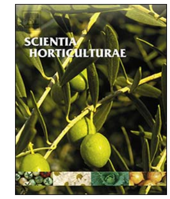
Scientia Horticulturae 225 (2017) 416–421

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti



Sample size determination in horticultural research: An empirical approach

Kouman S. Koumanov

Dept. of Water Regime and Mineral Nutrition, Fruitgrowing Institute, 12 Ostromila Str., Plovdiv 4004, Bulgaria

A R T I C L E I N F O

Keywords:
Study design
Statistics
Normal distribution
Pilot sample
Fruit

A B S T R A C T

Proper sample size determination is important for the representativeness of the research and the validity of the drawn conclusions. It is usually done iteratively because, under normal distribution, the number of trials N has a complex relationship with the mean $\overline{X}$, the standard deviation S of a sample, the chosen confidence probability p, the predefined standard error Δ and the Student's criterion t. The present paper offers a simplified sample size calculation based on the relationship between N and the ratio $(CV/\Delta)^2$, where CV is the variation coefficient of a pilot sample. It was established that the sample size might be calculated with a highly satisfying precision from the formula: $N = a(CV/\Delta)^2 + b$, the parameters a and b being dependent on p. Furthermore, it was concluded that in all experiments under normal distribution, irrespective of the studied subject, the size of the pilot sample might be evaluated through the nonlinear nonparametric equation: $n = 3.52(98.816 - y)/y$,where y is a predefined value of CV of CV of the series of measurements.

1. Introduction

Sample size is important for the representativeness of the research and the validity of the drawn conclusions. The sample has to be large enough in order to characterize the general population in a meaningful way. On the other hand, the increase in size is associated with an increased labour and time consumption and raises the cost of the investigation. The problem also has ethical aspects when related to testing of aggressive treatments and drugs on humans or animals.

Sample size is of fundamental significance for the horticultural research as well. In many cases the impact of applied treatments is estimated based on series of measurements or on sample measurements yielding numbers, i.e. estimates of various plant characteristics. The measurement units can be mass (yield; fruit/stone/seed mass), length (trunk circumference/diameter; shoot length; fruit diameter/height), area (cross section area), volume (tree crown), temperature (canopy; fruit), pressure (turgor; leaf/stem/soil water potential), concentration (water; mineral nutrients; sugars; acids), etc. The sample size may represent the number of replications in a treatment, of plants in a replication, of fruit/leaves/shoots in biometrical/biochemical estimations, of sensor or insect-trap density, etc.

Therefore, the interest in the methodology of sample size determination never languishes and there are numerous publications on the matter (e.g. Chadha, 2006; Chow et al., 2008; Cochran, 1977; Eyduran et al., 2006; Jones et al., 1957; Machin et al., 2009; Pérez et al., 2003; Van Belle., 2008; Vitanov and Vitanova, 1983). The methods of sample size determination are based on the principles of statistics, so they are applied regardless of the studied object and the scientific field.

It is generally accepted that with samples of the size that is common in practice, the sample estimates are approximately normally distributed (Cochran, 1977). Under this assumption, the size of a sample N, in case of a predefined relative standard error $\Delta_{p,k}$ with probability p, is estimated using the following equation (Mitkov and Minkov, 1989):

$$N = \frac{t_{p,k}^2 S^2 100^2}{\Delta_{p,k}^2 \overline{X}^2}; \tag{1}$$

where $\overline{X}$ is the sample mean; $t_{p;k}$ is determined from the Student's t-distribution for a confidence probability p and degrees of freedom $k = N - 1$; and S is the standard deviation. The equation is valid only if the sample data were drawn from a normally distributed population.

The estimation of the sample size using the above equation is usually done by iterations, because the way in which $t_{p;k}$ depends on the number of measurements N through the degrees of freedom k is too complicated. Moreover, $\overline{X}$ and S are unknown and one has to rely on his experience, to use historical data or to conduct a pilot study in order to obtain reliable values of these statistics (Israel, 1992). On the above stipulations, sample size can be determined using tables (Machin, 2009), nomograms (Carley et al., 2005; Mitkov and Minkov, 1989), graphs (Van Belle, 2008) or commercial sample-size determination software (Borenstein et al., 1997). Still, the procedure is conditional and, consequently, the results may be debaTable So there is the temptation to take some shortcuts (Lenth, 2001). In this context, the present article offers a simple empirical approach to successful and meaningful calculation of the sample size.

E-mail addresses: kskoumanov@hotmail.com, kskoumanov@abv.bg.

http://dx.doi.org/10.1016/j.scienta.2017.07.027
Received 30 March 2017; Received in revised form 14 July 2017; Accepted 20 July 2017
Available online 29 July 2017

