ΣΑΠΗΛΟguide

Θεωρία πιθανοτήτων και στατιστική

Σημείωση: Οι απαντήσεις μπορεί να μην είναι 100% σωστές

—Nontas

Λύσεις Σεπτεμβρίου 2022

Μηχανικών και Μηχανικών Υπολογιστών

Θέμα 1ο

Ένας ηλεκτρικός πίνακας περιέχει 25 ασφάλειες από τις οποίες οι 5 είναι καμμένες. Αφαιρούμε τυχαία μια ασφάλεια και στη συνέχεια μια άλλη. Να υπολογιστούν οι πιθανότητες:

(α) Τουλάχιστον μια από τις δύο ασφάλειες ήταν καμμένη.

(β) Ακριβώς μια από τις δύο ασφάλειες ήταν καμμένη.

Θέμα 2ο [2 μονάδες]

α) • Πιθανότητα <u>καμία</u> ασφάλεια να ήταν καμμένη:

$$\frac{20}{25}\cdot\frac{19}{24}=\frac{19}{30} \quad \left(\text{ή} \quad \frac{\binom{20}{2}}{\binom{25}{2}}=\frac{19}{30} \right)$$

Άρα τουλάχιστον μια καμμένη: $1-\frac{19}{30}=\frac{11}{30}$

β) • Πιθανότητα <u>και οι 2 καμμένες</u>: $\frac{5}{25}\cdot\frac{4}{24}=\frac{1}{30} \quad \left(\text{ή} \quad \frac{\binom{5}{2}}{\binom{25}{2}}=\frac{1}{30} \right)$

• Άρα ακριβώς μια καμμένη: $1-\frac{19}{30}-\frac{1}{30}=\frac{10}{30}=\frac{1}{3}$

Θέμα 2ο **[2 μονάδες]**

Ένας χρήστης υπολογιστικής υποδομής μπορεί να χρησιμοποιήσει 3 υπερυπολογιστές, A, B, C, για να τρέξει προγράμματα και η πιθανότητα να είναι ο καθένας από αυτούς διαθέσιμος (για οποιαδήποτε χρονική στιγμή) είναι P(A) = 0.5, P(B) = 0.7 και P(C) = 0.6. Υποθέτουμε πως οι διαθεσιμότητες των υπερυπολογιστών είναι ανεξάρτητες μεταξύ τους. Θεωρείστε την τυχαία μεταβλητή (τ.μ.) X να είναι το πλήθος των διαθέσιμων υπερυπολογιστών.

(α) Υπολογίστε την συνάρτηση μάζας πιθανότητας της τ.μ. X και κάνετε το γράφημα της.

(β) Υπολογίστε την πιθανότητα να είναι τουλάχιστον ένας υπερυπολογιστής διαθέσιμος.

(γ) Ποια από τις παρακάτω τρεις κατανομές μπορεί να περιγράψει ικανοποιητικά την κατανομή της τ.μ. X που βρήκατε στο (α): i) διακριτή ομοιόμορφη, 2) γεωμετρική, 3) διωνυμική; Αιτιολογείστε την απάντηση σας και δώστε προσεγγιστικά την τιμή της παραμέτρου ή τις τιμές των παραμέτρων της προτεινόμενης κατανομής.

Θέμα 3ο **[2 1/2 μονάδες]**

Η διάρκεια ζωής

α) • Κανένας διαθέσιμος: $0,5 \cdot 0,3 \cdot 0,4 = 0,06$ $(X=0)$

• A διαθέσιμος: $0,5 \cdot 0,3 \cdot 0,4 = 0,06$

• B διαθέσιμος: $0,5 \cdot 0,7 \cdot 0,4 = 0,14$

• Γ διαθέσιμος: $0,5 \cdot 0,3 \cdot 0,6 = 0,09$

$(X=1)$: $0,06 + 0,14 + 0,09 = 0,29$

• A, B διαθέσιμοι: $0,5 \cdot 0,7 \cdot 0,4 = 0,14$

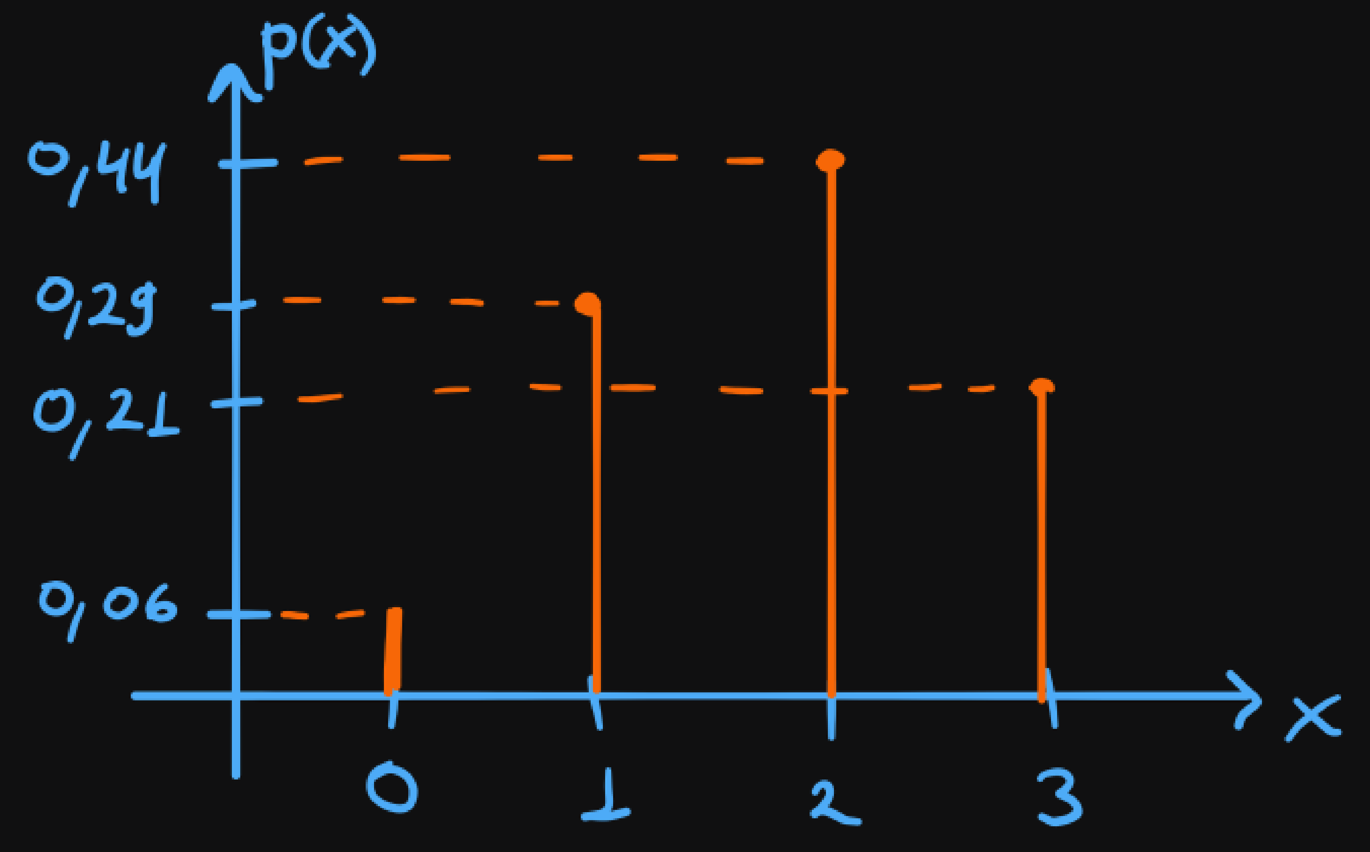
• B, Γ διαθέσιμοι: $0,5 \cdot 0,7 \cdot 0,6 = 0,21$

