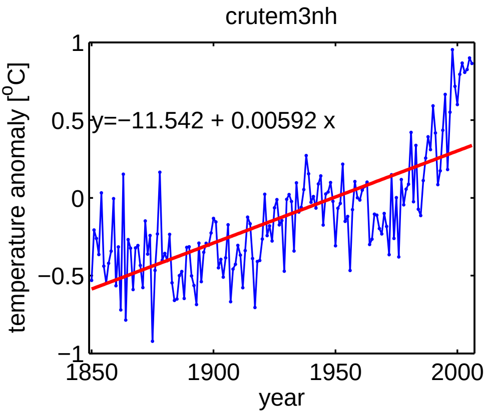
Συσχέτιση και Παλινδρόμηση

Δημήτρης Κουγιουμτζής

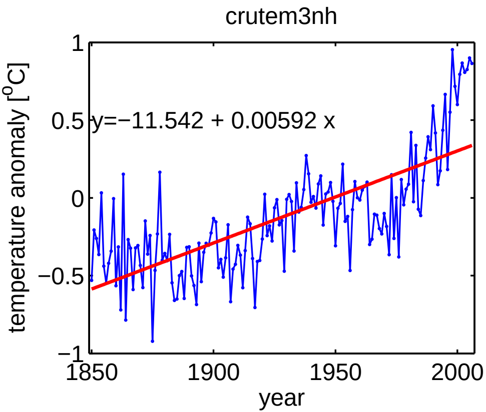
5 Νοεμβρίου 2019

Εισαγωγικά



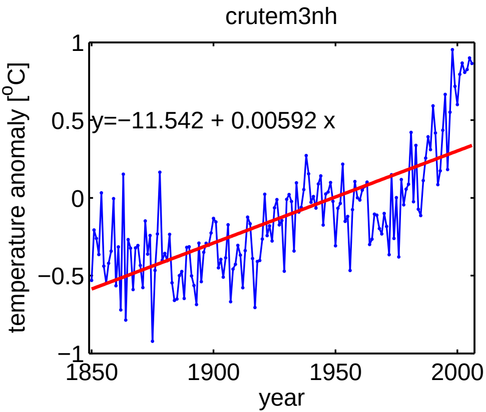
Συσχέτιση:

Εισαγωγικά



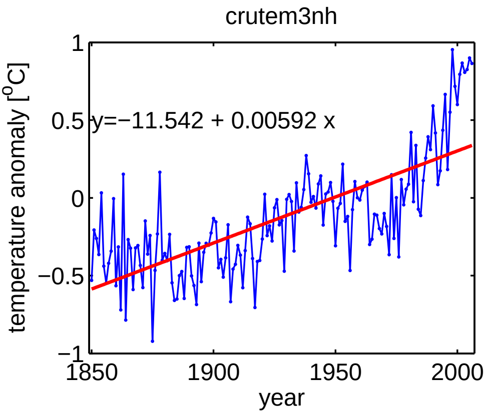
Συσχέτιση: γραμμική / μη-γραμμική,

Εισαγωγικά



Συσχέτιση: γραμμική / μη-γραμμική, μηδενική / ασθενής / ισχυρή

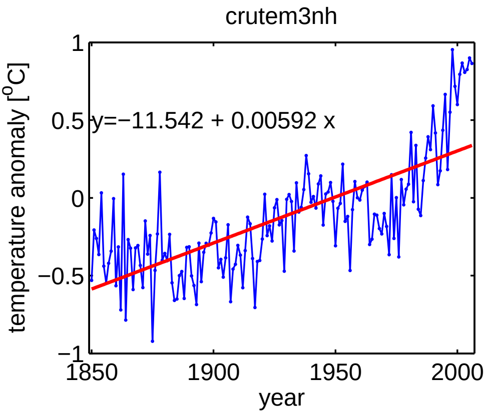
Εισαγωγικά



Συσχέτιση: γραμμική / μη-γραμμική, μηδενική / ασθενής / ισχυρή

Παλινδρόμηση:

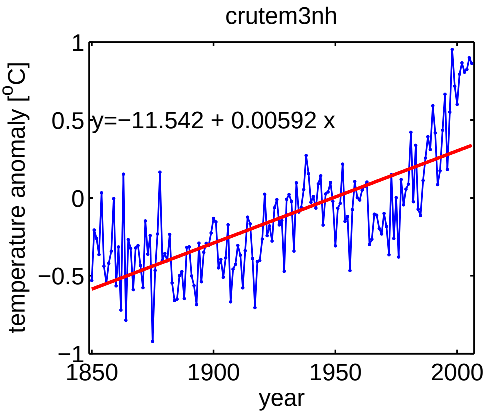
Εισαγωγικά



Συσχέτιση: γραμμική / μη-γραμμική, μηδενική / ασθενής / ισχυρή

Παλινδρόμηση: γραμμική / μη-γραμμική,

Εισαγωγικά



Συσχέτιση: γραμμική / μη-γραμμική, μηδενική / ασθενής / ισχυρή
Παλινδρόμηση: γραμμική / μη-γραμμική, απλή / πολλαπλή

Συσχέτιση

Δύο τ.μ. X και Y συσχετίζονται:

- Η μία επηρεάζει την άλλη

Συσχέτιση

Δύο τ.μ. X και Y συσχετίζονται:

- Η μία επηρεάζει την άλλη
- Επηρεάζονται και οι δύο από κάποια άλλη

Συσχέτιση

Δύο τ.μ. X και Y συσχετίζονται:

- Η μία επηρεάζει την άλλη
- Επηρεάζονται και οι δύο από κάποια άλλη

X: χρόνος ως την αποτυχία ενός στοιχείου κάποιας μηχανής
Y: ταχύτητα του κινητήρα της μηχανής

Συσχέτιση

Δύο τ.μ. X και Y συσχετίζονται:

- Η μία επηρεάζει την άλλη
- Επηρεάζονται και οι δύο από κάποια άλλη

X: χρόνος ως την αποτυχία ενός στοιχείου κάποιας μηχανής
Y: ταχύτητα του κινητήρα της μηχανής

X: χρόνος ως την αποτυχία ενός στοιχείου κάποιας μηχανής
Y: θερμοκρασία του στοιχείου της μηχανής

Συσχέτιση

Δύο τ.μ. X και Y συσχετίζονται:

- Η μία επηρεάζει την άλλη
- Επηρεάζονται και οι δύο από κάποια άλλη

X: χρόνος ως την αποτυχία ενός στοιχείου κάποιας μηχανής
Y: ταχύτητα του κινητήρα της μηχανής

X: χρόνος ως την αποτυχία ενός στοιχείου κάποιας μηχανής
Y: θερμοκρασία του στοιχείου της μηχανής

σ_X^2, σ_Y^2: διασπορά

Συσχέτιση

Δύο τ.μ. X και Y συσχετίζονται:

- Η μία επηρεάζει την άλλη
- Επηρεάζονται και οι δύο από κάποια άλλη

X: χρόνος ως την αποτυχία ενός στοιχείου κάποιας μηχανής
Y: ταχύτητα του κινητήρα της μηχανής

X: χρόνος ως την αποτυχία ενός στοιχείου κάποιας μηχανής
Y: θερμοκρασία του στοιχείου της μηχανής

σ_X^2, σ_Y^2: διασπορά

συνδιασπορά των X και Y
$\sigma_{XY} = \mathrm{Cov}(X, Y) = \mathrm{E}(X, Y) - \mathrm{E}(X)\mathrm{E}(Y)$,

Συντελεστής συσχέτισης

συνετελεστής συσχέτισης Pearson ρ

$$\rho \equiv \mathrm{Corr}(X, Y) = \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y}$$

