

Λυμένες Ασκήσεις στο Μάθημα Στατιστικής στο Τμήμα Πολιτικών Μηχανικών

Μέρος Α – Θεωρία Πιθανοτήτων

Άσκηση 1 [Θέμα στις εξετάσεις Φεβρουαρίου 2002]

Το νερό για μία κατοικημένη περιοχή πρώτα αντλείται από μία γεώτρηση, μετά περνά από ένα σύστημα χλωρίωσης, και τέλος μέσα από ένα φίλτρο νερού. Η πιθανότητα αποτυχίας της άντλησης, χλωρίωσης και φιλτραρίσματος μέσα σε ένα χρόνο είναι 0,1, 0,2 και 0,1 αντίστοιχα.

Η αποτυχία άντλησης προκαλεί έλλειψη ικανοποιητικής παροχής νερού, ενώ η αποτυχία της χλωρίωσης ή φιλτραρίσματος ελαττώνει την ποιότητα νερού κάτω από τα επιτρεπτά όρια. Το γεγονός αποτυχίας άντλησης νερού, χλωρίωσης, και φιλτραρίσματος θεωρούνται στατιστικά ανεξάρτητα.

(α) Ποια η πιθανότητα αυτόν τον χρόνο η κατοικημένη περιοχή να έχει ικανοποιητική παροχή νερού παραδεκτής ποιότητας;

(β) Όταν η κατοικημένη περιοχή πίνει νερό χαμηλής ποιότητας (κάτω από τα επιτρεπτά όρια), ποια η πιθανότητα να οφείλεται στην αποτυχία της χλωρίωσης;

(γ) Εάν το σύστημα χλωρίωσης αντικατασταθεί με ένα πιο αξιόπιστο, πιθανότητα αποτυχίας 0,1, κατά πόσο περιορίζεται το ποσοστό ευθύνης της χλωρίωσης της ερώτησης (β);

Λύση

Θεωρώ τα γεγονότα

$A = \{\text{ικανοποιητική παροχή νερού}\}$

$X = \{\text{επιτυχή χλωρίωση}\}$

$\Phi = \{\text{επιτυχή φιλτράρισμα}\}$

(α) Το γεγονός E του οποίου ζητώ την πιθανότητα γράφεται με την ακόλουθη άλγεβρα γεγονότων

$$E = A \cap (X \cap \Phi)$$

Η πιθανότητά του E υπολογίζεται από την τομή των ανεξάρτητων γεγονότων A, X, Φ .

$$\begin{aligned} P(E) &= P(A \cap (X \cap \Phi)) = P(A)P(X)P(\Phi) \\ &= 0,9 * 0,8 * 0,9 = \dots \end{aligned}$$

(β) Ζητώ την πιθανότητα

$$P(\bar{X} | (\bar{X} \cup \bar{\Phi}))$$

Εφαρμόζω τον κανόνα της υπό συνθήκη πιθανότητας,

$$\begin{aligned} P(\bar{X} | (\bar{X} \cup \bar{\Phi})) &= \frac{P(\bar{X} \cap (\bar{X} \cup \bar{\Phi}))}{P(\bar{X} \cup \bar{\Phi})} = \frac{P(\bar{X})}{P(\bar{X} \cup \bar{\Phi})} = \frac{0,2}{0,2 + 0,1 - 0,2 * 0,1} \\ &= \frac{0,2}{0,28} \end{aligned}$$

(γ) Βασικά υπολογίζω την πιθανότητα όπως και στην ερώτηση (β),

$$P(\bar{X} | (\bar{X} \cup \bar{\Phi})) = \frac{0,1}{0,1 + 0,1 - 0,1 * 0,1} = \frac{0,1}{0,19}$$

Άσκηση 2 [Θέμα στις εξετάσεις Φεβρουαρίου 2002]

(α) Η αντοχή συμπίεσης τσιμέντου έχει μία μέση τιμή 60.0 N/mm^2 και μία τυπική απόκλιση 5.0 N/mm^2 και υποτίθεται ότι ακολουθεί κανονική κατανομή.

(i) Ποια η πιθανότητα ότι σε ένα έλεγχο η αντοχή συμπίεσης τσιμέντου θα βρεθεί στο διάστημα από 50 μέχρι 75 N/mm^2 ;

(ii) Ποια η πιθανότητα ότι σε ένα έλεγχο 10 δοκιμών αντοχής συμπίεσης τσιμέντου όλες οι τιμές θα βρεθούν στο διάστημα από 50 μέχρι 75 N/mm^2 ;

(β) Η πιθανότητα ότι μία μηχανή κατασκευής πασσάλων θα πάθει βλάβη για κάθε 100 μέτρα πασσάλων που κάνει είναι 0,02.

(i) Ποια η πιθανότητα να παρουσιάσει βλάβη για πρώτη φορά μεταξύ της κατασκευής 1000 και 1100 μέτρων πασσάλων;

(ii) Αν η μηχανή κατασκευάζει 1000 μέτρα την εβδομάδα, κατά μέσο όρο κάθε πόσες εβδομάδες η μηχανή παρουσιάζει βλάβη;

Λύση

(α) Σε κάθε δοκιμή έχω μία σταθερή πιθανότητα p ότι η αντοχή του τσιμέντου θα πέσει μεταξύ 50 και 75 N/mm^2 .

$$\begin{aligned} p &= P(50 < X < 75) = \Phi\left(\frac{75 - 60}{5}\right) - \Phi\left(\frac{50 - 60}{5}\right) = \Phi(3) - \Phi(-2) \\ &= \Phi(3) - (1 - \Phi(2)) = 0,9987 - 1 + 0,9772 \\ &= 0,9759 \end{aligned}$$

διότι η τυχαία μεταβλητή X παριστάνει αντοχή και ακολουθεί **κανονική κατανομή με μέση τιμή 60 και τυπική απόκλιση 5**.

Η πιθανότητα του γεγονότος ότι και στις δέκα δοκιμές η αντοχή θα βρεθεί στο διάστημα 50 με 75 μπορεί να βρεθεί από την **Διωνυμική κατανομή** αν θεωρήσω

$$n = 10 \text{ (δοκιμές)}$$

$$p = 0,9759$$

και αν Y είναι η τυχαία μεταβλητή που παριστάνει τον αριθμό των επιτυχιών στις δέκα δοκιμές, $Y = \{0, 1, 2, 3, \dots, 10\}$, από τον τύπο της διωνυμικής έχω

$$P(Y = 10) = \binom{10}{10} p^{10} (1 - p)^{10 - 10} = 1 * 0,9759^{10} * 1 = 0,9759^{10}$$

(β) Εδώ θεωρώ ότι σε κάθε παραγωγή 100 μέτρων πασσάλων έχω μία δοκιμή Bernoulli με πιθανότητα βλάβης (**επιτυχία**) $p = 0,02$.

