Ασκήσεις Εβδομάδας 2

Διαλέξεις 4–6: ροπές, εκτιμήτριες και οικογένειες θέσης–κλίμακας

Συμβολισμοί. Για τυχαίο δείγμα $X_1, \dots, X_n$ γράφουμε, παραλείποντας τον δείκτη n όταν το μέγεθος παραμένει σταθερό,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad M_k' = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \qquad M_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k.$$

Θέτουμε $\mu = \mathbb{E}(X_1)$, $\mu_k' = \mathbb{E}(X_1^k)$, $\mu_k = \mathbb{E}[(X_1 - \mu)^k]$ και $\sigma^2 = \mu_2$, όταν οι απαιτούμενες απόλυτες ροπές είναι πεπερασμένες. Για $n \geq 2$,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{n}{n-1} M_2.$$

Για $0 < \sigma^2 < \infty$ και τις αντίστοιχες πεπερασμένες ροπές,

$$\lambda = \frac{\mu_3}{\sigma^3}, \qquad \kappa = \frac{\mu_4}{\sigma^4}, \qquad \kappa' = \kappa - 3.$$

Οι δειγματικές εκδοχές, για $M_2 > 0$, είναι

$$\widehat{\lambda} = \frac{M_3}{M_2^{3/2}}, \qquad \widehat{\kappa} = \frac{M_4}{M_2^2}, \qquad \widehat{\kappa'} = \widehat{\kappa} - 3.$$

Τα πεζά $\overline{x}, m_k', m_k, s^2$ δηλώνουν τιμές στα δεδομένα. Οι $\mathbb{P}, \mathbb{E}, \mathbb{V}, \mathrm{Cov}$ έχουν την ίδια σημασία με την 1η εβδομάδα.

Στην $\mathrm{Unif}[0, \theta]$ θα χρησιμοποιούμε σταθερά

$$\widehat{\theta} = X_{(n)}, \qquad \widetilde{\theta} = \frac{n+1}{n} X_{(n)}, \qquad \overline{\theta} = 2\overline{X}.$$

Στις οικογένειες θέσης–κλίμακας γράφουμε $X = \alpha + \beta Z$, $\beta > 0$, και $f(x; \alpha, \beta)$. Για συγκεκριμένη πυκνότητα ή συνάρτηση πιθανότητας f θέτουμε $S_f = \{x : f(x) > 0\}$. Όλα τα αριθμητικά δεδομένα του φύλλου είναι υποθετικά.

Αναφορές. Σ. Τρέβεζας, σημειώσεις *Μαθηματικής Στατιστικής*, σ. 15–30. G. Casella και R. L. Berger, *Στατιστική Συμπερασματολογία*, ελληνική μετάφραση της 2ης έκδοσης, §2.3, Παράδειγμα 2.2.4, §§3.5 και 7.2.1. Για τη γενικευμένη αντίστροφη και τα ποσοστημόρια: Σ. Τρέβεζας, *Μη Παραμετρική Στατιστική*, Κάλλιπος (2025), §3.12.

Α. Μέθοδος των ροπών και αξιολόγηση εκτιμητριών

✎ Άσκηση 1. Ποια ροπή δίνει πληροφορία;

Βασική–μέση

(α) Έστω $X_i \sim \mathrm{Unif}[-\theta, \theta]$, με $\theta > 0$. Γράψτε την εξίσωση της πρώτης ροπής και εξηγήστε γιατί δεν προσδιορίζει το θ. Χρησιμοποιήστε τη δεύτερη μη κεντρική ροπή για να κατασκευάσετε εκτιμήτρια ροπών.

(β) Έστω τώρα $X_i \sim N(\theta, 1)$, με $\theta \in \mathbb{R}$. Αν χρησιμοποιήσουμε μόνο την εξίσωση $\mathbb{E}_\theta(X_1^2) = M_2'$, πόσες λύσεις υπάρχουν όταν $M_2' > 1$, $M_2' = 1$ και $M_2' < 1$;

(γ) Στο κανονικό μοντέλο του (β), ποια εκτιμήτρια δίνει η πρώτη ροπή; Χρειάζεται να ικανοποιεί ταυτόχρονα και την εξίσωση του (β) για κάθε δείγμα;

✎ Άσκηση 2. Κανονικό μοντέλο με μία ή δύο άγνωστες παραμέτρους

Μέση

Έστω $n \geq 2$ και τυχαίο δείγμα από $N(\mu, \sigma^2)$, με $\sigma^2 > 0$.

