Τμήμα Επιστήμης Φυσικής Αγωγής & Αθλητισμού

Δημοκρίτειο Πανεπιστήμιο Θράκης

Στατιστική στη Φυσική Αγωγή

(Ν161)

9η & 10η Διάλεξη

Κοκκότης Χρήστος, PhD

Στόχοι Διάλεξης

Κατανόηση βασικών εννοιών:

- ✓ Παλινδρόμηση και συσχέτιση

Σε ορισμένες περιπτώσεις απαιτείται **η ανίχνευση της σχέσης μεταξύ δύο ποσοτικών μεταβλητών X και Y**

Άτομα	Μεταβλητή X	Μεταβλητή Y
1	X_1	Y_1
2	X_2	Y_2
3	X_3	Y_3
…	…	…
N	X_N	Y_N

Άτομα	Μεταβλητή Χ	Μεταβλητή Υ
1	X_1	Y_1
2	X_2	Y_2
3	X_3	Y_3
…	…	…
Ν	X_N	Y_N

α) ποια είναι η **σχέση μεταξύ των δύο** ποσοτικών μεταβλητών και

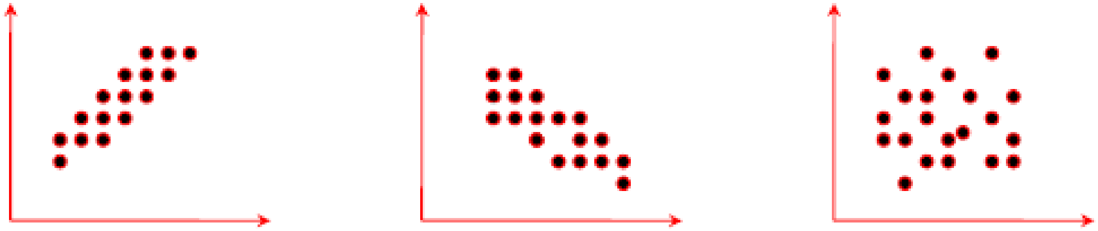
β) αν μπορούν να εκτιμηθούν, δηλαδή **να προβλεφθούν**,

οι τιμές της μιας μεταβλητής, γνωρίζοντας τις τιμές της άλλης μεταβλητής **(πρόβλημα παλινδρόμησης)**