Ζητώ την πιθανότητα να συμβεί η **επιτυχία** για πρώτη φορά στην ενδέκατη δοκιμή. Συμβολίζω με X την τυχαία μεταβλητή η οποία παριστάνει τον αριθμό δοκιμών μέχρι να φανεί για πρώτη φορά η **επιτυχία**. Η X ακολουθεί **Γεωμετρική κατανομή**,

$$P(X = 11) = p(1 - p)^{11 - 1} = 0,02(0,98)^{10}$$

1. Βασικά, ζητώ τον μέσο αριθμό δοκιμών μεταξύ δύο διαδοχικών βλαβών ή **περίοδο επαναφοράς** της βλάβης όπως αλλιώς λέγεται. Δηλαδή, θέλουμε την μέση τιμή της μεταβλητής X . Στην γεωμετρική κατανομή η μέση τιμή της X , $E(X)$ δίνεται από τον τύπο

$$E(X) = \frac{1}{p} = \frac{1}{0,02} = 50 \text{ δοκιμές}$$

Αφού την εβδομάδα έχουμε δέκα δοκιμές, συμπεραίνουμε ότι η περίοδος επαναφοράς της βλάβης είναι 5 εβδομάδες.

Άσκηση 3 [Θέμα στις εξετάσεις Φεβρουαρίου 2004]

Ένα φράγμα τεχνητής λίμνης βρίσκεται σε μία σεισμογενή περιοχή. Όταν συμβεί σεισμός, η πιθανότητα αστοχίας του φράγματος εξαρτάται από την ένταση του σεισμού και από το ύψος της επιφάνειας του νερού στη λίμνη την ώρα του σεισμού. Από παρατηρήσεις στο παρελθόν, μπορούμε να υποθέσουμε ότι το ύψος της επιφάνειας του νερού στην λίμνη μπορεί να είναι ψηλά ή χαμηλά με αντίστοιχες πιθανότητες 0,3 και 0,7. Ο σεισμός μπορεί να είναι ισχυρός, μέτριος ή ασθενής με αντίστοιχες πιθανότητες 0,2, 0,4 και 0,4.

Όταν συμβαίνει ισχυρός σεισμός, το φράγμα αστοχεί, άσχετα από το ύψος της επιφάνειας νερού στην λίμνη. Σε περίπτωση σεισμού μετρίου έντασης το φράγμα αστοχεί, αν το ύψος της επιφάνειας νερού βρίσκεται ψηλά, με πιθανότητα 70%. Στον ασθενή σεισμό το φράγμα δεν διατρέχει κανένα κίνδυνο.

(α) Ποια η πιθανότητα ότι το φράγμα θα υποχωρήσει στον επόμενο σεισμό;

(β) Γίνεται σεισμός και το φράγμα υποχωρεί. Ποια η πιθανότητα ότι η επιφάνεια του νερού της λίμνης ήταν ψηλά την ώρα του σεισμού;

Λύση

Όταν συμβαίνει σεισμός θεωρώ τα γεγονότα:

$$\begin{aligned} A1 &= \{ \text{η επιφάνεια του νερού είναι ψηλά} \} \\ A2 &= \{ \text{η επιφάνεια του νερού είναι χαμηλά} \} \\ I\Sigma &= \{ \text{ο σεισμός είναι ισχυρός} \} \\ M &= \{ \text{ο σεισμός είναι μέτριος} \} \\ A &= \{ \text{ο σεισμός είναι ασθενής} \} \end{aligned}$$

Επίσης, όταν συμβαίνει σεισμός θεωρώ το γεγονός

$$E = \{ \text{το φράγμα αστοχεί} \}.$$

(α) Ζητώ την πιθανότητα του γεγονότος E , $P(E)$. Αφού τα γεγονότα $A1$, $A2$ αποτελούν διαμέριση του δειγματοχώρου μπορώ να υπολογίσω την $P(E)$ με τον νόμο της ολικής πιθανότητας,

$$P(E) = P(E|A1)P(A1) + P(E|A2)P(A2)$$

Γνωρίζω τις πιθανότητες

$$P(A1) = 0,30$$

$$P(A2) = 0,70$$

Υπολογίζω τις πιθανότητες $P(E|A1)$ και $P(E|A2)$ ως ακολούθως.

Επειδή τα γεγονότα $I\Sigma$, M , A αποτελούν διαμέριση του δειγματικού χώρου για να υπολογίσω την πιθανότητα του γεγονότος $(E|A1)$ εφαρμόζοντας πάλι τον νόμο της ολικής πιθανότητας

$$\begin{aligned} P(E|A1) &= P((E|A1)|I\Sigma)P(I\Sigma) + P((E|A1)|M)P(M) + P((E|A1)|A)P(A) \\ &= P((E|(A1 \cap I\Sigma))P(I\Sigma) + P((E|(A1 \cap M))P(M) + P((E|(A1 \cap A))P(A) \\ &= 1(0,2) + 0,7(0,4) + 0(0,4) = 0,48 \end{aligned}$$

Η πιθανότητα του γεγονότος $(E|A2)$ βρίσκεται εύκολα αν σκεφτούμε ότι σε περίπτωση σεισμού με χαμηλά την επιφάνεια του νερού το φράγμα αστοχεί μόνον αν ο σεισμός είναι ισχυρός, δηλαδή

$$P(E|A2) = 0,2$$

Αντικαθιστώ τις πιθανότητες στον νόμο της ολικής πιθανότητας

$$P(E) = P(E|A1) P(A1) + P(E|A2) P(A2)$$

$$P(E) = 0,48 * 0,3 + 0,2 * 0,7 = 0,144 + 0,14 = 0,284$$

- (β) Ζητώ την πιθανότητα $P(A1 | E)$. Εφαρμόζω το Θεώρημα Bayes,
- $$P(A1 | E) = \frac{P(E|A1) P(A1)}{P(E)}$$
- $$= \frac{0,48 * 0,3}{0,284} = 0,5035$$

Άσκηση 4 [Θέμα στις εξετάσεις Φεβρουαρίου 2004]

Σε μία περιοχή ο αριθμός ισχυρών νεροποντών ακολουθεί Poisson κατανομή με μέσο αριθμό 1 νεροποντή στα 4 χρόνια.