Συντελεστής συσχέτισης

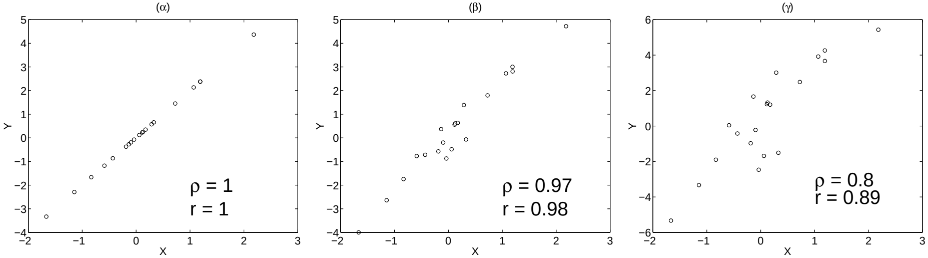
συνετελεστής συσχέτισης Pearson ρ

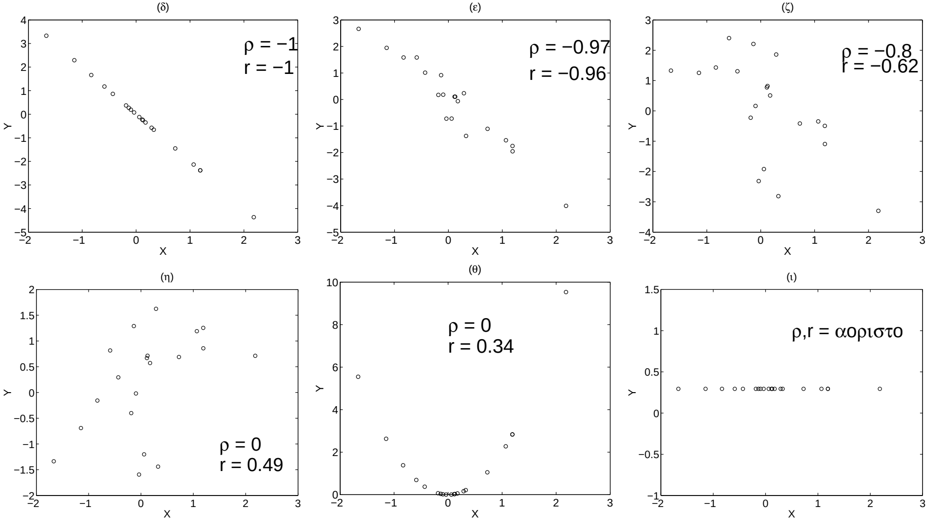
$$\rho \equiv \mathrm{Corr}(X, Y) = \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y}$$

- $\rho \in [-1, 1]$
- $\rho = 1$: *τέλεια θετική* συσχέτιση
- $\rho = 0$: καμιά (γραμμική) συσχέτιση
- $\rho = -1$: *τέλεια αρνητική* συσχέτιση
- ρ 'κοντά' στο -1 ή 1 $\rightarrow$ ισχυρή συσχέτιση
- ρ 'κοντά' στο 0 $\rightarrow$ οι τ.μ. είναι πρακτικά ασυσχέτιστες
- ρ δεν εξαρτάται από τη μονάδα μέτρησης των X και Y
- ρ είναι συμμετρικός ως προς τις X και Y.

Δειγματικός συντελεστής συσχέτισης

Παρατηρήσεις των δύο τ.μ. X και Y:
$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$
διάγραμμα διασποράς




(δ)
ρ = −1
r = −1
(ε)
ρ = −0.97
r = −0.96
(ζ)
ρ = −0.8
r = −0.62
(η)
ρ = 0
r = 0.49
(θ)
ρ = 0
r = 0.34
(ι)
ρ,r = αοριστο
X
Y

Σημειακή εκτίμηση του ρ

Εκτίμηση διασποράς

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

Σημειακή εκτίμηση του ρ

Εκτίμηση διασποράς

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

Εκτίμηση συνδιασποράς

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

Σημειακή εκτίμηση του ρ

Εκτίμηση διασποράς

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

Εκτίμηση συνδιασποράς

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

δειγματικός συντελεστής συσχέτισης (Pearson)

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad \rightarrow \quad \hat{\rho} \equiv r = \frac{s_{XY}}{s_X s_Y}$$

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}}$$

Συντελεστής προσδιορισμού r^2
(ή σε ποσοστά $100r^2\%$):
Δηλώνει το ποσοστό μεταβλητότητας που
μπορούμε να ερμηνεύσουμε για τη μια τ.μ. όταν γνωρίζουμε την άλλη.

Κατανομή του εκτιμητή r

Η κατανομή του εκτιμητή r εξαρτάται από:

Κατανομή του εκτιμητή r

Η κατανομή του εκτιμητή r εξαρτάται από:

- Την τιμή του ρ

Κατανομή του εκτιμητή r

Η κατανομή του εκτιμητή r εξαρτάται από:

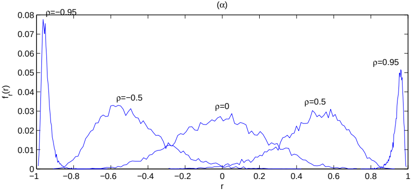
- Την τιμή του ρ
- Το μέγεθος του δείγματος n

Κατανομή του εκτιμητή r

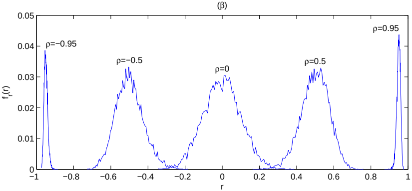
Η κατανομή του εκτιμητή r εξαρτάται από:

- Την τιμή του ρ
- Το μέγεθος του δείγματος n
- Την κατανομή των τ.μ. X και Y.

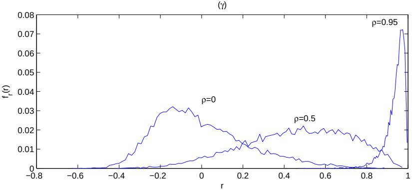
$(X, Y) \sim$ διμεταβλητή κανονική κατανομή
$n = 20$



$(X, Y) \sim$ διμεταβλητή κανονική κατανομή
$n = 100$



$X' = X^2$ και $Y' = Y^2$
$\rho' = \mathrm{Corr}(X', Y') = \rho^2$
$n = 20$



Fisher μετασχηματισμός του r

Fisher μετασχηματισμός

$$z = \tanh^{-1}(r) = 0.5 \ln \frac{1+r}{1-r}$$

Fisher μετασχηματισμός

$$z = \tanh^{-1}(r) = 0.5 \ln \frac{1+r}{1-r}$$

Όταν το δείγμα είναι μεγάλο και από διμεταβλητή κανονική κατανομή $\Longrightarrow z \sim \mathrm{N}(\mu_z, \sigma_z^2)$

$\mu_z \equiv \mathrm{E}(z) = \tanh^{-1}(\rho)$

$\sigma_z^2 \equiv \mathrm{Var}(z) = 1/(n-3)$.

Fisher μετασχηματισμός

$$z = \tanh^{-1}(r) = 0.5 \ln \frac{1+r}{1-r}$$

Όταν το δείγμα είναι μεγάλο και από διμεταβλητή κανονική κατανομή $\Longrightarrow z \sim \mathrm{N}(\mu_z, \sigma_z^2)$

$\mu_z \equiv \mathrm{E}(z) = \tanh^{-1}(\rho)$

$\sigma_z^2 \equiv \mathrm{Var}(z) = 1/(n-3)$.

Μπορούμε λοιπόν να υπολογίσουμε διάστημα εμπιστοσύνης και να κάνουμε έλεγχο υπόθεσης χρησιμοποιώντας την κανονική κατανομή του z.

Διάστημα εμπιστοσύνης για το συντελεστή συσχέτισης

$(1-\alpha)\%$ διάστημα εμπιστοσύνης για το ρ

Διάστημα εμπιστοσύνης για το συντελεστή συσχέτισης

$(1-\alpha)\%$ διάστημα εμπιστοσύνης για το ρ

1. Μετασχηματισμός του r στο z $(\tanh^{-1})$

Διάστημα εμπιστοσύνης για το συντελεστή συσχέτισης

$(1-\alpha)\%$ διάστημα εμπιστοσύνης για το ρ

1. Μετασχηματισμός του r στο z $\qquad (\tanh^{-1})$
2. $z \pm z_{1-\alpha/2}\sqrt{1/(n-3)}$ είναι το $(1-\alpha)\%$ διάστημα εμπιστοσύνης για ζ, $[\zeta_l, \zeta_u]$

Διάστημα εμπιστοσύνης για το συντελεστή συσχέτισης

$(1-\alpha)$% διάστημα εμπιστοσύνης για το ρ

1. Μετασχηματισμός του r στο z $(\tanh^{-1})$
2. $z \pm z_{1-\alpha/2}\sqrt{1/(n-3)}$ είναι το $(1-\alpha)$% διάστημα εμπιστοσύνης για ζ, $[\zeta_l, \zeta_u]$
3. Αντίστροφος μετασχηματισμός για τα άκρα του διαστήματος ζ_l και ζ_u

$$r_l = \tanh(\zeta_l) = \frac{\exp(2\zeta_l)-1}{\exp(2\zeta_l)+1}, \quad r_u = \frac{\exp(2\zeta_u)-1}{\exp(2\zeta_u)+1}$$

Έλεγχος από το διάστημα εμπιστοσύνης του ρ

Αν $[r_l, r_u]$ δεν περιέχει το 0 $\Longrightarrow$ οι δύο τ.μ. συσχετίζονται.