• A, Γ διαθέσιμοι: $0,5 \cdot 0,3 \cdot 0,6 = 0,09$

$(X=2)$: $0,14 + 0,21 + 0,09 = 0,44$

• Όλοι διαθέσιμοι: $0,5 \cdot 0,7 \cdot 0,6 = 0,21$ $(X=3)$

Άρα $$f_X(x) = \begin{cases} 0,06, & x=0 \\ 0,29, & x=1 \\ 0,44, & x=2 \\ 0,21, & x=3 \end{cases}$$ και



β) $P(X \geq 1) = 1 - P(X=0) = 1 - 0,06 = 0,94$

γ) • Η κατανομή που περιγράφει την τ.μ. X σε ικανοποιητικό βαθμό είναι η διωνυμική λόγω του σχήματος καμπάνα (έχουμε υψηλή πιθανότητα για $X=1,2$ και χαμηλότερη για $X=0,3$)

• Το συγκεκριμένο σχήμα μπορούμε να το πετύχουμε για μέσες τιμές του p (π.χ. $p=0,6$) και $N=3$

(δεν είμαι σίγουρος αν αυτό θέλει)

Θέμα 3° [2 1/2 μονάδες]

Η διάρκεια ζωής T ενός τρανζίστορ είναι τυχαία μεταβλητή (τ.μ.) με συνάρτηση πυκνότητας πιθανότητας (t σε μήνες)

$$f_T(t) = \begin{cases} ce^{-\frac{t}{8}}, & t > 0 \\ 0, & t \le 0 \end{cases}$$

Να υπολογιστούν:

(α) Η μέση τιμή και η τυπική απόκλιση της τ.μ. T.

(β) Η πιθανότητα ότι ο χρόνος ζωής του τρανζίστορ είναι μεταξύ 6 και 10 μηνών.

[2 μονάδες]

Θέμα 4°

α) • $\int_{-\infty}^{+\infty} f_T(t)\,dt = 1 \Leftrightarrow c\int_0^{+\infty} e^{-\frac{t}{8}}\,dt = 1$

$\Leftrightarrow -8c\left[e^{-\frac{t}{8}}\right]_0^{+\infty} = 1 \Leftrightarrow -8c(0-1) = 1$

$\Leftrightarrow 8c = 1 \Leftrightarrow \underline{c = \frac{1}{8}}$

• Έχουμε $f_T(t) = \begin{cases} \frac{1}{8}e^{-\frac{t}{8}}, & t>0 \\ 0, & t\le 0 \end{cases}$ εκθετική κατανομή με $\lambda = \frac{1}{8}$

οπότε $E(T) = \frac{1}{\lambda} \Leftrightarrow E(T) = 8$

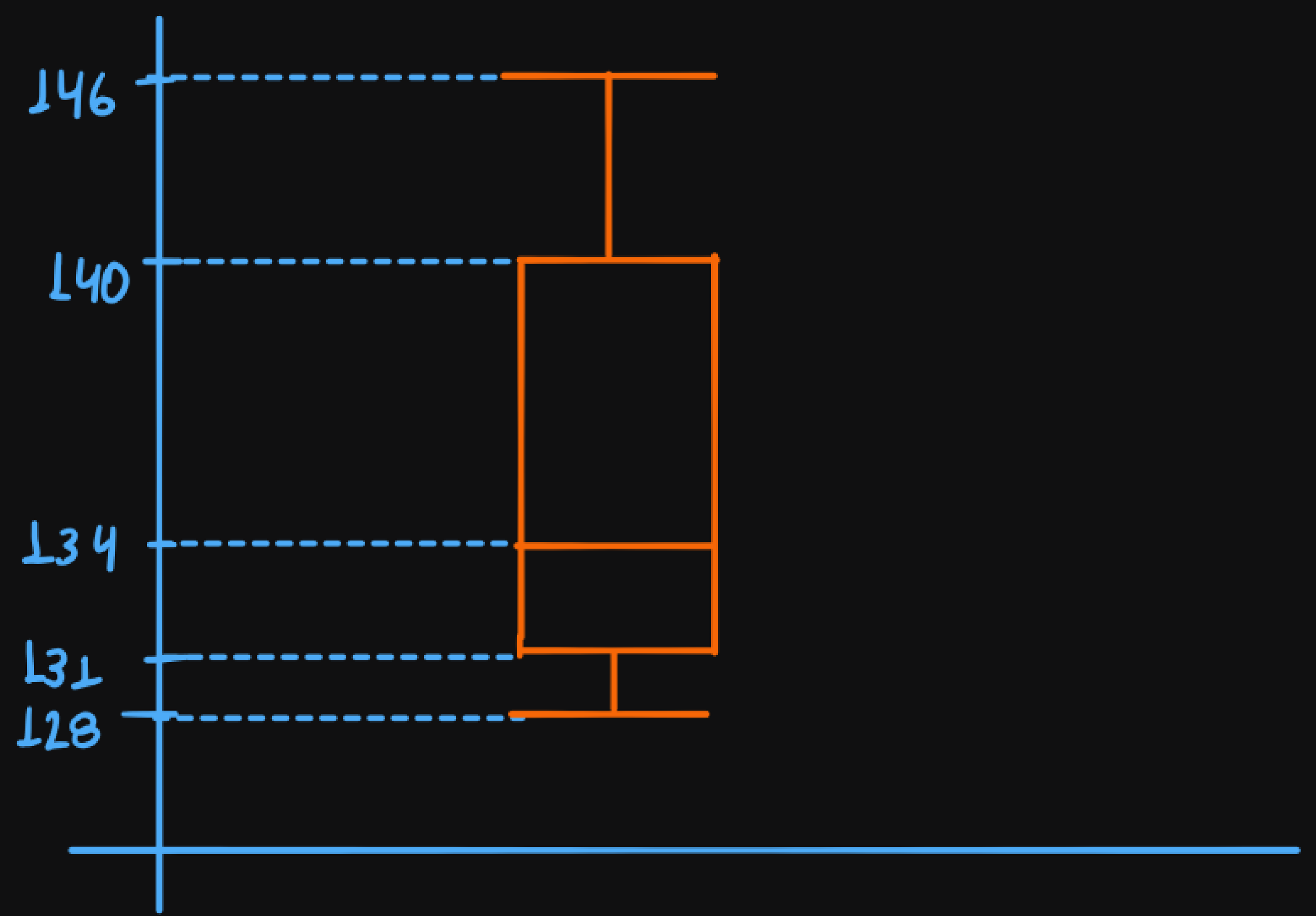
και $Var(T) = \frac{1}{\lambda^2} \Leftrightarrow Var(T) = 64 \Leftrightarrow \sigma = 8$

β) • $F_T(t) = \int_{-\infty}^{t} f_T(t)\,dt = \int_0^t \frac{1}{8}e^{-\frac{t}{8}} = \left[-e^{-\frac{t}{8}}\right]_0^t = -e^{-\frac{t}{8}} + 1 = 1 - e^{-\frac{t}{8}}$

• $P(6<T<10) = F_T(10) - F_T(6) = 1 - e^{-\frac{10}{8}} - \left(1 - e^{-\frac{6}{8}}\right) = e^{-\frac{3}{4}} - e^{-\frac{5}{4}} \simeq 0{,}185$

α) 128 130 131 133 134 138 140 142 146

128	130	131	133	134	138	140	142	146
min		Q_1		$\tilde{x}$		Q_3		max



[2 μονάδες]

Θέμα 4ο

Μετρήθηκε η χωρητικότητα (σε αμπερώρες) 9 μπαταριών και τα αποτελέσματα δίνονται ως εξής:

130	146	140	134	138	142	128	131	133

Από παλιότερες μετρήσεις γνωρίζουμε ότι η χωρητικότητα της μπαταρίας ακολουθεί κανονική κατανομή.

