Θεωρία Πιθανοτήτων και Στατιστική
Θέματα Εξετάσεων Ιουνίου 2024 και ενδεικτικές λύσεις

Θέμα 1
Ένας εργαζόμενος, για να πάει στη δουλειά του, χρησιμοποιεί το αυτοκίνητό του το 60% των ημερών του έτους, πάει με τα πόδια το 10% των ημερών και παίρνει το λεωφορείο το 30% του ημερών του έτους. Η πιθανότητα να καθυστερήσει είναι 3% όταν πηγαίνει με το αυτοκίνητο, 10% όταν πηγαίνει με τα πόδια και 7% όταν παίρνει το λεωφορείο.

(α) Ποια είναι η πιθανότητα ο εργαζόμενος να φτάσει αργοπορημένος στην εργασία του;

(β) Στις 365 ημέρες του χρόνου, πόσες ημέρες αναμένεται να καθυστερήσει ο εργαζόμενος;

(γ) Ποια είναι η πιθανότητα να πήρε το λεωφορείο αν έχει καθυστερήσει;

(δ) Ποια είναι η πιθανότητα να πήρε το λεωφορείο αν δεν έχει καθυστερήσει;

Βαθμολογία: (α) 0,8μ (β) 0,2μ (γ) 0,8μ (δ) 0,7μ

Ενδεικτική λύση

Α = {πήγε με το αυτοκίνητο}, Π = {πήγε με τα πόδια}, Λ = {πήγε με το λεωφορείο}
Κ = {καθυστέρησε στην άφιξή του}

Από τα δεδομένα έχουμε:

$P(A) = 0{,}6$, $P(\Pi) = 0{,}1$, $P(\Lambda) = 0{,}3$

$P(K \mid A) = 0{,}03$, $P(K \mid \Pi) = 0{,}1$, $P(K \mid \Lambda) = 0{,}07$.

(α) $P(K) = P(K \mid A)\,P(A) + P(K \mid \Pi)\,P(\Pi) + P(K \mid \Lambda)\,P(\Lambda)$
$= 0{,}03 \cdot 0{,}6 + 0{,}1 \cdot 0{,}1 + 0{,}07 \cdot 0{,}3$
$= 0{,}049 = 4{,}9\%$.

(β) Αναμενόμενο πλήθος: $365 \cdot 0{,}049 = 17{,}9$ ημέρες.

(γ) Κανόνας Bayes: $P(\Lambda \mid K) = P(K \mid \Lambda) \cdot P(\Lambda) / P(K)$
$= 0{,}07 \cdot 0{,}3 / 0{,}049$
$= 0{,}429$

(δ) Κανόνας Bayes: $P(\Lambda \mid K') = P(K' \mid \Lambda) \cdot P(\Lambda) / P(K')$
$= [1 - P(K \mid \Lambda)] \cdot P(\Lambda) / [1 - P(K)]$
$= (1 - 0{,}07) \cdot 0{,}3 / (1 - 0{,}049)$
$= 0{,}293$.

Θέμα 2
Για την τ.μ. $X \sim B(3, \theta)$ πήραμε ένα δείγμα μεγέθους 4 (δηλαδή 4 φορές επαναλήφθηκε η σειρά των 3 δοκιμασιών) και βρήκαμε τις τιμές 2, 3, 2, 0.
(α) Να δείξετε ότι η συνάρτηση πιθανοφάνειας των παραπάνω παρατηρήσεων είναι η

$$L(\theta) = 9\theta^7(1 - \theta)^5.$$

(β) Να δείξετε ότι ο εκτιμητής μέγιστης πιθανοφάνειας για την παράμετρο θ είναι ο 7/12.

Βαθμολογία: (α) 1,3μ (β) 1,2μ

Ενδεικτική Λύση

(α) Γνωρίζουμε ότι αν $X \sim B(3, \theta)$, τότε $P(X = k) = (3 \text{ ανά } k)\theta^{\kappa}(1 - \theta)^{3-k}$.