Έλεγχος από το διάστημα εμπιστοσύνης του ρ

Αν $[r_l, r_u]$ δεν περιέχει το 0 $\Longrightarrow$ οι δύο τ.μ. συσχετίζονται.

Έλεγχος υπόθεσης H_0: $\rho = 0$

κατανομή του r κάτω από την H_0

$$t = r\sqrt{\frac{n-2}{1-r^2}} \;\sim\; t_{n-2},$$

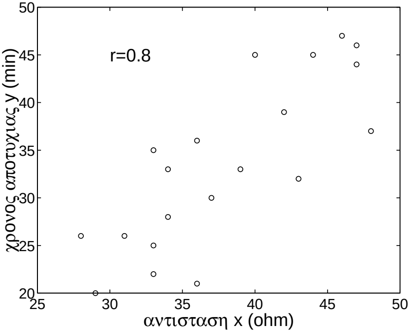
Απόφαση ελέγχου από το t

$p = 2 * (1 - F(t))$

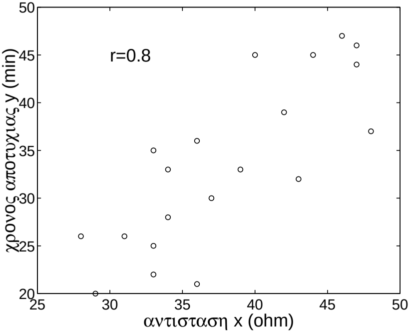
Παράδειγμα: Συσχέτιση αντίστασης / χρόνου αποτυχίας

Α/Α (i)	Αντίσταση x_i (ohm)	Χρόνος αποτυχίας y_i (min)
1	28	26
2	29	20
3	31	26
4	33	22
5	33	25
6	33	35
7	34	28
8	34	33
9	36	21
10	36	36
11	37	30
12	39	33
13	40	45
14	42	39
15	43	32
16	44	45
17	46	47
18	47	44
19	47	46
20	48	37

Παράδειγμα: Συσχέτιση αντίστασης / χρόνου αποτυχίας



Παράδειγμα: Συσχέτιση αντίστασης / χρόνου αποτυχίας



$$\bar{x} = 38 \qquad \bar{y} = 33.5$$

$$\sum_{i=1}^{20} x_i^2 = 29634 \qquad \sum_{i=1}^{20} y_i^2 = 23910 \qquad \sum_{i=1}^{20} x_i y_i = 26305.$$

Παράδειγμα: Συσχέτιση αντίστασης / χρόνου αποτυχίας

$$r = \frac{26305 - 20 \cdot 38 \cdot 33.5}{\sqrt{(29634 - 20 \cdot 38^2) \cdot (23910 - 20 \cdot 33.5^2)}} = 0.804.$$

Η μεταβλητότητα της μιας τ.μ. (αντίσταση ή χρόνος αποτυχίας) μπορεί να εξηγηθεί από τη συσχέτιση της με την άλλη κατά ποσοστό

$$r^2 \cdot 100 = 0.804^2 \cdot 100 = 64.64 \quad \rightarrow \quad \simeq 65\%.$$

Παράδειγμα: Συσχέτιση αντίστασης / χρόνου αποτυχίας

$(1-\alpha)\%$ διάστημα εμπιστοσύνης για το ρ

Παράδειγμα: Συσχέτιση αντίστασης / χρόνου αποτυχίας

$(1-\alpha)\%$ διάστημα εμπιστοσύνης για το ρ

1. Μετασχηματισμός του r στο z, $z = 1.110$

Παράδειγμα: Συσχέτιση αντίστασης / χρόνου αποτυχίας

$(1-\alpha)\%$ διάστημα εμπιστοσύνης για το ρ

1. Μετασχηματισμός του r στο z, $z = 1.110$
2. 95% διαστήματος εμπιστοσύνης του z: $\zeta_l = 0.634$, $\zeta_u = 1.585$

Παράδειγμα: Συσχέτιση αντίστασης / χρόνου αποτυχίας

$(1-\alpha)\%$ διάστημα εμπιστοσύνης για το ρ

1. Μετασχηματισμός του r στο z, $z = 1.110$
2. 95% διαστήματος εμπιστοσύνης του z: $\zeta_l = 0.634$, $\zeta_u = 1.585$
3. Αντίστροφος μετασχηματισμός

$$r_l = 0.561, \quad r_u = 0.919$$

- $r_l > 0, \implies$ η αντίσταση και ο χρόνος αποτυχίας συσχετίζονται
- Έλεγχος υπόθεσης H_0: $\rho = 0$
  στατιστικό $t = 5.736 \implies p = 2 * (1 - F(t)) = 0.0000194$

Συσχέτιση και γραμμικότητα

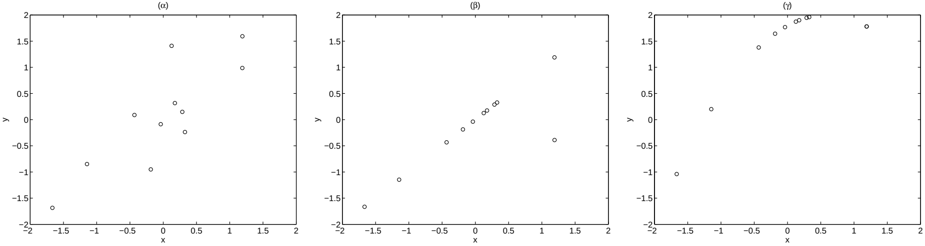
ρ και r κατάλληλα για γραμμική συσχέτιση και κανονικότητα

Τρία δείγματα των (X, Y) με $r = 0.84$.

(α) (X, Y) από διμεταβλητή κανονική κατανομή

(β) $Y = X$ για όλα εκτός από ένα ζευγάρι

(γ) $Y = 2 - 0.6(X - 0.585)^2$



Απλή Γραμμική Παλινδρόμηση

Συντελεστής συσχέτισης: γραμμική σχέση δύο τ.μ. X και Y

Απλή Γραμμική Παλινδρόμηση

Συντελεστής συσχέτισης: γραμμική σχέση δύο τ.μ. X και Y

παλινδρόμηση: εξάρτηση της τ.μ. Y από τη X

Απλή Γραμμική Παλινδρόμηση

Συντελεστής συσχέτισης: γραμμική σχέση δύο τ.μ. X και Y

παλινδρόμηση: εξάρτηση της τ.μ. Y από τη X

Y: εξαρτημένη μεταβλητή $\Longleftarrow$ είναι τ.μ.

Απλή Γραμμική Παλινδρόμηση

Συντελεστής συσχέτισης: γραμμική σχέση δύο τ.μ. X και Y

παλινδρόμηση: εξάρτηση της τ.μ. Y από τη X

Y: εξαρτημένη μεταβλητή $\Longleftarrow$ είναι τ.μ.

X: ανεξάρτητη μεταβλητή $\Longleftarrow$ δεν είναι τ.μ.

Απλή Γραμμική Παλινδρόμηση

Συντελεστής συσχέτισης: γραμμική σχέση δύο τ.μ. X και Y

παλινδρόμηση: εξάρτηση της τ.μ. Y από τη X

Y: εξαρτημένη μεταβλητή $\Longleftarrow$ είναι τ.μ.

X: ανεξάρτητη μεταβλητή $\Longleftarrow$ δεν είναι τ.μ.

Μια μόνο ανεξάρτητη μεταβλητή: απλή παλινδρόμηση

Απλή Γραμμική Παλινδρόμηση

Συντελεστής συσχέτισης: γραμμική σχέση δύο τ.μ. X και Y

παλινδρόμηση: εξάρτηση της τ.μ. Y από τη X

Y: εξαρτημένη μεταβλητή $\Longleftarrow$ είναι τ.μ.

X: ανεξάρτητη μεταβλητή $\Longleftarrow$ δεν είναι τ.μ.

Μια μόνο ανεξάρτητη μεταβλητή: απλή παλινδρόμηση

Παράδειγμα: σε μια μονάδα παραγωγής ηλεκτρικής ενέργειας από λιγνίτη, για να προσδιορίσουμε το κόστος της παραγωγής ενέργειας, μελετάμε την εξάρτηση του από το κόστος του λιγνίτη. X; Y

Συντελεστής συσχέτισης: γραμμική σχέση δύο τ.μ. X και Y

παλινδρόμηση: εξάρτηση της τ.μ. Y από τη X

Y: εξαρτημένη μεταβλητή $\Longleftarrow$ είναι τ.μ.

X: ανεξάρτητη μεταβλητή $\Longleftarrow$ δεν είναι τ.μ.

