Agri, Res. J. Kerala, 1982, 20 (1) 1—4

DISPERSION – A NEW APPROACH

E. J. Thomas
College of Agriculture,
Vellayani-695 522, Kerala.

Dispersion of a set of observations $x_1, x_2, x_3, \ldots\ldots x_n$ is measured by variance, $s^2 = (1/n-1)\sum_{i=1}^{n} (x_i - \bar{x})^2$

In this paper it is shown that s^2 can be expressed in terms of the sum of squares of differences between all possible pairs of observations. The implications of this new definition are studied in a number of situations.

Results

The expression for variance $s^2 = (1/n-1)\sum_{i=1}^{n} (x_i - x)^2$ can be simplified to the form,

$$s^2 = \left\{ \frac{1}{n(n-1)} \right\} \sum_{i>j}\sum (x_i - x_j)^2$$

Proof

$$s^2 = \frac{1}{(n-1)} \sum_i (x_i - \bar{x})^2$$

$$= \frac{1}{(n-1)} \left\{ \sum_i x_i^2 \quad \sum_i x_i^2 - \sum_i \sum_j x_i x_j \right\}$$

$$= \frac{1}{n(n-1)} \sum_{i>j}\sum (x_i - x_j)^2$$

Thus s^2 is $\frac{1}{n(n-1)}$ times the sum of squares between pairs of observations x_i, x_j such that i>j. Further, this can be partitioned into $(n-1)$ components $d_1^2, d_2^2, \ldots\ldots d_{n-1}^2$, where $d_j^2 = \sum_{i=1}^{n-j} (x_{i+j} - x_i)^2$, is the. sum of squares of differences between pairs of observations which are j units apart. Thus d_1^2 is the sum of squares $\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$. which on division by $(n-1)$ gives the mean square successive difference D_1^2. Thus $D_1^2 = \frac{d_1^2}{(n-1)}$ Similarly $D_2^2, D_3^2 \ldots . D_{n-1}^2$ can be defined as the mean square differences of orders

2, 3, ... (n—1), respectively, where

$$D_2^2 = \frac{d_2^2}{(n-2)}, \quad D_3^2 = \frac{d_3^2}{(n-3)}, \; \ldots\ldots\ldots \; D^2_{n-1} = d^2_{n-1}$$

These are different from the mean square successive j—th difference δ_j^2 where

$$\delta_1^2 = \frac{1}{(n-1)} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 \cdot D_1^2$$

$$\delta_2^2 = \frac{1}{(n-2)} \sum_{i=1}^{n-2} (x_{i+2} - 2x_{i+1} + x_i)^2$$

$$\delta_3^2 = \frac{1}{(n-3)} \sum_{i=1}^{n-3} (x_{i+3} - 3x_{i+2} + 3x_{i+1} - x_i)^2$$

$$\delta_4^2 = \frac{1}{(n-4)} \sum_{i=1}^{n-4} (x_{i+4} - 4x_{i+3} + 6x_{i+2} - 4x_{i+1} + x_i)^2$$

The measure of dispersion, s^2 can be represented in terms of D_i^2 as shown below:

$$s^2 = \frac{1}{n(n-1)} (d_1^2 + d_2^2 + \ldots\ldots + d^2_{n-1})$$

$$= \frac{1}{n(n-1)} \left[(n-1) D_1^2 + (n-2) D_2^2 + \ldots\ldots + D_{n-1}^2 \right]$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n-1} (n-i) \; D_i^2$$

Hence s^2 is $(1/n)$ times the simple average of d_i^2 values or is half of the weighted average of the D_i^2 values.

These lead us to a number of useful results which are stated below :

(a) The sum of squares of n observations $x_1, x_2, x_3, \ldots x_n$ namely $\sum (x_i - \bar{x})^2$ can be expressed as

(ii) $$\sum_{i>j}\sum (x_i - x_j)^2 = \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 + \sum_{i+1}^{n-2} (x_{i+2} - x_i)^2 + \ldots\ldots + (x_n - x_1)^2$$

$$= d_1^2 + = d_2^2 \ldots\ldots + d^2_{n-1}$$

(b) The sum of squares between v samplas of sizes $r_1, r_2, \ldots, r_v$ having sample totals $T_1, T_2, \ldots, T_v$ is given by

$$\left(\frac{1}{N}\right) \sum\sum_{i>j} (r_i T_j - r_j T_i)^2 / r_i r_j$$ where N is the sum $r_1 + r_2 + \ldots + r_v$

Obviously when sample sizes are equal [to r, this simplifies to

$$\left(\frac{1}{vr}\right) \sum\sum_{i>j} (T_i - T_j)^2$$ This shows that the SS between samples is derived as the sum of squares of differences between all possible pairs of treatment totals adjusted by dividing by the total number of plots. The contribution of each pair can be easily seen and in fact is indicative of the distances between samples in the one-dimensional space.