- (α) Ποια είναι η πιθανότητα ότι στα επόμενα 8 χρόνια

- δεν θα έχουμε ισχυρή νεροποντή;
- θα έχουμε πάνω από 2 ισχυρές νεροποντές;

Στην περιοχή αυτή βρίσκεται ένας ποταμός ο οποίος όταν συμβαίνει ισχυρή νεροποντή ξεχειλίζει με πιθανότητα 0,06.

- (β) Να υπολογιστεί η πιθανότητα πλημμύρας του ποταμού σε 4 ισχυρές νεροποντές.

- (γ) Να υπολογιστεί η πιθανότητα πλημμύρας του ποταμού στα επόμενα 4 χρόνια.

Λύση

Ο μέσος αριθμός νεροποντών ανά χρόνο είναι

$$\lambda = 1/4 = 0,25$$

(α) Θεωρώ σαν X_8 την τυχαία μεταβλητή η οποία παριστάνει τον αριθμό νεροποντών σε οκτώ χρόνια και η οποία ακολουθεί Poisson κατανομή

- $P(X_8=0) = e^{-\lambda t} \frac{(\lambda t)^0}{0!} = e^{-0,25(8)} \frac{(0,25(8))^0}{0!} = e^{-2} = \dots$
- $P(X_8 > 2) = 1 - P(X_8 \leq 2) = 1 - (P(X_8=0) + P(X_8=1) + P(X_8=2))$

$$= 1 - \left(e^{-\lambda t} \frac{(\lambda t)^0}{0!} + e^{-\lambda t} \frac{(\lambda t)^1}{1!} + e^{-\lambda t} \frac{(\lambda t)^2}{2!} \right)$$

$$= 1 - \left(e^{-2} \frac{(2)^0}{0!} + e^{-2} \frac{(2)^1}{1!} + e^{-2} \frac{(2)^2}{2!} \right) = \dots$$

(β) Θα υπολογίσω την πιθανότητα του συμπληρωματικού γεγονότος και θα την αφαιρέσω από την μονάδα. Δηλαδή θα υπολογίσω την πιθανότητα να μην υπάρξει πλημμύρα σε 4 νεροποντές και θα το αφαιρέσω από την μονάδα.

$$P(\{\text{πλημμύρα σε 4 νεροποντές}\}) = 1 - P(\{\text{όχι πλημμύρα σε 4 νεροποντές}\})$$

$$= 1 - 0,94 * 0,94 * 0,94 * 0,94 = \dots$$

Διότι σε κάθε νεροποντή η πιθανότητα πλημμύρας είναι ανεξάρτητη από τι συνέβη στις προηγούμενες νεροποντές (γεγονότα ανεξάρτητα).

- (γ) Συμβολίζω με E το γεγονός

$$E = \{\text{όχι πλημμύρα στα επόμενα 4 χρόνια}\}$$

Την πιθανότητα του E να την υπολογίσω αν ξέρω ότι στα επόμενα 4 χρόνια θα συμβούν

$$A_0 = \{0 \text{ νεροποντές}\}$$

$$A_1 = \{1 \text{ νεροποντή}\}$$

$$A_2 = \{2 \text{ νεροποντές}\}$$

$$A_3 = \{3 \text{ νεροποντές}\}$$

$$A_4 = \{4 \text{ νεροποντές}\}$$

$$A_5 = \{5 \text{ νεροποντές}\}$$

$$A_6 = \{6 \text{ νεροποντές}\}$$

.....
 Τις πιθανότητες των γεγονότων $A_0, A_1, A_2, A_3, A_4, A_5, A_6, \dots$ τα οποία αποτελούν διαμέριση του δειγματικού χώρου μπορώ να υπολογίσω όπως ακολουθεί

$$P(A_0) = P(X_4=0) = e^{-1} \frac{(1)^0}{0!} = e^{-1} = \dots$$

$$P(A_1) = P(X_4=1) = e^{-1} \frac{(1)^1}{1!} = e^{-1} = \dots$$

$$P(A_2) = P(X_4=2) = e^{-1} \frac{(1)^2}{2!} = 0,5e^{-1} = \dots$$

$$P(A_3) = P(X_4=3) = e^{-1} \frac{(1)^3}{3!} = (1/6)e^{-1} = \dots$$

$$P(A_4) = P(X_4=4) = e^{-1} \frac{(1)^4}{4!} = (1/24)e^{-1} = \dots$$

.....
 συνεχίζουμε μέχρι που η πιθανότητα $P(A_k)$ γίνεται αμελητέα.

Η πιθανότητα του E προκύπτει από τον νόμο της ολικής πιθανότητας

$$P(E) = P(E/A_0)P(A_0) + P(E/A_1)P(A_1) + P(E/A_2)P(A_2) + \dots$$

όπου $P(A_0), P(A_1), P(A_2), \dots$ υπολογίστηκαν παραπάνω, τα δε $P(E/A_0), P(E/A_1), P(E/A_2), \dots$ υπολογίζονται όπως στην απάντηση (β),

$$P(E/A_0) = 1$$

$$P(E/A_1) = 0,94$$

$$P(E/A_2) = 0,94 * 0,94$$

$$P(E/A_3) = 0,94 * 0,94 * 0,94$$

.....
 Τέλος, η ζητούμενη πιθανότητα βρίσκεται αν από την μονάδα αφαιρέσω την πιθανότητα του γεγονότος E ,

$$1 - P(E)$$

Μέρος Β – Στατιστική

Άσκηση 1 [Θέμα στις εξετάσεις Φεβρουαρίου 2002]

(α) Αν δύο τυχαίες μεταβλητές X_1 και X_2 έχουν κοινή διασπορά σ^2 , και s_1^2 , s_2^2 είναι οι αμερόληπτες δειγματικές διασπορές των X_1 και X_2 , αντίστοιχα, από δείγματα μεγέθους n_1 και n_2 , δείξτε ότι η εκτιμήτρια $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ της σ^2 είναι επίσης αμερόληπτη.