(α) Με βάση το θηκόγραμμα, που θα πρέπει να σχηματίσετε από τις παρατηρήσεις του δείγματος, σχολιάστε αν η χωρητικότητα της μπαταρίας ακολουθεί πράγματι κανονική κατανομή.

(β) Σε ποιο επίπεδο εμπιστοσύνης το διάστημα εμπιστοσύνης για τη μέση χωρητικότητα της μπαταρίας μπορεί να έχει ως τιμή του άνω άκρου του 140.5 αμπερώρες; Εξηγείστε πως φτάνετε στην απάντηση σας (για τις αριθμητικές πράξεις χρησιμοποιείστε στρογγυλοποίηση στο πρώτο δεκαδικό).

[2 μονάδες]

Παρατηρούμε ότι:
- Η διάμεσος αποκλίνει προς το 1ο τεταρτημόριο
- Τα μήκη των 2 μύστακων έχουν διαφορετικό μέγεθος
- Δεν υπάρχουν ακραίες τιμές

Με πολύ μικρή βεβαιότητα μπορούμε να θεωρήσουμε ότι ακολουθεί κανονική κατανομή.

Χρειαζόμαστε περισσότερες μετρήσεις για να γίνει εμφανής η κανονική κατανομή.

β)
- $\bar{x} = 135,8$
- $s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{9} x_i^2 - n\bar{x}^2\right) = \frac{1}{8}\left(166214 - 9 \cdot 135,8^2\right) = 29,905 \Leftrightarrow s = 5,46 \simeq 5,5$
- $\bar{x} \pm t_{1-\frac{a}{2},n-1} \cdot \frac{s}{\sqrt{n}} \xLeftrightarrow[=140,5]{\text{άνω άκρο}} \bar{x} + t_{1-\frac{a}{2},8} \cdot \frac{s}{\sqrt{n}} = 140,5 \Leftrightarrow t_{1-\frac{a}{2},8} = \frac{3(140,5-135,8)}{5,5} = 2,56 \simeq 2,6$
- Η τιμή 2,6 είναι μεταξύ 2,306 και 2,896 (περίπου στο μέσο), άρα θέλουμε $1-\frac{a}{2} = 0,9825 \left(0,975 + \frac{0,99-0,975}{2}\right)$

Οπότε $1-\frac{a}{2} = 0,9825 \Leftrightarrow \frac{a}{2} = 0,0175 \Leftrightarrow a = 0,035 \Leftrightarrow 1-a = 0,965 \rightarrow$ Διάστημα εμπιστοσύνης 96,5%

Θέμα 5°

Η απευθείας μέτρηση της ποσότητας πρωτεΐνης στο συκώτι είναι δύσκολη και απαιτεί πολύ χρόνο. Πιστεύεται ότι η ποσότητα της πρωτεΐνης σχετίζεται με την ποσότητα του φωτός που απορροφάται από το συκώτι. Για αυτό έγινε στο εργαστήριο το ακόλουθο πείραμα. Στάλθηκε από ένα φασματόμετρο φως σε διάλυμα που περιείχε δείγμα συκωτιού και μετρήθηκε το φως που απορροφήθηκε. Αυτή η διαδικασία εφαρμόσθηκε σε 6 δείγματα συκωτιού για τα οποία η ποσότητα της πρωτεΐνης ήταν γνωστή. Τα αποτελέσματα δίνονται στον παρακάτω πίνακα.

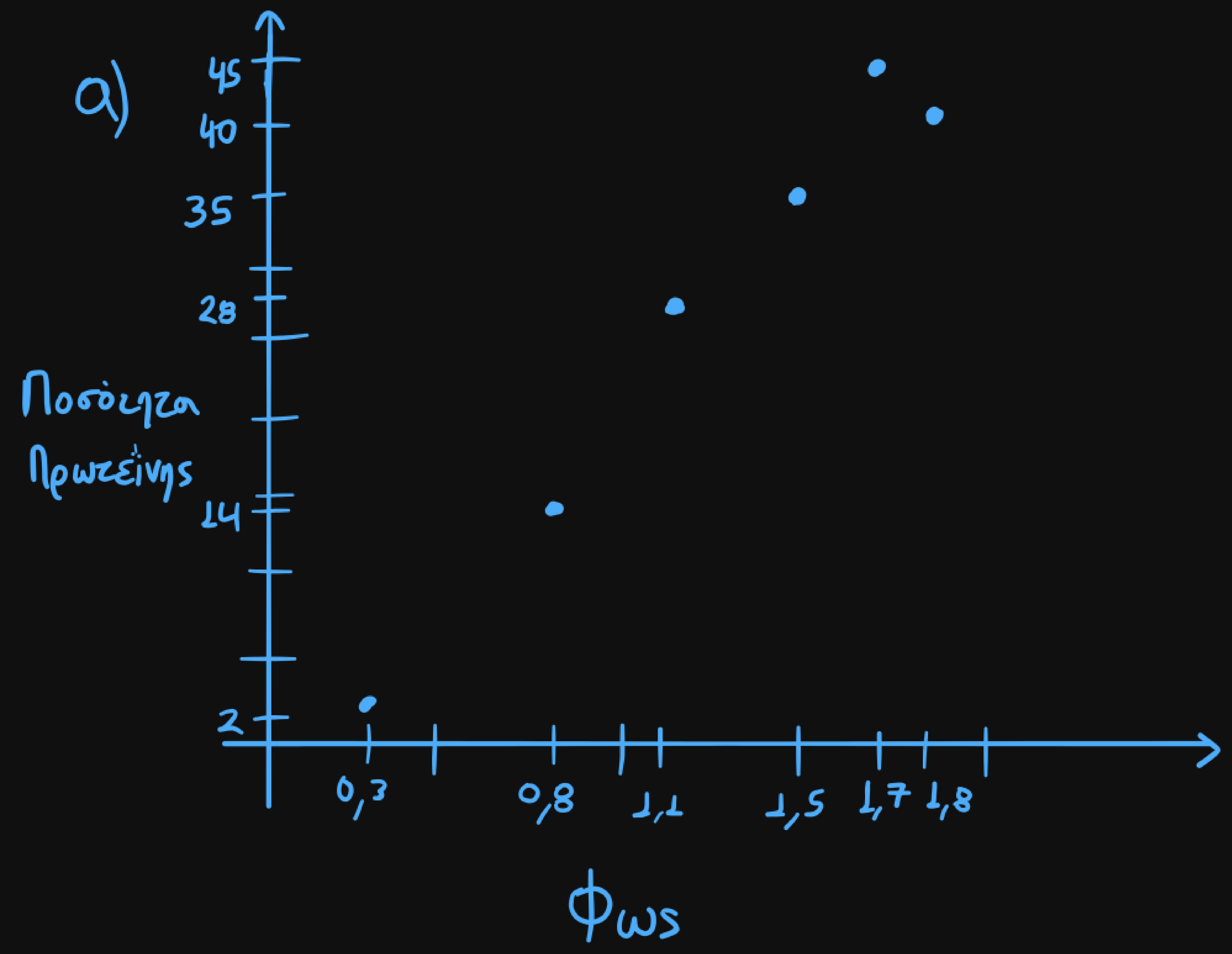
Φως που απορροφήθηκε	0.3	0.8	1.1	1.5	1.7	1.8
Ποσότητα πρωτεΐνης (mg)	2	14	28	35	45	40

(α) Σχηματίστε το κατάλληλο διάγραμμα διασποράς και σχολιάστε αν η υπόθεση, ότι η ποσότητα πρωτεΐνης σε δείγμα συκωτιού εξαρτάται γραμμικά από το αντίστοιχο φως που απορροφάται από το δείγμα, φαίνεται σωστή με βάση το δείγμα των παρατηρήσεων από τον πίνακα.

