



## ΣΠΟΥΔΑΙΟΤΗΤΑ ΚΑΙ ΙΔΙΟΤΗΤΕΣ ΤΟΥ ΣΥΝΤΕΛΕΣΤΗ ΠΑΡΑΛΛΑΚΤΙΚΟΤΗΤΑΣ

Γ. Μενεξές, Θ. Κουτσός

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Σχολή Γεωπονίας, Δασολογίας και Φυσικού Περιβάλλοντος, Τμήμα Γεωπονίας, Εργαστήριο Γεωργίας, Πανεπιστημιούπολη, 52124 Θεσσαλονίκη.

### Περίληψη

Ο συντελεστής παραλλακτικότητας ή μεταβλητότητας (coefficient of variation - CV) είναι δείκτης σχετικής διακύμανσης (ή διασποράς), ο οποίος εκφράζει την ομοιογένεια των τιμών μιας ποσοτικής τυχασίας μεταβλητής. Είναι “καθαρός” αριθμός που παρέχει τη δυνατότητα σύγκρισης ακόμη και της διακύμανσης των τιμών δύο ή περισσότερων ποσοτικών τυχαιών μεταβλητών που έχουν διαφορετικές μονάδες μέτρησης. Η σπουδαιότητα του δείκτη αναδεικνύεται από την ευρεία χρήση του σε ποικίλα επιστημονικά πεδία των Βιολογικών, Οικονομικών, Κοινωνικών και άλλων Επιστημών. Στην παρούσα εργασία παρουσιάζονται ορισμένες ιδιότητες του δείκτη και αναδεικνύεται η σημασία του.

Λέξεις κλειδιά: Ομοιογένεια τιμών τυχασίας ποσοτικής μεταβλητής, δείκτες σχετικής διακύμανσης, σύγκριση τιμών τυχαιών ποσοτικών μεταβλητών

### Εισαγωγή

Ο συντελεστής παραλλακτικότητας ή μεταβλητότητας (coefficient of variation - CV) είναι δείκτης σχετικής διακύμανσης (ή διασποράς), ο οποίος εκφράζει (α) την ομοιογένεια ενός συνόλου μετρήσεων τιμών μιας τυχασίας ποσοτικής μεταβλητής και (β) την ακρίβεια ενός Πειραματικού Σχεδίου. Ο συντελεστής παραλλακτικότητας είναι ένας σημαντικός και χρήσιμος δείκτης σε πολλές επιστήμες, όπως Βιολογικές, Οικονομικές, Κοινωνικές, Ανθρωπιστικές και Θετικές Επιστήμες. Στην εργασία αυτή παρουσιάζονται ορισμένες ιδιότητες του δείκτη CV και αναδεικνύεται η σημασία του.

### Υλικά και Μέθοδοι

#### Υπολογισμός του δείκτη CV

Για τον υπολογισμό του δείκτη CV για δεδομένα δείγματος χρησιμοποιείται η σχέση:

$$CV = \frac{S}{\bar{Y}} \times 100$$

όπου  $S$  η τυπική απόκλιση και  $\bar{Y}$  ο αριθμητικός μέσος όρος των μετρήσεων του δείγματος.

Αντίστοιχα, για τον υπολογισμό του δείκτη CV για δεδομένα πληθυσμού χρησιμοποιείται η σχέση:

$$CV = \frac{\sigma}{\mu} \times 100$$

όπου  $\sigma$  η τυπική απόκλιση και  $\mu$  ο αριθμητικός μέσος όρος των μετρήσεων του πληθυσμού. Ο δείκτης CV για δεδομένα πληθυσμού συμβολίζεται συνήθως με το γράμμα  $\gamma$ . Ο δείκτης CV εκφράζει και την ακρίβεια ενός Πειραματικού Σχεδίου σύμφωνα με τον τύπο:



$$CV = \frac{\sqrt{MSE}}{\bar{Y}_{..}}$$

όπου  $MSE$  το μέσο τετράγωνο των σφαλμάτων (από την κατάλληλη ANOVA) και  $\bar{Y}_{..}$  ο γενικός αριθμητικός μέσος όρος των μετρήσεων του πειράματος.