Μια μόνο ανεξάρτητη μεταβλητή: απλή παλινδρόμηση

Παράδειγμα: σε μια μονάδα παραγωγής ηλεκτρικής ενέργειας από λιγνίτη, για να προσδιορίσουμε το κόστος της παραγωγής ενέργειας, μελετάμε την εξάρτηση του από το κόστος του λιγνίτη. X; Y

Εξάρτηση είναι γραμμική: απλή γραμμική παλινδρόμηση

Το πρόβλημα της απλής γραμμικής παλινδρόμησης

Γενικά: $F_Y(y|X = x)$ για κάθε τιμή x της X

Το πρόβλημα της απλής γραμμικής παλινδρόμησης

Γενικά: $F_Y(y|X=x)$ για κάθε τιμή x της X

Περιορίζουμε το πρόβλημα σε $\mathrm{E}(Y|X=x)$

Γενικά: $F_Y(y|X=x)$ για κάθε τιμή x της X

Περιορίζουμε το πρόβλημα σε $\mathrm{E}(Y|X=x)$

Υπόθεση εργασίας:

$$\mathrm{E}(Y|X=x) = \beta_0 + \beta_1 x$$

απλή γραμμική παλινδρόμηση της Y στη X

Το πρόβλημα της απλής γραμμικής παλινδρόμησης

Γενικά: $F_Y(y|X=x)$ για κάθε τιμή x της X

Περιορίζουμε το πρόβλημα σε $\mathrm{E}(Y|X=x)$

Υπόθεση εργασίας:

$$\mathrm{E}(Y|X=x) = \beta_0 + \beta_1 x$$

απλή γραμμική παλινδρόμηση της Y στη X

$\beta_0 = ?$, $\beta_1 = ?$

Το πρόβλημα της απλής γραμμικής παλινδρόμησης

Γενικά: $F_Y(y|X = x)$ για κάθε τιμή x της X

Περιορίζουμε το πρόβλημα σε $\mathrm{E}(Y|X = x)$

Υπόθεση εργασίας:

$$\mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x$$

απλή γραμμική παλινδρόμηση της Y στη X

$\beta_0 =?$, $\beta_1 =?$

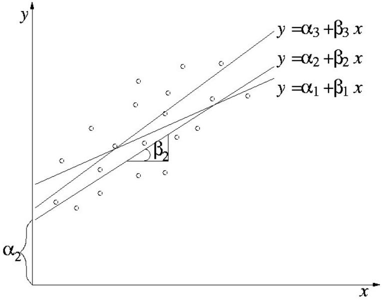
β_0: y για $x = 0$, **διαφορά ύψους**

β_1: συντελεστής του x, **κλίση** της ευθείας παλινδρόμησης ή **συντελεστής παλινδρόμησης**

Το πρόβλημα της απλής γραμμικής παλινδρόμησης

Δείγμα: $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

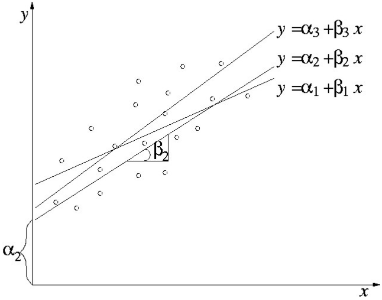
Πολλές ευθείες που προσαρμόζονται σε αυτό



Το πρόβλημα της απλής γραμμικής παλινδρόμησης

Δείγμα: $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

Πολλές ευθείες που προσαρμόζονται σε αυτό

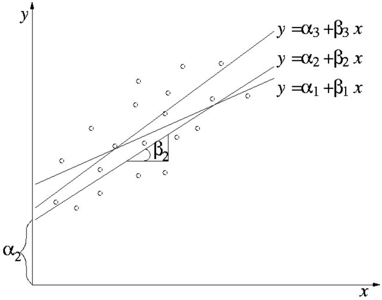


Για x_i της X αντιστοιχούν διαφορετικές τιμές y_i της Y

Το πρόβλημα της απλής γραμμικής παλινδρόμησης

Δείγμα: $\{(x_1, y_1), \dots, (x_n, y_n)\}$

Πολλές ευθείες που προσαρμόζονται σε αυτό



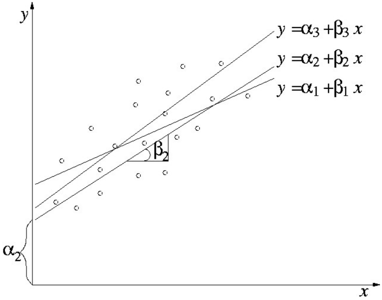
Για x_i της X αντιστοιχούν διαφορετικές τιμές y_i της Y

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Το πρόβλημα της απλής γραμμικής παλινδρόμησης

Δείγμα: $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

Πολλές ευθείες που προσαρμόζονται σε αυτό



Για x_i της X αντιστοιχούν διαφορετικές τιμές y_i της Y

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

σφάλμα παλινδρόμησης: $\epsilon_i = y_i - \mathrm{E}(Y|X = x_i)$

Υποθέσεις

- Η μεταβλητή X είναι *ελεγχόμενη*.

Υποθέσεις

- Η μεταβλητή X είναι *ελεγχόμενη*.
- Η σχέση είναι πράγματι γραμμική.

Υποθέσεις

- Η μεταβλητή X είναι *ελεγχόμενη*.
- Η σχέση είναι πράγματι γραμμική.
- $\mathsf{E}(\epsilon_i) = 0$ και $\mathsf{Var}(\epsilon_i) = \sigma_\epsilon^2$ για κάθε τιμή x_i της X

Υποθέσεις

- Η μεταβλητή X είναι *ελεγχόμενη*.
- Η σχέση είναι πράγματι γραμμική.
- $\mathrm{E}(\epsilon_i) = 0$ και $\mathrm{Var}(\epsilon_i) = \sigma_\epsilon^2$ για κάθε τιμή x_i της X ή ισοδύναμα

$$\mathrm{Var}(Y|X = x) \equiv \sigma_{Y|X}^2 = \sigma_\epsilon^2,$$

Υποθέσεις

- Η μεταβλητή X είναι *ελεγχόμενη*.
- Η σχέση είναι πράγματι γραμμική.
- $\mathrm{E}(\epsilon_i) = 0$ και $\mathrm{Var}(\epsilon_i) = \sigma_\epsilon^2$ για κάθε τιμή x_i της X ή ισοδύναμα

$$\mathrm{Var}(Y|X = x) \equiv \sigma_{Y|X}^2 = \sigma_\epsilon^2,$$

Υποθέτουμε κανονική κατανομή

$$Y|X = x \sim \mathrm{N}(\beta_0 + \beta_1 x, \sigma_\epsilon^2).$$

Υποθέσεις

- Η μεταβλητή X είναι *ελεγχόμενη*.
- Η σχέση είναι πράγματι γραμμική.
- $\mathrm{E}(\epsilon_i) = 0$ και $\mathrm{Var}(\epsilon_i) = \sigma_\epsilon^2$ για κάθε τιμή x_i της X ή ισοδύναμα

$$\mathrm{Var}(Y|X = x) \equiv \sigma^2_{Y|X} = \sigma^2_\epsilon,$$

Υποθέτουμε κανονική κατανομή

$$Y|X = x \sim \mathrm{N}(\beta_0 + \beta_1 x, \sigma^2_\epsilon).$$

όχι απαραίτητο για σημειακή εκτίμηση των παραμέτρων.

Σημειακή εκτίμηση παραμέτρων της απλής γραμμικής παλινδρόμησης

Το πρόβλημα: εκτίμηση των τριών παραμέτρων παλινδρόμησης:

1. β_0,
2. β_1,
3. σ_ϵ^2.

Σημειακή εκτίμηση παραμέτρων της απλής γραμμικής παλινδρόμησης

Το πρόβλημα: εκτίμηση των τριών παραμέτρων παλινδρόμησης:

1. β_0,
2. β_1,
3. σ_ϵ^2.

Εκτίμηση των β_0, β_1 με τη μέθοδο των **ελαχίστων τετραγώνων**:

το άθροισμα των τετραγώνων των κατακόρυφων αποστάσεων των σημείων από την ευθεία να είναι το ελάχιστο.

