THE RELATIONSHIP BETWEEN REGRESSION MODEL-SAMPLE SIZE AT THE ESTIMATION OF STAND VOLUME INCREMENT

Serdar Carus, Ersel Yilmaz

Faculty of Forestry University of Istanbul Department of Forest Yield and Biometry Istanbul/ TURKIYE

ABSTRACT

In this study, sample size had been dwelt upon by us that effective on closeness of relative values by founding with statistical methods for obtaining of increment value at the single tree. In this report, were used periodical measurements which belongs to artificial stand that has four provenance Cluster pine (*Pinus maritima* Ait.) in The Belgrade Forest Burunsuz region was planted. Linear models were formed by sample groups (8,16,24 and 30) which have different number of samples for obtaining relation of diameter at breast height - diameter increment. For linear regression estimation, a supplementary variable x_i, correlated with the primary variable y_i, is obtained from each unit in the sample. Frequency histograms of the sample means were constructed. Results from suggest that samples as small as 24 units can often be select. In young stand can use 30 units for reduce of standard error at the prediction.

Key words: sample size, regression, ratio of means, stand volume increment, Cluster pine.

1. INTRODUCTION

First Barrett and Goldsmith (1976) after Gove et al. (1982), examined the confidence intervals to determine how large n must be to generate approximately normal distributions of sample means when using simple random sampling. This research extends their investigation to ratio and linear regression estimation based on sample random sampling. For both ratio and linear regression estimation, a supplementary variable x , correlated with the primary variable y , is obtained from each unit in the sample. The population means m of the x must be known. The following formulae for ratio and regression estimation were adapted from Cochran (1977).

The ratio estimator and confidence interval are:

$$Y_R = (y/x)\,\mu_x$$

$$Y_R \pm t^*\sqrt{v(Y_R)} \qquad (1)$$

Where Y_R is the ratio estimator of the population mean, y and x are the sample means of the y_i and x_i, μ_x is the population mean of the x_i, t is Student's t- value and $v(Y_R)$ is the sample variance of the ratio estimator.

The linear regression estimator and confidence interval are:

$$y_r = y + b(m_x - x)$$

$$y_r \pm t^*\sqrt{v(y_r)} \qquad (2)$$

Where y_r is the linear regression estimator of the population mean, b is the sample estimator of the slope coefficient, $v(y_r)$ is the variance of the linear regression estimator, and the others are as defined for ratio estimation. The ratio estimator is unbiased when the relationship between y_i and x_i is a line through the origin; the linear regression estimator is unbiased when the relationship between x_i and y_i is linear. The ratio estimator is more efficient than the linear regression estimator is when the variance of y_i is proportional to x_i; the linear regression estimator is more efficient when the variance of y_i is the same for all levels of x (Cochran 1977). Actual populations often fail to fulfill exactly the assumptions for these two types of estimators. If the sample size is

sufficiently large, however, satisfactory estimates can be obtained even if the conditions are only approximately met. Not only must the sample size be large enough so that the bias by ratio or regression estimation is negligible, but it must be large enough so that the sampling means on repeated sampling closely approximate a normal distribution. Generally a sample size of 30 or more units is recommended.

In this study, for selected population are used to examine the suitability of ratio and linear regression estimators. Linear models were formed by sample groups (8-16-24-30) which have different number of samples for obtaining relation of diameter at breast height- diameter increment and model outputs were compared with real values.

2. MATERIAL AND METHOD

2.1 Material

Short description follows of the population examined to obtain a better idea about the sample size required. Sample plots are located Burunsuz region of Belgrade forest in Istanbul. Plantation at 1951 and 1953 years produced four provenance of the Cluster pine (Ispanya, Lambert, Gironde and Korsika) in Burunsuz region. Throughout 1965 and 1967 years, fixed application (permanent plots) areas were formatted in this plot, that planted provenance were 14-15 years old. Purpose of this study, is research wood productivity of cluster pines. Measurements of three provenance areas consist of four parcels, which are 25x25m, but Gironde's measurements consist of three parcels, which are 16x25m. Trees that present each plot of land has 15 pieces of parcels were numbered to start from one. Also four *Pinus maritima* Ait. provenance diameter at breast height of numbered trees in plot of land measured perpendicular two directions to each other. Thus obtained values were recorded.

Each parcel were accepted such as a different stand. Subsequently, random draw were made 11-18 times for 8,16,24 and 30 piece of samples in each parcel and this, mean diameter at breast height (x_i) and diameter increment (y_i) were obtained. Note that the diameter – diameter increment data tend to approximate the requirements for a ratio estimator. The relationship between diameter at breast height-diameter increment data is linear.

2.2 Method

In this study, used periodic measurements. These periodic measurements belong to 1967-1975-1981 and 1995. These measurements are shown with SP67 (1967 measurements of sample plots), SP75,SP81 and SP95. Periodical diameter increments divided by period time and were found annular diameter increments. Sample sizes for periodical measurements are taken SP:133, SP81:131 and SP95: 119.

As emphasize, in the material section, in this study, for various sample sizes and random sampling technique were made. In the selection of 119- 133 samples of a specified size. This process, were made the following quantities were computed.