(β) Δύο εργοστάσια Α και Β παραγωγής χάλυβα θέλουν να εκτιμήσουν την περιεκτικότητα του χάλυβα σε ραδιενέργεια και γι αυτό έκαναν τις παρακάτω μετρήσεις ραδιενέργειας (η ραδιενέργεια μετριέται σε Bq/g) σε τυχαία δοκίμια χάλυβα:

Δοκίμια	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A (Bq/g)	0.37	0.00	0.54	0.59	0.16	0.86	0.86	0.49	0.60	0.55					
B (Bq/g)	0.24	0.52	0.12	0.95	0.26	0.33	0.62	0.32	0.27	0.05	0.39	0.10	0.51	0.79	0.09

Θεωρούμε ότι η περιεκτικότητα του χάλυβα σε ραδιενέργεια ακολουθεί κανονική κατανομή και η διασπορά της ραδιενέργειας στο χάλυβα είναι ίδια για τα δύο εργοστάσια ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

(i) Εκτιμήστε τη μέση ραδιενέργεια στο χάλυβα για το εργοστάσιο Α και Β (σημειακή εκτίμηση και 95% διάστημα εμπιστοσύνης).

(ii) Το μέσο ανώτατο επιτρεπτό όριο μέσης ραδιενέργειας στο χάλυβα είναι 0.5Bq/g. Με βάση τα παραπάνω δείγματα και με εμπιστοσύνη σε επίπεδο 95% θα γινόταν αποδεκτός στην αγορά ο χάλυβας από το εργοστάσιο Α? Από το εργοστάσιο Β?

(iii) Ελέγξτε σε επίπεδο εμπιστοσύνης 95% αν η μέση ραδιενέργεια στο χάλυβα των δύο εργοστασίων είναι ίδια.

[Για την απάντησή σας στα ερωτήματα (ii) και (iii) μπορείτε να χρησιμοποιήσετε διάστημα εμπιστοσύνης ή στατιστικό έλεγχο]

Λύση

(α) Για να είναι η εκτιμήτρια s_p^2 της κοινής διασποράς σ^2 θα πρέπει $E(s_p^2) = \sigma^2$. Έχουμε

$$E(s_p^2) = E\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right) = \frac{(n_1-1)E(s_1^2) + (n_2-1)E(s_2^2)}{n_1+n_2-2} = \frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2} = \frac{(n_1-1)\sigma^2 + (n_2-1)\sigma^2}{n_1+n_2-2} = \sigma^2.$$

(β) Η ραδιενέργεια του χάλυβα στα δύο εργοστάσια είναι οι δύο τ.μ. X_1 και X_2 που ακολουθούν κανονική κατανομή και έχουν κοινή αλλά άγνωστη διασπορά ($\sigma_1^2 = \sigma_2^2 = \sigma^2$). Έχουμε $n_1 = 10$, $n_2 = 15$.

(i) Υπολογίζουμε τις δειγματικές μέσες τιμές, διασπορές και τυπικές αποκλίσεις για τα δύο δείγματα. (τυπολόγιο: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$)

$$\bar{x}_1 = 0.502, s_1^2 = 0.07377, s_1 = 0.272$$

$$\bar{x}_2 = 0.371, s_2^2 = 0.06979, s_2 = 0.264.$$

Οι δύο δειγματικές μέσες τιμές $\bar{x}_1 = 0.502$ και $\bar{x}_2 = 0.371$ αποτελούν τις σημειακές εκτιμήσεις της μέσης ραδιενέργειας χάλυβα στα δύο εργοστάσια. Το 95% διάστημα εμπιστοσύνης (δ.ε.) της μέσης ραδιενέργειας χάλυβα στα δύο εργοστάσια δίνεται από τον τύπο για μικρό δείγμα, και τ.μ. με άγνωστη διασπορά που ακολουθεί κανονική κατανομή, κάνοντας χρήση της κρίσιμης τιμής $t_{1-\alpha/2, n-1}$ της κατανομής student, δηλαδή

$$\text{είναι } \bar{x} \pm t_{1-\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \text{ (τυπολόγιο).}$$

Για το πρώτο εργοστάσιο, επίπεδο σημαντικότητας $\alpha = 0.05$, έχουμε από τον στατιστικό πίνακα για την κατανομή student $t_{0.975, 9} = 2.26$ και το 95% δ.ε. είναι

$$0.502 \pm 2.26 \frac{0.272}{\sqrt{10}} \longrightarrow 0.502 \pm 0.194 \longrightarrow [0.308, 0.696].$$

Αντίστοιχα για το δεύτερο εργοστάσιο έχουμε την κρίσιμη τιμή $t_{0.975, 14} = 2.14$ και το 95% δ.ε. είναι

$$0.371 \pm 2.14 \frac{0.264}{\sqrt{15}} \longrightarrow 0.371 \pm 0.146 \longrightarrow [0.225, 0.517].$$

Η μέση ραδιενέργεια του χάλυβα είναι πολύ πιθανό (με εμπιστοσύνη σε επίπεδο 95%) να βρίσκεται μεταξύ 0.308 Bq/g και 0.696 Bq/g για το εργοστάσιο A και μεταξύ 0.225 Bq/g και 0.517 Bq/g για το εργοστάσιο B.

(ii) Χρησιμοποιώντας διάστημα εμπιστοσύνης: Με βάση τα παραπάνω αποτελέσματα, το μέσο ανώτατο επιτρεπτό όριο μέσης ραδιενέργειας στο χάλυβα είναι 0.5 Bq/g περιέχεται στο 95% δ.ε. της μέσης ραδιενέργειας και για τα δύο εργοστάσια, που σημαίνει ότι η μέση ραδιενέργεια του χάλυβα μπορεί να ξεπεράσει το μέσο επιτρεπτό όριο. Άρα ο χάλυβας και από τα δύο εργοστάσια δε θα γίνει αποδεκτός.

Χρησιμοποιώντας έλεγχο υπόθεσης: Για να απαντήσουμε στο ερώτημα, μπορούμε εναλλακτικά να κάνουμε έλεγχο υπόθεσης για το αν η μέση ραδιενέργεια μπορεί να πάρει την τιμή $\mu_0 = 0.5$.

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0 \text{ [δίπλευρος έλεγχος]}$$

Για μικρό δείγμα και τ.μ. με άγνωστη διασπορά που ακολουθεί κανονική κατανομή, η

στατιστική για τον έλεγχο αυτό είναι $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$ (δεν υπάρχει στο τυπολόγιο,

προκύπτει άμεσα από το αντίστοιχο δ.ε. στο τυπολόγιο).

Η απορριπτική περιοχή R για επίπεδο σημαντικότητας $\alpha = 0.05$ σχηματίζεται από την κρίσιμη τιμή $t_{1-\alpha/2, n-1}$ της κατανομής student : $R = \{t \mid |t| > t_{1-\alpha/2, n-1}\}$.

