

Θεωρία Πιθανοτήτων και Στατιστική
Θέματα Εξετάσεων Ιουνίου 2025 και ενδεικτικές λύσεις

Θέμα 1

Γνωρίζουμε ότι το 1% των ατόμων σε έναν πληθυσμό έχει μία ασθένεια. Επιπλέον, γνωρίζουμε ότι το 90% από εκείνους που έχουν την ασθένεια εμφανίζουν χαμηλό επίπεδο λευκών αιμοσφαιρίων ενώ μόνο 3% από τους μη ασθενείς παρουσιάζουν το ίδιο εύρημα.

(α) Να δείξετε ότι το ποσοστό του πληθυσμού που έχει χαμηλό επίπεδο λευκών αιμοσφαιρίων είναι 3.87%.

(β) Να δείξετε ότι η πιθανότητα ένα τυχαίο άτομο του πληθυσμού που εμφανίζει το συγκεκριμένο εύρημα να έχει πράγματι την ασθένεια είναι ίση με 0.233 (ακριβής τιμή: 0.232558139534884).

(γ) Επιλέγουμε 430 άτομα με χαμηλό επίπεδο λευκών αιμοσφαιρίων από τον πληθυσμό. Να βρεθεί η πιθανότητα το πολύ 90 από αυτούς να έχουν την ασθένεια.

Ενδεικτική λύση

Έστω, $A = \{\text{ασθένεια}\}$, $X = \{\text{χαμηλό επίπεδο αιμοσφαιρίων}\}$. Από τα δεδομένα έχουμε:

$$P(A) = 0.01, P(X | A) = 0.9, P(X | A') = 0.03$$

$$(α) P(X) = P(X | A) P(A) + P(X | A') P(A') = 0.9 \cdot 0.01 + 0.03 \cdot 0.99 = 0.0387$$

$$(β) P(A | X) = P(X | A) \cdot P(A) / P(X) = 0.9 \cdot 0.01 / 0.0387 = 0.232558139534884 \approx 0.233$$

(γ) Αν $Y = \{\text{πλήθος ασθενών στα 430 άτομα}\}$ τότε είναι

$$Y \sim B(430, 0.233) \approx N(100, 76.74) \text{ (είναι } np = 100 > 5, npq = 76.74 > 5)$$

Αναζητούμε την $P_B(Y \leq 90) = P_N(Y \leq 90.5) = P_N(Z \leq (90.5 - 100) / \sqrt{76.74}) = P_N(-1.084) = \Phi(-1.084) = 1 - \Phi(1.084) = 1 - 0.85993 = 0.14007 \approx 14\%$.

Βαθμολογία: (α) 0,8μ (β) 0,7μ (γ) 1μ

Θέμα 2

Η τυχαία μεταβλητή (τ.μ.) X έχει μία σ.π.π. που δίνεται από την ακόλουθη σχέση

$$f(x) = \begin{cases} \alpha x(1 - x^2), & 0 \leq x \leq 1 \\ 0, & \text{αλλού} \end{cases}, \alpha > 0.$$

(α) Να δείξετε ότι $\alpha = 4$.

(β) Να δείξετε ότι η αθροιστική συνάρτηση κατανομής της τ.μ. X είναι η συνάρτηση

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x^2(2 - x^2), & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}.$$

(γ) Να υπολογιστεί η συνάρτηση κατανομής της τ.μ. $Y = X^2$.

Βαθμολογία: (α) 0,7μ (β) 0,8μ (γ) 1μ

Ενδεικτική λύση

(α) Πρέπει $\int_{-\infty}^{\infty} f(x) dx = 1 \Leftrightarrow \int_0^1 \alpha x(1-x^2) dx = 1 \Leftrightarrow \alpha = 4$.

(β) $F(x) = \int_{-\infty}^x f(x) dx$. Παίρνουμε τρεις περιπτώσεις για το x :

(i) $x < 0$: $F(x) = \int_{-\infty}^x f(x) dx = 0$, διότι $f(x) = 0$, για $x < 0$.

(ii) $0 \leq x \leq 1$: $F(x) = \int_{-\infty}^x f(x) dx = \int_0^x 4t(1-t^2) dt = 2x^2 - x^4 = x^2(2-x^2)$.

(iii) $x > 1$: $F(x) = \int_{-\infty}^x f(x) dx = 1$, καθώς $f(x) = 0$, για $x > 1$.

(γ) Η $Y = X^2$ λαμβάνει μόνο θετικές τιμές. Για τυχαίο $y \geq 0$, είναι:

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$\text{Όμως, } -\sqrt{y} < 0 \Rightarrow F_X(-\sqrt{y}) = 0 \text{ συνεπώς } F_Y(y) = F_X(\sqrt{y}) = \begin{cases} y(2-y) & , 0 \leq y \leq 1 \\ 1 & , y > 1 \end{cases}.$$

Θέμα 3

Γνωρίζουμε ότι Πανελλαδικά, η μέση επίδοση των μαθητών στα Μαθηματικά είναι 54 μονάδες (στα 100). Μία ομάδα 40 μαθητών από ένα σχολείο είχε μέση επίδοση 63 μονάδες με τυπική απόκλιση 3,2 μονάδες. Θέλουμε να ελέγξουμε τον ισχυρισμό ότι οι μαθητές του σχολείου έχουν σημαντικά διαφορετική επίδοση στα μαθηματικά από την αντίστοιχη πανελλαδική.