(β) Χρησιμοποιώντας την εκτίμηση της ευθείας παλινδρόμησης (με τη μέθοδο των ελάχιστων τετραγώνων), προβλέψτε την ποσότητα πρωτεΐνης σε δείγμα συκωτιού για το οποίο η απορρόφηση του φωτός είναι 1.6.

[1 1/2 μονάδες]

α)



• Υπάρχει ισχυρή γραμμική εξάρτηση με βάση τον πίνακα ($r > 0,9$).

β) • $S_{xy} = \frac{1}{n-1}\left(\sum_{i=1}^{6} x_i y_i - n\bar{x}\bar{y}\right) = \frac{1}{5}\left(243,6 - 6\cdot 1,2\cdot 27,3\right) \Leftrightarrow S_{xy} = 9,408$

• $S_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{6} x_i^2 - n\bar{x}^2\right) = \frac{1}{5}\left(10,32 - 6\cdot 1,2^2\right) \Leftrightarrow S_x^2 = 0,336$

• $b = \frac{S_{xy}}{S_x^2} \Leftrightarrow b = 28$

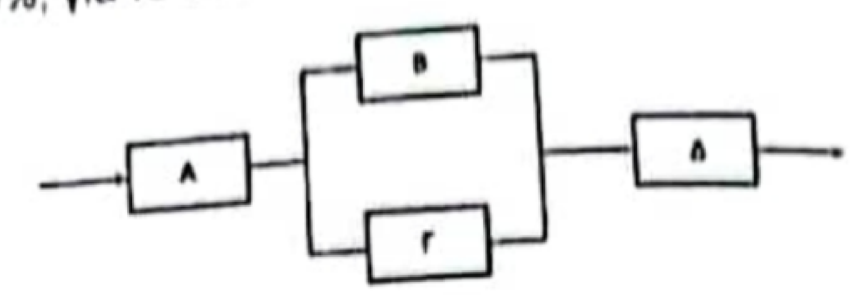
• $a = \bar{y} - b\bar{x} = 27,3 - 28\cdot 1,2 \Leftrightarrow a = -6,3$

Άρα $\boxed{y = -6,3 + 28x}$

• Για $x = 1,6$: $y = -6,3 + 28\cdot 1,6 \Leftrightarrow \boxed{y = 38,5}$

Λύσεις Ιουνίου 2022

Θέμα 1ο
Ένα ηλεκτρικό σύστημα αποτελείται από τα τμήματα Α, Β, Γ και Δ και λειτουργεί αν λειτουργούν το τμήμα Α, τουλάχιστον ένα από τα Β και Γ, και το τμήμα Δ, όπως φαίνεται στο παρακάτω σχήμα. Τα τέσσερα τμήματα λειτουργούν ανεξάρτητα και η αξιοπιστία τους είναι για το Α 90%, για το Β 85%, για το Γ 90% και για το Δ 95%, όπου αξιοπιστία είναι η πιθανότητα να λειτουργεί.



Να υπολογιστούν:
(α) Η πιθανότητα να λειτουργεί το σύστημα.
(β) Η πιθανότητα το τμήμα Γ να μην λειτουργεί όταν το σύστημα λειτουργεί.
[2 μονάδες]

α) • $P(A)=0,9 \quad P(B)=0,85 \quad P(\Gamma)=0,9 \quad P(\Delta)=0,95$

• $P(\Lambda) = P\big(A\cap(B\cup\Gamma)\cap\Delta\big) =$

$= P(A)\cdot P(B\cup\Gamma)\cdot P(\Delta) =$

$= P(A)\cdot\big(P(B)+P(\Gamma)-P(B\cap\Gamma)\big)\cdot P(\Delta) =$

$= P(A)\cdot\big(P(B)+P(\Gamma)-P(B)\cdot P(\Gamma)\big)\cdot P(\Delta) = 0,9\cdot(0,85+0,9-0,85\cdot 0,9)\cdot 0,95$

$\Leftrightarrow P(\Lambda)=0,842$

μπορείτε να το ψάξετε στο google ως absorption law

Γ λειτουργεί → δεν χρειάζεται

β) • $P(\bar{\Gamma}|\Lambda) = 1 - P(\Gamma|\Lambda)$

• $P(\Gamma|\Lambda) = P\big(\Gamma\,|\,A\cap(B\cup\Gamma)\cap\Delta\big) = \dfrac{P(\Gamma\cap A\cap\cancel{(B\cup\Gamma)}\cap\Delta)}{P(A\cap(B\cup\Gamma)\cap\Delta)} = \dfrac{P(\Gamma\cap A\cap\Delta)}{P(A\cap(B\cup\Gamma)\cap\Delta)} = \dfrac{P(\Gamma)\cdot\cancel{P(A)}\cdot\cancel{P(\Delta)}}{\cancel{P(A)}\cdot P(B\cup\Gamma)\cdot\cancel{P(\Delta)}} =$

$= \dfrac{P(\Gamma)}{P(B\cup\Gamma)} = \dfrac{P(\Gamma)}{P(B)+P(\Gamma)-P(B\cap\Gamma)} = \dfrac{P(\Gamma)}{P(B)+P(\Gamma)-P(B)\cdot P(\Gamma)} = \dfrac{0,9}{0,85+0,9-0,85\cdot 0,9} \Leftrightarrow P(\Gamma|\Lambda)=0,913$

Άρα, $P(\bar{\Gamma}|\Lambda) = 1-0,913 \Leftrightarrow P(\bar{\Gamma}|\Lambda)=0,087$

Β' τρόπος

• $P(\bar{\Gamma}|\Lambda) = \dfrac{P(\bar{\Gamma})\cdot P(\Lambda|\bar{\Gamma})}{P(\Lambda)} = \dfrac{P(\bar{\Gamma})\cdot\big(P(A)\cdot P(B)\cdot P(\Delta)\big)}{P(\Lambda)} = \dfrac{0,1\cdot(0,9\cdot 0,85\cdot 0,95)}{0,842} \Leftrightarrow \boxed{P(\bar{\Gamma}|\Lambda)=0,0863}$

Θέμα 2°

Έστω ότι η πιθανότητα να παρουσιάσει τεχνικό πρόβλημα μια ηλεκτρική μηχανή κατά της διάρκεια συνεχούς λειτουργίας της σε μια ημέρα είναι 0.01. Μας ενδιαφέρει αν η μηχανή παρουσιάζει τεχνικό πρόβλημα κατά τη διάρκεια συνεχούς λειτουργίας σε ένα μήνα (γίνεται επιθεώρηση για εντοπισμό τεχνικών προβλημάτων με τη συμπλήρωση του μήνα). Αν βρεθεί η μηχανή να έχει τεχνικό πρόβλημα αντικαθίσταται από μια άλλη (την πρώτη ημέρα του μήνα). Υποθέτοντας ότι ο μήνας έχει 30 ημέρες να βρεθούν τα ακόλουθα:

(α) Η πιθανότητα η μηχανή να μην παρουσιάσει τεχνικό πρόβλημα σε ένα μήνα συνεχούς καθημερινής λειτουργίας.

(β) Η πιθανότητα η μηχανή να παρουσιάσει για πρώτη φορά τεχνικό πρόβλημα τον τέταρτο μήνα συνεχούς καθημερινής λειτουργίας (εντοπίζεται τεχνικό πρόβλημα στην επιθεώρηση που γίνεται με τη συμπλήρωση του τέταρτου μήνα).

(γ) Η πιθανότητα να πρέπει να γίνει αντικατάσταση μηχανής ακριβώς για δεύτερη φορά στους επτά μήνες (δηλαδή να παρουσιαστεί τεχνικό πρόβλημα για δεύτερη φορά στην έβδομη επιθεώρηση).

