this is not feasible.

In the past few years, several mathematical procedures have been suggested for determining the number of cycles to be observed in a stop watch time study in order to accurately estimate the true average task or cycle time.[2,3,4,5,6,7] These various

---

1. Barnes, R. M., Motion and Time Study. New York, New York: John Wiley and Sons, Inc., 4th edition, 1958.

2. Mundel, M. W., Motion and Time Study. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1960.

methods differ to some degree; however, each utilizes statistical methods as an aid in the determination of the number of cycles, N, to study in arriving at a satisfactory estimate of the true average time, μ , required for completion of the task. Usually, a small sample of observations, n, are made, and sample mean and sample variance are calculated; substituting these values in the given formula for determining N', the estimate of N, the total number of observations to be made in order to provide the desired confidence level on the estimate of μ is determined. All of the formulae are of the form,

$$N = \left[\frac{A\sigma}{k\mu} \right]^2$$

where μ = population mean,

σ = population standard deviation,

A = confidence interval constant, and

k = an acceptance per cent of μ .

---

3. Nadler, G., Motion and Time Study. New York: McGraw-Hill Book Co., Inc., 1955.

4. Niebel, B. W., Motion and Time Study. Homewood, Illinois: Richard D. Irwin, Inc., Revised Edition, 1958.

5. Radkins, A. P., "Calculating the Required Number of Time Study Readings Using Moving Ranges." Unpublished paper, Purdue University.

6. Allderige, J. H., "Statistical Procedures in Stop Watch Work Measurements." The Journal of Industrial Engineering, July-August, 1956, Vol. VII, No. 4, pp. 154-163.

7. Lifson, K. A., "Number of Observations for a Statistical Average." Time Study Engineer and Time Study Engineering, Vol. 6, August, 1951, pp. 247-248.

whereas N', the estimate of N, may be determined by the formula

$$N' = \left[\frac{A}{k}\,\frac{\sigma_x}{\bar{X}}\right]^2$$

where $\bar{X}$ = estimate of population mean

σ_x = estimate of population standard deviation

Each of the above formulae was derived under the assumption that either the population of elemental time values are normally distributed or the sample means are approximately normally distributed. In the area of work measurement, a number of investigations have been conducted on the distribution characteristics of time study data in order to verify the validity of some of the assumptions underlying the methods of determining the number of observations to record.[8] Recently a number of investigations[9,10] have been conducted to compare several methods for determining N' and to test by a process of simulation, their reliability and superiority.

---

8. Lehrer, R. N., and Moder, J. J., "Mathematical Characteristics of Performance Times." Time and Motion Study, October, 1955, Vol. 4.

9. Schrader, G. F., "A Critical Analysis of the Reliability and Relative Superiority of the Various Methods Recommended for Use in Determining the Number of Cycles to Record During a Time Study." Unpublished Doctoral Thesis, University of Illinois, 1960.

10. Hicks, C. R., and Young, H. H., "A Comparison of Several Methods for Determining the Number of Readings in a Time Study." The Journal of Industrial Engineering, Vol. XII, No. 2, March-April, 1962, pp. 93-96.

In his investigation, Dr. Schrader posed a question, Given sample size n and coefficient of variation $V = \sigma/\mu$, what would be the distribution of sample sizes required in a time study? Certainly N' must have a statistical distribution. Thus, the purpose of this investigation is to investigate the distribution of N', sample sizes required in a time study, and to calculate the parameters for the distribution of N', making use of the simulated data obtained by Dr. Schrader in his investigation. However, in this investigation, only formulae 1, 2, and 3 in Dr. Schrader's thesis are considered.

Table 1. List of formulae for determining N'.

Designation	Formula	Author
N_1'	$\left[\dfrac{40\sqrt{(n\Sigma X^2) - (\Sigma X)^2}}{\Sigma X}\right]^2$	Mundel
N_2'	$\left[\dfrac{39.2\sqrt{(n\Sigma X^2) - (\Sigma X)^2}}{\Sigma X}\right]^2$	Niebel
N_3'	$\left[\dfrac{42.6\sqrt{(n\Sigma X^2) - (\Sigma X)^2}}{\Sigma X}\right]^2$	Niebel

Note: Subscripts on N' are used later on in discussions to identify a particular N' in relation to a given formula. All formulae are developed for use with a preliminary sample size n of 16, and are based on 95 per cent confidence on plus or minus 5 per cent precision of estimate.

ANALYTIC APPROACH

Before going into the investigation of the distribution of N', it was proposed to compare the simulated average values $\bar{N}'$ of N' and the expected theoretical average values of N' for a given value of sample size n and coefficient of variation V. The theoretical average value of N' was calculated as shown below.[11]

σ_x is used as an estimator of σ and

$$\sigma_x^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n} \qquad (3)$$

Hence as in (2), and substituting C for $\left[\frac{A}{\bar{X}}\right]$, we have

$$N' = C^2\left[\frac{\sigma_x}{\bar{X}}\right]^2 \qquad (4)$$

Under the assumption that sampling is done from a normal population, it can be shown that σ_x and $\bar{X}$ are independent random variables. Thus, taking expectations

$$E(N') = C^2\, E\left[\frac{\sigma_x}{\bar{X}}\right]^2$$

$$= C^2\, E(\sigma_x)^2\, E\left[\frac{1}{\bar{X}^2}\right]^2$$

$$= C^2\, \frac{(n-1)}{n}\, \sigma^2\, E\left[\frac{1}{(\bar{X} + \mu - \mu)^2}\right]$$

$$= C^2\, \frac{(n-1)}{n}\, \sigma^2\, E\left[(\bar{X} + \mu - \mu)^{-2}\right]$$

$$= C^2\, \frac{(n-1)}{n}\, \sigma^2\, \frac{1}{\mu^2}\, E\left[1 + \frac{(-2)}{1!}\, \frac{(\bar{X} - \mu)}{} + \frac{6(\bar{X} - \mu)^2}{2!} + \cdots\right]$$

