SAMPLE SIZE REQUIRED FOR ESTIMATING THE VARIANCE WITHIN d UNITS OF THE TRUE VALUE[1]

By Franklin A. Graybill and Terrence L. Connell

Colorado State University

1. Introduction. The problem of estimating the variance (σ^2) of a normal density arises in many experimental situations. J. A. Greenwood and M. M. Sandomire [3] have presented a means of obtaining the sample size required to estimate the variance of a normal density within a given per cent of its true value. An investigator may prefer instead to estimate the variance within a given number of units. This paper will provide a two step sampling procedure to solve that problem.

Assume a preliminary sample of size m; $z_1, z_2, \cdots, z_m$, is taken from a normal density with variance σ^2. The unbiased estimator of the variance s_m^2 is computed by the formula $s_m^2 = (m-1)^{-1}\sum (z_i - \bar{z})^2$, and d and $1-\alpha$ are specified in advance. It is desired to determine n, on the basis of the preliminary sample, such that

$$P[|s_{n+1}^2 - \sigma^2| < d] > 1 - \alpha \tag{1.1}$$

where s_{n+1}^2 is equal to $(1/n)\sum_{i=1}^{n+1}(y_i - \bar{y})^2$ and where $y_1, y_2, \cdots, y_{n+1}$ is a random sample of size $n+1$, from a normal density with variance σ^2.

Table I in Section 3 provides the sample size $n+1$, such that (1.1) is true, for $1-\alpha = .90, .95, .99$; $m = 5, 10, 15, 20, 50, 100, 200, 500, 1000$. The only other known method for solving this problem is given in [1], which requires the use of Tchebycheff's inequality. It can be shown that the method presented in this paper provides a significantly smaller second sample size than does [1]. For some comparisons with [1], see Table III.

2. Solution. Equation (1.1) may be written as

$$P[|s_{n+1}^2 - \sigma^2| < d] = E_n\{P[(1-a) < v < (1+a) \mid n]\}$$
$$= \int_1^\infty g(n) \int_{1-a}^{1+a} f_1(v \mid n)\, dv\, dn$$

where E_n is expectation with respect to n; $a = d/\sigma^2$; $v = s_{n+1}^2/\sigma^2$; $g(\cdot)$ is the density of n, and $f_1(\cdot \mid n)$ is the density of a chi-square variable divided by n, its degrees of freedom. We shall restrict n such that $n \geqq 1$. By definition

$$f_1(v \mid n) = [(n/2)^{(n/2)}/\Gamma(n/2)]v^{(n/2-1)}e^{-(n/2)v}, \qquad 0 < v < \infty$$
$$= 0 \qquad\qquad , \quad -\infty < v \leqq 0.$$

Received 22 July 1963.

[1] This research was supported by the Air Force Office of Scientific Research.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org