Άρα,
$P(X = 0) = (3 \text{ ανά } 0)(1 - \theta)^3 = (1 - \theta)^3$,
$P(X = 2) = (3 \text{ ανά } 2)\theta^2(1 - \theta) = 3\,\theta^2(1 - \theta)$,
$P(X = 3) = (3 \text{ ανά } 3)\theta^3 = \theta^3$.

Η συνάρτηση πιθανοφάνειας είναι $L(\theta) = P(X = 0)P(X = 2)^2P(X = 3) = 9\theta^7(1 - \theta)^5$,

(β) Είναι $l(\theta) = \ln(L(\theta)) = \ln(9) + 7\ln(\theta) + 5\ln(1 - \theta)$.

Επιπλέον, $l'(\theta) = 7/\theta - 5/(1-\theta)$ και $l'(\theta) = 0 \leftrightarrow \theta = 7/12$.

Παρατηρώντας τον πίνακα προσήμου της 1ης παραγώγου διαπιστώνουμε πως το 7/12 είναι μέγιστο. Άρα η τιμή 7/12 είναι ο εκτιμητής μέγιστης πιθανοφάνειας για την παράμετρο θ.

Θέμα 3
Είναι γνωστό ότι το μέσο πλήθος αριθμητικών σφαλμάτων σε μία φορολογική δήλωση είναι 1,2 και πως η εμφάνιση ενός σφάλματος δεν συσχετίζεται με την εμφάνιση ενός άλλου.
(α) Να δείξετε ότι η πιθανότητα μία δήλωση να έχει τουλάχιστον 2 αριθμητικά λάθη είναι ίση με 0,337.

(β) Στην εφορεία της Ξάνθης υποβάλλονται κάθε χρόνο 27.000 φορολογικές δηλώσεις. Να βρείτε το αναμενόμενο πλήθος και την τυπική απόκλιση του πλήθους των δηλώσεων με τουλάχιστον 2 αριθμητικά λάθη.

(γ) Κάθε εφοριακός αναλαμβάνει να ελέγξει το πολύ 2.000 δηλώσεις. Να βρείτε την πιθανότητα μεταξύ 2.000 φορολογικών δηλώσεων να υπάρξουν περισσότερες από 700 δηλώσεις με τουλάχιστον 2 αριθμητικά λάθη.

Βαθμολογία: (α) 0,7μ (β) 0,7μ (γ) 1,1μ

Ενδεικτική λύση
(α) Έστω Σ = {πλήθος σφαλμάτων σε μία φορολογική δήλωση}. Η εμφάνιση ενός σφάλματος δεν συσχετίζεται με την εμφάνιση ενός άλλου, άρα Σ ~ Poisson(1,2).

Υπολογίζουμε:

$$P(\Sigma \geq 2) = 1 - P(\Sigma < 2) = 1 - P(\Sigma = 0) - P(\Sigma = 1) = 1 - e^{-1,2} - 1,2e^{-1,2} = 0,337.$$

(β)
Αν Χ = {το πλήθος των δηλώσεων με τουλάχιστον 2 αριθμητικά λάθη στις 27.000} τότε

$$X \sim B(27.000,\ 0,337)$$

οπότε $EX = n \cdot p = 27.000 \cdot 0,337 = 9.099$ και $VarX = n \cdot p \cdot q = 27.000 \cdot 0,337 \cdot 0,663 = 6.032,6$, άρα η τυπική απόκλιση είναι $6.032,6^{0,5} = 77,7$ δηλώσεις.

(γ)Έστω τώρα
Y = {πλήθος δηλώσεων με τουλάχιστον 2 αριθμητικά λάθη στις 2.000 δηλώσεις}.