(α) Για γνωστή σ^2 και άγνωστη $\mu \in \mathbb{R}$, βρείτε εκτιμήτρια με την πρώτη ροπή.

(β) Για γνωστή $\mu = \mu_0$, χρησιμοποιήστε χωριστά τις εξισώσεις $\mathbb{E}[(X_1 - \mu_0)^2] = \sigma^2$ και $\mathbb{E}(X_1^2) = \mu_0^2 + \sigma^2$. Συμβολίστε τις αλγεβρικές λύσεις με T και U. Βρείτε τη διαφορά τους και τις μέσες τιμές τους. Δώστε παράδειγμα όπου η U είναι αρνητική.

(γ) Για άγνωστες και τις δύο παραμέτρους, λύστε τα συστήματα

$$\mu = \overline{X}, \quad \mu^2 + \sigma^2 = M_2', \qquad \text{και} \qquad \mu = \overline{X}, \quad \sigma^2 = M_2.$$

Είναι ισοδύναμα; Τι συμβαίνει αν $M_2 = 0$;

(δ) **Εμβάθυνση.** Στην περίπτωση (β), χρησιμοποιώντας τις κεντρικές ροπές της κανονικής, συγκρίνετε τις $\mathbb{V}(T)$ και $\mathbb{V}(U)$.

Άσκηση 3. Διωνυμική με άγνωστα N και p

Μέση–εμβάθυνση

Έστω $X_i \sim \mathrm{Bin}(N, p)$ με $\Theta = \mathbb{N}^* \times (0,1)$, όπου $\mathbb{N}^* = \{1, 2, \ldots\}$. Το n είναι το μέγεθος του δείγματος και το N το άγνωστο πλήθος δοκιμών μιας παρατήρησης.

(α) Χρησιμοποιώντας $\mathbb{E}(X_1) = Np$ και $\mathbb{V}(X_1) = Np(1-p)$, λύστε το σύστημα ροπών αρχικά ως προς πραγματικό $N > 0$. Ποιος περιορισμός στις $\overline{X}, M_2$ εξασφαλίζει $0 < \widetilde{p} < 1$;

(β) Εξηγήστε τα δύο διαφορετικά ζητήματα: N ακέραιο και $N \geq X_{(n)}$. Για μη σταθερό δείγμα μη αρνητικών ακεραίων με $0 < M_2 < \overline{X}$, εξετάστε τον κανόνα

$$\widehat{N} = \max\left\{1, X_{(n)}, \left\lfloor \widetilde{N} + \frac{1}{2} \right\rfloor\right\}, \qquad \widehat{p} = \frac{\overline{X}}{\widehat{N}}.$$

Ποια εξίσωση διατηρείται μετά τη διόρθωση;

(γ) Εφαρμόστε τους υπολογισμούς στα δείγματα

$$A = (1,2,2,3,3,4), \qquad B = (0,0,0,0,0,6), \qquad C = (0,0,0,0,0,0),$$

και στο D, που περιέχει εννέα τιμές ίσες με 3 και μία ίση με 6.

(δ) Μπορεί η αποτυχία της συνθήκης $M_2 < \overline{X}$ σε ένα δείγμα να αποδείξει ότι το διωνυμικό μοντέλο είναι λανθασμένο; Αιτιολογήστε με το δείγμα B υπό $N = 6$ και $0 < p < 1$.

Άσκηση 4. Εκτιμήτριες του άγνωστου άκρου

Μέση

Έστω $X_1, \ldots, X_n \sim \mathrm{Unif}[0, \theta]$ ανεξάρτητες, με $\theta > 0$. Συγκρίνουμε

$$\overline{\theta} = 2\overline{X}, \qquad \widehat{\theta} = X_{(n)}, \qquad \widetilde{\theta} = \frac{n+1}{n} X_{(n)}.$$

Για εκτιμήτρια A του θ ορίζουμε

$$b_A(\theta) = \mathbb{E}_\theta(A) - \theta, \qquad \mathrm{MSE}_A(\theta) = \mathbb{E}_\theta[(A-\theta)^2].$$

(α) Βρείτε τη μέση τιμή, τη διασπορά και το μέσο τετραγωνικό σφάλμα της $\overline{\theta}$.