**Παρατηρήσεις:** α) Αν οι μετρήσεις είναι αρνητικοί αριθμοί τότε στις παραπάνω σχέσεις χρησιμοποιείται η απόλυτη τιμή του μέσου όρου, β) Σε πολλά επιστημονικά πεδία χρησιμοποιείται το τετράγωνο του δείκτη  $CV$  ( $CV^2$ ), γ) Σε πολλά επιστημονικά πεδία η τυπική απόκλιση εκφράζει Σταθερότητα (Stability) ή Αβεβαιότητα (Uncertainty), δ) Υπάρχουν οικογένειες κατανομών με μεγάλη τυπική απόκλιση σε σχέση με το μέσο όρο και επομένως σε αυτές τις περιπτώσεις αναμένονται υψηλές τιμές του δείκτη  $CV$ .

#### Ιδιότητες του δείκτη $CV$

Οι σημαντικότερες ιδιότητες του συντελεστή παραλλακτικότητας είναι:

- Είναι “καθαρός” αριθμός (δεν έχει μονάδα μέτρησης)
- Εκφράζει την τυπική απόκλιση ως ποσοστό % του μέσου όρου
- Είναι δείκτης σχετικής μεταβλητότητας
- Επιτρέπει τη σύγκριση της μεταβλητότητας δύο ή περισσότερων συνόλων μετρήσεων που έχουν διαφορετικές μονάδες μέτρησης
- Επιτρέπει τη σύγκριση της μεταβλητότητας δύο ή περισσότερων συνόλων μετρήσεων που έχουν την ίδια μονάδα μέτρησης αλλά αρκετά διαφορετικούς μέσους όρους ή/και τυπικές αποκλίσεις
- Η μέγιστη τιμή του δείκτη μπορεί να φθάσει την τιμή  $\sqrt{N-1}$ , όπου  $N$  είναι το μέγεθος δείγματος ή την τιμή  $\sqrt{N}$  όπου  $N$  είναι το μέγεθος πληθυσμού (όταν όλες οι τιμές πλην μίας είναι ίσες με μηδέν)
- Χρησιμοποιείται στον καθορισμό του ελάχιστου μεγέθους δείγματος (επαναλήψεις ανά επέμβαση) σε πειράματα
- Δεν παραμένει αμετάβλητος σε μετασχηματισμούς των αρχικών μετρήσεων ( $\log(Y)$ ,  $1/Y$ ,  $Y^{1/2}$ )
- Δεν πρέπει να χρησιμοποιείται όταν οι μετρήσεις περιλαμβάνουν αρνητικούς και θετικούς αριθμούς
- Δεν πρέπει να χρησιμοποιείται όταν ο μέσος όρος είναι κοντά στο μηδέν
- Δεν μπορεί να υπολογιστεί όταν ο μέσος όρος είναι ίσος με μηδέν
- Δεν έχει νόημα (ερμηνεία) όταν οι τιμές του δείγματος δεν είναι μετρημένες σε κλίμακα αναλογίας (ratio scale)

#### Επιθυμητές τιμές του συντελεστή παραλλακτικότητας $CV$

Οι τιμές του συντελεστή πρέπει να είναι, εν γένει, μικρότερες από 10% ( $CV \leq 10\%$ ). Οι αποδεκτές τιμές του δείκτη  $CV$  καθορίζονται ανάλογα:

- **Με τη “φύση” των μετρήσεων** (π.χ. απόδοση, ύψος, βάρος, συγκέντρωση)
- **Με το είδος των πειραματικών ή δειγματοληπτικών μονάδων** (π.χ. ρύζι, σιτάρι, αγγελάδα, άνθρωπος)
- **Με τις συνθήκες πειραματισμού** (π.χ. αγρός, θερμοκήπιο, εργαστήριο)



- **Με την πειραματική μονάδα** (π.χ. πειραματικό τεμάχιο ή μεμονωμένο φυτό)

Στους τοπικούς **πειραματικούς σχεδιασμούς**, όπου οι μετρήσεις γίνονται ανά πειραματικό τεμάχιο και όχι ανά φυτό, οι τιμές του συντελεστή παραλλακτικότητας εκφράζουν την ακρίβεια του πειράματος σύμφωνα με τις παρακάτω νόρμες:

- Τιμές CV: 0-10% δηλώνουν υψηλή ακρίβεια
- Τιμές CV: 10-20% δηλώνουν μέτρια ακρίβεια
- Τιμές CV: 20-30% δηλώνουν μικρή ακρίβεια
- Τιμές CV: >30% δηλώνουν πολύ μικρή ακρίβεια

Να σημειωθεί, ότι σε πειράματα βελτίωσης οι τιμές του δείκτη CV ενδεχομένως να είναι επιθυμητά υψηλές (Kargiotidou et al., 2016).