Εκτίμηση των παραμέτρων της ευθείας παλινδρόμησης

Ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \epsilon_i^2 \qquad \text{ή} \qquad \min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

Ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \epsilon_i^2 \qquad \text{ή} \qquad \min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .$$

Λύση:

$$\left.\begin{array}{l} \frac{\partial \sum (y_i - \beta_0 - \beta_1 x_i)^2}{\partial \beta_0} = 0 \\ \frac{\partial \sum (y_i - \beta_0 - \beta_1 x_i)^2}{\partial \beta_1} = 0 \end{array}\right\} \begin{array}{l} \sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 \end{array}$$

Εκτίμηση των παραμέτρων της ευθείας παλινδρόμησης

Ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \epsilon_i^2 \qquad \text{ή} \qquad \min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

Λύση:

$$\left.\begin{array}{l} \frac{\partial \sum (y_i - \beta_0 - \beta_1 x_i)^2}{\partial \beta_0} = 0 \\ \frac{\partial \sum (y_i - \beta_0 - \beta_1 x_i)^2}{\partial \beta_1} = 0 \end{array}\right\} \begin{array}{l} \sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 \end{array}$$

Εκτίμηση για την κλίση

$$\hat{\beta}_1 \equiv b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}.$$

$s_{XY} = \frac{S_{xy}}{n-1}$ και $s_X^2 = \frac{S_{xx}}{n-1}$

Εκτίμηση των παραμέτρων της ευθείας παλινδρόμησης

Εκτίμηση του σταθερού όρου ως

$$\hat{\beta}_0 \equiv b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n}$$

Εκτίμηση των παραμέτρων της ευθείας παλινδρόμησης

Εκτίμηση του σταθερού όρου ως

$$\hat{\beta}_0 \equiv b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n}$$

$$b_1 = \frac{s_{XY}}{s_X^2}, \qquad b_0 = \bar{y} - b_1 \bar{x}.$$

Εκτίμηση των παραμέτρων της ευθείας παλινδρόμησης

Εκτίμηση του σταθερού όρου ως

$$\hat{\beta}_0 \equiv b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n}$$

$$b_1 = \frac{s_{XY}}{s_X^2}, \qquad b_0 = \bar{y} - b_1 \bar{x}.$$

ευθεία ελαχίστων τετραγώνων

$$\hat{y} = b_0 + b_1 x$$

Εκτίμηση της διασποράς των σφαλμάτων παλινδρόμησης

y_i: παρατηρούμενη τιμή για x_i

$\hat{y}_i$: εκτιμούμενη τιμή από την ευθεία ελαχίστων τετραγώνων για x_i.

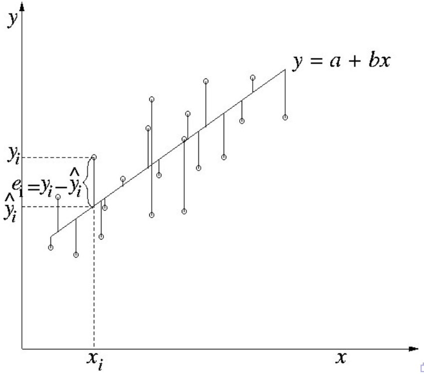
Εκτίμηση της διασποράς των σφαλμάτων παλινδρόμησης

y_i: παρατηρούμενη τιμή για x_i

$\hat{y}_i$: εκτιμούμενη τιμή από την ευθεία ελαχίστων τετραγώνων για x_i.

σφάλμα ελαχίστων τετραγώνων ή **υπόλοιπο**

$$e_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i$$



Εκτίμηση της διασποράς των σφαλμάτων παλινδρόμησης

e_i: εκτίμηση του σφάλματος παλινδρόμησης $\epsilon_i = y_i - \beta_0 - \beta_1 x_i$.

Εκτίμηση της διασποράς των σφαλμάτων παλινδρόμησης

e_i: εκτίμηση του σφάλματος παλινδρόμησης $\epsilon_i = y_i - \beta_0 - \beta_1 x_i$.

Εκτίμηση της διασποράς του e_i:

$$s_\epsilon^2 \equiv \hat{\sigma_\epsilon^2} = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

$n-2$: βαθμοί ελευθερίας

Εκτίμηση της διασποράς των σφαλμάτων παλινδρόμησης

e_i: εκτίμηση του σφάλματος παλινδρόμησης $\epsilon_i = y_i - \beta_0 - \beta_1 x_i$.

Εκτίμηση της διασποράς του e_i:

$$s_\epsilon^2 \equiv \hat{\sigma_\epsilon^2} = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

$n-2$: βαθμοί ελευθερίας

$$s_\epsilon^2 = \frac{n-1}{n-2}\left(s_Y^2 - \frac{s_{XY}^2}{s_X^2}\right) = \frac{n-1}{n-2}(s_Y^2 - b_1^2 s_X^2)$$

Παρατηρήσεις

1. Η ευθεία ελαχίστων τετραγώνων περνάει από το σημείο $(\bar{x}, \bar{y})$

$$b_0 + b_1\bar{x} = \bar{y} - b_1\bar{x} + b_1\bar{x} = \bar{y}.$$

Παρατηρήσεις

1. Η ευθεία ελαχίστων τετραγώνων περνάει από το σημείο $(\bar{x}, \bar{y})$

$$b_0 + b_1\bar{x} = \bar{y} - b_1\bar{x} + b_1\bar{x} = \bar{y}.$$

Η ευθεία ελαχίστων τετραγώνων μπορεί να οριστεί ως

$$y_i - \bar{y} = b_1(x_i - \bar{x}).$$

Παρατηρήσεις

1. Η ευθεία ελαχίστων τετραγώνων περνάει από το σημείο $(\bar{x}, \bar{y})$
$$b_0 + b_1\bar{x} = \bar{y} - b_1\bar{x} + b_1\bar{x} = \bar{y}.$$
Η ευθεία ελαχίστων τετραγώνων μπορεί να οριστεί ως
$$y_i - \bar{y} = b_1(x_i - \bar{x}).$$

2. Η σημειακή εκτίμηση των β_0 και β_1 με τη μέθοδο των ελαχίστων τετραγώνων δεν προϋποθέτει σταθερή διασπορά και κανονική κατανομή της $Y|X$.

Παρατηρήσεις

1. Η ευθεία ελαχίστων τετραγώνων περνάει από το σημείο $(\bar{x}, \bar{y})$

$$b_0 + b_1\bar{x} = \bar{y} - b_1\bar{x} + b_1\bar{x} = \bar{y}.$$

   Η ευθεία ελαχίστων τετραγώνων μπορεί να οριστεί ως

$$y_i - \bar{y} = b_1(x_i - \bar{x}).$$

2. Η σημειακή εκτίμηση των β_0 και β_1 με τη μέθοδο των ελαχίστων τετραγώνων δεν προϋποθέτει σταθερή διασπορά και κανονική κατανομή της $Y|X$.
3. Για κάθε x της X η πρόβλεψη της y της Y:

$$\hat{y} = b_0 + b_1 x$$

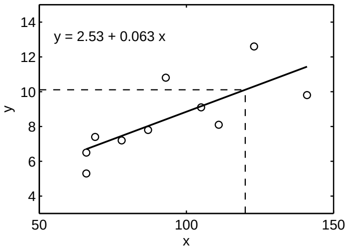
   x πρέπει να ανήκει στο εύρος τιμών της X από το δείγμα.

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ

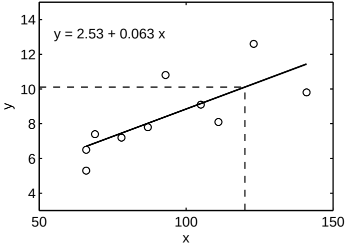
Θέλουμε να μελετήσουμε σε ένα ολοκληρωμένο κύκλωμα την εξάρτηση της απολαβής ρεύματος κρυσταλλολυχνίας από την αντίσταση του στρώματος της κρυσταλλολυχνίας.

A/A (i)	Αντίσταση στρώματος x_i (ohm/cm)	Απολαβή ρεύματος y_i
1	66	5.3
2	66	6.5
3	69	7.4
4	78	7.2
5	87	7.8
6	93	10.8
7	105	9.1
8	111	8.1
9	123	12.6
10	141	9.8