---

11. Chaddha, R. L., "Determination of the Total Sample Size from a Preliminary Small Sample." Unpublished paper, Kansas State University, 1962.

and neglecting higher power terms of $(\frac{\sigma}{\mu})$,

$$E(N') \doteq \frac{(n - 1)}{n} C^2 \frac{\sigma^2}{\mu^2} \left[1 + \frac{3\sigma^2}{n\mu^2}\right] \qquad (5)$$

But from (1), $$N = C^2 \left[\frac{\sigma}{\mu}\right]^2$$

Hence, $$E(N') \doteq N \left[(1 - \frac{1}{n})(1 + \frac{3\sigma^2}{n\mu^2})\right]$$

$$\doteq N\left[1 - \frac{1}{n}(1 - \frac{3\sigma^2}{\mu^2})\right] \qquad (6)$$

Equation (6) indicates that N' is a biased estimator of N for finite n, and on the average, underestimates N. It is, however, an asymptotically unbiased estimator of N. From (6) it could be seen that for a given n, $E(N')$ increases as V increases, and for a given V, $E(N')$ increases as n increases.

Using (6), the expected theoretical values of N' were calculated for the three formulae under consideration, for a sample size of n = 16 and coefficient of variation V = 0.25. Table 2 compares the values of $\overline{N}'$, the average N_i' based on 1000 simulations, with that of the theoretical expected values of $E(N')$. In the Appendix, a larger table (Table 9) is included showing the expected values for the average N' for various values of sample size n (= 2(1)16(2)32) and coefficient of variation V (= 0.15(0.05)0.40). It can be seen from Table 2 that all of the three values of $\overline{N}'$ compare well with $E(N')$, and that $\overline{N}'$ from simulated sampling is less than $E(N')$ in all cases as would be indicated from (6).

Table 2. Average sample size.

Formula	$E(N')$	$\overline{N}'$
N_1'	94.848	92.461
N_2'	91.080	88.825
N_3'	107.570	105.570

A similar procedure was followed for calculating expected standard deviation of N', and these standard deviations were compared with the standard deviations obtained by simulation. In order to evaluate the standard deviation of N', the following formulae[12] for the variances were used:

$$\text{(a)} \quad \mathrm{Var}\left[g(x)\right] \doteq \left[\frac{dg}{dm_1}\right]^2 \mathrm{Var}(x_1), \tag{7}$$

where $g(x_1)$ is the density function of the random variable x_1 with mean m_1.

$$\text{(b)} \quad \mathrm{Var}(X_1/X_2) \doteq \left[\frac{m_1}{m_2}\right]^2 \left[\frac{\mathrm{Var}(X_1)}{m_1^2} + \frac{\mathrm{Var}(X_2)}{m_2^2} - \frac{2\mathrm{Cov}(X_1, X_2)}{m_1 m_2}\right] \tag{8}$$

= variance of the ratio of two random variables.

From (4), we have,

$$\mathrm{Var}(N') = \mathrm{Var}\left[C^2 \frac{\sigma_x}{\overline{X}}\right]^2 \quad \text{or,}$$

$$\mathrm{Var}(N') = C^4 \mathrm{Var}\left[\frac{\sigma_x}{\overline{X}}\right]^2$$

$$= C^4 \mathrm{Var}(V'^2) \tag{9}$$

12. Kendall, M. G., and Stuart, A., Advanced Theory of Statistics. New York: Hafner Publishing Co., 1958.

where V' is the sample coefficient of variation. For a sample from a normal population with coefficient of variation V, it can be shown[13] that

$$\mathrm{Var}(V'^2) \doteq \frac{V^2}{n}(\tfrac{1}{2} + V^2)$$

Hence, using (7), we have

$$\mathrm{Var}(N') = C^4\left(\frac{n-1}{n}\right)^2 \frac{4V^2}{n}(\tfrac{1}{2} + V^2)$$

Therefore,

$$\text{Standard deviation}(N') \doteq \frac{2(n-1)}{n}(CV)^2\left[\frac{1}{2n} + \frac{V^2}{n}\right]^{\frac{1}{2}} \qquad (10)$$

From equation (10) we can say that for a given n, variance(N') increases as V increases and also, for a given V, var(N') decreases as n increases.

Using (10), the expected values for standard deviations of N' for the three formulae under consideration were calculated for a sample size of $n = 16$ and coefficient of variation $V = 0.25$. In Table 3, the values of standard deviation from simulated data and the expected values of standard deviation of N' are shown. In the Appendix, a larger table (Table 10) is included, showing the expected values for standard deviations of N_1' for various values of sample size n (=2(1)16(2)32) and coefficient of variation V (=0.15(0.05)0.40). It can be seen from Table 3 that all the three values of simulated standard deviations compare well with the theoretical standard deviations.

---

13. Chaddha, R. L., op. cit., p. 5.

Table 3. Standard deviations.