Για το εργοστάσιο A είναι $R = \{t \mid |t| > 2.26\}$. Η δειγματική στατιστική είναι

$$\tilde{t} = \frac{0.502 - 0.5}{0.272/\sqrt{10}} = 0.023. \text{ Ισχύει } \tilde{t} \notin R \text{ και δεν απορρίπτεται η } H_0.$$

Για το εργοστάσιο A είναι $R = \{t \mid |t| > 2.14\}$. Η δειγματική στατιστική είναι

$$\tilde{t} = \frac{0.371 - 0.5}{0.264/\sqrt{15}} = -1.04. \text{ Ισχύει και πάλι } \tilde{t} \notin R \text{ και δεν απορρίπτεται η } H_0. \text{ Άρα και}$$

για τα δύο εργοστάσια η μέση ραδιενέργεια του χάλυβα μπορεί να είναι 0.5 Bq/g, δηλαδή να ξεπεράσει το μέσο επιτρεπτό όριο και ο χάλυβας και από τα δύο εργοστάσια δε θα γίνει αποδεκτός.

Θα μπορούσαμε να κάνουμε μονόπλευρο έλεγχο, δηλαδή να εξετάσουμε για την εναλλακτική υπόθεση $H_1: \mu < \mu_0$, όποτε και θα άλλαζε η απορριπτική περιοχή,

δηλαδή θα ήταν $R = \{t \mid t < -t_{1-\alpha, n-1}\}$ (θα εξετάζαμε μόνο αν η δειγματική στατιστική

\tilde{t} μπορεί να βρίσκεται στην αριστερή ουρά της κατανομής student). Τα συμπεράσματα θα ήταν τα ίδια αφού οι δειγματικές στατιστικές και για τα δύο δείγματα δεν είναι κοντά στην αριστερή ουρά της κατανομής student.

(iii) Υπολογίζουμε πρώτα τη διαφορά των δειγματικών μέσων τιμών $\bar{x}_1 - \bar{x}_2 = 0.131$ και την εκτίμηση της κοινής διασποράς

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{9 \cdot 0.07377 + 14 \cdot 0.06979}{23} = 0.07135.$$

(δεν υπάρχει στο τυπολόγιο, προκύπτει ως σταθμισμένος μέσος των δύο δειγματικών διασπορών σταθμίζοντας με τους βαθμούς ελευθερίας κάθε δείγματος (μέγεθος δείγματος - 1))

Χρησιμοποιώντας διάστημα εμπιστοσύνης: Χρησιμοποιούμε το 95% διάστημα εμπιστοσύνης (δ.ε.) για τη διαφορά της μέσης ραδιενέργειας χάλυβα στα δύο εργοστάσια $\mu_1 - \mu_2$. Εδώ έχουμε ότι τα δύο δείγματα είναι μικρά και οι τ.μ. ακολουθούν κανονική κατανομή αλλά με άγνωστη και κοινή διασπορά. Γι αυτό κάνουμε χρήση του τύπου που βασίζεται στην κατανομή student και η κρίσιμη τιμή είναι $t_{0.975, 23} = 2.07$. Το 95% δ.ε. είναι

$$(\bar{x}_1 - \bar{x}_2) \pm t_{1-\alpha/2, n_1+n_2-2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (\text{τυπολόγιο})$$

και έχουμε

$$0.131 \pm 2.07 \cdot 0.267 \cdot \sqrt{\frac{1}{10} + \frac{1}{15}} \longrightarrow 0.131 \pm 0.226 \longrightarrow [-0.094, 0.357].$$

Παρατηρούμε ότι το δ.ε. περιέχει το 0, έστω και οριακά, δηλαδή η διαφορά $\mu_1 - \mu_2$ μπορεί να είναι και 0, άρα οι δύο μέσες ραδιενέργειες σε χάλυβα του εργοστασίου A και B δε φαίνεται να διαφέρουν (σε επίπεδο εμπιστοσύνης 95%).

Χρησιμοποιώντας έλεγχο υπόθεσης: Ελέγχουμε την υπόθεση η μέση ραδιενέργεια να είναι ίδια στους χάλυβες των δύο εργοστασίων.

$$H_0: \mu_1 = \mu_2 \text{ ή } \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 \neq \mu_2 \text{ [δ्वίπλευρος έλεγχος]}$$

Η στατιστική για τον έλεγχο αυτόν ακολουθεί κατανομή student και είναι

$$t \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

(δεν υπάρχει στο τυπολόγιο, προκύπτει άμεσα από το αντίστοιχο δ.ε. στο τυπολόγιο)

Η απορριπτική περιοχή R για επίπεδο σημαντικότητας $\alpha = 0.05$ σχηματίζεται από την κρίσιμη τιμή $t_{0.975,23} = 2.07$ της κατανομής student : $R = \{t \mid |t| > 2.07\}$. Η δειγματική στατιστική είναι

$$\tilde{t} = \frac{0.131}{0.267\sqrt{1/10+1/15}} = 0.201. \text{ Ισχύει } \tilde{t} \notin R \text{ και δεν απορρίπτεται η } H_0, \text{ οριακά}$$

όμως καθώς η τιμή της δειγματικής στατιστικής \tilde{t} είναι πολύ κοντά στην κρίσιμη τιμή για τη δεξιά ουρά.

Έτσι παρ' όλο που φαίνεται (με δ.ε. και έλεγχο υπόθεσης) η μέση ραδιενέργεια στο χάλυβα του εργοστασίου Α να είναι μεγαλύτερη από αυτή του εργοστασίου Β, η διαφορά αυτή δε βρέθηκε σημαντική σε επίπεδο εμπιστοσύνης 95%.

Άσκηση 2 [Θέμα στις εξετάσεις Φεβρουαρίου 2002]

Στον παρακάτω πίνακα δίνεται για 10 σταθμούς ο αριθμός των ημερών σ' ένα χρόνο που η θερμοκρασία έπεσε κάτω από 0°C και το υψόμετρο τους.