#### Όρια Εμπιστοσύνης

Αν οι μετρήσεις ακολουθούν Κανονική Κατανομή τότε ένα  $(1-a)\%$  διάστημα εμπιστοσύνης για το δείκτη CV δίνεται από τη σχέση (Abdi, 2010):

$$\tilde{c}_v \pm t_{\frac{a}{2}, \nu} \tilde{s}_{c_v}$$

όπου:  $\tilde{s}_{c_v} = \frac{\tilde{c}_v}{\sqrt{2N}}$  και  $\tilde{c}_v = \left(1 + \frac{1}{4N}\right) CV$

και

$\tilde{c}_v$ : αμερόληπτος εκτιμητής του δείκτη CV

$N$  = το μέγεθος του δείγματος

$\nu = N-1$  οι βαθμοί ελευθερίας της  $t$ -Κατανομής

$t_{\alpha/2, \nu}$ : η κρίσιμη τιμή της  $t$ -Κατανομής για  $\nu$  βαθμούς ελευθερίας σε επίπεδο σημαντικότητας  $\alpha/2$ .

**Παρατήρηση:** Υπάρχουν και άλλες μεθοδολογικές προσεγγίσεις στην κατασκευή διαστημάτων εμπιστοσύνης για το δείκτη (π.χ. μέθοδοι που βασίζονται στην μη κεντρική  $t$ -Κατανομή).

#### Στατιστική σύγκριση δεικτών CV

Για τη σύγκριση δύο δεικτών CV που προέρχονται από ανεξάρτητα δείγματα έχει προταθεί ο παρακάτω στατιστικός έλεγχος:

$$H_0: \gamma_1 = \gamma_2$$

$$H_1: \gamma_1 \neq \gamma_2$$

Σε επίπεδο σημαντικότητας  $\alpha$  (συνήθως  $\alpha=0,05$ ).

Το στατιστικό του ελέγχου υπολογίζεται με βάση την παρακάτω σχέση (Forkman, 2009):

$$F = (c_1^2 / (1 + c_1^2) (n_1 - 1) / n_1) / (c_2^2 / (1 + c_2^2) (n_2 - 1) / n_2)$$

$$c_1 = CV_1$$

$$c_2 = CV_2$$

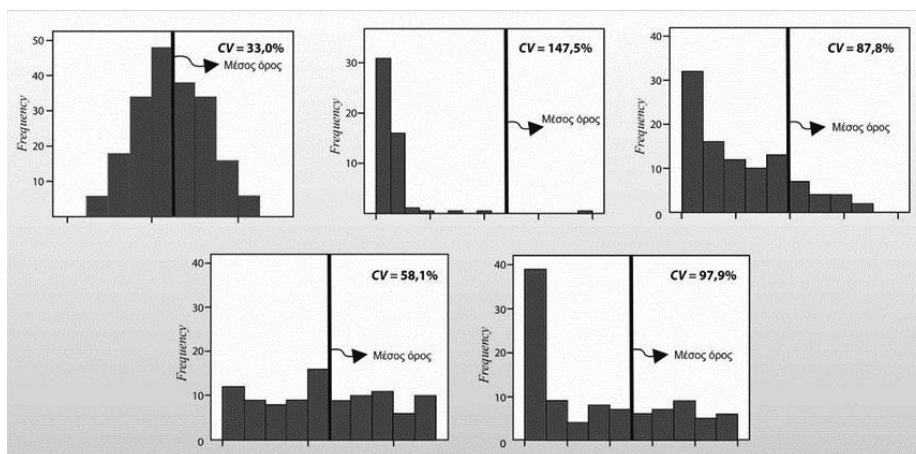
$n_1$ : μέγεθος πρώτου δείγματος

$n_2$ : μέγεθος δεύτερου δείγματος

Η απόφαση για την απόρριψη ή όχι της μηδενικής υπόθεσης  $H_0$  βασίζεται στην κρίσιμη τιμή της  $F$ -Κατανομής.