(β) Βρείτε τη συνάρτηση κατανομής και την πυκνότητα του $\widehat{\theta}$. Υπολογίστε τις $\mathbb{E}_\theta(\widehat{\theta})$ και $\mathbb{E}_\theta(\widehat{\theta}^2)$.

(γ) Συμπληρώστε πίνακα μεροληψίας, διασποράς και μέσου τετραγωνικού σφάλματος για τις τρεις εκτιμήτριες.

(δ) Για ποιες ακέραιες τιμές του n η μεροληπτική $\widehat{\theta}$ έχει μικρότερο μέσο τετραγωνικό σφάλμα από την $\overline{\theta}$; Ποια είναι η καλύτερη από τις τρεις με αυτό το κριτήριο;

Β. Κεντρικές ροπές, λοξότητα και κύρτωση

Άσκηση 5. Από τις μη κεντρικές στις κεντρικές ροπές

Βασική–μέση

(α) Για $\mathbb{E}|X_1|^4 < \infty$, εκφράστε τις μ_2, μ_3, μ_4 μέσω των $\mu, \mu'_2, \mu'_3, \mu'_4$.

(β) Γράψτε τις αντίστοιχες αλγεβρικές ταυτότητες για τις M_2, M_3, M_4. Δείξτε επίσης ότι $M_1 = 0$.

(γ) Επιστρέψτε στους χρόνους της Άσκησης 1 της 1ης εβδομάδας, $x = (1, 2, 2, 3, 4, 6)$. Υπολογίστε m_2, m_3, m_4 και τις τιμές των $\widehat{\lambda}, \widehat{\kappa}, \widehat{\kappa'}$.

(δ) Από αυτές τις τιμές προσδιορίστηκε η λοξότητα και η κύρτωση της άγνωστης κατανομής; Ποιες ποσότητες παραμένουν σταθερές και ποιες μπορεί να αλλάξουν όταν ληφθεί νέο δείγμα από την ίδια κατανομή;

Άσκηση 6. Κεντροποίηση, τυποποίηση και αλλαγή μονάδας

Βασική–μέση

Έστω $\mathbb{E}|X|^4 < \infty$ και $0 < \mathbb{V}(X) = \sigma^2$. Θέτουμε $Y = \alpha + \beta X$, με $\alpha \in \mathbb{R}$ και $\beta \neq 0$.

(α) Βρείτε τη μέση τιμή, τη διασπορά και τις τρίτη και τέταρτη κεντρικές ροπές της Y.

(β) Δείξτε πώς μεταβάλλονται οι λ, κ, κ' για $\beta > 0$ και για $\beta < 0$.

(γ) Συγκρίνετε τις $X - \mu$, X/σ και $(X - \mu)/\sigma$ ως προς μέση τιμή και διασπορά.

(δ) Για δεδομένα $x_1, \ldots, x_n$ με $m_2 > 0$, θέτουμε $z_i = (x_i - \overline{x})/\sqrt{m_2}$. Βρείτε τον μέσο και τη διασπορά με παρονομαστή n των z_i. Τι θα προέκυπτε αν διαιρούσαμε με s αντί με $\sqrt{m_2}$;

Άσκηση 7. Λοξότητα και κύρτωση σε τέσσερις κατανομές

Μέση

Θεωρούμε: $U \sim \text{Unif}[-a, a]$ με $a > 0$, $Z \sim N(0,1)$, E με πυκνότητα $e^{-x}\mathbf{1}_{(0,\infty)}(x)$, και L με πυκνότητα $\frac{1}{2}e^{-|x|}$ στο $\mathbb{R}$.

(α) Υπολογίστε λ, κ, κ' σε καθεμία από τις τέσσερις περιπτώσεις. Μπορείτε να χρησιμοποιήσετε $\mathbb{E}(Z^4) = 3$ και $\int_0^\infty x^k e^{-x}\,dx = k!$ για $k = 0, 1, 2, 3, 4$.

