Προαιρετικό Quiz — Εβδομάδα 1

Εισαγωγή, τυχαίο δείγμα και εμπειρική προσέγγιση

Κάθε ερώτηση έχει τέσσερις επιλογές. Μπορεί να είναι σωστή μία ή περισσότερες. Επιλέξτε **όλες και μόνο τις σωστές**.

Οι παρατηρήσεις που χαρακτηρίζονται ως τυχαίο δείγμα είναι ανεξάρτητες και ισόνομες. Θεωρούμε $n \geq 2$. Οι μέσες τιμές, διασπορές και ροπές που χρησιμοποιούνται υποτίθενται πεπερασμένες, εκτός αν δηλώνεται διαφορετικά. Γράφουμε $\mu = \mathbb{E}(X_1)$ και $\sigma^2 = \mathbb{V}(X_1)$. Οι $M_k = n^{-1}\sum_i (X_i - \overline{X})^k$ είναι δειγματικές κεντρικές ροπές.

1. Ποια από τα παρακάτω μπορούν να αποτελέσουν δεδομένα σε ένα στατιστικό πρόβλημα;
   - ☐ **α.** Η άγνωστη παράμετρος θ πριν παρατηρηθεί οποιοδήποτε δείγμα.
   - ☐ **β.** Μια χρονοσειρά ημερήσιων τιμών μιας μετοχής.
   - ☐ **γ.** Οι χρόνοι ζωής λαμπτήρων μιας συγκεκριμένης μάρκας.
   - ☐ **δ.** Η άγνωστη κατανομή F του πληθυσμού.

2. Για τα δεδομένα $(1, 2, 3, 4)$, θέτουμε $\widehat{F}^{-}(u) = \inf\{t : \widehat{F}(t) \geq u\}$. Ποιες προτάσεις είναι σωστές;
   - ☐ **α.** Η συνήθης δειγματική διάμεσος είναι 2.5.
   - ☐ **β.** Ο αριθμός 1.5 είναι ποσοστημόριο τάξης 0.3 της εμπειρικής κατανομής.
   - ☐ **γ.** $\widehat{F}^{-}(0.5) = 2.5$.
   - ☐ **δ.** $\widehat{F}^{-}(0.3) = 2$.

3. Ποιες προτάσεις αποδίδουν σωστά την κεντρική ιδέα της στατιστικής μοντελοποίησης;
   - ☐ **α.** Το στατιστικό μοντέλο ταυτίζεται με τον πίνακα των παρατηρημένων τιμών.
   - ☐ **β.** Οι παρατηρήσεις $x_1, \ldots, x_n$ θεωρούνται πραγματοποιήσεις τυχαίων μεταβλητών $X_1, \ldots, X_n$.
   - ☐ **γ.** Ένα στατιστικό μοντέλο αποτελεί υπόθεση για την άγνωστη κατανομή του τυχαίου διανύσματος $(X_1, \ldots, X_n)$.
   - ☐ **δ.** Στη Στατιστική η κατανομή παραγωγής των δεδομένων θεωρείται πάντοτε πλήρως γνωστή.

4. Ποιες ιδιότητες αποτελούν μέρος του ορισμού ενός τυχαίου δείγματος $X_1, \ldots, X_n$;
   - ☐ **α.** Οι X_i είναι ανεξάρτητες.
   - ☐ **β.** Οι X_i έχουν μηδενική συνδιακύμανση, χωρίς άλλη απαίτηση.
   - ☐ **γ.** Οι X_i είναι ισόνομες.
   - ☐ **δ.** Η κοινή κατανομή τους είναι υποχρεωτικά κανονική.

5. Σε ένα παραμετρικό μοντέλο $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$, τι είναι το Θ;
   - ☐ **α.** Το σύνολο όλων των δυνατών δειγμάτων.
   - ☐ **β.** Το σύνολο των επιτρεπτών τιμών της παραμέτρου.
   - ☐ **γ.** Μία παράμετρος που αλλάζει από παρατήρηση σε παρατήρηση μέσα στο ίδιο δείγμα.
   - ☐ **δ.** Μία στατιστική συνάρτηση του δείγματος.

6. Έστω $X_i \sim N(\mu, \sigma_0^2)$, όπου η $\sigma_0^2 > 0$ είναι γνωστή και μόνο η μ είναι άγνωστη. Ποιες προτάσεις είναι σωστές;

   ☐ **α.** Κάθε $\mu \in \mathbb{R}$ προσδιορίζει μία υποψήφια κατανομή $N(\mu, \sigma_0^2)$.

   ☐ **β.** Ο παραμετρικός χώρος είναι $\Theta = \mathbb{R}$.

   ☐ **γ.** Το μέγεθος δείγματος n είναι άγνωστη παράμετρος της κανονικής κατανομής.