Υψόμετρο (μ)	1000	1050	1110	1220	1320	1380	1420	1560	1670	1950
Αριθμός ημερών	32	29	36	38	43	53	52	63	73	100

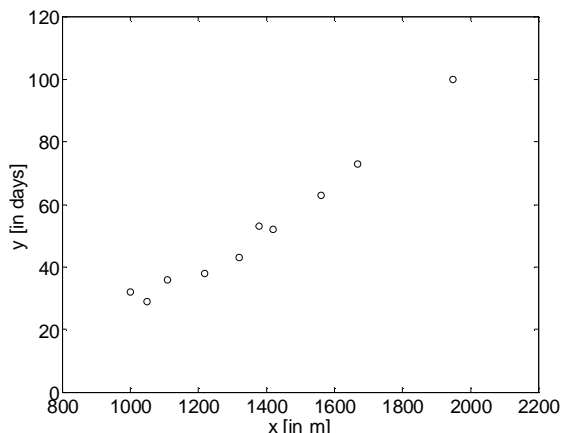
(α) Υποθέτουμε ότι ο αριθμός των ημερών Y εξαρτάται γραμμικά από το υψόμετρο X ($E(Y \mid X=x) = a + \beta x$). Σχηματίστε το κατάλληλο διάγραμμα διασποράς και σχολιάστε αν αυτή η υπόθεση φαίνεται σωστή με βάση το δείγμα των παρατηρήσεων του πίνακα.

(β) Υπολογίστε τις σημειακές εκτιμήσεις a και b των παραμέτρων a και β της ευθείας παλινδρόμησης (με τη μέθοδο των ελαχίστων τετραγώνων).

(γ) Με βάση το δείγμα, μπορείτε να εκτιμήσετε το μέσο αριθμό ημερών το χρόνο που η θερμοκρασία πέφτει κάτω από 0°C σε υψόμετρο 1500μ; Σε υψόμετρο 2200μ;

Λύση

(α) Η ανεξάρτητη μεταβλητή X είναι το υψόμετρο του σταθμού και η εξαρτημένη μεταβλητή Y είναι ο αριθμός των ημερών σ' ένα χρόνο που η θερμοκρασία έπεσε κάτω από 0°C. Σχηματίζουμε το διάγραμμα διασποράς.



Από το διάγραμμα διασποράς φαίνεται να υπάρχει γραμμική θετική εξάρτηση γιατί όταν μεγαλώνει το υψόμετρο πληθαίνουν αναλογικά οι μέρες που η θερμοκρασία πέφτει κάτω από 0°C . Φαίνεται επίσης η εξάρτηση αυτή να είναι ισχυρή γιατί μπορούμε να καθορίσουμε με αρκετή ακρίβεια των αριθμό των ημερών ανά έτος που η θερμοκρασία πέφτει κάτω από 0°C όταν γνωρίζουμε το υψόμετρο (τα σημεία βρίσκονται πολύ κοντά σε μια νοητή ευθεία).

(β) Έχουμε δείγμα μεγέθους $n = 10$. Υπολογίζουμε τα παρακάτω:

$$\bar{x} = 1368 \quad \bar{y} = 51.9 \quad \sum_{i=1}^{10} x_i^2 = 19511200 \quad \sum_{i=1}^{10} x_i \cdot y_i = 767700$$

και βρίσκουμε τη δειγματική διασπορά της X καθώς και τη δειγματική συνδιασπορά των X και Y :

$$s_X^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) = \frac{1}{9} (19511200 - 10 \cdot 1368^2) = 88551.11 \quad (\text{τυπολόγιο})$$

$$s_{XY} = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right) = \frac{1}{9} (767700 - 10 \cdot 1368 \cdot 51.9) = 6412. \quad (\text{τυπολόγιο})$$

Στη συνέχεια εκτιμούμε τις παραμέτρους της ευθείας ελαχίστων τετραγώνων, δηλαδή του μοντέλου γραμμικής παλινδρόμησης:

$$b = \frac{s_{XY}}{s_X^2} = \frac{6412}{88551.11} = 0.0724 \quad (\text{τυπολόγιο})$$

$$a = \bar{y} - b \cdot \bar{x} = 51.9 - 0.0724 \cdot 1368 = -47.16 \quad (\text{τυπολόγιο})$$

και η ευθεία ελαχίστων τετραγώνων είναι $y = -47.16 + 0.0724 \cdot x$.

(γ) Κάνουμε προβλέψεις χρησιμοποιώντας την ευθεία ελαχίστων τετραγώνων για υψόμετρα μέσα στο εύρος του δείγματος από 1000μ μέχρι 1950μ.

Για υψόμετρο $x_0 = 1500$, έχουμε $y = -47.16 + 0.0724 \cdot 1500 = 61.46$ και άρα περιμένουμε 62 μέρες το χρόνο να πέφτει η θερμοκρασία κάτω από 0°C .

Για υψόμετρο $x_0 = 2200$ δε μπορούμε να κάνουμε πρόβλεψη γιατί δεν είναι μέσα στο εύρος γνωστών υψομέτρων για τα οποία ισχύει το γραμμικό μοντέλο.

Άσκηση 3 [Θέμα στις εξετάσεις Φεβρουαρίου 2004]

Δίνονται οι παρακάτω μετρήσεις

5 8 10 3 4 10 7 32 9 7

(α) Σχεδιάστε το θηκόγραμμα αφού εξηγήσετε πως προέκυψαν οι 5 αριθμοί που χρησιμοποιήσατε για να το σχεδιάσετε.

(β) Σχολιάστε αν η κατανομή της μεταβλητής στην οποία αναφέρονται οι μετρήσεις φαίνεται να είναι κανονική.

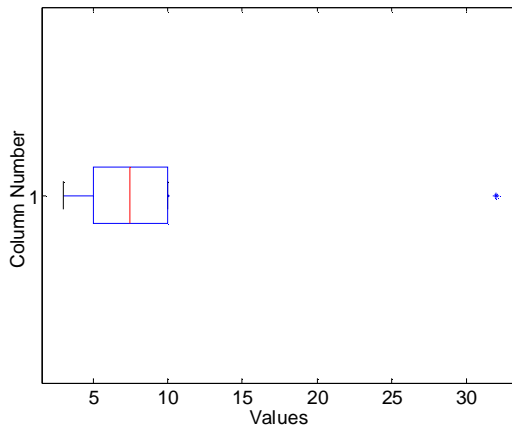
Λύση

(α) Παραθέτουμε τις παρατηρήσεις σε αύξουσα σειρά

βρίσκουμε:

3	4	5	7	7	8	9	10	10	32
↓		↓		↓			↓		↓
$x_{\min} = 3$		$Q_1 = 5$		$\tilde{x} = 7.5$			$Q_3 = 10$		$x_{\max} = 32$

και σχηματίζουμε το θηκόγραμμα, όπως στο παρακάτω σχήμα (σε κατακόρυφη θέση).



(β) Το δείγμα περιέχει μια απόμακρη τιμή, την τιμή 32. Η ύπαρξη μιας τόσο ακραίας τιμής σε ένα μικρό δείγμα 10 παρατηρήσεων δημιουργεί κάποια ανησυχία για το αν η κατανομή της τυχαίας μεταβλητής είναι κανονική.