Συνέπειες μεγάλων τιμών του συντελεστή CV

Γενικά, οι υψηλές τιμές του συντελεστή  $CV$  (>50%), δηλώνουν μη κανονικότητα των δεδομένων ή πρόβλημα στη διαδικασία παραγωγής των δεδομένων. Στα παρακάτω διαγράμματα (Σχήμα 1), παρουσιάζονται κατανομές τιμών δειγμάτων και οι αντίστοιχοι δείκτες  $CV$ .



Σχήμα 1. Διαγράμματα κατανομών και αντίστοιχοι δείκτες  $CV$

Διαπιστώνεται ότι οι κατανομές είτε είναι ασύμμετρες είτε προσεγγίζουν την ομοιόμορφη κατανομή (αλλά με μεγάλο εύρος τιμών). Δεν παρουσιάζουν κανονικότητα και η ύπαρξη ακραίων τιμών (outliers) προκαλεί ασυμμετρία και μεγάλη παραλλακτικότητα. Συνεπώς, στις περιπτώσεις αυτές ο αριθμητικός μέσος όρος δεν είναι κατάλληλος δείκτης κεντρικής τάσης και δεν μπορεί να αξιοποιηθεί περαιτέρω σε άλλες στατιστικές αναλύσεις. Με άλλα λόγια, ο μέσος όρος δεν “αντιπροσωπεύει” τα δεδομένα. Η περαιτέρω στατιστική επεξεργασία των δεδομένων θα πρέπει να γίνει με βάση Μη Παραμετρικές στατιστικές μεθόδους ή με χρήση Γενικευμένων Γραμμικών Υποδειγμάτων.

**Παρατήρηση:** Αν η τιμή του δείκτη  $CV$  είναι 100%, π.χ. 120%, τότε είναι προτιμότερο ο δείκτης να γράφεται ως 1,2.

2.6. Παραδείγματα άλλων δεικτών σχετικής μεταβλητότητας

A) Μη παραμετρικός δείκτης παραλλακτικότητας (Non Parametric  $CV_{np}$ )

$$CV_{np} = (\text{Διάμεση τιμή} / \text{Ημιενδοτεταρτημοριακό εύρος}) \times 100$$

$$CV_{np} = \frac{Q_{50}}{\frac{Q_{75} - Q_{25}}{2}} \times 100 = \frac{2Q_{50}}{Q_{75} - Q_{25}} \times 100$$



$$CV_{np} = \frac{2Q_{50}^{\eta}}{Q_{75} - Q_{25}} \times 100$$

Όπου  $Q_{50}$ : Διάμεση τιμή,  $Q_{25}$ : πρώτο τεταρτημόριο,  $Q_{75}$ : τρίτο τεταρτημόριο της κατανομής των μετρήσεων.

B) Αναλογική μεταβλητότητα (Proportional Variability-PV)

$$PV = \frac{1}{c} \sum_{i \neq j} \left( 1 - \frac{\min(y_i, y_j)}{\max(y_i, y_j)} \right)$$

όπου  $i, j = 1, \dots, N$  και  $c = \binom{N}{2}$  και  $y_i$  οι μετρήσεις του δείγματος (Heath and Borowski, 2013).

#### Ένας «έξυπνος» μετασχηματισμός

Με βάση την παρακάτω σχέση [1] μπορούμε να προσθέσουμε στις αρχικές τιμές ενός συνόλου μετρήσεων (με τυπική απόκλιση  $s$  και συντελεστή  $CV=V$  ως ποσοστό) έναν κατάλληλο αριθμό  $\lambda$  ώστε οι νέες μετασχηματισμένες μετρήσεις να έχουν συγκεκριμένη επιθυμητή τιμή συντελεστή μεταβλητότητας  $\delta$  (ως ποσοστό).

$$\lambda \geq s \left( \frac{1}{\delta} - \frac{1}{V} \right) \quad [1]$$

**Παράδειγμα:** Έστω  $\bar{Y}=40$  και  $s=30$ ,  $V=0,75$  και  $\delta=0,20$ . Συνεπώς:  $\lambda \geq 30(5-1,333\dots) \Rightarrow \lambda \geq 110$

Επομένως, αν στις αρχικές τιμές προστεθεί η ποσότητα 110, τότε ο δείκτης  $CV$  για τις μετασχηματισμένες τιμές θα είναι περίπου 0,20 ή 20%.