1- the mean of the sample means,

2- the percentage of times the confidence interval missed the population means on the low and on the high side,

In addition, frequency histograms of the sample means were constructed. The distribution of means in all the histograms seems to approximate a normal distribution.

3. FINDINGS AND DISCUSSION

3.1 Bias

In this study, bias is defined as follows;

$$\text{Bias} = E(y) - \mu \qquad (3)$$

where $E(y)$ is the estimated expected mean of the population based on an average of the sample means, and μ is the true population mean. Estimates of the population mean for each population are shown in Table 1. Even for a sample size of eight (SP75), the means for ratio estimation 8.50 and the population mean of 8.663.

Table 1: Ratio and regression estimation of the population means of 8.663, 4.135 and 3.190 for the diameter at breast height-diameter increment population.

	Sample size 8		Sample size 16		Sample size 24		Sample size 30	
	Ratio	Regression	Ratio	Regression	Ratio	Regression	Ratio	Regression
SP75 (8.663)*	8.50	8.480	8.62	8.604	8.61	8.621	8.56	8.570
SP81 (4.135)	4.25	4.103	4.31	4.185	4.28	4.162	4.24	4.112
SP95 (3.190)	3.28	3.244	3.23	3.212	3.22	3.198	3.19	3.178

* In the parenthesis were given the mean of population means.

3.2 Percentage of times the confidence interval missed the population means

A value from Student's t- distribution was used to compute a 95% confidence interval that is, we should fail to cover the population mean only 5% of the time, provided the estimates are normally distributed and the bias of the estimates is negligible. The percentage of times the confidence interval missed the population mean is shown in table 2. As an example, ratio estimates of the population based on a sample (SP81) of 24 resulted in the confidence interval missing the population mean 5.34% of the time. Generally, confidence intervals by ratio estimation failed to include the population mean less frequently than those by regression estimation.

Table 2: Percentage of times the confidence interval missed the population mean. Each value in the table is based on 119-133 samples, and given in % (ratio and regression).

Ratio	Sample size 8			Sample size 16			Sample size 24			Sample size 30		
	Low	High	Total	Low	High	Total	Low	High	Total	Low	High	Total
SP75	2.26	3.01	5.27	0.0	3.01	3.01	2.26	0.0	2.26	0.75	0.0	0.75
SP81	0.0	4.58	4.58	0.0	5.34	5.34	0.0	5.34	5.34	0.0	6.11	6.11
SP95	0.84	3.36	4.20	1.68	3.36	5.04	0.84	1.68	2.52	0.84	1.68	2.52
Regressi on	Sample size 8			Sample size 16			Sample size 24			Sample size 30		
	Low	High	Total	Low	High	Total	Low	High	Total	Low	High	Total
SP75	2.26	3.01	5.27	3.01	0.75	3.76	2.26	1.50	3.76	1.50	0.75	2.25
SP81	0.0	4.58	4.58	0.0	5.34	5.34	0.0	5.34	5.34	0.0	6.87	6.87
SP95	0.0	3.36	3.36	1.68	3.36	5.04	0.84	2.52	3.36	0.84	1.68	2.52

The mean confidence intervals indicate that ratio estimation is slightly more precise than regression estimation for the population; on the other hand, regression estimation proved to be more precise for diameter at breast height increment.

The mean confidence interval width for those that covered the population mean is smaller than for those that are above.

Analysis regression was made by sample data that is valid for population. Number of sample units must take enough large far the small standard errors. Because structure of population reflects to the sample and sample can represent the population well.

4. RESULT AND PURPOSALS

Table 3 shows the relationship of bias to the percentage of times the confidence interval catches the population mean at $\alpha = 0.05$ level assuming the estimates follow a normal distribution (BARRETT and NUTT 1979).

Table 3: The relationship of bias to the percentage of times the confidence interval catches the population mean.

Percentage Bias	Percentage Catch
2	95.0
10	94.9
20	94.5
40	93.2
60	90.8
80	87.4
100	83.0
150	67.7

Here, bias is expressed as a percentage of the population standard error. The table indicates that, as bias exceeds 20%, the percentage catch drops well below 95% and bias may no longer be considered negligible. For a population, which fits the constraints of a ratio estimator such as the population, a sample size of 24 units is sufficient. The confidence interval catches the population mean about 94% of the time.

Results from this study suggest that ratio and regression estimation can be used for sample sizes as small as 24 units, provided the plotted data only approximately fulfill the conditions for either method of estimation. But in young stands for reduce of standard error in prediction can use 30 units.

REFERENCES

1. Barrett, J.P.- Goldsmith,L. (1976): When is n sufficiently large ? Am. Stat. Vol. 30, pp.67-70.
2. Barrett, J.P.–Nutt,M.E. (1979): Survey Sampling in the Environmental Sciences: A Computer Approach. New Hampshire: COM Press, Inc.
3. Cochran,W.(1977): Sampling Techniques. New York: John Wiley and Sons.
4. Gove,H.J.-Barrett,J.P.-Gregorie,G.T. (1982): When is n Sufficiently Large for Regression Estimation?. Journal of Environmental Management Vol.15, pp.229-237.