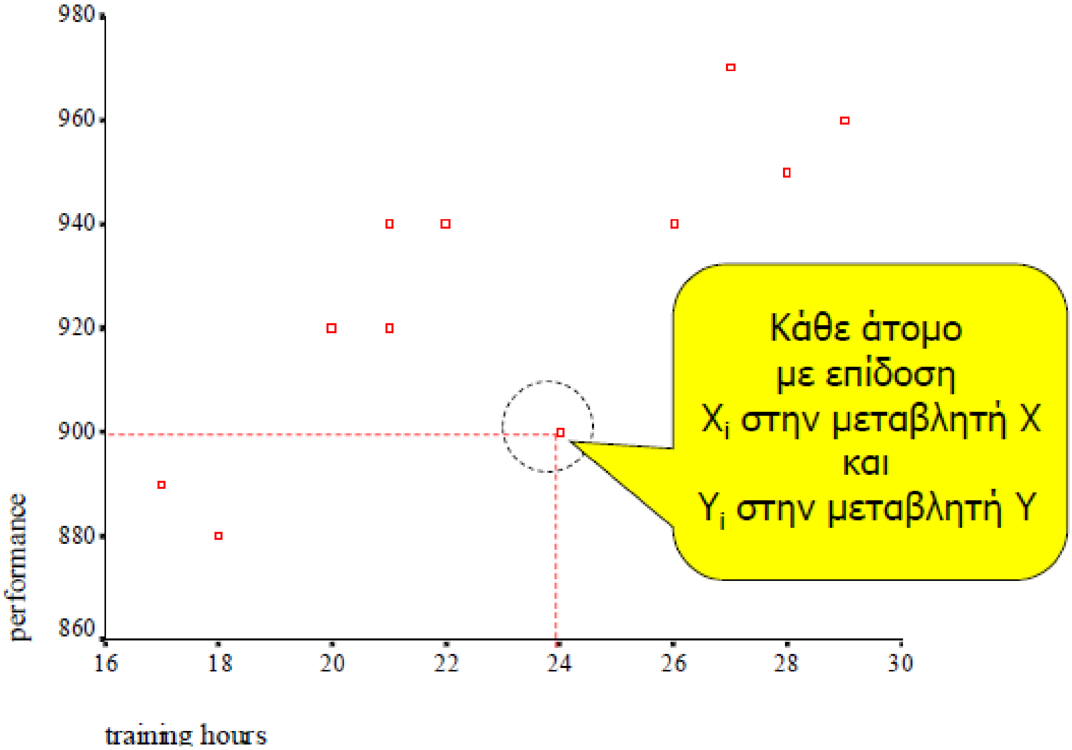
Γραμμική παλινδρόμηση του Χ στο Υ

Άτομα	Μεταβλητή X	Μεταβλητή Y
1	X_1	Y_1
2	X_2	Y_2
3	X_3	Y_3
…	…	…
N	X_N	Y_N

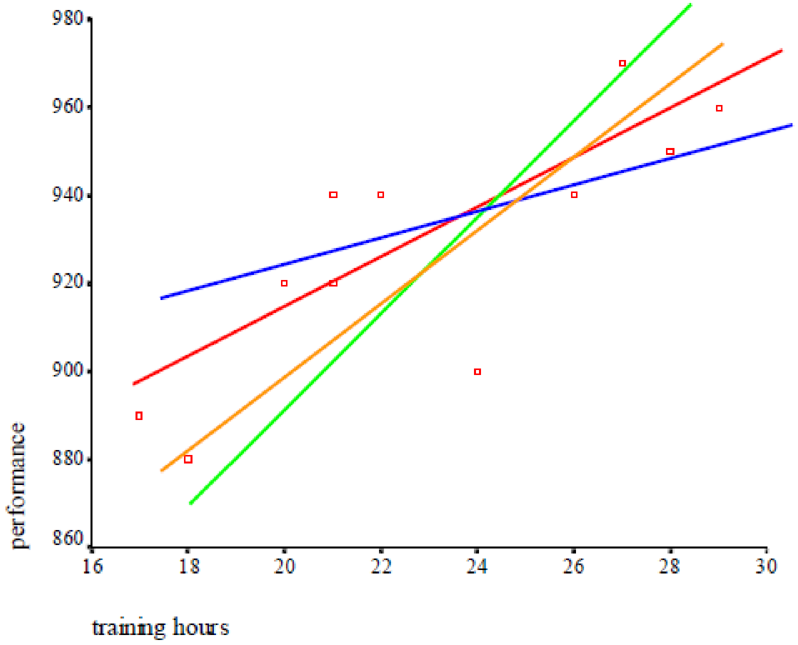
κάθε ζεύγος τιμών (X_i, Y_i) μπορεί να αναπαρασταθεί ως ένα σημείο με συντεταγμένες τις δύο τιμές του ζεύγους, δηλαδή την επίδοση του κάθε ατόμου στις μεταβλητές X και Y



Αναπαράσταση του κάθε ατόμου ως σημείο με συντεταγμένες (X_i, Y_i) τις τιμές του κάθε ατόμου στις μεταβλητές X και Y



Η αναπαράσταση του «νέφους των σημείων» μπορεί να γίνει με πολλές ευθείες γραμμές οι οποίες να έχουν διάφορες διευθύνσεις



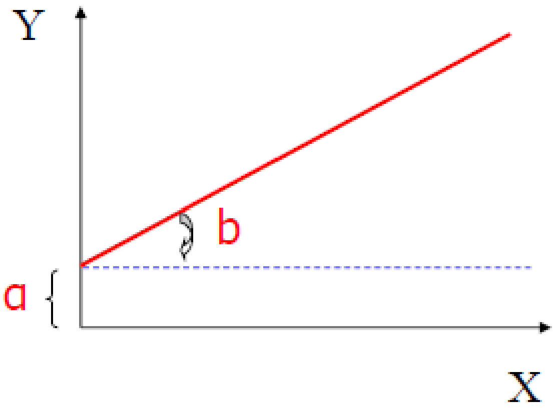
Ποια ευθεία (γραμμή παλινδρόμησης)
εκφράζει καλύτερα (με μεγαλύτερη ακρίβεια) τα δεδομένα;

Πώς καθορίζεται η διεύθυνση
αυτής της γραμμής παλινδρόμησης;