#### Αποτελέσματα και συζήτηση

Η τιμή του δείκτη μπορεί να μειωθεί με κατάλληλο μετασχηματισμό των αρχικών δεδομένων (π.χ. log μετασχηματισμός). Όμως, τότε, η φυσική του ερμηνεία δεν είναι ξεκάθαρη. Η “βελτιστοποίηση” της τιμής του δείκτη  $CV$  εξαρτάται από το σκοπό της έρευνας. Διότι:  $CV = f(s, \bar{Y})$ , αλλά και  $s = g(\bar{Y}, N)$ , συνεπώς ισχύει:  $CV = \varphi(s, \bar{Y}, N)$ .

Με βάση τις παραπάνω σχέσεις, μπορεί τρία δείγματα να έχουν ίδια τιμή του δείκτη  $CV$ , αλλά με τελείως διαφορετική “σύνθεση” σε σχέση με την τυπική απόκλιση και το μέσο όρο (και το μέγεθος δείγματος).

#### **Παράδειγμα:**

$CV=40\%$ , με τυπική απόκλιση 40 και μέσο όρο 100 ( $N=50$ )

$CV=40\%$ , με τυπική απόκλιση 20 και μέσο όρο 50 ( $N=500$ )

$CV=40\%$ , με τυπική απόκλιση 80 και μέσο όρο 200 ( $N=20$ )

Συνεπώς, η βελτιστοποίηση της τιμής του δείκτη  $CV$  εξαρτάται από το εάν, για παράδειγμα, επιδίωξη της έρευνας είναι η επίτευξη υψηλών μέσων όρων (π.χ. απόδοσης) με



μικρή μεταβλητότητα (π.χ. σταθερότητα, ομοιογένεια) ή η επίτευξη μεγάλης παραλλακτικότητας με υψηλούς μέσους όρους.

### Συμπεράσματα

Σε πολλά επιστημονικά πεδία ο δείκτης *CV* είναι χρήσιμος και σημαντικός δείκτης σχετικής μεταβλητότητας ή/και ακρίβειας. Πολύ υψηλές τιμές του δείκτη *CV* δηλώνουν μη κανονικότητα των δεδομένων ή πρόβλημα στη διαδικασία παραγωγής των δεδομένων. Σε δεδομένα που προέρχονται από μη κανονικές κατανομές, με μεγάλη ασυμμετρία και μεγάλη παραλλακτικότητα, ο αριθμητικός μέσος όρος δεν είναι “καλός” δείκτης κεντρικής τάσης. Μέσω κατάλληλων μετασχηματισμών η τιμή του δείκτη *CV* μπορεί να μεταβληθεί σε επιθυμητά ή αποδεκτά επίπεδα. Ο δείκτης *CV* δεν θα πρέπει να χρησιμοποιείται όταν η σχέση των δύο “συστατικών” του, δηλαδή του μέσου όρου και της τυπικής απόκλισης, είναι “ανταγωνιστική” (π.χ. απόδοση *vs* σταθερότητα).

### Βιβλιογραφία

- Abdi H. 2010. Coefficient of Variation. In Neil Salkind (Ed.), Encyclopedia of Research Design. Thousand Oaks, CA: Sage. 2010.
- Forkman, J. 2009. Estimator and Tests for Common Coefficients of Variation in Normal Distributions. Communications in Statistics - Theory and Methods. Volume: 38 Number: 2, pp 233-251.
- Heath, J.P. and Borowski, P. 2013. Quantifying Proportional Variability. PLoS ONE 8(12): e84074. doi:10.1371/journal.pone.0084074.
- Kargiotidou, A., Vlachostergios, D.N., Tzantarmas, C., Mylonas, I., Foti, C., Menexes, G., Polidoros, A., Tokatlidis, I.S. 2016. Addressing huge spatial heterogeneity induced by virus infections in lentil breeding trials. Journal of Biological Research-Thessaloniki, 23(2). DOI 10.1186/s40709-016-0039-6.

## IMPORTANCE AND PROPERTIES OF THE COEFFICIENT OF VARIANCE

**Georgios Menexes, Thomas Koutsos**

The coefficient of variance (*CV*) is a relative index of dispersion, which expresses the homogeneity of values of a quantitative random variable. It is a "clear" number that allows comparison of the variation of two or more quantitative random variables having different measurement units. The importance of the coefficient of variation is highlighted by its widespread use in a variety of scientific fields such as biological, economic, social and other sciences. This paper presents some properties of the index and highlights the importance of the index.