ΑΣΚΗΣΗ: # ΣΤΑΤΙΣΤΙΚΗ ΣΗΜΕΙΩΣΕΙΣ [#1]

Αν $(x_1, x_2, \ldots, x_n)$ είναι δείγμα από πληθυσμό μεγέθους N $(0<n<N)$ με μέσο (μ) και διακύμανση (σ^2), τότε αν οι x_i θεωρηθούν εξαρτημένες τότε θα ισχύει ότι:

- $\forall i \in [n] \quad E[x_i] = \mu$
- $\forall i \in [n] \quad Var[x_i] = \sigma^2$

$$(\alpha): \; E[\bar{x}] = \mu \quad \text{και} \quad (\beta): \; Var[\bar{x}] = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

ΑΠΟΔΕΙΞΗ: Έστω S_n το μερικό άθροισμα που αντιστοιχεί στις τιμές του δείγματος:

$$S_n = x_1 + x_2 + \cdots + x_n \Rightarrow$$

$$S_n = n \cdot \bar{x} \Rightarrow \boxed{\bar{x} = \frac{S_n}{n}} \quad (1)$$

- $E[\bar{x}] = E\left[\frac{1}{n} S_n\right] = \frac{1}{n} E[S_n] \Rightarrow$

$$E[\bar{x}] = \frac{1}{n} E\left[\sum_{k=1}^{n} x_k\right] = \frac{1}{n} \sum_{k=1}^{n} E[x_k] = \frac{1}{n} \cdot n \cdot \mu = \mu .$$

s.o.s

- Οι τυχαίες μεταβλητές x_i δεν είναι ανεξάρτητες αλλά είναι ισοκατανεμημένες (identically distributed).

- Αυτό σημαίνει πως για κάθε ζεύγος (x_i, x_j) με $i \neq j$ θα έχουμε την ίδια από κοινού κατανομή και κατά συνέπεια την ίδια συνδιακύμανση.

- Δηλαδή, θα έχουμε ότι:

$$\boxed{\forall i \neq j, \; Cov[x_i, x_j] = c} \quad (2)$$

• Θα θεωρήσουμε τις βοηθητικές τ.μ. Y_i με $i \in [n]$ με $Y_i = X_i - \mu$ οπότε θα έχουμε ότι:

$$E[Y_i] = E[X_i - \mu] \Rightarrow E[Y_i] = E[X_i] - \mu \Rightarrow$$

$$E[Y_i] = \mu - \mu \Rightarrow$$

$$\boxed{E[Y_i] = 0 \,, \forall i \in [n] \quad (3).2}$$

και

$$\boxed{Var[Y_i] = Var[X_i] \,, \forall i \in [n] \quad (3).1}$$

• Μπορούμε επίσης να γράψουμε ότι:

$$\boxed{E[S_n] = E[n\bar{x}] = n\mu \quad (4)}$$

$$Var[S_n] = E[(S_n - E[S_n])^2] \Rightarrow$$

$$Var[S_n] = E[\{X_1 + X_2 + \cdots + X_n - n\mu\}^2] \Rightarrow$$

$$Var[S_n] = E[\{X_1 - \mu + X_2 - \mu + \cdots + X_n - \mu\}^2] \Rightarrow$$

$$\boxed{Var[S_n] = E[(Y_1 + Y_2 + \cdots + Y_n)^2] \quad (5)}$$

(*) Μπορούμε επίσης να γράψουμε ότι:

$$\boxed{(X_1 + X_2 + \cdots + X_n)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j \quad (6)}$$

(*) Η εξίσωση (6) μπορεί να γραφεί στην ισοδύναμη μορφή:

$$\boxed{S_n^2 = \sum_{i=1}^{n} x_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j \quad (7)}$$

Summing over the upper-triangular part of the rectangular $[n] \times [n]$ grid.

- Οι εξισώσεις (5) και (7) δίνουν ότι:

$$\boxed{Var[S_n] = \sum_{i=1}^{n} E[Y_i^2] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} E[Y_i Y_j] \quad (8)}$$

- Θα κάνουμε χρήση των παρακάτω ταυτοτήτων:

(i): $Var[Y_i] = E[Y_i^2] - E[Y_i]^2, \ \forall i \in [n] \ \{Με\ E[Y_i]=0\} \longrightarrow$

$$\boxed{Var[Y_i] = E[Y_i^2]} \quad (9.i)$$

(ii): $Cov[Y_i, Y_j] = E[Y_i \cdot Y_j] - E[Y_i]E[Y_j] \xrightarrow{\forall i\ E[Y_i]=0}$

$$\boxed{E[Y_i \cdot Y_j] = Cov[Y_i, Y_j]}$$

$$\boxed{Cov[Y_i, Y_j] = E[(X_i - \mu)(X_j - \mu)] = Cov[X_i, X_j]} \quad (9.ii)$$