   ☐ **δ.** Και οι δύο ποσότητες μ, σ_0^2 είναι άγνωστες και μεταβάλλονται ελεύθερα στο μοντέλο.

7. Στο μοντέλο Bernoulli με παράμετρο $\theta \in (0, 1)$, ποια πρόταση ισχύει για κάθε θ;

   ☐ **α.** $\theta = \mathbb{P}_\theta(X_i = 1) = \mathbb{E}_\theta(X_i)$.

   ☐ **β.** Ο δειγματικός μέσος δεν είναι στατιστική συνάρτηση.

   ☐ **γ.** $\theta = \mathbb{P}_\theta(X_i = 0)$.

   ☐ **δ.** Η άγνωστη παράμετρος είναι το μέγεθος δείγματος n.

8. Ποιες προτάσεις για το δείγμα, το μέγεθος δείγματος και τον πληθυσμό είναι σωστές;

   ☐ **α.** Το n είναι το πλήθος των παρατηρήσεων.

   ☐ **β.** Ο πληθυσμός είναι πάντοτε ένας πλήρως καταγεγραμμένος πεπερασμένος κατάλογος ατόμων.

   ☐ **γ.** Στο πλαίσιο του μαθήματος, ο πληθυσμός μπορεί να ταυτίζεται με την άγνωστη κατανομή από την οποία προέρχεται το δείγμα.

   ☐ **δ.** Μετά την παρατήρηση, τα $x_1, \ldots, x_n$ είναι συγκεκριμένες τιμές.

9. Όταν θ είναι άγνωστη παράμετρος και U τυχαία μεταβλητή που δεν είναι συνάρτηση του δείγματος, ποιες ποσότητες είναι στατιστικές συναρτήσεις;

   ☐ **α.** $\displaystyle\sum_{i=1}^{n} X_i^2 - U$.

   ☐ **β.** $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

   ☐ **γ.** $\overline{X} - \theta$.

   ☐ **δ.** $\displaystyle\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

10. Ποιες προτάσεις διακρίνουν σωστά την εκτιμήτρια από την εκτίμηση;

    ☐ **α.** Η εκτιμήτρια $T(\mathbf{X})$ μπορεί να περιέχει την άγνωστη παράμετρο θ.

    ☐ **β.** Πριν παρατηρηθεί το δείγμα, η $T(\mathbf{X})$ είναι τυχαία μεταβλητή.

    ☐ **γ.** Μετά την παρατήρηση $\mathbf{X} = \mathbf{x}$, η $T(\mathbf{x})$ είναι μία συγκεκριμένη αριθμητική εκτίμηση.

    ☐ **δ.** Η τιμή $T(\mathbf{x})$ ονομάζεται πάντοτε παράμετρος του μοντέλου.

11. Για τα δεδομένα $(0, 1, 1, 2)$, ποια πρόταση για την εμπειρική κατανομή είναι σωστή;

    ☐ **α.** Η εμπειρική κατανομή είναι συνεχής επειδή οι παρατηρήσεις είναι αριθμοί.

    ☐ **β.** $\widehat{F}(1) = 1/2$.

    ☐ **γ.** Η εμπειρική μάζα στο 1 είναι $1/4$.

    ☐ **δ.** Η εμπειρική μάζα στο 1 είναι $1/2$ και $\widehat{F}(1) = 3/4$.

12. Ποιες αντιστοιχίσεις θεωρητικού και εμπειρικού χαρακτηριστικού είναι σωστές;

    ☐ **α.** $\mathbb{V}(X) \longrightarrow S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ ως διασπορά της εμπειρικής κατανομής.

    ☐ **β.** $\mathbb{P}(X \in A) \longrightarrow \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_A(X_i)$.

    ☐ **γ.** $\mathbb{E}(X) \longrightarrow \overline{X}$.

    ☐ **δ.** Θεωρητική διάμεσος $\longrightarrow$ δειγματική διάμεσος.

13. Για τυχαίο δείγμα από $N(\mu, \sigma^2)$ με άγνωστα μ και σ^2, ποιες προτάσεις είναι σωστές;

    ☐ **α.** Ο παραμετρικός χώρος είναι $(0, \infty)^2$.

    ☐ **β.** Ο παραμετρικός χώρος είναι $\mathbb{R} \times (0, \infty)$.

    ☐ **γ.** Η διασπορά της εμπειρικής κατανομής είναι ακριβώς η S^2.

    ☐ **δ.** Φυσικές εμπειρικές εκτιμήσεις είναι η $\overline{X}$ για τη μ και η $M_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ για τη σ^2.

14. Ζητάμε εκτιμήτρια του θ με τιμές στον παραμετρικό χώρο Θ. Ποια πρόταση είναι σωστή;

    ☐ **α.** Κάθε αλγεβρική συνάρτηση των δεδομένων είναι αποδεκτή εκτιμήτρια, ανεξάρτητα από το πεδίο τιμών της.