(α) Να αναφέρετε (i) το στατιστικό έλεγχο που θα χρησιμοποιήσετε και (ii) τις H_0 και H_1 .

(β) Να βρείτε αν οι μαθητές του σχολείου έχουν σημαντικά διαφορετική μέση επίδοση στα μαθηματικά από την αντίστοιχη πανελλαδική ($\alpha = 0,05$).

Βαθμολογία: (α) 0,5μ (β) 1,5μ

Ενδεικτική λύση

(α) (i) Εφαρμόζουμε τη δοκιμασία t – test για ένα δείγμα. (ii) $H_0: \mu = 54$ έναντι $H_1: \mu \neq 54$

(β) Υπολογίζουμε $t_0 = \frac{\bar{x} - 54}{s/\sqrt{40}} = \frac{63 - 54}{3.2/\sqrt{40}} = 17.787$

Είναι $t \sim t(39) \approx N(0, 1)$ και $p = P(|t| > 17.787) < 0.001$, άρα η H_0 απορρίπτεται.

Συμπεραίνουμε, ότι οι μαθητές του σχολείου έχουν σημαντικά διαφορετική μέση επίδοση στα μαθηματικά από την αντίστοιχη πανελλαδική.

Θέμα 4

Για τα δεδομένα του πίνακα:

X	0	2	4
Y	6	5	1

- α) Να γίνει το διάγραμμα διασποράς.
β) Να βρεθεί η συνδιακύμανση των X, Y.
γ) Να βρεθεί ο συντελεστής συσχέτισης των X, Y.
δ) Να βρεθεί η εξίσωση της ευθείας γραμμικής παλινδρόμησης της Y πάνω στη X.
ε) Να εκτιμηθεί η τιμή του Y για την τιμή 3 του X.
στ) Να υπολογιστούν τα ESS, TSS, RSS.
ζ) Να βρεθεί ο συντελεστής προσδιορισμού R^2 .

Βαθμολογία: (α) 0,3μ (β) 0,5μ (γ) 0,5μ (δ) 0,5μ (ε) 0,2μ (στ) 0,5μ (ζ) 0,5μ

Ενδεικτική λύση

$$(\alpha) \quad \bar{x} = \frac{6}{3} = 2, \quad \bar{y} = \frac{12}{3} = 4, \quad s_{xy} = \frac{1}{3-1} \left(\sum_{i=1}^3 x_i y_i - 3 \bar{x} \bar{y} \right) = -5.$$

$$(\beta) \quad s_x^2 = \frac{1}{3-1} \left(\sum_{i=1}^3 x_i^2 - 3 \bar{x}^2 \right) = 4, \quad s_y^2 = \frac{1}{3-1} \left(\sum_{i=1}^3 y_i^2 - 3 \bar{y}^2 \right) = 7.$$

$$(\gamma) \quad r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = -0.944.$$

$$(\delta) \quad \hat{b} = \frac{s_{xy}}{s_x^2} = -1.25 \quad \text{και} \quad \hat{a} = \bar{y} - \hat{b} \bar{x} = 6.5.$$

Η εξίσωση της ευθείας γραμμικής παλινδρόμησης είναι η

$$Y = -1.25 \cdot X + 6.5$$

(ε) Για $x = 3$, είναι $y = -1.25 \cdot 3 + 6.5 = 2.75$

(στ)

X	0	2	4
Y	6	5	1
$\hat{Y}$	6,5	4	1,5
$Y - \hat{Y}$	-0,5	1	-0,5

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 = 14, \quad ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 12.5, \quad RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 1.5$$

$$(\zeta) \quad R^2 = \frac{ESS}{TSS} = \frac{12.5}{14} = 0.893$$

Τυπολόγιο

A. Πίνακας κανονικής κατανομής: $\Phi(z) = P(Z < z)$, για $1 \leq z \leq 1,99$ όπου $Z \sim N(0, 1)$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670

Σημείωση: Για τιμές του z ανάμεσα σε δύο τιμές του πίνακα υπολογίστε το ημιάθροισμα των αντίστοιχων πιθανοτήτων.

B. Κανόνας του Bayes: $P(A | B) = P(B | A)P(A) / P(B)$

Γ. Κατανομές

$X \sim N(\mu, \sigma^2) \rightarrow EX = \mu, \text{Var}X = \sigma^2$.

$X \sim B(n, p), P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, EX = np, \text{Var}X = npq$ και $X \sim N(np, npq)$ ($np > 5, npq > 5$).

Δ. Στατιστική

Συνδιακύμανση $s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right)$

Διακύμανση $s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right)$

Συντελεστής Συσχέτισης Pearson $r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$

Συντελεστές εξίσωσης $\hat{a} = \bar{y} - \hat{b} \bar{x}, \hat{b} = \frac{s_{xy}}{s_x^2} = r_{xy} \frac{s_y}{s_x}$

$TSS = \sum_{i=1}^n (y_i - \bar{y})^2, ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$