- Η σχέση (8) μπορεί να γραφτεί ως:

$$\boxed{Var[S_n] = \sum_{i=1}^{n} Var[Y_i] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Cov[Y_i, Y_j] \quad (10)}$$

(*) Οι σχέσεις (3.1) και (9.ii) οδηγούν στην σχέση:

$$\boxed{Var[S_n] = \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Cov[X_i, X_j] \quad (11)}$$

→ Γνωρίζουμε ότι: $\begin{cases} \forall i \in [n], \ Var[X_i] = \sigma^2 \\ \forall i \in [n], \ Cov[X_i, X_j] = c \end{cases}$

Επομένως, η σχέση (11) παίρνει την τελική μορφή:

$$\boxed{Var[S_n] = \sum_{i=1}^{n} \sigma^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} c \quad (12)}$$

(*) Μας ενδιαφέρει να υπολογίσουμε το πλήθος των όρων που εμφανίζονται στο διπλό άθροισμα της σχέσης (12):

Για $i=1$ έχουμε $j \in \{2, \dots, n\}$ $[n-1]$ όροι

Για $i=2$ έχουμε $j \in \{3, \dots, n\}$ $[n-2]$ όροι

⋮

Για $i=n-1$ έχουμε $j \in \{n\}$ $[1]$ όρος

Για $i=n$ έχουμε $j \in \{\emptyset\}$ $[\phi]$ όροι

(*) Επομένως, έχουμε συνολικά:

$$\boxed{1+2+\dots+(n-1) = \frac{(A+T)\cdot \Pi}{2} = \frac{[1+(n-1)]\cdot \Pi}{2} = \frac{n(n-1)}{2} \quad (13)}$$

(*) Άρα, η σχέση (12) γράφεται ως:

$$Var[S_n] = n\sigma^2 + 2n(n-1)\cdot\frac{1}{2}\cdot c \Rightarrow$$

$$\boxed{Var[S_n] = n\sigma^2 + n(n-1)c \quad (14)}$$

⊛ Η σχέση (14) ισχύει $\forall n$ με $0 < n < N$.

Επομένως, θα ισχύει και για $n = N$.

Σε αυτή την περίπτωση θα έχουμε ότι:

$$\boxed{Var[S_N] = N\sigma^2 + N(N-1)c \quad (15)}$$

S.O.S: Για $n = N$ η παράσταση $S_N = x_1 + x_2 + \dots + x_N$ αποτιμάται για το σύνολο του πληθυσμού N (το μέγεθος του δείγματος είναι ίσο με το μέγεθος του πληθυσμού) και επομένως είναι σταθερά. Πάυει να είναι μια Τυχαία Μεταβλητή. Ως σταθερά, επομένως, θα έχουμε ότι:

$$\boxed{Var[S_N] = 0 \quad (16)}$$

⊛ Έτσι, θα έχουμε ότι: $N\sigma^2 + N(N-1)c = 0 \Rightarrow$

$$N(N-1)c = -N\sigma^2 \Rightarrow$$

$$\boxed{c = -\frac{\sigma^2}{N-1} \quad (17)}$$

⊛ Αντικαθιστώντας την (17) στην (14) θα έχουμε ότι:

$$Var[S_n] = n\sigma^2 + n(n-1)\left(-\frac{\sigma^2}{N-1}\right) \Rightarrow$$

$$Var[S_n] = n\sigma^2 - \frac{n(n-1)}{N-1}\sigma^2 \Rightarrow$$

$$Var[S_n] = n\sigma^2\left[1 - \frac{(n-1)}{N-1}\right] = n\sigma^2 \cdot \frac{N-1-(n-1)}{N-1} \Rightarrow$$

$$\boxed{Var[S_n] = n\sigma^2 \cdot \frac{N-n}{N-1} \quad (18)}$$

Τελικά, θα έχουμε ότι:

$$Var[\bar{x}] = Var\left[\frac{S_n}{n}\right] \Rightarrow$$

$$Var[\bar{x}] = \frac{1}{n^2} \cdot Var[S_n] \Rightarrow$$

$$Var[\bar{x}] = \frac{1}{n^2} \cdot n \cdot \sigma^2 \cdot \frac{N-n}{N-1} \Rightarrow$$

$$\boxed{Var[\bar{x}] = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \quad (1a)}$$