Formula	Standard deviation(N')	From Schrader's simulated data
N_1'	35.156	35.000
N_2'	33.764	33.440
N_3'	39.875	39.630

The results tabulated in Tables 2 and 3 indicate that the simulated results provide relatively accurate estimators for the parameters of the distribution of N', except for the slight bias in the case of $\bar{N}'$. To make further analysis of the formulae and to gain an insight into the characteristics of the distribution of N', higher moments of N' were calculated. In this section, to be more general, the assumption that sampling was done from a normal population was not made; however, it was assumed that $\bar{X}$, sample mean has a normal distribution. Under this assumption, first, second, third and fourth moments for N' were derived thus: From this assumption, it can be shown[14] that the sample variance

$$\sigma_x^2 = \left(\frac{\sigma^2}{n-1}\right) \chi^2_{n-1}$$

or denoting $v = n-1$,

$$\sigma_x^2 = \left(\frac{\sigma^2}{v}\right) \chi^2_v \qquad (11)$$

14. Fraser, D. A. S., Statistics: An Introduction. New York: John Wiley and Sons, Inc., 1960.

where χ_v^2 with $v = n-1$ degrees of freedom is defined as

$$f(\chi^2)d\chi^2 = \frac{(\frac{1}{2})^{v/2}(\chi^2)^{(v-2)/2}}{\Gamma(\frac{n-1}{n})} \quad e^{-\frac{1}{2}\chi^2} \, d\,\chi^2, \; 0 \le \chi^2 \tag{12}$$

From the properties of chi-square distribution[15], it can further be shown that, the first four moments, around zero, for σ_x^2 are

$$\mu_1' = \sigma^2 \tag{13}$$

$$\mu_2' = \sigma^4 \frac{(v+2)}{v} \tag{14}$$

$$\mu_3' = \frac{\sigma^6 (v+2)(v+4)}{v^2} \tag{15}$$

$$\mu_4' = \frac{\sigma^8 (v+2)(v+4)(v+6)}{v^3} \tag{16}$$

From (4) $$N' = C^2 \left[\frac{\sigma_x}{\bar{X}}\right]^2 \tag{17}$$

Hence $$E(N') = C^2 E\left[\frac{\sigma_x}{\bar{X}^2}\right]^2$$

$$= C^2 \frac{\sigma^2}{\mu^2} \left[1 + \frac{3\sigma^2}{n\,\mu^2}\right]$$

or $$\mu_1'(N') = E(N') = C^2 \frac{\sigma^2}{\mu^2}\left[1 + \frac{3}{n} V^2\right] \tag{18}$$

and $$\mu_2'(N') = E(N')^2 = C^4 \mu_2'(\sigma_x^2)\, \mu_2'(1/\bar{X}^2)$$

$$= C^4 \sigma^4 \frac{(v+2)}{v} \frac{1}{\mu^4} \left[1 + \frac{10\,\sigma^2}{n\,\mu^2}\right]$$

---

15. Kendall, op. cit., p. 7.

or $$\mu_2'(N') = C^4 \frac{\sigma^4}{\mu^4}\frac{(v+2)}{v}\left[1 + \frac{10}{n}\frac{\sigma^2}{\mu^2}\right] \tag{19}$$

therefore, $\mu_2(N')$, second moment around mean

$$= \mu_2' - (\mu_1')^2$$

or, $$\mu_2(N') = C^4\frac{\sigma^4}{\mu^4}\left[\frac{(v+2)}{v}\left(1 + \frac{10}{n}V^2\right) - \left(1 + \frac{3}{n}V^2\right)\right] \tag{20}$$

In a similar fashion, $\mu_3'(N')$ and $\mu_4'(N')$ were derived,

$$\mu_3'(N') = C^6\frac{\sigma^6}{\mu^6}\frac{(v+2)(v+4)}{v^2}\left[1 + \frac{21}{n}\frac{\sigma^2}{\mu^2}\right] \tag{21}$$

$$\mu_4'(N') = C^8\frac{\sigma^8}{\mu^8}\frac{(v+2)(v+4)(v+6)}{v^3}\left[1 + \frac{36}{n}\frac{\sigma^2}{\mu^2}\right] \tag{22}$$

from which $\mu_3(N')$ and $\mu_4(N')$ were calculated. For the special case of $V = 0.25$, and $n = 16$, the values of μ_2, μ_3, and μ_4 for (N') must be calculated by using the above equations, for all of the three formulae. The moments μ_2, μ_3, and μ_4 are used to give an indication of the type of distribution of N'. The type of Pearson curve to be used is determined[16] by the size of β_1, β_2, and k where

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad\qquad \beta_2 = \frac{\mu_4}{\mu_2^2} \tag{23}$$

and $$k = \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)} \tag{24}$$

---

16. Elderton, W. P., Frequency Curves and Correlation. London, Great Britain: Cambridge University Press, 1938.

if k is negative, this will indicate that type I curve will be used; if k is greater than zero but less than 1, type IV curve will be used, and if k is greater than 1, type VI curve will be used.

For the formulae under consideration, β_1, β_2, and k values were calculated. Due to their similar nature and due to the fact that when calculating β_1, and β_2, the Cs cancel out, and hence As, β_1, β_2, and k values were identical for three formulae:

$$\beta_1 = 0.498304, \qquad \beta_2 = 3.829524$$

$$\text{and} \quad k = 2.56876$$

which shows that Pearson type VI curve might be used to describe the distribution of N', for the special condition when $V = 0.25$ and $n = 16$.