(c) The formula for computation of variance in frequency tables can be expressed as $$\frac{1}{N(N-1)} \sum\sum_{i>j} (f_i . f_j) (x_i - x_j)^2$$

Variance is defined as $$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} f_i (x_i - \bar{x})^2$$

This can be brought to the form $$s^2 = \frac{1}{N(N-1)} \sum\sum_{i>j} f_i . f_j (x_i - x_j)^2$$

When the classes are of uniform width c, variance can be computed by using the formula,

$$s^2 = \frac{c^2}{N(N-1)} \sum\sum_{i>j} f_i . f_j (i-j)^2$$

(d) Similar to the simplified form of s^2, the covariance between x and y or the product moment can be put in the form,

$$P_{xy} = \left[\frac{1}{n-1}\right] \sum\sum_{i=j} (x_i - x_j)(y_i - y_j)$$

(e) In the analysis of variance of data relating to v samples of sizes r each, the SS for total is $$\left(\frac{1}{\ }\right) \sum\sum_{i=j} (y_{ij} - y_{i'j'})^2$$

and SS between samples is $$\frac{1}{rv} \sum\sum_{i>j} (y_{.j'})^2 . j>j .$$

(f) In randomised block experiments, block SS will be,

$$\frac{1}{rv} \sum\sum_{i>j} (y_{i.} - y_{i'.})^2$$

(g) The new method of defining the **SS** makes the method of estimation of missing plot yields simple.

Consider a randomised block experiment with v treatments and r replications. Where y_{ij} is missing the method of estimation of y_{ij} is by minimising

error SS $\frac{1}{rv}\left[\sum_{i'}\sum_{j'} (y_{ij} - y_{i'j'})^2 - \sum_{i'} (y_i - y_{i'})^2 - \sum_{j'} (y_j - y_{j'})^2 \right]$

This leads to the equation,

$$\sum_{i'j'} (y_{ij} - y_{i'j'}) - \sum_{i'} (y_i - y_{i'}) + \sum_{j'} (y_j - y_{j'})$$

$$(rv-1)y_{ij} - G' = (r-1)y_i - (G' - y_i) + (v-1)y_j - (G' - y_j)$$
$$= r.y_i - G' + v.y_j - G'$$

Where G' = total of y's except from the missing plot.

Thus the estimate of y_{ij} is $\frac{r.y_i + v.y_j - G'}{(r-1)(v-1)}$

Summary

The variance s^2 can be expressed in terms of the sum of squares of differences between all possible pairs of observations. This new approach to the definition of s^2 is useful in understanding many concepts in statistics and also provides simpler computational formulae. Implications of this new definition are studied in a number of situations.

സംഗ്രഹം

ഒരു നിരീക്ഷണ സമുച്ചയത്തിന്റെ വ്യതിയാനം നിർദ്ദിഷ്ട സമുച്ചയത്തിലെ അംഗങ്ങളായ നിരീക്ഷണങ്ങളുടെ സാധ്യമായ എല്ലാ ജോഡികളും അന്യോന്യം പ്രകടിപ്പിക്കുന്ന വ്യത്യാസങ്ങളുടെ വർഗ്ഗയോഗത്തിന്റെ ഒരു ഫലനമായി വ്യംജിപ്പിക്കാവുന്നതാണ്. ഈ പുതിയ ആശയം സാംഖികത്തിലെ പല സംഹിതകളും മനസ്സിലാക്കുന്നതിനും കൂടുതൽ ലഘുവായ പരികമന സൂത്രങ്ങളും ആവിഷ്ക്കരിക്കുന്നതിനും സഹായിക്കുന്നു. ഈ പുതിയ നിർവചനത്തിന്റെ പ്രതികരണങ്ങളും പല സാഹചര്യങ്ങളിലും പഠനവിധേയമാക്കുകയുണ്ടായി.

[M. S. received on 3-11-1380]