

NOTE ON THE USE OF COEFFICIENT OF VARIATION FOR DATA FROM AGRICULTURAL FACTORIAL EXPERIMENTS

M. KOZAK¹, J. BOCIANOWSKI^{2*} and W. RYBINSKI³

¹ *Warsaw University of Life Sciences, Department of Botany, 02-766 Warsaw, Poland*

² *Poznań University of Life Sciences, Department of Mathematical and Statistical Methods, 60-637 Poznan, Poland*

³ *Polish Academy of Science, Institute of Plant Genetics, 60-479 Poznan, Poland*

Abstract

KOZAK, M., J. BOCIANOWSKI and W. RYBINSKI, 2013. Note on the use of coefficient of variation for data from agricultural factorial experiments. *Bulg. J. Agric. Sci.*, 19: 644-646

Coefficient of variation is very often used to describe variability in a variable. However, when the variable is studied in a factorial experiment, it can be used to determine pooled coefficient of variation, based on the variance of the residual term from an analysis of variance model. Such a coefficient disregards the variability that comes from the sources of variation from the experiment. Because this issue is not mentioned in any of textbooks we know, this paper aims to underline such a use of coefficient of variation, and presents it for a two-year plant breeding experiment on grass pea.

Key words: analysis of variance, estimation, factorial experiments, variability

Introduction

Coefficient of variation (CV) is a measure commonly applied to present variation in agricultural traits. Its merits are well known, most important being one that CV deals with what we could call the scale-invariant variability in the traits; it is easier to understand than variance it is based upon. As Sokal and Rohlf (1995) point out, coefficient of variation can be used to compare the variation of a trait in two (or more) populations or, more commonly, the variation of different traits in a population of study. Let us recall, however, that one should be aware that not always CV can be freely applied; as Webster (2001) explains, this measure must be restricted to the variables that are measured on scales with absolute zero.

Despite the drawback mentioned, CV is a very useful measure. Especially important is that for students its idea is quite easy to grasp and the coefficient is easier to interpret than variance and standard deviation. The problem with interpreting variance in the teaching statistics context has been recently raised (Rumsey, 2009, 2010 and Granaas, 2010), and the call for interpretative coefficients of Rumsey (2009) is more than valid. In our opinion, CV is one such interpretative measure, and from our experience it follows that students

and researchers prefer interpreting variability in continuous variables by means of CV to any other coefficient.

It is worth noting that CV reflects the general variation of a trait, so for structured populations (as is the case with factorial experiments) it also takes account of the variation originating from this structure, in this way overestimating the trait's actual variation that does not come from the additional (incorporated by a researcher) sources of variation.

When data originate from factorial experiments, there are three possibilities to use CV. First, one can simply ignore the factorial structure of data and estimate the CV for the trait. This is not the best idea, however, because the variation in the variable will come from both the sources of variation studied in the experiment, and random variation within the trait. We assume here that this random variation is similar for all treatments, which is a standard assumption of analysis of variance. Second, one can estimate CV for each factor(s) level. Such estimation will be little precise under small samples from factor levels. However, because the within-group variation is assumed the same, we can use this assumption to estimate CV for the whole experiment, which will be based upon the variation in the residual term of the linear model appropriate for the experimental design considered.

* *Corresponding author:* jboc@up.poznan.pl, nyggus@gmail.com, wryb@igr.poznan.pl

The third approach is the best option if the assumption of homogeneous within-group variances holds. As far as we know, it is used in plant breeding companies for data coming from breeding experiments. Nevertheless, we have not been able to find any mention of such a use of CV in statistics textbooks even though clearly it can be very efficient and useful. Hence, in this note we wish to point the readers' attention to this use of CV. We will also present the usefulness of taking advantage of the residual term while estimating within-group CV for a two-year plant breeding experiment with grass pea.

Coefficient of variation

Let us consider a one-way analysis of variance model from a completely randomized experiment. Let S^2 be the estimated residual variance for the experiment; it does not matter whether the factor is assumed fixed or random, since in either case S^2 describes the same quantity. Webster (2007) has recently given a thorough explanation of this measure of variability in a trait. Representing trait variability within treatments, it describes the variance of the trait disregarding the variability incorporated by the treatments. For this reason, S^2 can be thought of as an intrinsic measure of the variability in the traits in the experiment, describing sampling and measurement fluctuation as well as variation over the site (Webster, 2007). Because analysis of variance assumes that the variance of the study trait be the same within each treatment, S^2 is a measure of variability in the whole experiment.

Coming back to CV, let us recall that it disregards the variability incorporated by the experiment treatments. Knowing what we said in the previous paragraph, however, we can easily estimate coefficient of variation for the trait in the experiment (experiment CV, CV_{exp} , say) based on S^2 , using the following formula:

$$CV_{\text{exp}} = \frac{S}{\bar{X}}, \quad (1)$$

where \bar{X} is the trait grand mean estimated from the experiment, and $S = \sqrt{S^2}$. Note that S^2 will be smaller than the classical variance estimator of the trait if only the treatments differentiate the trait (by the classical variance we understand the estimator that disregards the factorial structure of the experimental data), a fact that in turn makes the variability incorporated by treatments be included within the value of the trait variance.

It easily follows that CV_{exp} is a useful measure of trait variability in the experiment, providing information on different variability than the classical CV. It can be easily applied for experiments with more factors from fixed, mixed and linear models.