Pearson type VI curve, which has beta distribution is defined as

$$y = y_o(x - a)^{q_2}\, x^{-q_1} \qquad a \le x < \infty \qquad (25)$$

where

q_1, q_2 = constants defined in terms of β_1 and β_2

y_o = a constant dependent upon q_1 and q_2.

Assuming a beta distribution and making use of the simulated data and properties of beta distribution, the following distribution constants were calculated:

$$q_1 = 731.971 \qquad q_2 = 722.066$$

This was found difficult to handle on a computer; hence the whole distribution was scaled down by 10, and the new constants for the curve were:

$$q_1 = 67.794 \qquad q_2 = 58.570$$

Still this presented a problem in using a computer to fit a curve for the distribution of N' the procedure for which is described in the discussion on Curve Fitting.

Hence, it was felt by the writer that a gamma distribution (Pearson type V curve) which bears some similarity to beta distribution should be tried. Gamma distribution is defined as

$$y = C\, x^{a}\, e^{-bx} \qquad x > 0 \qquad (26)$$

where a and b are constants and

$$C = \frac{(b)^{a+1}}{\Gamma(a+1)}$$

CURVE FITTING

Having made a decision about a possible probability distribution of N', it was desired to calculate the theoretical frequencies for various values of N' and compare these with those frequencies obtained by simulation in Dr. Schrader's thesis. The range of the variable N' varied from 0 to 260. Since it is not practical, for goodness of fit purposes, to calculate the frequencies at every integer value of the variable N', the range of the variable was divided into a number of class intervals. However, in the literature only rules of thumb are found as to the choice of the number and lengths of the class intervals. One author suggests grouping of 8 to 22 whereas one paper[17]

17. Mann, H. B., and Wald, A., "On the Choice of the Number of Class Intervals in the Application of the Chi-square Test." Annals of Mathematical Statistics, Vol. XIII, No. 3, September, 1942, pp. 306-317.

suggests a procedure by which the lengths of the class intervals are determined so that the probability of each class under null hypothesis is equal to 1/k where k is the number of class intervals. It was decided to group the variable N' such that class intervals are 0-10, 11-20, 21-30, etc. Under these class intervals the simulated results from Dr. Schrader's thesis were taken and the frequencies for N' were determined for each class interval and for each formula.

To determine the theoretical frequencies for each class interval, the properties of the proposed curve were made use of: for a gamma distribution with parameters (a) and (b), it can be shown that

$$\text{Mean} = \frac{a+1}{b} \tag{27}$$

$$\text{Variance} = \frac{a+1}{b^2} \tag{28}$$

Substituting the values of mean and variance for N_1' in (27) and (28), a and b were calculated. Then the distribution of N_1' can be written as

$$y = \frac{(b)^{a+1}}{\Gamma(a+1)}(N_1')^a \, e^{-bN_1'} \qquad N_1' > 0 \tag{29}$$

This equation must integrate to 1 between limits $N_1' = 0$ and $N_1' = \infty$. Also the total area under this curve must equal 1000, which was the total number of simulations. To calculate the frequency between any class interval, say $N_1' = 121$ to $N_1' = 130$, the following procedure was followed: $(N_1')^a \, e^{-bN_1'}$ was integrated from N_1' = 0 to 300, at which point the curve almost coincides with N_1' axis, to represent the total area (A300) and in turn 1000 observations. Next, the same function was integrated from N_1' = 0 to 130, and this area (130) calculated. Integration is again repeated for limits N_1' = 0 to 121, and this area (A121) also calculated. Then (A130 - A121) represents the area between N_1' = 121 to 130; and (A130 - A121)/(A300) represents the fraction of the area under the class interval N_1' = 121 to 130, and the quantity (A130 - A121)(1000)/(A300) represents the theoretical frequency or total number of observations in the class interval N_1' = 121 and N_1' = 130. By a similar process, theoretical frequencies were calculated for each class interval for N_1'.

In a similar manner, making use of the respective values for mean and variance from simulated data, expected frequencies for the class intervals were calculated for N_2' and N_3'.

Due to the repetitive nature of the calculations, a digital computer, IBM 1620, was used.

RESULTS AND DISCUSSION

In order to test the hypothesis that N' has a gamma distribution, the theoretical frequencies for the class intervals were computed for each of the three formulae, as described in the section on Curve Fitting. Since this test is concerned with the agreement between the distribution of a set of sample values and a theoretical distribution, we call it a test for goodness of fit.

Attempts have been made to find test statistics whose sampling distribution does not depend upon either the explicit form of, or the value of certain parameters in, the distribution of the population. Such tests have been called non-parametric or distribution-free tests. Probably the most widely used of such tests is the chi-square test. However, an alternative distribution-free test for goodness of fit, called Kolmogorov-Smirnov test for goodness of fit, was suggested by Kolmogorov and Smirnov, and some evidence[18,19] was presented indicating that when it is applicable it may be better all-round test than the chi-square test. This test, denoted as d-test, can be explained briefly as follows:

Suppose that a population is thought to have some specified distribution function, say $F_o(x)$. That is, for any specified value of x, the value of $F_o(x)$ is the proportion of individuals in the population having measurements less than or equal to x. The cumulative step-function of a random sample of N observations is expected to be fairly close to this specified distribution function. If it is not close enough, this is evidence that the hypothetical distribution is not the correct one.

If $F_o(x)$ is the population cumulative distribution, and $S_N(x)$ the observed cumulative step-function of a sample N, then the d-test involves the determination of d = maximum $F_o(x)$ -

---

18. Siegel, Sidney, Nonparametric Statistics For the Behavioral Sciences. New York: McGraw-Hill Book Company, Inc., 1956.