(β) Κατατάξτε τις κατανομές σε πλατύκυρτη, μεσόκυρτη ή λεπτόκυρτη, με βάση τη σύγκριση της κ με το 3. Ποια από τις λεπτόκυρτες είναι συμμετρική;

(γ) Ποιους συντελεστές έχει η $-E$; Πώς σχετίζεται αυτό με ένα ιστόγραμμα με μακριά αριστερή ουρά;

(δ) Ένα σφάλμα στρογγυλοποίησης μοντελοποιείται ως $Q \sim \text{Unif}[-\Delta/2, \Delta/2]$. Βρείτε $\mathbb{V}(Q)$ και κ'_Q. Αρκεί το γεγονός ότι το σφάλμα βρίσκεται στο συγκεκριμένο διάστημα για να συμπεράνουμε ότι είναι ομοιόμορφο;

Άσκηση 8. Συμμετρία και ύπαρξη ροπών

Βασική–μέση

Έστω X με την πυκνότητα της τυπικής Cauchy

$$f(x) = \frac{1}{\pi(1 + x^2)}, \qquad x \in \mathbb{R}.$$

(α) Για $X^+ = \max(X, 0)$ και $X^- = \max(-X, 0)$, δείξτε ότι $\mathbb{E}(X^+) = \mathbb{E}(X^-) = +\infty$. Γιατί αυτό σημαίνει ότι η $\mathbb{E}(X)$ δεν ορίζεται, αντί να ισούται με $+\infty$ ή $-\infty$;

(β) Δείξτε ότι $\int_{-A}^{A} x f(x)\,dx = 0$ για κάθε $A > 0$. Γιατί δεν προκύπτει $\mathbb{E}(X) = 0$;

(γ) Δείξτε ότι $\mathbb{E}(X^2) = \mathbb{E}(X^4) = +\infty$. Ορίζονται οι λ, κ, κ' μέσω των τύπων των κεντρικών ροπών;

(δ) Σε ένα πεπερασμένο δείγμα από την Cauchy μπορούν να υπολογιστούν οι δειγματικοί συντελεστές, αν $M_2 > 0$. Αποδεικνύει αυτό ότι υπάρχουν οι αντίστοιχοι πληθυσμιακοί συντελεστές;

Άσκηση 9. Τι δεν συνεπάγονται οι συντελεστές

Προαιρετική

(α) Μια τυχαία μεταβλητή A παίρνει τις τιμές $-2, 1, 3$ με πιθανότητες $2/5, 1/2, 1/10$, αντίστοιχα. Υπολογίστε $\mathbb{E}(A)$, $\mu_3(A)$ και λ_A. Είναι η κατανομή συμμετρική;

(β) Για $0 < \varepsilon \leq 1$, μια B παίρνει τις τιμές $-1, 0, 1$ με πιθανότητες $\varepsilon/2, 1-\varepsilon, \varepsilon/2$. Βρείτε τις $\lambda_B, \kappa_B, \kappa'_B$. Πότε η πλεονάζουσα κύρτωση είναι θετική; Αρκεί το φραγμένο στήριγμα για αρνητική πλεονάζουσα κύρτωση;

(γ) Για ποια τιμή του ε η B έχει την ίδια λοξότητα και κύρτωση με μια κανονική; Είναι τότε κανονική;

Γ. Δειγματική διασπορά και μέσο τετραγωνικό σφάλμα

Άσκηση 10. Το λήμμα των διαφορών ανά ζεύγη

Βασική–μέση

Έστω $n \geq 2$, $\mathbb{E}(X_1) = \mu$ και $\mathbb{V}(X_1) = \sigma^2 < \infty$.

(α) Αποδείξτε για οποιουσδήποτε πραγματικούς $x_1, \ldots, x_n$ ότι

$$\sum_{i \neq j} (x_i - x_j)^2 = 2n \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Το άθροισμα αριστερά περιλαμβάνει *διατεταγμένα* ζεύγη. Ποιος συντελεστής προκύπτει αν αθροίζουμε μόνο για $i < j$;

(β) Δείξτε ότι $\mathbb{E}(\overline{X}) = \mu$ και $\mathbb{V}(\overline{X}) = \sigma^2/n$.