Άσκηση 4 [Θέμα στις εξετάσεις Φεβρουαρίου 2004]

Ένας δείκτης της κυκλοφορίας οχημάτων είναι ο αριθμός χιλιομέτρων που κάνει ένα όχημα το χρόνο. Για μια περιοχή A συλλέξαμε ένα τυχαίο δείγμα 200 αυτοκινήτων και καταγράψαμε για κάθε αυτοκίνητο τον αριθμό χιλιομέτρων που διένυσε τον τελευταίο χρόνο. Δίνονται τα παρακάτω αποτελέσματα για το δείγμα:

$$\text{μέση τιμή } \bar{x} = 14500 \text{ km,} \quad \text{τυπική απόκλιση } s = 4000 \text{ km}$$

(α) Υπολογίστε το 95% διάστημα εμπιστοσύνης για το μέσο αριθμό χιλιομέτρων που διανύει το χρόνο ένα αυτοκίνητο της περιοχής A. Κάνετε το ίδιο για 99% επίπεδο εμπιστοσύνης και συγκρίνετε τα δύο διαστήματα εμπιστοσύνης.

(β) Σε ίδια μελέτη που έγινε πριν 10 χρόνια είχε βρεθεί πως η τυπική απόκλιση ήταν 3000 km. Εξετάστε με βάση το νέο δείγμα και σε επίπεδο εμπιστοσύνης 90% αν μπορούμε να δεχτούμε ότι η τυπική απόκλιση δεν άλλαξε σημαντικά [μπορείτε να χρησιμοποιήσετε διάστημα εμπιστοσύνης ή έλεγχο υπόθεσης].

Λύση

(α) Έχουμε $n = 200$, $\bar{x} = 14500$ km και $s = 4000$ km. Άρα το $(1-\alpha)\%$ διάστημα εμπιστοσύνης (δ.ε.) για το μέσο αριθμό χιλιομέτρων που διανύει το χρόνο ένα αυτοκίνητο της περιοχής A θα δίνεται από τον τύπο για μεγάλο δείγμα και άγνωστη διασπορά, κάνοντας χρήση της κρίσιμης τιμής $z_{1-\alpha/2}$ της τυπικής κανονικής

κατανομής, δηλαδή είναι $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$ (τυπολόγιο).

Για $\alpha = 0.05$ έχουμε από τον στατιστικό πίνακα για την τυπική κανονική κατανομή $z_{0.975} = 1.96$ και το 95% δ.ε. είναι

$$14500 \pm 1.96 \frac{4000}{\sqrt{200}} \longrightarrow 14500 \pm 554.36 \longrightarrow [13945.69, 15054.36].$$

Για $\alpha = 0.01$ έχουμε από τον στατιστικό πίνακα για την τυπική κανονική κατανομή $z_{0.995} = 2.58$ και το 99% δ.ε. είναι

$$14500 \pm 2.58 \frac{4000}{\sqrt{200}} \longrightarrow 14500 \pm 728.55 \longrightarrow [13771.45, 15228.55].$$

Παρατηρούμε ότι το 99% δ.ε. για το μέσο αριθμό χιλιομέτρων που διανύει το χρόνο ένα αυτοκίνητο της περιοχής Α είναι μεγαλύτερο από αυτό για 95% επίπεδο εμπιστοσύνης, όπως αναμένεται αφού αυξάνουμε την εμπιστοσύνη (πιθανότητα) το διάστημα αυτό να περιέχει το πραγματικό μέσο αριθμό χιλιομέτρων.

(β) Χρησιμοποιώντας διάστημα εμπιστοσύνης: Βρίσκουμε πρώτα το 90% δ.ε. για τη διασπορά του αριθμού χιλιομέτρων που διανύει το χρόνο ένα αυτοκίνητο της περιοχής Α. Αυτό δίνεται κάνοντας χρήση της χ^2 κατανομής και είναι

$$\left(\frac{(n-1)s^2}{\chi_{n-1,1-\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1,\alpha/2}^2} \right) \quad (\text{τυπολόγιο}).$$

Για $\alpha = 0.1$ έχουμε από τον στατιστικό πίνακα για την χ^2 κατανομή ότι η αριστερή κρίσιμη τιμή είναι $\chi_{0.05,199}^2 = 167.36$ και η δεξιά $\chi_{0.95,199}^2 = 232.91$ (οι τιμές αυτές δεν συμπεριλαμβάνονται στον πίνακα που έχει ως 100 βαθμούς ελευθερίας). Το 90% δ.ε. για τη διασπορά είναι

$$\left[\frac{199 \cdot 4000^2}{232.91}, \frac{199 \cdot 4000^2}{167.36} \right] \longrightarrow [13670409.6, 19024743.5].$$

Το αντίστοιχο δ.ε. για την τυπική απόκλιση προκύπτει παίρνοντας τη τετραγωνική ρίζα των ορίων του παραπάνω διαστήματος και άρα το 90% δ.ε. για την τυπική απόκλιση του αριθμού χιλιομέτρων που διανύει το χρόνο ένα αυτοκίνητο της περιοχής Α είναι [3697.4, 4361.7]. Το διάστημα αυτό δεν περιέχει την εμπειρική τιμή 3000 km και άρα η τυπική απόκλιση άλλαξε σημαντικά από αυτήν που είχαμε εκτιμήσει πριν 10 χρόνια.

Χρησιμοποιώντας έλεγχο υπόθεσης: Κάνουμε έλεγχο υπόθεσης για το αν η διασπορά μπορεί να πάρει την τιμή $\sigma_0^2 = 3000^2 = 9 \cdot 10^6$.

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 \neq \sigma_0^2 \quad [\text{δίπλευρος έλεγχος}]$$

Η απορριπτική περιοχή R για επίπεδο σημαντικότητας $\alpha = 0.1$ σχηματίζεται από τις δύο κρίσιμες τιμές της χ^2 κατανομής: $R = \{ \chi^2 \mid \chi^2 < 167.36 \vee \chi^2 > 232.91 \}$.

Η δειγματική στατιστική είναι

$$\tilde{\chi}^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2} = \frac{199 \cdot 4000^2}{3000^2} = 354.$$

(δεν υπάρχει στο τυπολόγιο, προκύπτει άμεσα από το αντίστοιχο δ.ε. στο τυπολόγιο)

Ισχύει $\tilde{\chi}^2 \in R$ και άρα απορρίπτεται η H_0 και συμπεραίνουμε ότι με 90% εμπιστοσύνη (πιθανότητα) δε δεχόμαστε ότι η εμπειρική τιμή 3000 km που είχαμε εκτιμήσει πριν 10 χρόνια για την τυπική απόκλιση του αριθμού χιλιομέτρων μπορεί να ισχύει και τώρα.