19. Massey, Frank J., Jr., "The Kolmogorov-Smirnov Test for Goodness of Fit," Journal of American Statistical Association, Vol. 46, No. 253, March, 1951, pp. 68-78.

$S_N(x)$, calculation of d/N and comparing this value with the tabulated critical value $d_\alpha(N)$ for a desired level of significance, α.

Grouping observations into class intervals tends to lower the value of d. For grouped data, therefore, the appropriate d (N) values and hence significance levels are different than those tabulated. However, for large samples, grouping usually will cause little change in the appropriate significance levels.[20]

As an example of the application of this test of goodness of fit, the procedure followed by the writer is explained: The cumulative frequencies, along with individual frequencies in class intervals, obtained by Dr. Schrader's simulation are recorded in Tables 4, 5, and 6, respectively for the three formulae under consideration. The same tables also include the calculated theoretical frequencies for the respective formulae. Referring to Table 4, the maximum deviation in the absolute frequencies, which occurs at class interval 50, is 30.569 which represents a difference in the proportion of 30.569/1000 = 0.030569. The 5 per cent significant point as taken from standard tables is 1.36/ 1000 = 1.36/31.263 = 0.043. The observed value of d/N is less than the critical value; so we would accept, at the 5 per cent level of significance, the hypothesis that the population distribution was that recorded in Table 4.

By similar procedures, the hypothesis that N_2' and N_3', respectively follow gamma distribution were accepted at 5 per cent

---

20. Massey, op. cit., p. 16.

Table 4. Comparison of observed and theoretical frequencies for N_1'. $N_1' \sim$ Gamma ($a = 5.978$; $b = 0.0755$.)

Upper boundary of class interval	Observed	Theoretical	Cumulative frequency to upper boundary of class: Observed	Cumulative frequency to upper boundary of class: Theoretical	Absolute difference
300	0	0.028	1000	1000.000	0.000
250	0	0.382	1000	999.972	0.028
240	1	0.635	1000	999.590	0.410
230	1	1.042	999	998.955	0.045
220	1	1.689	998	997.913	0.087
210	1	2.702	997	996.224	0.776
200	8	4.262	996	993.522	2.478
190	5	6.616	988	989.260	1.260
180	13	10.091	983	982.644	0.356
170	14	15.092	970	972.553	2.553
160	17	22.079	956	957.461	1.461
150	29	31.503	939	935.382	3.618
140	51	43.683	910	903.879	6.121
130	60	58.610	859	860.196	1.196
120	64	75.672	799	801.586	2.586
110	101	93.364	735	725.914	9.086
100	131	109.088	634	632.550	1.450
90	117	119.261	503	523.462	20.463
80	104	119.997	386	404.201	8.201
70	91	108.542	282	284.204	2.204
60	70	85.231	191	175.662	15.338
50	81	54.975	121	90.431	30.569*
40	29	26.497	40	35.456	4.544
30	11	7.964	11	8.959	2.041
20	0	0.995	0	0.995	0.995
10	0	0.000	0	0.000	0.000

* Maximum absolute difference = 30.569
Hence d/N = 30.569/1000 = 0.030569
Critical value for d/N at 5 per cent significant level =
1.36/ 1000 = 1.36/31.623 = 0.043.

Table 5. Comparison of observed and theoretical frequencies for N_2'. $N_2' \sim$ Gamma ($a = 6.059$; $b = 0.0795$.)

Upper boundary of class interval	Observed	Theoretical	Cumulative frequency to upper boundary of class: Observed	Cumulative frequency to upper boundary of class: Theoretical	Absolute difference
300	0	0.035	1000	1000.000	0.000
250	0	0.227	1000	999.965	0.035
240	0	0.387	1000	999.741	0.259
230	1	0.659	1000	999.354	0.646
220	1	1.108	999	998.695	0.305
210	1	1.839	998	997.587	0.413
200	1	3.007	997	995.748	1.252
190	8	4.837	996	992.741	3.259
180	10	7.645	988	987.904	0.096
170	12	11.843	978	980.259	2.259
160	18	17.940	966	968.416	2.416
150	14	26.497	948	950.476	2.476
140	41	38.017	934	923.979	10.021
130	49	52.753	893	885.962	7.038
120	66	70.404	844	833.209	10.021
110	95	89.732	778	762.805	15.195
100	133	108.223	653	673.073	20.073
90	128	122.013	550	564.850	14.850
80	114	126.456	422	442.837	0.837
70	108	117.646	308	316.381	8.381
60	70	94.825	200	198.735	1.265
50	83	62.608	130	103.910	26.090*
40	34	30.764	47	41.302	5.698
30	13	9.366	13	10.538	2.462
20	0	1.172	0	1.172	1.172
10	0	0.000	0	0.000	0.000

* Maximum absolute value = 26.09
Hence d/N = 0.02609
Critical value for d/N at 5 per cent significant level = 0.043.

Table 6. Comparison of observed and theoretical frequencies for N_3. $N_3' \sim$ Gamma (a = 6.086; b = 0.0672.)