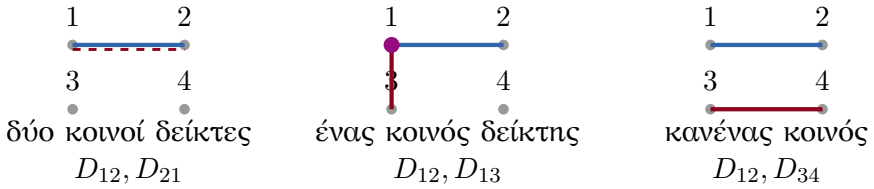
(γ) Χρησιμοποιήστε το λήμμα για να υπολογίσετε την $\mathbb{E}(M_2)$. Συμπεράνετε την αμεροληψία του S^2.

(δ) Απαιτήθηκε κανονικό μοντέλο ή πεπερασμένη τέταρτη ροπή στα προηγούμενα;

Άσκηση 11. Η διασπορά των M_2 και S^2

Μέση-εμβάθυνση

Έστω $n \geq 2$ και τυχαίο δείγμα με $\mathbb{E}|X_1|^4 < \infty$, $\mathbb{E}(X_1) = \mu$, $\mathbb{V}(X_1) = \sigma^2$ και $\mu_4 = \mathbb{E}[(X_1 - \mu)^4]$. Θέτουμε $D_{ij} = (X_i - X_j)^2$ για $i \neq j$.



Σχήμα 1. Η συνδιακύμανση εξαρτάται από τους κοινούς δείκτες. Το διάγραμμα αφορά $n \geq 4$· για μικρότερα δείγματα απουσιάζουν οι μη διαθέσιμες περιπτώσεις.

(α) Εξηγήστε γιατί μπορούμε, για τους υπολογισμούς με τις διαφορές, να αντικαταστήσουμε τα X_i με $Y_i = X_i - \mu$.

(β) Υπολογίστε την $\mathbb{V}(D_{12})$. Για $n \geq 3$, δείξτε ότι $\mathrm{Cov}(D_{12}, D_{13}) = \mu_4 - \sigma^4$.

(γ) Υπολογίστε τη διασπορά του $H = \sum_{i \neq j} D_{ij}$ χωρίζοντας τα ζεύγη δεικτών σε: δύο κοινούς δείκτες, ακριβώς έναν κοινό δείκτη, κανέναν κοινό δείκτη. Μετρήστε προσεκτικά και τα $(i,j),(j,i)$.

(δ) Χρησιμοποιήστε το λήμμα της Άσκησης 10 για να αποδείξετε

$$\mathbb{V}(S^2) = \frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4,$$

και βρείτε τη $\mathbb{V}(M_2)$. Ελέγξτε χωριστά την περίπτωση $n = 2$.

(ε) Για $\sigma^2 > 0$, γράψτε τον τύπο με την κ. Τι δείχνει όταν συγκρίνουμε κατανομές με ίδια σ^2 και ίδιο n, αλλά διαφορετική κύρτωση;

Άσκηση 12. Αμεροληψία ή μικρότερο συνολικό σφάλμα;

Προαιρετική

Έστω $n \geq 2$ και $X_i \sim N(\mu, \sigma^2)$, με άγνωστα μ, σ^2. Θέτουμε $Q = \sum_i (X_i - \overline{X})^2$ και $A_d = Q/d$ για γνωστό $d > 0$.

(α) Από την Άσκηση 11 και $\mu_4 = 3\sigma^4$, βρείτε τη μέση τιμή και τη διασπορά του Q.

(β) Βρείτε τη μεροληψία και το μέσο τετραγωνικό σφάλμα της A_d για τη σ^2.

(γ) Συγκρίνετε τις περιπτώσεις $d = n-1$, $d = n$ και $d = n+1$.