Είναι

$$Y \sim B(2000,\ 0.337) \approx N(674,\ 446.9) \text{ (είναι } np = 674 > 5,\ npq = 446,9 > 5)$$

Αναζητούμε την $P_B(Y > 700) = P_N(Y > 699,5) = P_N(Z > (699,5 - 674)/446,9^{0,5}) = P_N(Z > 1,206) = 1 - \Phi(1,206) = 1 - 0.885 = 0,115 = 11,5\%$.

Θέμα 4

Μία έρευνα, είχε ως στόχο τον εντοπισμό ενός τύπου που θα εκτιμά την ηλικία (y σε έτη) των μεγάλων αιωνόβιων δέντρων από την περιφέρειά τους (x σε μέτρα). Τα δεδομένα από 15 δέντρα συνοψίζονται με τις ακόλουθες πληροφορίες:

$$\sum_{i=1}^{15} x_i = 138, \quad \sum_{i=1}^{15} x_i^2 = 1.314, \quad \sum_{i=1}^{15} y_i = 4.095, \quad \sum_{i=1}^{15} y_i^2 = 1.124.949, \quad \sum_{i=1}^{15} x_i y_i = 38.176.$$

(α) Να βρεθεί η συνδιακύμανση των x, y.

(β) Να βρεθεί ο συντελεστής συσχέτισης των x, y.

(γ) Να βρεθεί γραμμικό μοντέλο πρόβλεψης της y από την x.

(δ) Να εκτιμηθεί η ηλικία ενός δέντρου με περίμετρο 15 μέτρα

Βαθμολογία: (α) 0,7μ (β) 0,7μ (γ) 0,7μ (δ) 0,4μ

Ενδεικτική λύση

(α) $\bar{x} = \frac{138}{15} = 9{,}2$, $\bar{y} = \frac{4.095}{15} = 273$, $s_{XY} = \frac{1}{15-1}\left(\sum_{i=1}^{15} x_i y_i - 15\,\bar{x}\,\bar{y}\right) = 35{,}9.$

(β) $s_X^2 = \frac{1}{15-1}\left(\sum_{i=1}^{15} x_i^2 - 15\,\bar{x}^2\right) = 3{,}17$, $s_Y^2 = \frac{1}{15-1}\left(\sum_{i=1}^{15} y_i^2 - 15\,\bar{y}^2\right) = 501.$

$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y} = 0{,}9.$

(γ) $\hat{b} = \frac{s_{XY}}{s_X^2} = \frac{35{,}86}{3{,}17} = 11{,}3$ και $\hat{\alpha} = \bar{y} - \hat{b}\,\bar{x} = 169.$

Η εξίσωση της ευθείας γραμμικής παλινδρόμησης είναι η

$$Y = 11{,}3 \cdot X + 169$$

(δ) Για x = 15, είναι y = 11,3*15 + 169 = 338,5 έτη.

Τυπολόγιο

Α. Πίνακας κανονικής κατανομής: $\Phi(z) = P(Z < z)$, για $1 \le z \le 1{,}99$ 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.

Β. Κανόνας του Bayes: $P(A \mid B) = P(B \mid A)P(A) / P(B)$

Γ. Κατανομές

$X \sim N(\mu, \sigma^2) \rightarrow EX = \mu$, $VarX = \sigma^2$.

$X \sim Poisson(\lambda)$, $P(X = k) = e^{-\lambda} \lambda^k/k!$, $EX = VarX = \lambda$ και για λ αρκετά μεγάλο $X \sim N(\lambda, \lambda)$.

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

Δ. Στατιστική

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

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

Συντελεστής Συσχέτισης Pearson $r_{XY} = \dfrac{s_{XY}}{s_X \cdot s_Y}$

Συντελεστές εξίσωσης $\hat{\alpha} = \bar{y} - \hat{b}\bar{x}$, $\hat{b} = \dfrac{s_{XY}}{s_X^2} = r_{XY}\dfrac{s_Y}{s_X}$