(δ) Σε ένα σύστημα λειτουργούν 10 τέτοιες ηλεκτρικές μηχανές. Ποια είναι η πιθανότητα να παρουσιάσουν τεχνικό πρόβλημα δύο από αυτές στο τέλος του μήνα (στην επιθεώρηση των 10 μηχανών που γίνεται με τη συμπλήρωση ενός μήνα συνεχούς καθημερινής λειτουργίας).

Βοήθημα: 1. Η συνάρτηση μάζας πιθανότητας (σμπ) της γεωμετρικής κατανομής της τ.μ. *X* που μετράει τις αποτυχίες πριν την πρώτη επιτυχία είναι $P(X=x)=(1-p)^x p, \quad x\in\{0,1,2,\dots\}$. Αντίστοιχα η σμπ της γεωμετρικής κατανομής της τ.μ. *Y=X+1* που μετράει τις προσπάθειες ως και την πρώτη επιτυχία είναι $P(Y=y)=(1-p)^{y-1}p, \quad y\in\{1,2,\dots\}$.

2. Η συνάρτηση μάζας πιθανότητας (σμπ) της αρνητικής διωνυμικής κατανομής της τ.μ. *X* που μετράει τις αποτυχίες πριν την *r*-στή επιτυχία είναι $P(X=x)=\binom{x+r-1}{r-1}p^r(1-p)^x, \quad x\in\{0,1,2,\dots\}$. Αντίστοιχα η σμπ της αρνητικής διωνυμικής κατανομής της τ.μ. *Y=X+r* που μετράει τις προσπάθειες ως και την *r*-στή επιτυχία είναι

$$P(Y=y)=\binom{y-1}{r-1}p^r(1-p)^{y-r}, \quad y\in\{r,r+1,r+2,\dots\}.$$

[2 1/2 μονάδες]

200 ώρες. Εάν ένας

α) • P(Να παρουσιάσει πρόβλημα σε 1 μέρα) $= p = 0{,}01$

• P(Να μην παρουσιάσει πρόβλημα σε 30 μέρες) =

$$= 0{,}99^{30} = 0{,}739$$

Σημείωση: στα ερώτημα β, γ, δ θεωρήσαμε ότι μπορεί να εμφανιστεί πρόβλημα μόνο την $\underbrace{\text{1η ημέρα κάθε μήνα}}_{=\text{στην επιθεώρηση}}$. (μπορεί να μην είναι σωστό)

β) • P(Πρόβλημα στον 4ο μήνα) $= P(X=120) =$

$$= (1-0{,}01)^{120}\cdot 0{,}01 = 3\cdot 10^{-3}$$

γ) P(2η φορά πρόβλημα στους 7 μήνες) $= P(X=210-1)$

$$= P(X=209) = \binom{209+2-1}{2-1} p^2 (1-p)^{209}$$

$$= \frac{210!}{1!\,(210-1)!}\,0{,}01^2\cdot 0{,}99^{209} = 210\cdot 0{,}01^2\cdot 0{,}99^{209} = 2{,}57\cdot 10^{-3}$$

πιθανότητα να εμφανίσει πρόβλημα την 31η μέρα
= μέρα που γίνεται έλεγχος

δ) • Αρχικά θα πάρουμε γεωμετρική με $X=30$: $P(X=30) = (1-p)^{30}\cdot p = 0{,}99^{30}\cdot 0{,}01 = 7{,}39\cdot 10^{-3}$

• Διωνυμική: $P(X=2) = \binom{10}{2}(7{,}39\cdot 10^{-3})^2\cdot(1-7{,}39\cdot 10^{-3})^8 = 2{,}315\cdot 10^{-3}$

$P(Y = y) = \binom{\;\;}{r-1} p \,(1-p)$

Θέμα 3ο

(α) Η διάρκεια ζωής X ενός ηλεκτρικού εξαρτήματος ακολουθεί κανονική κατανομή με μέση τιμή 200 ώρες. Εάν ένας αγοραστής απαιτεί τουλάχιστον το 90% των εξαρτημάτων του να έχουν διάρκεια ζωής πάνω από 180 ώρες, ποια είναι η μέγιστη δυνατή τιμή της τυπικής απόκλισης σ της μεταβλητής X ώστε να ικανοποιείται αυτή η απαίτηση;

(β) Ένας φωτοβολταϊκός σταθμός παράγει ημερήσια ρεύμα με μέση τιμή 40 kWh (κιλοβατώρες) και τυπική απόκλιση 4 kWh, και υποτίθεται ότι ακολουθεί κανονική κατανομή. Ποια η πιθανότητα, σε 10 διαφορετικές ημέρες που ελέγχθηκε η ημερήσια παραγωγή ρεύματος, 6 τιμές να βρεθούν στο διάστημα από 36 kWh μέχρι 44 kWh; [2 μονάδες]

α) • $P(X > 180) \geq 0{,}9$

• $P(X > 180) = P\left(\frac{X-\mu}{\sigma} > \frac{180-\mu}{\sigma}\right)$

$= P\left(Z > \frac{-20}{\sigma}\right) = 1 - P\left(Z < -\frac{20}{\sigma}\right)$

οπότε $1 - P\left(Z < -\frac{20}{\sigma}\right) \geq 0{,}9 \Leftrightarrow P\left(Z < -\frac{20}{\sigma}\right) \leq 0{,}1 \Leftrightarrow \Phi\left(-\frac{20}{\sigma}\right) \leq 0{,}1 \Leftrightarrow 1 - \Phi\left(\frac{20}{\sigma}\right) \leq 0{,}1$

$\Leftrightarrow \Phi\left(\frac{20}{\sigma}\right) \geq 0{,}9 \xLeftrightarrow[\Phi(1{,}28)=0{,}9]{\text{από πίνακα}} \frac{20}{\sigma} \geq 1{,}28 \Leftrightarrow \sigma \leq \frac{20}{1{,}28} \Leftrightarrow \sigma \leq 15{,}625 \Leftrightarrow \sigma_{max} = 15{,}625$