Κάθε ευθεία γραμμή σε ένα σύστημα συντεταγμένων, μπορεί να αναπαρασταθεί μαθηματικά μέσω μιας εξίσωσης που είναι της μορφής:

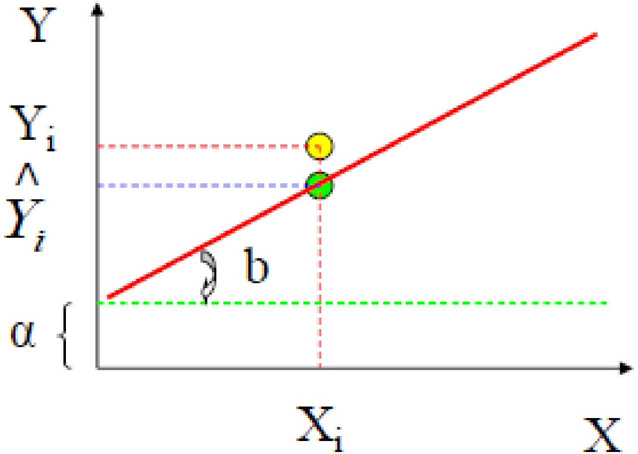
$$\mathbf{Y = \mathit{b}X + \mathit{a}}$$

- $\boldsymbol{a}$ είναι η απόσταση του σημείου τομής των αξόνων του συστήματος συντεταγμένων, δηλαδή του μηδενός, από το σημείο τομής της ευθείας με τον άξονα των Y και
- $\boldsymbol{b}$ αναφέρεται στην κλίση της ευθείας.



Μεγαλύτερη ακρίβεια γραμμής παλινδρόμησης όταν

$$\sum\left(Y - \hat{Y}\right)^2 = \text{ελάχιστο}$$



$$Y = a + b \cdot X_1 + e$$

$$\hat{Y} = a + b \cdot X_1$$

a = σταθερά (Constant)

b= κλίση ευθείας (slope)

$$b_{yx} = \frac{N\sum_{i=1}^{N}(X_i \cdot Y_i) - \sum_{i=1}^{N}X_i \cdot \sum_{i=1}^{N}Y_i}{N\sum_{i=1}^{N}X_i^2 - (\sum_{i=1}^{N}X_i)^2}$$

$$a_{yx} = \frac{\sum_{i=1}^{N}Y_i - b_{yx} \cdot \sum_{i=1}^{N}X_i}{N}$$

$\sum_{i=1}^{N} X_i$ το άθροισμα των τιμών της μεταβλητής X

$\sum_{i=1}^{N} Y_i$ το άθροισμα των τιμών της μεταβλητής Y

$\sum_{i=1}^{N} X_i^2$ το άθροισμα των τετραγώνων των τιμών της μεταβλητής X

$\sum_{i=1}^{N} (X_i \cdot Y_i)$ το άθροισμα των γινομένων των ζευγαρωτών τιμών στην μεταβλητή X και στην μεταβλητή Y

Παράδειγμα:

Ένας προπονητής καταγράφει τις επιδόσεις των αθλητών του (N= 7) στο άλμα σε μήκος (μεταβλητή Χ) και τις επιδόσεις τους στο τριπλούν (μεταβλητή Υ), έχοντας ως στόχο να μπορεί

να προβλέπει τις επιδόσεις των αθλητών του στο τριπλούν, βάσει των επιδόσεών τους στο άλμα σε μήκος.

α/α	Άλμα σε μήκος (X_i)	Άλμα τριπλούν (Y_i)
1	5.70	14.20
2	5.50	14.10
3	5.90	14.35
4	5.60	14.25
5	5.90	14.30
6	6.10	14.50
7	6.20	14.65

Για να υπολογιστεί η εξίσωση παλινδρόμησης θα πρέπει να καθοριστούν οι τιμές των

b_{yx} και a_{yx}

$$b_{yx} = \frac{N\sum_{i=1}^{N}(X_i \cdot Y_i) - \sum_{i=1}^{N} X_i \cdot \sum_{i=1}^{N} Y_i}{N\sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2}$$

$$a_{yx} = \frac{\sum_{i=1}^{N} Y_i - b_{yx} \cdot \sum_{i=1}^{N} X_i}{N}$$

$$\sum_{i=1}^{N} \boldsymbol{X_i} \qquad \sum_{i=1}^{N} \boldsymbol{Y_i} \qquad \sum_{i=1}^{N} \mathrm{X}_i^2 \qquad \sum_{i=1}^{N} \boldsymbol{X_i} \cdot \boldsymbol{Y_i}$$