Άσκηση 5 [Θέμα στις εξετάσεις Φεβρουαρίου 2004]

Σε μια έρευνα στις Η.Π.Α. για την επίδραση του πληθυσμού της πόλης στη συγκέντρωση του όζοντος συγκεντρώθηκαν τα παρακάτω στοιχεία. Ο πληθυσμός των πόλεων δίνεται σε εκατομμύρια και η συγκέντρωση του όζοντος που μετρήθηκε σε κάθε πόλη δίνεται σε ppb [parts per billion] ανά ώρα.

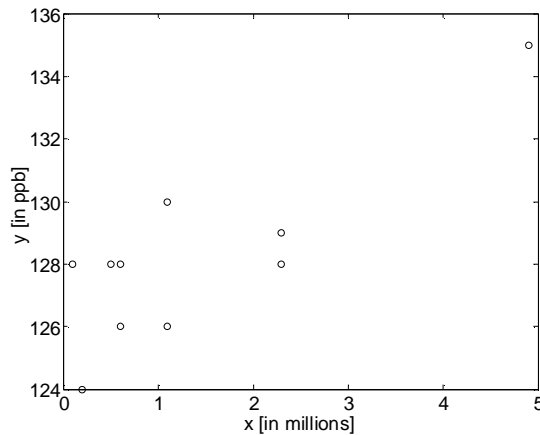
Πληθυσμός πόλης	0.1	0.2	0.5	0.6	0.6	1.1	1.1	2.3	2.3	4.9
Συγκέντρωση όζοντος	128	124	128	126	128	126	130	128	129	135

(α) Σχηματίστε το κατάλληλο διάγραμμα διασποράς. Εκτιμήστε το συντελεστή συσχέτισης μεταξύ της συγκέντρωσης του όζοντος και του πληθυσμού της πόλης. Με βάση αυτά τα αποτελέσματα σχολιάστε αν φαίνεται να υπάρχει εξάρτηση της συγκέντρωσης του όζοντος από τον πληθυσμό της πόλης.

(β) Υπολογίστε τις σημειακές εκτιμήσεις a και b των παραμέτρων a και b της ευθείας παλινδρόμησης (με τη μέθοδο των ελαχίστων τετραγώνων) για το πρόβλημα της γραμμικής εξάρτησης της συγκέντρωσης του όζοντος από τον πληθυσμό της πόλης. Σχηματίστε την ευθεία ελαχίστων τετραγώνων στο διάγραμμα διασποράς που σχηματίσατε στο (α).

Λύση

(α) Σχηματίζουμε το διάγραμμα διασποράς (X : πληθυσμός πόλης, Y : συγκέντρωση όζοντος)



Από το διάγραμμα διασποράς φαίνεται να υπάρχει γραμμική θετική συσχέτιση (η αύξηση του πληθυσμού της πόλης δημιουργεί αύξηση της συγκέντρωσης όζοντος), χωρίς όμως να φαίνεται πολύ ισχυρή (δεν εξηγείται σε μεγάλο βαθμό η μεταβολή της μιας τ.μ. όταν γνωρίζουμε τη μεταβολή της άλλης, τα σημεία απλώνονται αρκετά γύρω από μια νοητή ευθεία).

Έχουμε δείγμα μεγέθους $n = 10$. Υπολογίζουμε τα παρακάτω:

$$\bar{x} = 1.37 \quad \bar{y} = 128.2 \quad \sum_{i=1}^{10} x_i^2 = 38.03 \quad \sum_{i=1}^{10} y_i^2 = 164430 \quad \sum_{i=1}^{10} x_i \cdot y_i = 1788.2$$

και βρίσκουμε τις δειγματικές διασπορές και τυπικές αποκλίσεις των X και Y καθώς και τη δειγματική συνδιασπορά τους: (οι τύποι δίνονται στο τυπολόγιο)

$$s_x^2 = \frac{1}{9} (38.03 - 10 \cdot 1.37^2) = 2.140 \quad \longrightarrow \quad s_x = \sqrt{2.140} = 1.46$$

$$s_y^2 = \frac{1}{9} (164430 - 10 \cdot 128.2^2) = 8.622 \quad \longrightarrow \quad s_y = \sqrt{8.622} = 2.94$$

$$s_{xy} = \frac{1}{9} (1788.2 - 10 \cdot 1.37 \cdot 128.2) = 3.54.$$

Η εκτίμηση του συντελεστή συσχέτισης μεταξύ πληθυσμού πόλης και συγκέντρωση όζοντος είναι

$$r = \frac{3.54}{1.46 \cdot 2.94} = 0.82.$$

Η εκτίμηση του συντελεστή συσχέτισης επιβεβαιώνει ότι η συσχέτιση δεν είναι ισχυρή ($r < 0.9$).

(β) Η ανεξάρτητη μεταβλητή X είναι ο πληθυσμός πόλης και η εξαρτημένη μεταβλητή Y είναι η συγκέντρωση όζοντος. Εκτιμούμε τις παραμέτρους του μοντέλου γραμμικής παλινδρόμησης:

$$b = \frac{s_{XY}}{s_X^2} = \frac{3.54}{2.140} = 1.654 \quad (\text{τυπολόγιο})$$

$$a = \bar{y} - b \cdot \bar{x} = 128.2 - 1.654 \cdot 1.37 = 125.94 \quad (\text{τυπολόγιο})$$

και η ευθεία ελαχίστων τετραγώνων είναι $y = 125.94 + 1.654 \cdot x$.

Για να σχηματίσουμε την ευθεία υπολογίζουμε δύο σημεία που ανήκουν σε αυτήν (καλύτερα για τη μικρότερη και μεγαλύτερη τιμή της X στο δείγμα), π.χ.

$$x = 0.1 \longrightarrow y = 125.94 + 1.654 \cdot 0.1 = 126.10$$

$$x = 4.9 \longrightarrow y = 125.94 + 1.654 \cdot 4.9 = 134.04$$

και χαράζουμε το ευθύγραμμο τμήμα που περνά από αυτά τα δύο σημεία και προεκτείνεται μόνο για το εύρος των γνωστών τιμών του πληθυσμού πόλης X .