β) • $\mu = 40$, $\sigma = 4$

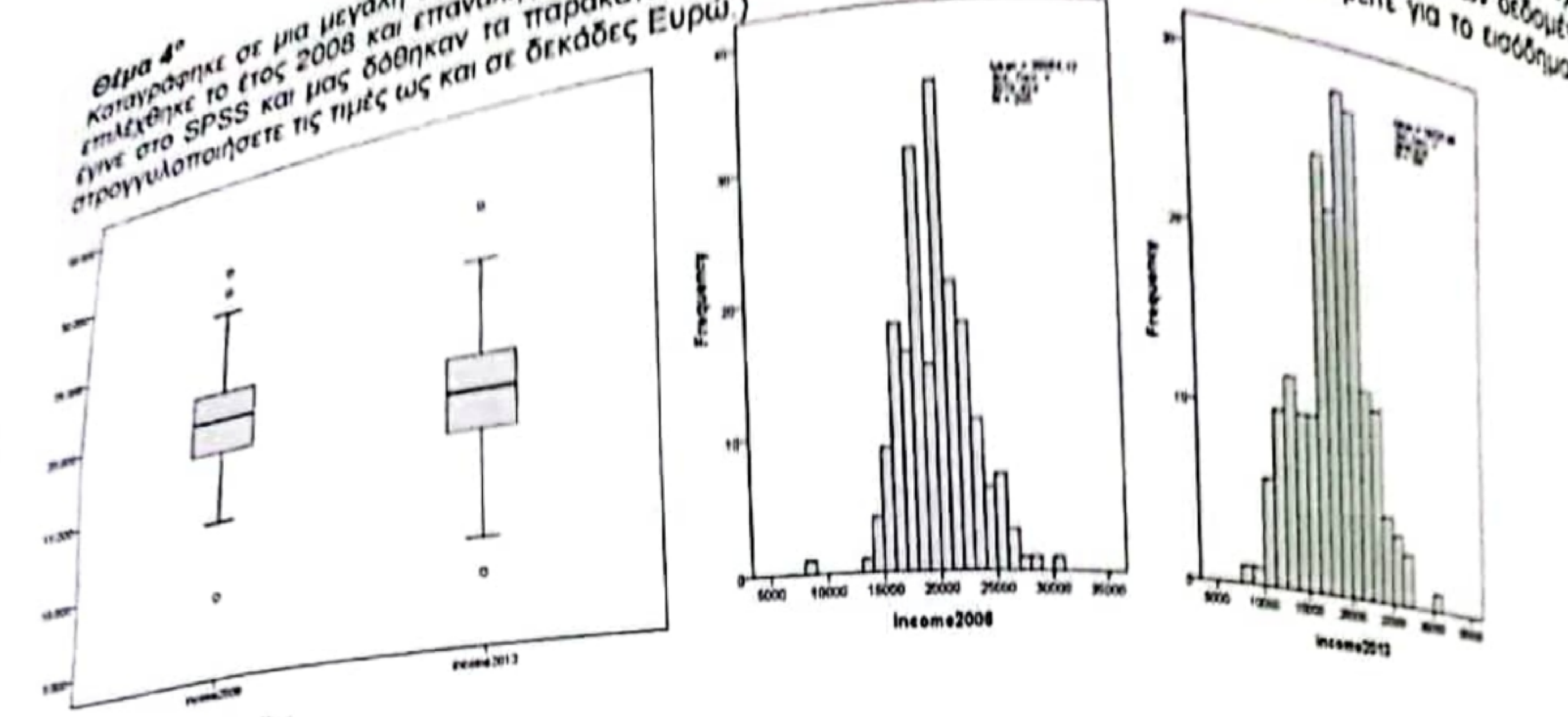
• $P(36 < X < 44) = F_X(44) - F_X(36) = \Phi\left(\frac{44-40}{4}\right) - \Phi\left(\frac{36-40}{4}\right) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1))$

$= 2\Phi(1) - 1 = 2 \cdot 0{,}8413 - 1 \Leftrightarrow \underline{P(36 < X < 44) = 0{,}6826}$

• Άρα $P(6 \text{ τιμές από } 10 \text{ τιμή } 36 < X < 44) = \binom{10}{6} \cdot 0{,}6826^6 (1 - 0{,}6826)^4 = 0{,}2155$

Θέμα 4°

Καταγράφηκε σε μια μεγάλη εταιρία το ετήσιο εισόδημα (σε Ευρώ) από ένα μεγάλο πλήθος εργαζομένων που τυχαία επιλέχθηκε το έτος 2008 και επαναλήφθηκε το ίδιο σε άλλο τυχαίο δείγμα το έτος 2013. Η καταγραφή των δεδομένων έγινε στο SPSS και μας δόθηκαν τα παρακάτω περιγραφικά στοιχεία για τα δεδομένα. (Μπορείτε για το εισόδημα να στρογγυλοποιήσετε τις τιμές ως και σε δεκάδες Ευρώ.)



Statistics

		income2008	income2013
N	Valid	200	190
	Missing	0	10
Mean		20054,13	18227,66
Std. Deviation		3079,424	3948,816
Variance		9482853,386	15593148,12

(α) Με βάση τα συνοπτικά μέτρα, τα ιστογράμματα και τα θηκογράμματα για τα δύο δείγματα σχολιάστε αν η κατανομή του ετήσιου εισοδήματος για το έτος 2008 και το έτος 2013 φαίνεται να είναι κανονική, καθώς επίσης και αν φαίνεται να διαφέρει η κατανομή του ετήσιου εισοδήματος στα δύο έτη (ως προς το κέντρο και τη μεταβλητότητα τους).

(β) Ελέγξετε με κατάλληλο διάστημα εμπιστοσύνης σε επίπεδο εμπιστοσύνης 95% αν το μέσο εισόδημα εργαζομένου στην εταιρία για το 2008 μπορεί να είναι 22000 Ευρώ.

(γ) Ελέγξετε με κατάλληλο διάστημα εμπιστοσύνης σε επίπεδο εμπιστοσύνης 99% αν το μέσο εισόδημα εργαζομένου στην εταιρία έχει αλλάξει από το έτος 2008 στο έτος 2013. [2 μονάδες]

α) • Παρατηρούμε ότι η διάμεσος δεν αποκλίνει σημαντικά προς το 1° ή 3° τεταρτημόριο (Q_1 και Q_3) και ότι τα μήκη των μυστάκων είναι περίπου ίσα. Όμως υπάρχουν ακραίες τιμές και στα 2 θηκογράμματα.

• Απ' το ιστόγραμμα παρατηρούμε ότι έχουμε πολλές παρατηρήσεις στο μέσο και οι υπόλοιπες τιμές απλώνονται με κάποια συμμετρία γύρω από το μέσο. Η εικόνα του ιστογράμματος συμφωνεί με τη γραφική παράσταση της κανονικής κατανομής (σχήμα καμπάνα).

• Επίσης απ' το πινακάκι παρατηρούμε ότι έχουμε μεγάλο n ($n>30$)

Οι κατανομές φαίνεται να είναι κανονικές.

• Το 2008 το κέντρο είναι υψηλότερο απ' του 2013 (κατά μικρό βαθμό), ενώ η μεταβλητότητα των τιμών είναι μεγαλύτερη το 2013. Γενικά έχουμε μικρή (αλλά εμφανή) διαφορά στις 2 κατανομές.

β) • $1-a=0,95 \Leftrightarrow a=0,05 \Leftrightarrow \frac{a}{2}=0,025 \Leftrightarrow 1-\frac{a}{2}=0,975$

• $Z_{1-\frac{a}{2}} = Z_{0,975} = 1,96$ (παίρνουμε Normal και όχι Student λόγω μεγάλου n)

• $\bar{X} \pm Z_{0,975}\frac{S_1}{\sqrt{n_1}} = 20050 \pm 1,96\frac{3080}{\sqrt{200}} \rightarrow [19623 , 20476]$ το μέσο εισόδημα δεν γίνεται να είναι 22.000€

γ) • $1-a=0,99 \Leftrightarrow a=0,01 \Leftrightarrow \frac{a}{2}=0,005 \Leftrightarrow 1-\frac{a}{2}=0,995$

• $Z_{1-\frac{a}{2}} = Z_{0,995} = 2,57$

• $(\bar{X_1}-\bar{X_2}) \pm Z_{0,995}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} = (20050-18230) \pm 2,57\sqrt{\frac{3080^2}{200}+\frac{3950^2}{190}}$

$\rightarrow [894 , 2745]$, δηλαδή οι μισθοί έχουν μειωθεί κατά μεγάλο βαθμό.

Θέμα 5°

Στον παρακάτω πίνακα δίνεται η φορολογική κλίμακα για τον ΕΝιαίο Φόρο Ιδιοκτησίας Ακινήτων (ΕΝΦΙΑ). Για τις 7 φορολογικές ζώνες ακινήτων (στήλη 1) δίνεται το διάστημα τιμών ζώνης (από / εώς) σε Ευρώ ανά τετραγωνικό μέτρο (τ.μ.) (στήλη 2 και 3) και ο αντίστοιχος φόρος σε Ευρώ ανά τετραγωνικό μέτρο (τ.μ.) (στήλη 4).