Example

To present the application of the proposed coefficient, we used the experimental data studied and described by Kozak et al. (2008); the reader is referred to that article for details of the experiment. In summary, the experiment was arranged in a randomized complete design with three replications and two factors, namely genotype (2 cultivars of grasspea (*L. sativus* L.), Derek and Krab, along with 17 mutants derived from the former and 3 from the latter) and year (2002 and 2003). Eight traits are of interest, namely number of branches per plant (BP); height of the lowest pod (HLP); pod length (PL); number of pods per plant (PP); plant height (PH); number of seeds per plant (SP); 1000-seed weight (TSW); and weight of seeds per plant (WSP). See Kozak et al. (2008) for trait means with the corresponding standard errors as well as the relationships among the traits.

We conducted three separate analyses. First of them referred to a full ANOVA model with main effects of year and genotype, interaction between them as well as block nested within years, while second and third to year-wise models (for 2002 and 2003) with genotype and block effects. Based on the first analysis we will compare variation of the traits in the whole two-year experiment, while based on the second and third ones we will compare (a) each trait's variability in the two years of experiment and (b) the traits' variability in each year.

The classical CV and CV_{exp} for these three analyses are given in Table 1. First of all, for the pooled analysis (for both years) the large differences in the two measures are noticeable, indicating the large contribution of sources of variation within the traits variability; this was especially visible for PH and SP. Interestingly, WSP, the most interesting trait in the experiment, had the largest CV_{exp} , a result that should be thought of as unfavourable as indicating that this trait is much variable over the site as well as replications. On the other pole is PL, CV_{exp} of which was the smallest among the traits.

Quite similar conclusions can be drawn from the separate analyses for 2002 and 2003, although in the latter TSW had the highest (17.6%) and PH the lowest (0.6%) CV_{exp} . Worth noting is that in 2003 BP's CV was just slightly greater than CV_{exp} , showing that the genotypes did not contribute too much to variability in this trait.

Comparison of the two years shows that the performance of some of the traits was different in the years, indicating an influence of environment (year) on the variability of the trait. One good example is TSW, for which the classical CV was similar in both years but CV_{exp} was rather small in 2002 (4.1%) but high in 2003 (17.6%). A similar result was obtained for BP.

That the differences between CV and CV_{exp} for the pooled analysis were greater than those for the year-wise analyses for most of the traits could be explained by the fact that year

Table 1
Classical (CV) and experiment-wise (CV_{exp}) coefficient of variation (%) for the grasspea traits (both given in %)

| | PH | HLP | BP | PP | PL | SP | TSW | WSP |
|-----------------|------|------|------|------|-----|------|------|------|
| Pooled analysis | | | | | | | | |
| CV | 45.2 | 27.8 | 27.3 | 37.5 | 8 | 39.3 | 34.7 | 20.4 |
| CV_{exp} | 3.7 | 4.9 | 8.7 | 3.5 | 1.4 | 6.7 | 10 | 11.9 |
| 2002 | | | | | | | | |
| CV | 8.9 | 13.4 | 18.2 | 19.5 | 6.8 | 22.1 | 22.6 | 21.6 |
| CV_{exp} | 3.7 | 4 | 2.6 | 2.2 | 0.9 | 6.9 | 4.1 | 9.8 |
| 2003 | | | | | | | | |
| CV | 12.7 | 11.8 | 16.8 | 23 | 5.9 | 28.6 | 23.3 | 17.9 |
| CV_{exp} | 0.6 | 6.5 | 14.9 | 5.7 | 1.7 | 4.5 | 17.6 | 13.3 |

influenced these traits, in this way contributing to the variability of the trait, shown by CV; this variability is not considered in CV_{exp} .

Conclusions

From the above it follows that CV_{exp} can be very efficient for data from factorial experiments. Note that under the assumption of variance homogeneity for the linear model, each treatment is assumed to have the same variance, and so we can estimate treatment-wise coefficients of variation based on the very same idea of using the pooled variance. Interestingly, only treatment means will affect the treatment-wise CVs, which is an interesting way of showing that similar variances do not have to mean similar variability in groups.

This coefficient may be especially important for agricultural students, simply because most agricultural statistics deals with factorial experiments, but also because we heard that CV_{exp} is used in plant breeding companies in at least several countries. Nonetheless, its use is not limited to such experiments: it can be used to any one-way experiment in which CV of a dependent variable is of interest. Thus, we believe it is important that researchers remember about the possibility

of using coefficient of variation for variables observed in factorial experiments in the way we discuss in the paper.

References

- Granaas, M. M.**, 2010. Keep teaching the variance. *Journal of Statistics Education*, **18** (1): <http://www.amstat.org/publications/jse/v18n1/pointcounterpoint.pdf>.
- Kozak, M., J. Bocianowski and W. Rybiński**, 2008. Selection of promising genotypes based on path and cluster analyses. *Journal of Agricultural Science Cambridge*, **146**: 85-92.
- Rumsey, D. J.**, 2010. Rejoinder to "Let's just eliminate the variance". *Journal of Statistics Education*, **18** (2): <http://www.amstat.org/publications/jse/v18n2/rumseytb.pdf>.
- Rumsey, D. J.**, 2009. Let us just eliminate the variance. *Journal of Statistics Education*, **17** (3): <http://www.amstat.org/publications/jse/v17n3/rumsey.pdf>.
- Sokal, R. R. and F. J. Rohlf**, 1995. Biometry: the principles and practice of statistics in biological research. 3rd edition. *W. H. Freeman and Co.* New York, 887 pp.
- Webster, R.**, 2007. Analysis of variance, inference, multiple comparisons and sampling effects in soil research. *European Journal of Soil Science*, **58**: 74-82.
- Webster, R.**, 2001. Statistics to support soil research and their presentation. *European Journal of Soil Science*, **52**: 331-340.

Received August, 2, 2012; accepted for printing February, 2, 2013.