Upper boundary of class interval	Observed	Theoretical	Cumulative frequency to upper boundary of class: Observed	Cumulative frequency to upper boundary of class: Theoretical	Absolute difference
300	2	0.014	1000	1000.000	0.000
250	1	1.432	998	999.986	1.986
240	1	2.178	997	998.554	1.554
230	2	3.273	996	996.376	0.376
220	9	4.859	994	993.103	0.897
210	9	7.119	985	988.244	3.244
200	8	10.278	976	981.125	5.125
190	14	14.602	968	970.847	2.847
180	17	20.377	954	956.245	2.245
170	19	27.876	937	935.868	1.132
160	44	37.287	918	907.992	10.008
150	59	48.623	874	870.705	3.295
140	35	61.590	815	822.082	7.082
130	83	75.444	780	760.492	19.508
120	100	88.873	697	685.048	11.952
110	113	99.972	597	596.175	0.825
100	105	106.399	484	496.203	12.203
90	91	105.840	379	389.804	10.804
80	87	96.769	288	283.964	4.036
70	61	79.406	201	187.195	13.805
60	73	56.442	140	107.789	32.211*
50	46	32.860	67	51.347	15.653
40	14	14.239	21	18.487	2.513
30	7	3.825	7	4.248	2.752
20	0	0.423	0	0.423	0.423
10	0	0.000	0	0.000	0.000

* Maximum absolute difference = 32.211
Hence d/N = 0.032211
Critical value for d/N at 5 per cent significant level = 0.043.

significant level.

Thus there is evidence, based upon the above simulated data for formulae 1, 2, and 3, to indicate and suggest that the distribution of N' follows a gamma distribution.

FURTHER RESULTS

The results in the last section did not reject the hypothesis of gamma distribution for N_1', N_2', and N_3'. To extend further the confirmation of finding the distribution of N', the procedure followed was applied for the two formulae[21] N_9' and N_{10}' suggested by Dr. Schrader in his thesis. Using the Kolmogorov-Smirnov test, the hypothesis that N_9' has a gamma distribution was accepted at 5 per cent level of significance. A similar conclusion also was reached in the case of N_{10}'.

However, when the above results in Tables 4, 5, and 6 were subjected to the chi-square test, the conclusions drawn were mixed. For N_1', the calculated chi-square, after appropriate subgrouping, was 31.557. The degrees of freedom were 17. At 5 per cent level of significance, $\chi^2_{17(0.05)}$ is 27.5871. Therefore, at 5 per cent significant level, we would reject the hypothesis that N_1' follows gamma distribution. But $\chi^2_{17(0.01)} = 33.4087$, which might make us accept the hypothesis at 1 per cent level of

---

21.

$$N_9' = \left[\frac{42.6\sqrt{\frac{n}{n-1}\Sigma X^2 - (\Sigma X)^2}}{\Sigma X}\right]^2$$

$$N_{10}' = \left[\frac{239.064(\bar{R}_8)}{\Sigma X}\right]^2$$

where $\bar{R}_8$ is the average range of a subgroup size of 8.

Note: Both formulae were developed for use with a preliminary sample size n of 16, and were based on 95 per cent confidence on plus or minus 5 per cent precision of estimate.

significance. The large calculated chi-square value was partially due to the fact that at class interval 50, the expected value of frequency was 55 whereas the observed value was 81, this alone contributing a chi-square value of 12.319. The same trend also was noticed in the case of N_2' and N_3'.

However, by applying the chi-square test to N_9' and N_{10}', respectively, it was found that the hypothesis would be accepted at 5 per cent level of significance in both cases. Table 7 gives a summary of results of chi-square tests conducted on N_1', N_2', N_3', N_9', and N_{10}'.

Table 7. Summary of results of chi-square test for goodness of fit for distributions of N_1', N_2', N_3', N_9', and N_{10}'.

Formula	Degrees of freedom	Calculated chi-square	Tabulated chi-square at 5 per cent significant level	Decision about hypothesis
N_1'	17	31.557	27.587	Reject
N_2'	16	29.640	26.296	Reject
N_3'	19	42.412	30.143	Reject
N_9'	21	25.200	32.670	Accept
N_{10}'	21	27.659	32.670	Accept

This procedure of determining the distribution of N' was subjected to another application. The data from Hicks and Young[22], for a method using N_1', were taken and for these data and

22. Hicks, op. cit., p. 3.

class intervals, theoretical frequencies were calculated. These results were subjected to Kolmogorov-Smirnov and chi-square tests. The hypothesis was accepted in both the tests at 5 per cent level of significance. Table 8 shows the comparison of the calculated theoretical and simulated individual frequencies in the class intervals obtained by Hicks and Young. Further, the calculated chi-square value was 7.805 compared to the tabulated chi-square value, with 9 degrees of freedom = 16.919 at 5 per cent level of significance.

Table 8. Comparison of observed and theoretical frequencies for N_1' (Hick and Young data [a = 5.268; b = 0.0661]).

Upper boundary of class interval	Observed	Theoretical	Cumulative frequency to upper boundary of class: Observed	Cumulative frequency to upper boundary of class: Theoretical	Absolute difference
314	0	0.159	500	500.000	0.000
289	1	0.233	500	499.841	0.159
264	0	0.738	499	499.608	0.608
239	1	2.210	499	498.870	0.130
214	9	6.190	498	496.660	1.340
189	11	15.949	489	490.470	1.470
164	36	36.865	478	474.521	3.479
139	73	73.510	442	437.656	4.354
114	121	118.554	369	364.146	4.854
89	145	137.550	248	245.592	2.408
64	84	89.982	103	108.042	5.042*
39	19	18.060	19	18.060	0.940
14	0	0.000	0	0.000	0.000

* Maximum absolute difference = 5.042
Hence d/N = 5.042/500 = 0.010084
Critical value for d/N at 5 per cent significant level
= 1.36/ 500 = 1.36/22.361 = 0.06082.