    ☐ **β.** Οι περιορισμοί του παραμετρικού χώρου δεν επηρεάζουν ποτέ την κατασκευή εκτιμητριών.

    ☐ **γ.** Πρέπει να ελέγχουμε αν η στατιστική συνάρτηση λαμβάνει τιμές στον επιτρεπτό χώρο της παραμέτρου.

    ☐ **δ.** Η δειγματική αναλογία Bernoulli μπορεί να είναι μικρότερη από 0 ή μεγαλύτερη από 1.

15. Ποιες προτάσεις είναι σωστές για την τυχαιότητα μιας εκτιμήτριας $\widehat{\theta} = T(\mathbf{X})$;

    ☐ **α.** Η τιμή της μπορεί να αλλάζει από δείγμα σε δείγμα.

    ☐ **β.** Παίρνει την ίδια αριθμητική τιμή σε κάθε δυνατό δείγμα, επειδή ο τύπος της είναι προκαθορισμένος.

    ☐ **γ.** Η κατανομή της μπορεί να εξαρτάται από την άγνωστη παράμετρο θ.

    ☐ **δ.** Μετά την παρατήρηση του δείγματος, η πραγματοποιημένη τιμή $T(\mathbf{x})$ παραμένει άγνωστη τυχαία μεταβλητή.

16. Αν πάρουμε επανειλημμένα διαφορετικά δείγματα ίδιου μεγέθους από το ίδιο μοντέλο, ποια πρόταση είναι σωστή;

    ☐ **α.** Η θεωρητική μέση τιμή της κατανομής αλλάζει από δείγμα σε δείγμα.

    ☐ **β.** Η παράμετρος του μοντέλου αλλάζει κάθε φορά που αλλάζει το δείγμα.

    ☐ **γ.** Οι δειγματικοί μέσοι είναι υποχρεωτικά ίσοι επειδή το n είναι ίδιο.

    ☐ **δ.** Οι παρατηρήσεις και οι τιμές των στατιστικών συναρτήσεων μπορούν να διαφέρουν από δείγμα σε δείγμα.

17. Ποιες προτάσεις για την προεπεξεργασία και τη μοντελοποίηση δεδομένων είναι σωστές;

    ☐ **α.** Η αλληλεπίδραση με τα δεδομένα μπορεί να οδηγήσει σε αναθεώρηση του αρχικού μοντέλου.

    ☐ **β.** Η μοντελοποίηση αρχίζει μόνο αφού ολοκληρωθεί οριστικά κάθε συμπέρασμα.

    ☐ **γ.** Ο καθαρισμός και ο χειρισμός χαμένων τιμών μπορούν να επηρεάσουν το τελικό αποτέλεσμα.

    ☐ **δ.** Η ίδια προεπεξεργασία είναι κατάλληλη για κάθε εφαρμογή, ανεξάρτητα από τη φύση των δεδομένων.

18. Έχουμε θετικούς χρόνους ζωής εξαρτημάτων και επιλέγουμε ως μοντέλο την εκθετική κατανομή με άγνωστη παράμετρο. Ποιες προτάσεις είναι σωστές;

    ☐ **α.** Η επιλογή εγγυάται ότι η προσαρμογή στα δεδομένα είναι τέλεια.

    ☐ **β.** Η επιλογή της εκθετικής κατανομής επιβάλλεται λογικά μόνο από το ότι οι παρατηρήσεις είναι θετικές.

    ☐ **γ.** Η επιλογή αποτελεί υπόθεση μοντελοποίησης για τον μηχανισμό παραγωγής των δεδομένων.

    ☐ **δ.** Η καταλληλότητα της υπόθεσης πρέπει να εξεταστεί σε σχέση με τα δεδομένα και με εναλλακτικά μοντέλα.

19. Για το δείγμα Bernoulli $(1, 0, 1, 1, 0)$, ποιες προτάσεις είναι σωστές;

    ☐ **α.** $\overline{x} = 3/5$.

    ☐ **β.** Η φυσική εμπειρική εκτίμηση της πιθανότητας επιτυχίας είναι $3/5$.

    ☐ **γ.** Η διασπορά της εμπειρικής κατανομής είναι $2/5$.

    ☐ **δ.** Η εμπειρική πιθανότητα της τιμής 0 είναι $2/5$.

20. Για τα δεδομένα $(1, 2, 2, 3, 4, 6)$, θέτουμε $m_2' = 6^{-1}\sum_i x_i^2$ και $m_2 = 6^{-1}\sum_i (x_i - \overline{x})^2$. Ποιες προτάσεις είναι σωστές;

    ☐ **α.** $\overline{x} = 18$.

    ☐ **β.** $m_2 = 8/3$.

    ☐ **γ.** $m_2 = m_2'$, επειδή και τα δύο υπολογίζονται από το ίδιο δείγμα.

    ☐ **δ.** $m_2' = 35/3$.