α/α	**Άλμα σε μήκος (Xi)**	**Άλμα τριπλούν (Yi)**
1	5.70	14.20
2	5.50	14.10
3	5.90	14.35
4	5.60	14.25
5	5.90	14.30
6	6.10	14.50
7	6.20	14.65

$$b_{yx} = \frac{N\sum_{i=1}^{N}(X_i \cdot Y_i) - \sum_{i=1}^{N}X_i \cdot \sum_{i=1}^{N}Y_i}{N\sum_{i=1}^{N}X_i^2 - (\sum_{i=1}^{N}X_i)^2}$$

$$a_{yx} = \frac{\sum_{i=1}^{N}Y_i - b_{yx} \cdot \sum_{i=1}^{N}X_i}{N}$$

$$\sum_{i=1}^{N} \boldsymbol{X_i} \qquad \sum_{i=1}^{N} \boldsymbol{Y_i} \qquad \sum_{i=1}^{N} \mathrm{X}_i^2 \qquad \sum_{i=1}^{N} \boldsymbol{X_i \cdot Y_i}$$

α/α	Άλμα σε μήκος (X_i)	Άλμα τριπλούν (Y_i)	X_i^2
1	5.70	14.20	32.49
2	5.50	14.10	30.25
3	5.90	14.35	34 81
4	5.60	14.25	31.36
5	5.90	14.30	34.81
6	6.10	14.50	37.21
7	6.20	14.65	38.44

$$b_{yx} = \frac{N\sum_{i=1}^{N}(X_i \cdot Y_i) - \sum_{i=1}^{N} X_i \cdot \sum_{i=1}^{N} Y_i}{N\sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2}$$

$$a_{yx} = \frac{\sum_{i=1}^{N} Y_i - b_{yx} \cdot \sum_{i=1}^{N} X_i}{N}$$

$$\sum_{i=1}^{N} \boldsymbol{X_i} \quad \sum_{i=1}^{N} \boldsymbol{Y_i} \quad \sum_{i=1}^{N} X_i^2 \quad \sum_{i=1}^{N} \boldsymbol{X_i \cdot Y_i}$$

α/α	Άλμα σε μήκος (X_i)	Άλμα τριπλούν (Y_i)	X_i^2	$X_i \cdot Y_i$
1	5.70	14.20	32.49	80.94
2	5.50	14.10	30.25	77.55
3	5.90	14.35	34 81	84 665
4	5.60	14.25	31.36	79.8
5	5.90	14.30	34.81	84.37
6	6.10	14.50	37.21	88.45
7	6.20	14.65	38.44	90.83

$$b_{yx} = \frac{N\sum_{i=1}^{N}(X_i \cdot Y_i) - \sum_{i=1}^{N} X_i \cdot \sum_{i=1}^{N} Y_i}{N\sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2}$$

$$a_{yx} = \frac{\sum_{i=1}^{N} Y_i - b_{yx} \cdot \sum_{i=1}^{N} X_i}{N}$$

$$\sum_{i=1}^{N} \boldsymbol{X_i} \quad \sum_{i=1}^{N} \boldsymbol{Y_i} \quad \sum_{i=1}^{N} X_i^2 \quad \sum_{i=1}^{N} \boldsymbol{X_i \cdot Y_i}$$

α/α	Άλμα σε μήκος (X_i)	Άλμα τριπλούν (Y_i)	X_i^2	$X_i \cdot Y_i$
1	5.70	14.20	32.49	80.94
2	5.50	14.10	30.25	77.55
3	5.90	14.35	34 81	84 665
4	5.60	14.25	31.36	79.8
5	5.90	14.30	34.81	84.37
6	6.10	14.50	37.21	88.45
7	6.20	14.65	38.44	90.83
	$= \sum_{i=1}^{N} \boldsymbol{X_i} = 40.9$			

$$b_{yx} = \frac{N\sum_{i=1}^{N}(X_i \cdot Y_i) - \sum_{i=1}^{N} X_i \cdot \sum_{i=1}^{N} Y_i}{N\sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2}$$

$$a_{yx} = \frac{\sum_{i=1}^{N} Y_i - b_{yx} \cdot \sum_{i=1}^{N} X_i}{N}$$

$$\sum_{i=1}^{N} \boldsymbol{X_i} \quad \sum_{i=1}^{N} \boldsymbol{Y_i} \quad \sum_{i=1}^{N} X_i^2 \quad \sum_{i=1}^{N} \boldsymbol{X_i \cdot Y_i}$$