Further, the simulated results for N_1', N_2', and N_3' were subjected to another procedure for determining the distribution, called the method of moments.[23] This analysis indicated that the distribution of N' might follow a Pearson type IV curve, thereby throwing clouds on justification that N' has a gamma distribution or a Pearson type V curve.

SUMMARY AND CONCLUSIONS

In determining the distribution of N', an analytical approach was used which indicated that a Pearson type VI curve might be fitted. However, due to the difficulty in using the computer for fitting a beta distribution, the gamma distribution (Pearson type V curve) was tried for the distribution of N'. Kolmogorov-Smirnov test results indicated that, at 5 per cent significant level, the hypothesis that N' has a gamma distribution would be accepted. However, when the above hypothesis was subjected to the chi-square test, the conclusions reached were mixed and varied. Further, another procedure for determining the distribution of N', viz., method of moments, indicated that N' might follow a Pearson type IV curve. These facts caused the writer to believe that even though there is evidence to show that N' has a gamma distribution, it cannot be said forcefully. It is therefore suggested that more research coupled with analysis could possibly lead to a stronger conclusion about the distribution of N'.

---

23. Elderton, op. cit., p. 11.

BIBLIOGRAPHY

Books

Barnes, R. M., Motion and Time Study. New York: John Wiley and Sons, Inc., 1958.

Elderton, W. P., Frequency Curves and Correlation. London, Great Britain: Cambridge University Press, 1938.

Fraser, D. A. S., Statistics: An Introduction. New York: John Wiley and Sons, Inc., 1960.

Kendall, M. G., and A. Stuart, Advanced Theory of Statistics. New York: Hafner Publishing Co., 1958.

Mundel, M. W., Motion and Time Study. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1960.

Nadler, G., Motion and Time Study. New York: McGraw-Hill Book Co., Inc., 1955.

Niebel, B. W., Motion and Time Study. Homewood, Illinois: Richard Irwin, Inc., Revised Edition, 1958.

Siegel, Sidney, Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Co., 1956.

Theses, Papers, and Articles

Allderige, J. H., "Statistical Procedures in Stop Watch Work Measurements." The Journal of Industrial Engineering, July-August, 1956, Vol. VII, No. 4, pp. 154-163.

Chaddha, R. L., "Determination of the Total Sample Size from a Preliminary Small Sample." Unpublished paper, Kansas State University, 1962.

Hicks, C. R., and H. H. Young, "A Comparison of Several Methods for Determining the Number of Readings in a Time Study." The Journal of Industrial Engineering, Vol. XII, No. 2, March-April, 1962, pp. 93-96.

Lehrer, R. N., and J. J. Moder, "Mathematical Characteristics of Performance Times." Time and Motion Study, October, 1955, Vol. 4, No. 10, pp. 44-52.

Lifson, K. A., "Number of Observations for a Statistical Average." _Time Study Engineer and Time Study Engineering_, Vol. 6, August, 1951, pp. 247-248.

Mann, H. B., and A. Wald, "On the Choice of the Number of Class Intervals in the Application of the Chi-square Test." _Annals of Mathematical Statistics_, Vol. XIII, No. 3, September, 1942.

Massey, Frank J., Jr., "The Kolmogorov-Smirnov Test for Goodness of Fit." _Journal of American Statistical Association_, Vol. 46, No. 253, March, 1951, pp. 68-78.

Radkins, A. P., "Calculating the Required Number of Time Study Readings Using Moving Ranges." Unpublished paper, Purdue University.

Schrader, G. F., "A Critical Analysis of the Reliability and Relative Superiority of the Various Methods Recommended for Use in Determining the Number of Cycles to Record during a Time Study." Unpublished Doctoral Thesis, University of Illinois, 1960.

APPENDIX

Table 9. Expected values for averages of N_1'.

n	V							
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
2	2.007	8.120	18.607	33.920	54.687	81.720	116.007	158.719
3	2.673	10.773	24.540	44.373	70.833	104.640	146.673	197.973
4	3.005	12.090	27.455	49.440	78.515	115.290	160.505	215.039
5	3.204	12.876	29.188	52.428	83.000	121.420	168.324	224.460
6	3.337	13.400	30.337	54.400	85.937	125.400	173.337	230.399
7	3.432	13.773	31.154	55.797	88.010	128.189	176.819	234.475
8	3.503	14.052	31.765	56.840	89.550	130.252	179.378	237.439
9	3.558	14.269	32.240	57.647	90.740	131.840	181.336	239.691
10	3.602	14.443	32.618	58.291	91.687	133.099	182.882	241.459
11	3.638	14.585	32.928	58.816	92.458	134.122	184.134	242.882
12	3.668	14.703	33.185	59.253	93.098	134.970	185.168	244.053
13	3.694	14.803	33.403	59.622	93.639	135.683	186.037	245.032
14	3.716	14.888	33.589	59.937	94.100	136.293	186.777	245.864
15	3.735	14.963	33.751	60.211	94.500	136.819	187.415	246.579
16	3.751	15.028	33.892	60.450	94.848	137.278	187.970	247.199
18	3.779	15.136	34.127	60.847	95.428	138.039	188.890	248.225
20	3.801	15.222	34.315	61.164	95.890	138.646	189.621	249.036
22	3.819	15.293	34.469	61.424	96.268	139.141	190.216	249.695
24	3.834	15.352	34.597	61.640	96.582	139.552	190.709	250.239
26	3.847	15.402	34.705	61.822	96.847	139.899	191.125	250.698
28	3.858	15.445	34.797	61.978	97.074	140.196	191.480	251.088
30	3.867	15.482	34.878	62.114	97.270	140.452	191.787	251.426
32	3.875	15.514	34.948	62.232	97.442	140.677	192.055	251.719