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ



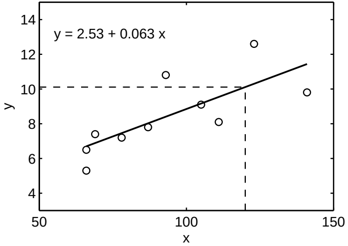
Παράδειγμα: Απολαβή ρεύματος τρανζίστορ



$$\bar{x} = 93.9 \qquad \bar{y} = 8.46$$

$$\sum_{i=1}^{10} x_i^2 = 94131 \qquad \sum_{i=1}^{10} y_i^2 = 757.64 \qquad \sum_{i=1}^{10} x_i y_i = 8320.2$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ



$$\bar{x} = 93.9 \qquad \bar{y} = 8.46$$

$$\sum_{i=1}^{10} x_i^2 = 94131 \qquad \sum_{i=1}^{10} y_i^2 = 757.64 \qquad \sum_{i=1}^{10} x_i y_i = 8320.2$$

$$s_{XY} = 41.81 \quad s_X^2 = 662.1 \quad s_Y^2 = 4.66.$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ

$$
\begin{aligned}
b_1 &= \frac{41.81}{662.1} = 0.063 \\
b_0 &= 8.46 - 0.063 \cdot 93.9 = 2.53.
\end{aligned}
$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ

$$
\begin{aligned}
b_1 &= \frac{41.81}{662.1} = 0.063 \\
b_0 &= 8.46 - 0.063 \cdot 93.9 = 2.53.
\end{aligned}
$$

$$
s_\epsilon^2 = \frac{9}{8}(4.66 - 0.063^2 \cdot 41.81) = 2.271.
$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ

$$\begin{aligned} b_1 &= \frac{41.81}{662.1} = 0.063 \\ b_0 &= 8.46 - 0.063 \cdot 93.9 = 2.53. \end{aligned}$$

$$s_\epsilon^2 = \frac{9}{8}(4.66 - 0.063^2 \cdot 41.81) = 2.271.$$

1. $b_1 = 0.063$: απολαβή ρεύματος για αύξηση της αντίστασης στρώματος κατά 1 ohm/cm

$$
\begin{aligned}
b_1 &= \frac{41.81}{662.1} = 0.063 \\
b_0 &= 8.46 - 0.063 \cdot 93.9 = 2.53.
\end{aligned}
$$

$$
s_\epsilon^2 = \frac{9}{8}(4.66 - 0.063^2 \cdot 41.81) = 2.271.
$$

1. $b_1 = 0.063$: απολαβή ρεύματος για αύξηση της αντίστασης στρώματος κατά 1 ohm/cm
2. $b_0 = 2.53$: απολαβή ρεύματος όταν δεν υπάρχει αντίσταση στρώματος ($x = 0$)

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ

$$\begin{aligned} b_1 &= \frac{41.81}{662.1} = 0.063 \\ b_0 &= 8.46 - 0.063 \cdot 93.9 = 2.53. \end{aligned}$$

$$s_\epsilon^2 = \frac{9}{8}(4.66 - 0.063^2 \cdot 41.81) = 2.271.$$

1. $b_1 = 0.063$: απολαβή ρεύματος για αύξηση της αντίστασης στρώματος κατά 1 ohm/cm
2. $b_0 = 2.53$: απολαβή ρεύματος όταν δεν υπάρχει αντίσταση στρώματος $(x = 0)$
3. $s_\epsilon^2 = 2.271 \Longrightarrow s_\epsilon = 1.507$: το τυπικό σφάλμα της εκτίμησης της παλινδρόμησης .

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ

Πρόβλεψη απολαβή ρεύματος μπορεί να γίνει για κάθε αντίσταση στρώματος κρυσταλλολυχνίας στο διάστημα $[66, 141]$ ohm/cm.

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ

Πρόβλεψη απολαβή ρεύματος μπορεί να γίνει για κάθε αντίσταση στρώματος κρυσταλλολυχνίας στο διάστημα $[66, 141]\,\mathrm{ohm/cm}$.

Για αντίσταση στρώματος $x = 120\,\mathrm{ohm/cm}$

$$\hat{y} = 2.53 + 0.063 \cdot 120 = 10.11.$$

Σχέση του συντελεστή συσχέτισης και παλινδρόμησης

Παλινδρόμηση: X ελεγχόμενη και Y τυχαία

Συσχέτιση: X και Y τυχαίες, αλλά υπολογίζουμε το r και για X ελεγχόμενη.

Σχέση του συντελεστή συσχέτισης και παλινδρόμησης

Παλινδρόμηση: X ελεγχόμενη και Y τυχαία

Συσχέτιση: X και Y τυχαίες, αλλά υπολογίζουμε το r και για X ελεγχόμενη.

$r = \frac{s_{XY}}{s_X s_Y}$ και $b_1 = \frac{s_{XY}}{s_X^2} \Longrightarrow$

$$r = b_1 \frac{s_X}{s_Y} \quad \text{ή} \quad b_1 = r \frac{s_Y}{s_X}.$$

r και b_1 εκφράζουν ποσοτικά τη γραμμική συσχέτιση των X και Y

Σχέση του συντελεστή συσχέτισης και παλινδρόμησης

Παλινδρόμηση: X ελεγχόμενη και Y τυχαία

Συσχέτιση: X και Y τυχαίες, αλλά υπολογίζουμε το r και για X ελεγχόμενη.

$r = \frac{s_{XY}}{s_X s_Y}$ και $b_1 = \frac{s_{XY}}{s_X^2} \Longrightarrow$

$$r = b_1 \frac{s_X}{s_Y} \quad \text{ή} \quad b_1 = r \frac{s_Y}{s_X}.$$

r και b_1 εκφράζουν ποσοτικά τη γραμμική συσχέτιση των X και Y

b_1 εξαρτάται από τη μονάδα μέτρησης των X και Y
Σχέση των r και b_1

- $r > 0 \Leftrightarrow b_1 > 0$
- $r < 0 \Leftrightarrow b_1 < 0$
- $r = 0 \Leftrightarrow b_1 = 0$

Σχέση του συντελεστή συσχέτισης και παλινδρόμησης

Σχέση r^2 και s_ϵ^2

$$s_\epsilon^2 = \frac{n-1}{n-2} s_Y^2 (1 - r^2) \quad \text{ή} \quad r^2 = 1 - \frac{n-2}{n-1} \frac{s_\epsilon^2}{s_Y^2}.$$

Όσο μεγαλύτερο είναι το r^2 (ή το $|r|$) τόσο μειώνεται το s_ϵ^2

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Συντελεστής συσχέτισης της απολαβής ρεύματος και της αντίστασης στρώματος

$$r = \frac{s_{XY}}{s_X \, s_Y} = \frac{41.81}{\sqrt{662.1 \cdot 4.66}} = 0.753$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Συντελεστής συσχέτισης της απολαβής ρεύματος και της αντίστασης στρώματος

$$r = \frac{s_{XY}}{s_X \, s_Y} = \frac{41.81}{\sqrt{662.1 \cdot 4.66}} = 0.753$$

Το r δηλώνει την σχετικά ασθενή θετική συσχέτιση

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Συντελεστής συσχέτισης της απολαβής ρεύματος και της αντίστασης στρώματος

$$r = \frac{s_{XY}}{s_X \, s_Y} = \frac{41.81}{\sqrt{662.1 \cdot 4.66}} = 0.753$$

Το r δηλώνει την σχετικά ασθενή θετική συσχέτιση

$b_1 = 0.063$ εξηγεί το βαθμό εξάρτησης;

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Συντελεστής συσχέτισης της απολαβής ρεύματος και της αντίστασης στρώματος

$$r = \frac{s_{XY}}{s_X \, s_Y} = \frac{41.81}{\sqrt{662.1 \cdot 4.66}} = 0.753$$

Το r δηλώνει την σχετικά ασθενή θετική συσχέτιση

$b_1 = 0.063$ εξηγεί το βαθμό εξάρτησης;

$s_\epsilon^2 = 2.249$ εξηγεί το βαθμό εξάρτησης;

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

b_1 και b_0 είναι εκτιμητές των β_1 και β_0

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

b_1 και b_0 είναι εκτιμητές των β_1 και β_0

κατανομή των b_1 και b_0;

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

b_1 και b_0 είναι εκτιμητές των β_1 και β_0

κατανομή των b_1 και b_0;

b_1 γραμμικός συνδυασμός των τ.μ. $y_1, \ldots, y_n$

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

b_1 και b_0 είναι εκτιμητές των β_1 και β_0

κατανομή των b_1 και b_0;

b_1 γραμμικός συνδυασμός των τ.μ. $y_1, \ldots, y_n$

$$\mu_{b_1} \equiv \mathrm{E}(b_1) = \beta_1$$

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

b_1 και b_0 είναι εκτιμητές των β_1 και β_0

κατανομή των b_1 και b_0;

b_1 γραμμικός συνδυασμός των τ.μ. $y_1, \ldots, y_n$

$$\mu_{b_1} \equiv \mathrm{E}(b_1) = \beta_1$$

$$\sigma^2_{b_1} \equiv \mathrm{Var}(b_1) = \frac{\sigma^2_\epsilon}{S_{xx}} \quad \Longrightarrow \quad \sigma_{b_1} = \sigma_\epsilon / \sqrt{S_{xx}}$$

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

b_1 και b_0 είναι εκτιμητές των β_1 και β_0

κατανομή των b_1 και b_0;

b_1 γραμμικός συνδυασμός των τ.μ. $y_1, \ldots, y_n$

$$\mu_{b_1} \equiv \mathrm{E}(b_1) = \beta_1$$

$$\sigma^2_{b_1} \equiv \mathrm{Var}(b_1) = \frac{\sigma^2_\epsilon}{S_{xx}} \quad \Longrightarrow \quad \sigma_{b_1} = \sigma_\epsilon / \sqrt{S_{xx}}$$