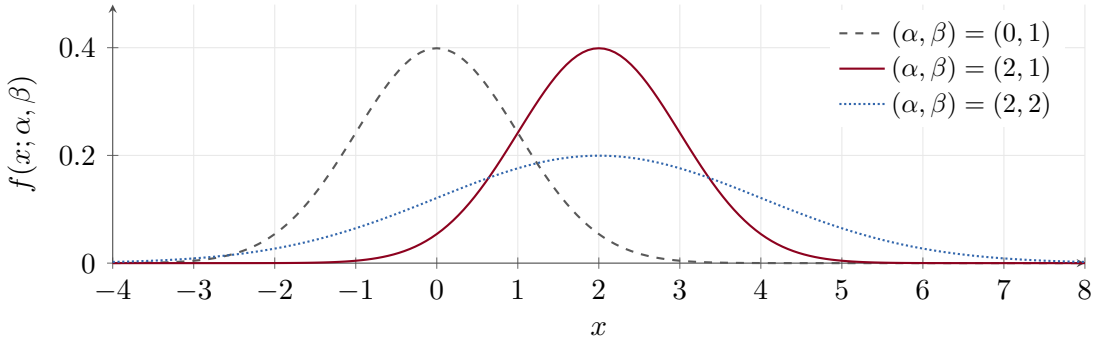
(δ) Βρείτε το $d > 0$ που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα μέσα στη συγκεκριμένη κλάση. Αυτό αποδεικνύει βελτιστότητα μεταξύ όλων των δυνατών εκτιμητριών;

Δ. Οικογένειες θέσης–κλίμακας

Άσκηση 13. Το μέλος αναφοράς δεν είναι πάντοτε τυποποιημένο

Βασική–μέση

Έστω $Z \sim K$, όπου K είναι δεδομένη κατανομή με συνάρτηση κατανομής F και πυκνότητα f. Για $\alpha \in \mathbb{R}$ και $\beta > 0$ ορίζουμε $X = \alpha + \beta Z$.



Σχήμα 2. Οικογένεια θέσης–κλίμακας με μέλος αναφοράς $N(0,1)$. Η μετατόπιση αλλάζει τη θέση, ενώ η αύξηση της κλίμακας απλώνει την πυκνότητα διατηρώντας εμβαδόν 1.

(α) Βρείτε τη συνάρτηση κατανομής $F(x;\alpha,\beta)$, την πυκνότητα $f(x;\alpha,\beta)$ και το σύνολο $S_{f(\cdot;\alpha,\beta)}$.

(β) Ποιες υποοικογένειες προκύπτουν για $\beta = 1$ και για $\alpha = 0$;

(γ) Όταν $m_0 = \mathbb{E}(Z)$ και $v_0 = \mathbb{V}(Z)$ είναι πεπερασμένα, με $v_0 > 0$, εκφράστε τις $\mu = \mathbb{E}(X)$ και $\sigma^2 = \mathbb{V}(X)$. Πότε οι α, β είναι η μέση τιμή και η τυπική απόκλιση;

(δ) Συγκρίνετε την αναγωγή $(X-\alpha)/\beta$ με την τυποποίηση $(X-\mu)/\sigma$. Εξετάστε ως μέλη αναφοράς την $N(0,1)$, την εκθετική με μέση τιμή 1 και την Cauchy της Άσκησης 8.

(ε) Όταν υπάρχουν οι αναγκαίες ροπές, μπορούν να μεταβάλλονται οι λ, κ, κ' μέσα στην ίδια οικογένεια θέσης–κλίμακας;

Άσκηση 14. Μετατοπισμένη εκθετική: παράμετροι και παρατηρήσεις

Μέση–εμβάθυνση

Έστω $X_i = \alpha + \beta Z_i$, όπου τα Z_i είναι ανεξάρτητα με πυκνότητα $e^{-z}\mathbf{1}_{[0,\infty)}(z)$. Οι άγνωστες παράμετροι ανήκουν στο $\Theta = \mathbb{R} \times (0,\infty)$.

(α) Γράψτε την πυκνότητα και βρείτε τις $\mathbb{E}(X_1), \mathbb{V}(X_1)$ ως συναρτήσεις των α, β.

(β) Για $M_2 > 0$, βρείτε τη λύση του συστήματος που εξισώνει μέση τιμή και διασπορά με $\overline{X}$ και M_2.

(γ) Για $x = (1, 2, 2, 3, 4, 6)$, υπολογίστε τη λύση. Ανήκει στο Θ; Είναι όλες οι παρατηρήσεις στο στήριγμα της προτεινόμενης πυκνότητας;

(δ) Προτείνετε διόρθωση που εξασφαλίζει ότι καμία παρατήρηση δεν είναι κάτω από το εκτιμημένο κατώτερο άκρο και διατηρεί την εξίσωση της μέσης τιμής. Διατηρείται υποχρεωτικά και η εξίσωση της διασποράς;