Table 10. Expected values for standard deviations of N_1'.

n	V							
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
2	2.004	8.079	18.400	33.255	53.033	78.212	109.347	147.060
3	2.182	8.795	20.031	36.203	57.735	85.146	119.042	160.099
4	2.126	8.569	19.516	35.272	56.250	82.956	115.981	155.981
5	2.028	8.175	18.620	33.652	53.665	79.144	110.652	148.815
6	1.929	7.774	17.705	32.000	51.031	75.259	105.220	141.509
7	1.837	7.403	16.860	30.472	48.595	71.667	100.198	134.755
8	1.754	7.069	16.100	29.098	46.403	68.435	95.679	128.678
9	1.680	6.771	15.420	27.869	44.444	65.545	91.639	123.244
10	1.613	6.503	14.812	26.770	42.690	62.959	88.023	118.381
11	1.554	6.263	14.265	25.782	41.115	60.635	84.774	114.012
12	1.500	6.047	13.771	24.890	39.692	58.538	81.842	110.068
13	1.451	5.850	13.324	24.080	38.402	56.634	79.181	106.489
14	1.407	5.671	12.915	23.343	37.225	54.899	76.755	103.226
15	1.366	5.507	12.542	22.667	36.147	53.310	74.532	100.238
16	1.329	5.356	12.197	22.045	35.156	51.847	72.488	97.488
18	1.262	5.087	11.585	20.938	33.391	49.244	68.848	92.593
20	1.204	4.854	11.055	19.980	31.863	46.992	65.699	88.358
22	1.154	4.650	10.591	19.142	30.526	45.019	62.942	84.358
24	1.109	4.470	10.180	18.399	29.342	43.274	60.501	81.367
26	1.069	4.309	9.814	17.737	28.285	41.715	58.322	78.437
28	1.033	4.164	9.484	17.140	27.334	40.313	56.361	75.799
30	1.000	4.033	9.185	16.600	26.473	39.042	54.584	73.410
32	0.971	3.913	8.912	16.108	25.687	37.883	52.965	71.232

DETERMINATION OF THE DISTRIBUTION OF SAMPLE SIZES
REQUIRED IN TIME STUDY

by

JAGADEESAN V. POOLA

B. E., University of Madras, 1960

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1963

Stop watch study is the most commonly used method of measuring work in industry today. The time required to perform the elements of an operation may be expected to vary slightly from cycle to cycle. In work measurement, the determination of an acceptable time value to be assigned to an activity often is based on average task or cycle time obtained from timing several cycles of the activity with a stop watch.

In the past few years, several mathematical formulae have been suggested for determining N, the number of cycles to be observed in a stop watch time study in order to accurately estimate the true average task or cycle time. Usually, a small sample of observations, n, are made and sample mean $\bar{X}$ and sample variance σ_x^2 are calculated; substituting these values in the given formula for determining N', the estimate of N, the total number of observations to be made in order to provide the desired confidence level on the estimate of population mean is determined. All of the formulae are of the form, $N = (A\sigma_x/k\bar{X})^2$, where A is the confidence interval constant and k is an acceptance percentage of population mean (usually 5 per cent).

Recently Dr. Schrader conducted an investigation to compare several methods for determining N' and to test by a process of simulation, their reliability and relative superiority. In his investigation, he raised a question, given sample size n and coefficient of variation V which is the ratio between population standard deviation and population mean, what would be the distribution of N'? The purpose of this investigation was to investigate the distribution of N' and to calculate the parameters for the distribution of N', making use of the simulated data obtained by Dr. Schrader in his investigation.

Before going into the investigation of the distribution of N', the simulated average values $\bar{N}'$ of N' were compared with the expected theoretical average values $E(N')$ for the three formulae considered. In all three cases the values of $\bar{N}'$ compared well with $E(N')$. A similar comparison was made, and a similar conclusion reached in the case of standard deviations also. This indicated that simulated results provide relatively accurate estimators for the parameters of the distribution of N'.

Further analysis was made, and higher moments for N' were calculated. The second, third, and fourth moments were used to give an indication of the type of distribution of N'. For the special case of $n = 16$ and $V = 0.25$, it was found that a beta distribution (Pearson type VI curve) might be used to describe the distribution of N'. However, due to the trouble experienced by the writer in using a computer to fit a beta distribution, it was decided to try a gamma distribution (Pearson type V curve) which bears some similarity to beta distribution.

The results of curve fitting showed that the hypothesis that N' follows a gamma distribution was accepted in all cases, at 5 per cent level of significance, under Kolmogorov-Smirnov goodness of fit test. However, chi-square test rejected the same hypothesis at 5 per cent significant level; but would have accepted at 1 per cent significant level.

Another set of data simulated by Hick and Young was subjected to the above procedure. The results in this case indicated that the hypothesis that N' follows a gamma distribution would be accepted under both the tests at 5 per cent significant level.

Another method suggested by Elderton, and called method of moments for determining the distribution of N', indicated that the distribution of N' might be described by a Pearson type IV curve.

The above facts cause the writer to believe that even though there is evidence to show that N' has a gamma distribution, it cannot be said forcefully.