Εκτίμηση:

$$s_{b_1} = \frac{s_\epsilon}{\sqrt{S_{xx}}}.$$

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

Y ακολουθεί κανονική κατανομή $\Longrightarrow$ b_1 ακολουθεί κανονική κατανομή

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

Y ακολουθεί κανονική κατανομή $\Longrightarrow$ b_1 ακολουθεί κανονική κατανομή

$(1-\alpha)\%$ διάστημα εμπιστοσύνης του β_1:

$$b_1 \pm t_{n-2,1-\alpha/2} s_{b_1} \quad \text{ή} \quad b_1 \pm t_{n-2,1-\alpha/2} \frac{s_\epsilon}{\sqrt{S_{xx}}}.$$

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

Y ακολουθεί κανονική κατανομή $\Longrightarrow$ b_1 ακολουθεί κανονική κατανομή

$(1-\alpha)\%$ διάστημα εμπιστοσύνης του β_1:

$$b_1 \pm t_{n-2,1-\alpha/2} s_{b_1} \quad \text{ή} \quad b_1 \pm t_{n-2,1-\alpha/2} \frac{s_\epsilon}{\sqrt{S_{xx}}}.$$

Αντίστοιχα για b_0

$$\sigma_{b_0} = s_\epsilon \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

Διάστημα εμπιστοσύνης των παραμέτρων της απλής γραμμικής παλινδρόμησης

Y ακολουθεί κανονική κατανομή $\Longrightarrow$ b_1 ακολουθεί κανονική κατανομή

$(1-\alpha)$% διάστημα εμπιστοσύνης του β_1:

$$b_1 \pm t_{n-2,1-\alpha/2} s_{b_1} \quad \text{ή} \quad b_1 \pm t_{n-2,1-\alpha/2} \frac{s_\epsilon}{\sqrt{S_{xx}}}.$$

Αντίστοιχα για b_0

$$\sigma_{b_0} = s_\epsilon \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

$(1-\alpha)$% διάστημα εμπιστοσύνης του β_0:

$$b_0 \pm t_{n-2,1-\alpha/2} s_\epsilon \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}.$$

Έλεγχος υπόθεσης για τις παραμέτρους της απλής γραμμικής παλινδρόμησης

H_0: $\beta_1 = \beta_1^0$

Έλεγχος υπόθεσης για τις παραμέτρους της απλής γραμμικής παλινδρόμησης

H$_0$: $\beta_1 = \beta_1^0$

Στατιστικό παραμετρικού ελέγχου

$$t = \frac{b_1 - \beta_1^0}{s_b} = \frac{(b_1 - \beta_1^0)\sqrt{S_{xx}}}{s_\epsilon}, \quad t \sim t_{n-2}$$

Έλεγχος υπόθεσης για τις παραμέτρους της απλής γραμμικής παλινδρόμησης

H_0: $\beta_1 = \beta_1^0$

Στατιστικό παραμετρικού ελέγχου

$$t = \frac{b_1 - \beta_1^0}{s_b} = \frac{(b_1 - \beta_1^0)\sqrt{S_{xx}}}{s_\epsilon}, \quad t \sim t_{n-2}$$

Ιδιαίτερο ενδιαφέρον έχει H_0: $\beta_1 = 0$ ή η Y δεν εξαρτάται από την X.

Έλεγχος υπόθεσης για τις παραμέτρους της απλής γραμμικής παλινδρόμησης

H$_0$: $\beta_1 = \beta_1^0$

Στατιστικό παραμετρικού ελέγχου

$$t = \frac{b_1 - \beta_1^0}{s_b} = \frac{(b_1 - \beta_1^0)\sqrt{S_{xx}}}{s_\epsilon}, \quad t \sim t_{n-2}$$

Ιδιαίτερο ενδιαφέρον έχει H$_0$: $\beta_1 = 0$ ή η Y δεν εξαρτάται από την X.

H$_0$: $\beta_0 = \beta_0^0$

Έλεγχος υπόθεσης για τις παραμέτρους της απλής γραμμικής παλινδρόμησης

H_0: $\beta_1 = \beta_1^0$

Στατιστικό παραμετρικού ελέγχου

$$t = \frac{b_1 - \beta_1^0}{s_b} = \frac{(b_1 - \beta_1^0)\sqrt{S_{xx}}}{s_\epsilon}, \quad t \sim t_{n-2}$$

Ιδιαίτερο ενδιαφέρον έχει H_0: $\beta_1 = 0$ ή η Y δεν εξαρτάται από την X.

H_0: $\beta_0 = \beta_0^0$

Στατιστικό παραμετρικού ελέγχου

$$t = \frac{b_0 - \beta_0^0}{s_\epsilon\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}}, \quad t \sim t_{n-2}$$

Διαστήματα πρόβλεψης

$\hat{y} = b_0 + b_1 x$: εκτιμητής της $\mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x$ για κάποιο x

Διαστήματα πρόβλεψης

$\hat{y} = b_0 + b_1 x$: εκτιμητής της $\mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x$ για κάποιο x

$\hat{y}$: γραμμικός συνδυασμός των τ.μ. $y_1, \ldots, y_n$, των $x_1, \ldots, x_n$ και x.

Διαστήματα πρόβλεψης

$\hat{y} = b_0 + b_1 x$: εκτιμητής της $\mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x$ για κάποιο x

$\hat{y}$: γραμμικός συνδυασμός των τ.μ. $y_1, \ldots, y_n$, των $x_1, \ldots, x_n$ και x.

$$\mu_{\hat{y}} \equiv \mathrm{E}(\hat{y}) = \mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x.$$

Διαστήματα πρόβλεψης

$\hat{y} = b_0 + b_1 x$: εκτιμητής της $\mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x$ για κάποιο x

$\hat{y}$: γραμμικός συνδυασμός των τ.μ. $y_1, \ldots, y_n$, των $x_1, \ldots, x_n$ και x.

$$\mu_{\hat{y}} \equiv \mathrm{E}(\hat{y}) = \mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x.$$

$$\sigma^2_{\hat{y}} \equiv \mathrm{Var}(\hat{y}) = \sigma^2_{\epsilon} \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right)$$

Διαστήματα πρόβλεψης

$\hat{y} = b_0 + b_1 x$: εκτιμητής της $\mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x$ για κάποιο x

$\hat{y}$: γραμμικός συνδυασμός των τ.μ. $y_1, \ldots, y_n$, των $x_1, \ldots, x_n$ και x.

$$\mu_{\hat{y}} \equiv \mathrm{E}(\hat{y}) = \mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x.$$

$$\sigma^2_{\hat{y}} \equiv \mathrm{Var}(\hat{y}) = \sigma^2_{\epsilon} \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right)$$

Εκτίμηση της τυπικής απόκλισης του $\hat{y}$

$$s_{\hat{y}} = s_{\epsilon} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}.$$

Διαστήματα πρόβλεψης

$\hat{y}$ για κάποιο x ακολουθεί κανονική κατανομή

Διαστήματα πρόβλεψης

$\hat{y}$ για κάποιο x ακολουθεί κανονική κατανομή

$(1-\alpha)$% διάστημα εμπιστοσύνης της μέσης τιμής του Y για κάποιο x

$$\hat{y} \pm t_{n-2,1-\alpha/2} s_{\hat{y}} \quad \text{ή} \quad (b_0 + b_1 x) \pm t_{n-2,1-\alpha/2} s_{\epsilon} \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}.$$

$\hat{y}$ για κάποιο x ακολουθεί κανονική κατανομή

$(1-\alpha)$% διάστημα εμπιστοσύνης της μέσης τιμής του Y για κάποιο x

$$\hat{y} \pm t_{n-2,1-\alpha/2} s_{\hat{y}} \quad \text{ή} \quad (b_0 + b_1 x) \pm t_{n-2,1-\alpha/2} s_{\epsilon} \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}.$$

Λέγεται και **διάστημα της μέσης πρόβλεψης**: τα όρια της πρόβλεψης για τη μέση (αναμενόμενη) τιμή της Y για κάποιο x

Διαστήματα πρόβλεψης

$\hat{y}$ για κάποιο x ακολουθεί κανονική κατανομή

$(1-\alpha)\%$ διάστημα εμπιστοσύνης της μέσης τιμής του Y για κάποιο x

$$\hat{y} \pm t_{n-2,1-\alpha/2} s_{\hat{y}} \quad \text{ή} \quad (b_0 + b_1 x) \pm t_{n-2,1-\alpha/2} s_{\epsilon} \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}.$$