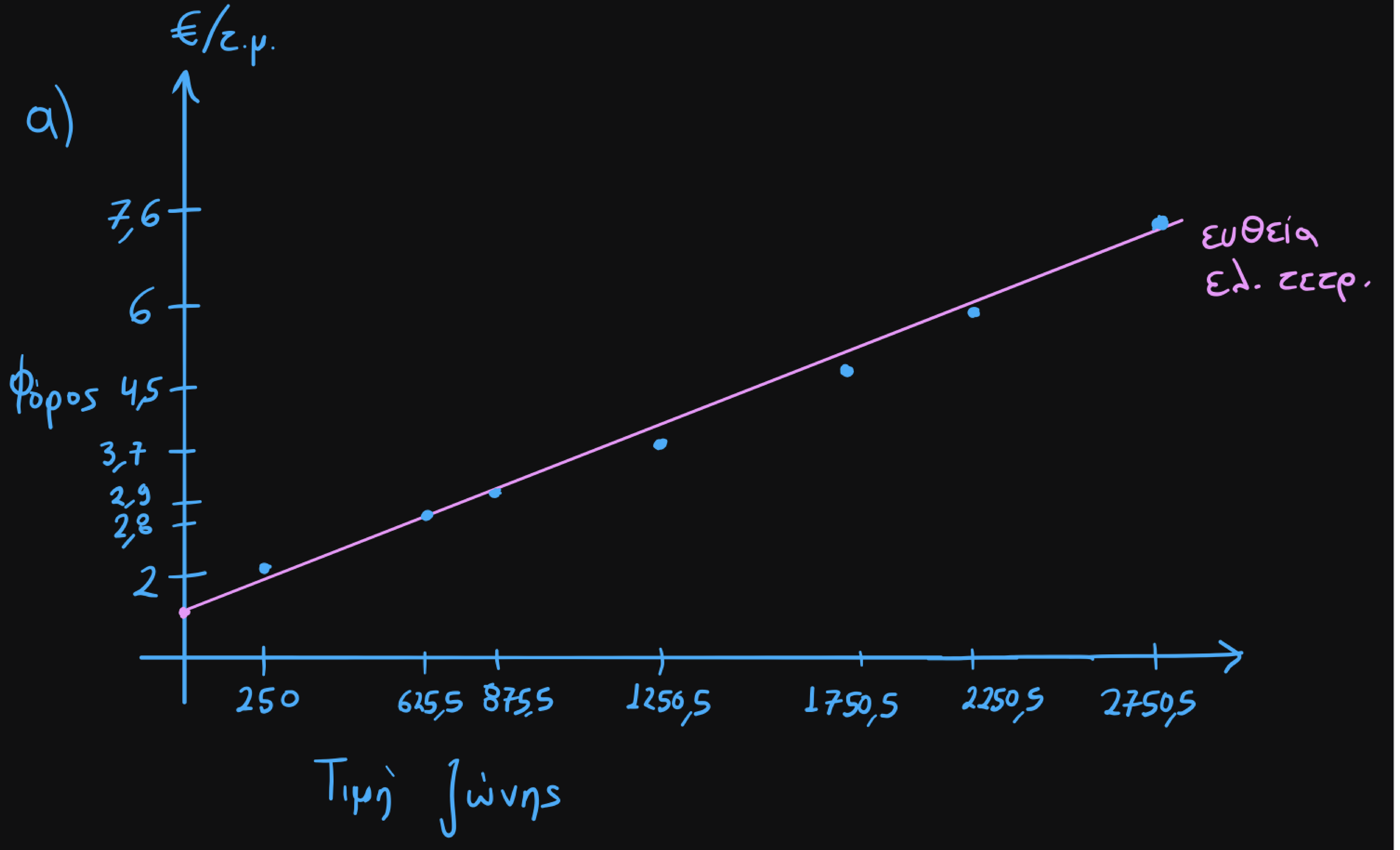
Φορολογική Ζώνη	Τιμή Ζώνης (Ευρώ/τμ) από	εώς	Φόρος (Ευρώ/τμ)
1	0	500	2
2	501	750	2,8
3	751	1000	2,9
4	1001	1500	3,7
5	1501	2000	4,5
6	2001	2500	6
7	2501	3000	7,6

(α) Κάνετε ένα κατάλληλο διάγραμμα (διάγραμμα διασποράς) που δείχνει την αύξηση του φόρου με την τιμή ζώνης. Για την τιμή ζώνης χρησιμοποιείστε το μέσο όρο της τιμής ζώνης (ημιάθροισμα τιμής «από» και «εώς»). Φαίνεται η αύξηση να είναι γραμμική;

(β) Σε συνέχεια του (α), βρείτε κατάλληλο μοντέλο γραμμικής παλινδρόμησης που να υπολογίζει το φόρο για κάθε τιμή ζώνης (με τη μέθοδο ελάχιστων τετραγώνων). Σχηματίστε την ευθεία ελαχίστων τετραγώνων στο σχήμα του (α).

(γ) Υπολογίστε, αν γίνεται, το φόρο που θα πρέπει να πληρώσει κάποιος για ακίνητο σε περιοχή με τιμή ζώνης 1200 Ευρώ/τμ σύμφωνα με το μοντέλο που βρήκατε στο (β). Το ίδιο για τιμής ζώνης 4000 Ευρώ/τμ. [1 1/2 μονάδες]

α)



Φαίνεται πως υπάρχει ισχυρή γραμμική εξάρτηση $(r>0,9)$.

β) $\bullet \sum_{i=1}^{7} x_i y_i = 51699,8 \qquad \bullet \bar{x} = 1393,28 \qquad \bullet \bar{y} = 4,214$

$\bullet\ S_{XY} = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right) = \frac{1}{6}\left(51699,8 - 7 \cdot 1393,28 \cdot 4,214\right) \Leftrightarrow S_{XY} = 1766,8$

$\bullet\ S_X^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \frac{1}{6}\left(18478251,5 - 7 \cdot 1941229,158\right) \Leftrightarrow S_X^2 = 814941,23$

$\bullet\ b = \frac{S_{XY}}{S_X^2} \Leftrightarrow b = 2,16 \cdot 10^{-3}$

$\bullet\ a = \bar{y} - b\bar{x} = 4,214 - 2,16 \cdot 10^{-3} \cdot 1393,28 \Leftrightarrow a = 1,204$

Άρα $\boxed{y = 1,204 + 2,16 \cdot 10^{-3} x}$

γ) • Για $x = 1200$: $y = 1,204 + 2,16 \cdot 10^{-3} \cdot 1200 \Leftrightarrow \boxed{y = 3,796}$

• Δεν μπορούμε να υπολογίσουμε για $x = 4000$ επειδή έχουμε δεδομένα για x από 0 έως 3000.
(δεν θα είναι καλή η εκτίμηση)

Λύσεις Φεβρουαρίου 2020

(για τις υπόλοιπες λύσεις δες το αρχείο "Συλλογή Λυμένων Θεμάτων")

Θέμα 2ο

(Α)
Υποθέστε ότι κάθε τηλεφωνική κλήση σας προς ένα δημοφιλή ραδιοφωνικό σταθμό έχει πιθανότητα 0,20 για σύνδεση, δηλαδή να μην είναι απασχολημένη η γραμμή. Υποθέτουμε, ακόμη, ότι οι τηλεφωνικές κλήσεις είναι στατιστικά ανεξάρτητες.

(α) Ποια η πιθανότητα να συνδεθείτε με το ραδιοφωνικό σταθμό στην πέμπτη τηλεφωνική σας κλήση;
(β) Ποιος ο μέσος αριθμός τηλεφωνικών κλήσεων για να συνδεθείτε με τον ραδιοφωνικό σταθμό;

(Β)
Σε μία συσκευή λήψης σημάτων, ο αριθμός σημάτων ακολουθεί Poisson κατανομή με μέση συχνότητα πέντε σήματα ανά λεπτό.
(α) Ποια η πιθανότητα ότι σε διάστημα μισού λεπτού η συσκευή δεν λαμβάνει κανένα σήμα;
(β) Ποια η πιθανότητα ότι η πρώτη λήψη συμβαίνει μεταξύ 10 και 20 δευτερόλεπτα μετά την έναρξη λειτουργίας της συσκευής;
(γ) Με προσέγγιση Κανονικής Κατανομής να εκτιμηθεί η πιθανότητα ότι σε διάστημα 5 λεπτών η συσκευή θα λάβει το πολύ 25 σήματα.