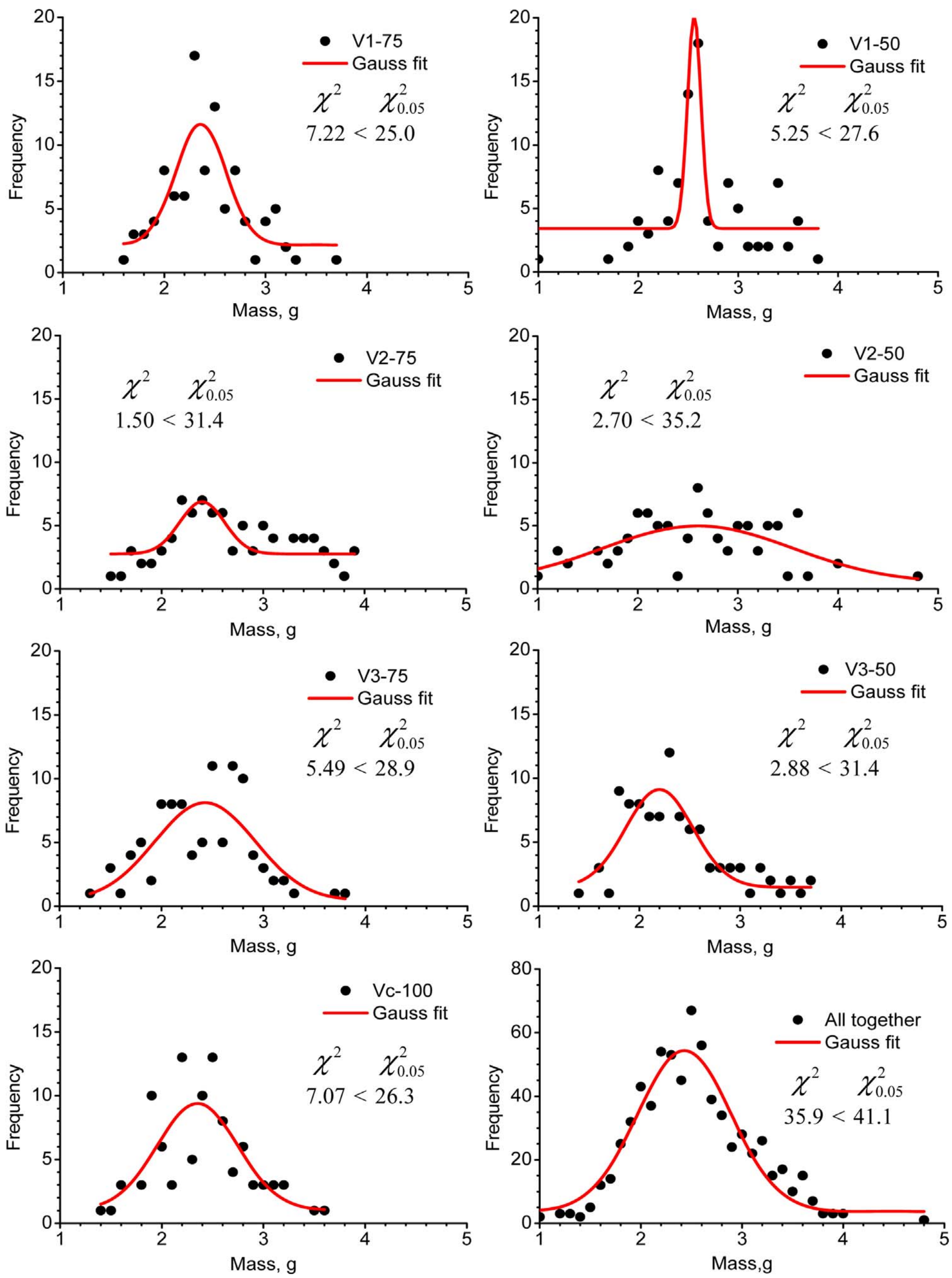


Fig. 1. Frequency distributions of the raspberry fruit mass under different irrigation regimes and their conformity with the normal distribution at 0.05 confidence level.

2. Material and methods

Taking in account that:

$$\frac{S^2}{\overline{X}^2} = CV^2; \tag{2}$$

Eq. (1) can be rewritten in the following form:

$$N = t^2\left(\frac{CV}{\Delta}\right)^2; \tag{3}$$

where CV is expressed as percentage. For convenience, in Eq. (3) and hereafter $\Delta_{p,k}$ is written only as Δ; it is expressed as percentage, too.

It becomes clear that the sample size depends in some way on $(CV/\Delta)^2$. This relationship was subjected to correlation and regression analyses using data of the mass of a single fruit from samples of raspberry, cherry and apple fruits. Regulated deficit irrigation and cultivars with small, medium and large size of the fruits were included in the study in order to account for eventual impact on the sample size of both stress factors and cultivar specificity. With the same intention, towards extending the validity of the results, length characteristics such as the diameter and the height of the apple fruits were also included in these analyses. All fruit samples were collected from treatments in ongoing experiments with purposes and subjects of investigation differing from the present study ones. The experimental setups are practically irrelevant to the current analysis and, therefore, their detailed consideration is not in the scope of the present paper.

Raspberry fruits were collected in the course of an experiment on regulated deficit drip irrigation with the ‘Lyulin’ primocane-fruiting cultivar. Seven samples of 100 raspberry fruits each were randomly collected from seven different variants of drip irrigation, respectively recovering 100%, 75% and 50% of the crop evapotranspiration during the main phenological phases: (1) intensive growth, (2) blossom, and (3) fruit ripening. Additionally, an integral sample of 700 fruits was analyzed, too.

Three samples of 400 cherry fruits each were formed from the ‘Hudson’, ‘Nalina’ and ‘Summit’ cultivars, with an average mass of one