Λέγεται και **διάστημα της μέσης πρόβλεψης**: τα όρια της πρόβλεψης για τη μέση (αναμενόμενη) τιμή της Y για κάποιο x

$(1-\alpha)\%$ διάστημα πρόβλεψης για μια παρατήρηση y της Y για κάποιο x

$$\hat{y} \pm t_{n-2,1-\alpha/2} \sqrt{s_{\epsilon}^2 + s_{\hat{y}}^2} \quad \text{ή} \quad (b_0 + b_1 x) \pm t_{n-2,1-\alpha/2} s_{\epsilon} \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Υπολογίσαμε $b_1 = 0.063$, $b_0 = 2.53$, $s_\epsilon = 1.507$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Υπολογίσαμε $b_1 = 0.063$, $b_0 = 2.53$, $s_\epsilon = 1.507$

Ακρίβεια / σημαντικότητα εκτιμήσεων και πρόβλεψης;

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Υπολογίσαμε $b_1 = 0.063$, $b_0 = 2.53$, $s_\epsilon = 1.507$

Ακρίβεια / σημαντικότητα εκτιμήσεων και πρόβλεψης;

β_1:

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Υπολογίσαμε $b_1 = 0.063$, $b_0 = 2.53$, $s_\epsilon = 1.507$

Ακρίβεια / σημαντικότητα εκτιμήσεων και πρόβλεψης;

β_1: $s_{b_1} = 0.0195$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Υπολογίσαμε $b_1 = 0.063$, $b_0 = 2.53$, $s_\epsilon = 1.507$

Ακρίβεια / σημαντικότητα εκτιμήσεων και πρόβλεψης;

β_1: $s_{b_1} = 0.0195$

95% διάστημα εμπιστοσύνης για β_1

$$0.063 \pm 2.306 \cdot 0.0195 \quad \Rightarrow \quad [0.018, 0.108].$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Υπολογίσαμε $b_1 = 0.063$, $b_0 = 2.53$, $s_\epsilon = 1.507$

Ακρίβεια / σημαντικότητα εκτιμήσεων και πρόβλεψης;

β_1: $s_{b_1} = 0.0195$

95% διάστημα εμπιστοσύνης για β_1

$$0.063 \pm 2.306 \cdot 0.0195 \quad \Rightarrow \quad [0.018, 0.108].$$

β_1 σημαντικά διάφορο του 0

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

Υπολογίσαμε $b_1 = 0.063$, $b_0 = 2.53$, $s_\epsilon = 1.507$

Ακρίβεια / σημαντικότητα εκτιμήσεων και πρόβλεψης;

β_1: $s_{b_1} = 0.0195$

95% διάστημα εμπιστοσύνης για β_1

$$0.063 \pm 2.306 \cdot 0.0195 \quad \Rightarrow \quad [0.018, 0.108].$$

β_1 σημαντικά διάφορο του 0

Στατιστικό ελέγχου για H_0: $\beta_1 = 0$

$$t = \frac{0.063}{0.0195} = 3.235$$

Υπολογίσαμε $b_1 = 0.063$, $b_0 = 2.53$, $s_\epsilon = 1.507$

Ακρίβεια / σημαντικότητα εκτιμήσεων και πρόβλεψης;

β_1: $s_{b_1} = 0.0195$

95% διάστημα εμπιστοσύνης για β_1

$$0.063 \pm 2.306 \cdot 0.0195 \quad \Rightarrow \quad [0.018, 0.108].$$

β_1 σημαντικά διάφορο του 0

Στατιστικό ελέγχου για H_0: $\beta_1 = 0$

$$t = \frac{0.063}{0.0195} = 3.235$$

$t > t_{0.975,8} = 2.306 \Longrightarrow \mathrm{H}_0$ απορρίπτεται.

Υπολογίσαμε $b_1 = 0.063$, $b_0 = 2.53$, $s_\epsilon = 1.507$

Ακρίβεια / σημαντικότητα εκτιμήσεων και πρόβλεψης;

β_1: $s_{b_1} = 0.0195$

95% διάστημα εμπιστοσύνης για β_1

$$0.063 \pm 2.306 \cdot 0.0195 \quad \Rightarrow \quad [0.018, 0.108].$$

β_1 σημαντικά διάφορο του 0

Στατιστικό ελέγχου για H_0: $\beta_1 = 0$

$$t = \frac{0.063}{0.0195} = 3.235$$

$t > t_{0.975,8} = 2.306 \Longrightarrow \mathrm{H}_0$ απορρίπτεται.

Η τιμή $t = 3.235$ αντιστοιχεί σε $p = 0.012$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

β_0:

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

β_0: $s_{b_0} = 1.894$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

β_0: $s_{b_0} = 1.894$

95% διάστημα εμπιστοσύνης για β_0

$$2.53 \pm 2.306 \cdot 1.894 \quad \Rightarrow \quad [-1.837, 6.898]$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

β_0: $s_{b_0} = 1.894$

95% διάστημα εμπιστοσύνης για β_0

$$2.53 \pm 2.306 \cdot 1.894 \quad \Rightarrow \quad [-1.837, 6.898]$$

β_0 μπορεί να είναι 0

β_0: $s_{b_0} = 1.894$

95% διάστημα εμπιστοσύνης για β_0

$$2.53 \pm 2.306 \cdot 1.894 \quad \Rightarrow \quad [-1.837, 6.898]$$

β_0 μπορεί να είναι 0

Στατιστικό ελέγχου για H_0: $\beta_1 = 0$

$$t = \frac{2.53}{1.894} = 1.336$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

β_0: $s_{b_0} = 1.894$

95% διάστημα εμπιστοσύνης για β_0

$$2.53 \pm 2.306 \cdot 1.894 \quad \Rightarrow \quad [-1.837, 6.898]$$

β_0 μπορεί να είναι 0

Στατιστικό ελέγχου για H_0: $\beta_1 = 0$

$$t = \frac{2.53}{1.894} = 1.336$$

$t < t_{0.975,8} = 2.306 \Longrightarrow H_0$ δεν απορρίπτεται.

β_0: $s_{b_0} = 1.894$

95% διάστημα εμπιστοσύνης για β_0

$$2.53 \pm 2.306 \cdot 1.894 \quad \Rightarrow \quad [-1.837, 6.898]$$

β_0 μπορεί να είναι 0

Στατιστικό ελέγχου για H_0: $\beta_1 = 0$

$$t = \frac{2.53}{1.894} = 1.336$$

$t < t_{0.975,8} = 2.306 \Longrightarrow H_0$ δεν απορρίπτεται.

Η τιμή $t = 1.336$ αντιστοιχεί σε $p = 0.218$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

διάστημα πρόβλεψης για αντίσταση στρώματος
$x = 120\,\mathrm{ohm/cm}$

διάστημα πρόβλεψης για αντίσταση στρώματος

$x = 120\,\mathrm{ohm/cm}$

$\hat{y}$:

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

διάστημα πρόβλεψης για αντίσταση στρώματος
$x = 120\,\text{ohm/cm}$

$\hat{y}$:

$$s_{\hat{y}} = 1.507\sqrt{\frac{1}{10} + \frac{(120 - 93.9)^2}{9 \cdot 662.1}} = 0.698$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

διάστημα πρόβλεψης για αντίσταση στρώματος
$x = 120\,\text{ohm/cm}$

$\hat{y}$:

$$s_{\hat{y}} = 1.507\sqrt{\frac{1}{10} + \frac{(120-93.9)^2}{9 \cdot 662.1}} = 0.698$$

95% διάστημα πρόβλεψης της $\hat{y}$ για $x = 120$

$$10.108 \pm 2.306 \cdot 0.698 \quad \Rightarrow \quad [8.499, 11.717]$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)

διάστημα πρόβλεψης για αντίσταση στρώματος
$x = 120\,\text{ohm/cm}$

$\hat{y}$:

$$s_{\hat{y}} = 1.507\sqrt{\frac{1}{10} + \frac{(120-93.9)^2}{9 \cdot 662.1}} = 0.698$$

95% διάστημα πρόβλεψης της $\hat{y}$ για $x = 120$

$$10.108 \pm 2.306 \cdot 0.698 \quad \Rightarrow \quad [8.499, 11.717]$$

παρατήρηση y:

95% διάστημα πρόβλεψης για μια (μελλοντική) παρατήρηση y
για $x = 120$

$$10.108 \pm 2.306 \cdot 1.507\sqrt{1 + \frac{1}{10} + \frac{(120-93.9)^2}{9 \cdot 662.1}} \quad \Rightarrow \quad [6.279, 13.937]$$

Παράδειγμα: Απολαβή ρεύματος τρανζίστορ (συνέχεια)