[2,5 μονάδες]

A

α) $P=(1-0,2)^4\cdot 0,2 = 0,8^4\cdot 0,2 \Leftrightarrow P=0,081$

(Τύπος γεωμετρικής)

β) $E(x)=\frac{q}{p}=\frac{1-p}{p}=\frac{0,8}{0,2}=4$ (μέσος αριθμός αποτυχιών)

άρα $4+1=5$ (μέσος αριθμός κλήσεων)

B

α) • $E(x)=\lambda=5$

• $P(x=0)=e^{-\lambda t}\cdot\frac{(\lambda t)^x}{x!} \xrightarrow[\text{και } t=0,5]{x=0} e^{-5\cdot 0,5}\cdot\frac{(5\cdot 0,5)^0}{0!}=e^{-2,5}\simeq 0,082$

(β' τρόπος, χρήση εκθετικής $P(x\geq 0,5)=1-P(x<0,5)=1-(1-e^{-5\cdot 0,5})=e^{-5\cdot 0,5}=e^{-2,5}\simeq 0,082$)

β) • $\underbrace{P(x=0)}_{\text{καμία λήψη για } t=\frac{1}{6}}\cdot\underbrace{P(x\geq 1)}_{\text{πρώτη λήψη για } t=\frac{1}{6}}=e^{-5\cdot\frac{1}{6}}\cdot\frac{(5\cdot\frac{1}{6})^0}{0!}\cdot(1-P(x=0))=e^{-\frac{5}{6}}\cdot(1-e^{-\frac{5}{6}})\simeq 0,245$

(β' τρόπος, χρήση εκθετικής $P(\frac{1}{6}<X<\frac{2}{6})=P(\frac{2}{6})-P(\frac{1}{6})=1-e^{-5\cdot\frac{2}{6}}-(1-e^{-5\cdot\frac{1}{6}})=-e^{-\frac{10}{6}}+e^{-\frac{5}{6}}\simeq 0,245$)

γ) • $P(X\leq 25)$ με $\lambda t=5\cdot 5=25$

• Προσέγγιση κανονικής κατανομής: $Y\sim N(\overset{\lambda t}{\overset{\shortparallel}{25}},\overset{\lambda t}{\overset{\shortparallel}{25}})$

οπότε $P(X\leq 25)\simeq P(Y\leq 25+0,5)=P(Y\leq 25,5)=\Phi\left(\frac{25,5-25}{\sqrt{25}}\right)=\Phi(0,1)=0,5398$

continuity correction (εφαρμόζεται όταν έχουμε προσέγγιση διακριτής σε συνεχή)

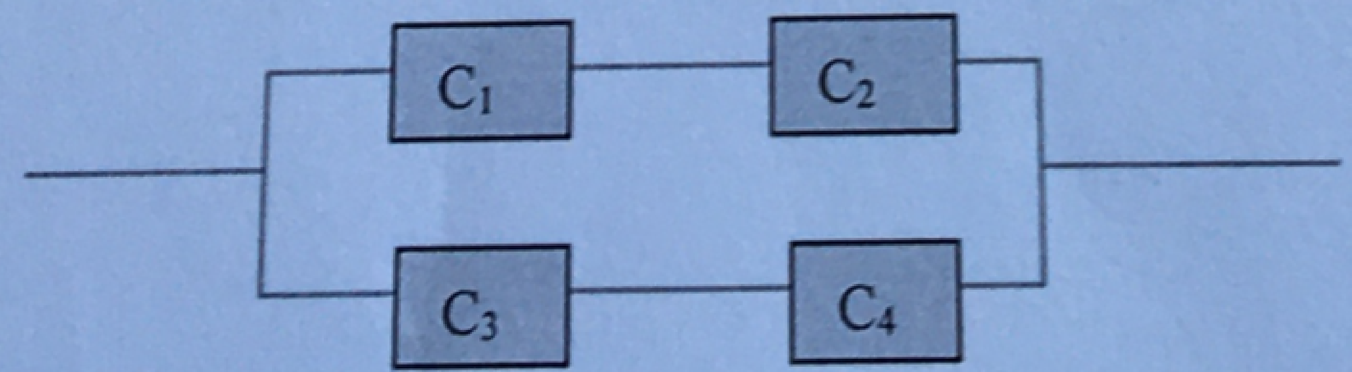
- If $P(X=n)$ use $P(n-0.5<X<n+0.5)$
- If $P(X>n)$ use $P(X>n+0.5)$
- If $P(X\leq n)$ use $P(X<n+0.5)$
- If $P(X<n)$ use $P(X<n-0.5)$
- If $P(X\geq n)$ use $P(X>n-0.5)$

Θέμα 3ο

Στο παρακάτω σύστημα ο χρόνος καλής λειτουργίας, T, των διακοπτών C_1, C_2, C_3, C_4 είναι τυχαία μεταβλητή με σ.π.π. εκθετικής μορφής, $f(x)=(1/50)e^{-(1/50)x}$. Οι διακόπτες λειτουργούν στατιστικά ανεξάρτητα μεταξύ τους.

(α) Ποια η πιθανότητα ότι το σύστημα δεν έχει σταματήσει μετά από 60 ώρες συνεχούς λειτουργίας του;

(β) Να ευρεθεί η σ.π.π. $f_T(t)$ του χρόνου λειτουργίας του συστήματος.



[2 μονάδες]

α) • $F(x)=1-e^{-\lambda x}=1-e^{-\frac{x}{50}}$

• Για $X=60$: $F(60)=1-e^{-\frac{60}{50}}\simeq 0,698$

• $P(\Lambda)=P\big((C_1\cap C_2)\cup(C_3\cap C_4)\big)$

$=P(C_1\cap C_2)+P(C_3\cap C_4)-P(C_1\cap C_2\cap C_3\cap C_4)$

$=P(C_1)\cdot P(C_2)+P(C_3)\cdot P(C_4)-P(C_1)\cdot P(C_2)\cdot P(C_3)\cdot P(C_4) \Longleftrightarrow$ $P(C_1)=P(C_2)=P(C_3)=P(C_4)=P(X\geq 60)$ $=1-P(X<60)=1-0,698=0,302$ (1)

$\Leftrightarrow P(\Lambda)=0,302^2+0,302^2-0,302^4 \Leftrightarrow \boxed{P(\Lambda)=0,174}$

β) • Έχουμε $P(X>x)=1-P(X<x)=1-(1-e^{-\lambda x})=e^{-\frac{x}{50}}$ (X = χρόνος λειτουργίας διακόπτη) ($P(C_1)=P(C_2)=P(C_3)=P(C_4)=e^{-\frac{x}{50}}$)

• Οπότε (1) $\Leftrightarrow F(x)=2\cdot e^{-\frac{2x}{50}}-e^{-\frac{4x}{50}}$

και $f(x)=\frac{dF}{dx}=2\cdot\left(-\frac{2}{50}\right)\cdot e^{-\frac{2x}{50}}+\frac{4}{50}\cdot e^{-\frac{4x}{50}} \Leftrightarrow f(x)=\frac{4}{50}\cdot e^{-\frac{2x}{50}}\left(e^{-\frac{2x}{50}}-1\right) \Leftrightarrow$

$\Leftrightarrow \boxed{f(x)=\frac{2}{25}e^{-\frac{x}{25}}\left(e^{-\frac{x}{25}}-1\right)}$