α/α	Άλμα σε μήκος (X_i)	Άλμα τριπλούν (Y_i)	X_i^2	$X_i \cdot Y_i$
1	5.70	14.20	32.49	80.94
2	5.50	14.10	30.25	77.55
3	5.90	14.35	34 81	84 665
4	5.60	14.25	31.36	79.8
5	5.90	14.30	34.81	84.37
6	6.10	14.50	37.21	88.45
7	6.20	14.65	38.44	90.83
	$= \sum_{i=1}^{N} X_i = 40.9$	$= \sum_{i=1}^{N} Y_i = 110.35$		

$$b_{yx} = \frac{N\sum_{i=1}^{N}(X_i \cdot Y_i) - \sum_{i=1}^{N} X_i \cdot \sum_{i=1}^{N} Y_i}{N\sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2}$$

$$a_{yx} = \frac{\sum_{i=1}^{N} Y_i - b_{yx} \cdot \sum_{i=1}^{N} X_i}{N}$$

$$\sum_{i=1}^{N} \boldsymbol{X_i} \quad \sum_{i=1}^{N} \boldsymbol{Y_i} \quad \sum_{i=1}^{N} X_i^2 \quad \sum_{i=1}^{N} \boldsymbol{X_i \cdot Y_i}$$

α/α	Άλμα σε μήκος (X_i)	Άλμα τριπλούν (Y_i)	X_i^2	$X_i \cdot Y_i$
1	5.70	14.20	32.49	80.94
2	5.50	14.10	30.25	77.55
3	5.90	14.35	34 81	84 665
4	5.60	14.25	31.36	79.8
5	5.90	14.30	34.81	84.37
6	6.10	14.50	37.21	88.45
7	6.20	14.65	38.44	90.83
	$= \sum_{i=1}^{N} \boldsymbol{X_i} = 40.9$	$= \sum_{i=1}^{N} \boldsymbol{Y_i} = 110.35$	$= \sum_{i=1}^{N} X_i^2 = 239.37$	

$$b_{yx} = \frac{N\sum_{i=1}^{N}(X_i \cdot Y_i) - \sum_{i=1}^{N}X_i \cdot \sum_{i=1}^{N}Y_i}{N\sum_{i=1}^{N}X_i^2 - (\sum_{i=1}^{N}X_i)^2}$$

$$a_{yx} = \frac{\sum_{i=1}^{N}Y_i - b_{yx} \cdot \sum_{i=1}^{N}X_i}{N}$$

$$\sum_{i=1}^{N}\boldsymbol{X_i} \quad \sum_{i=1}^{N}\boldsymbol{Y_i} \quad \sum_{i=1}^{N}\mathrm{X}_i^2 \quad \sum_{i=1}^{N}\boldsymbol{X_i \cdot Y_i}$$

α/α	Άλμα σε μήκος (X_i)	Άλμα τριπλούν (Y_i)	X_i^2	$X_i \cdot Y_i$
1	5.70	14.20	32.49	80.94
2	5.50	14.10	30.25	77.55
3	5.90	14.35	34 81	84 665
4	5.60	14.25	31.36	79.8
5	5.90	14.30	34.81	84.37
6	6.10	14.50	37.21	88.45
7	6.20	14.65	38.44	90.83
	$= \sum_{i=1}^{N}X_i = 40.9$	$= \sum_{i=1}^{N}Y_i = 110.35$	$= \sum_{i=1}^{N}X_i^2 = 239.37$	$= \sum_{i=1}^{N}X_i \cdot Y_i = 586.605$

$$b_{yx} = \frac{N\sum_{i=1}^{N}(X_i \cdot Y_i) - \sum_{i=1}^{N}X_i \cdot \sum_{i=1}^{N}Y_i}{N\sum_{i=1}^{N}X_i^2 - (\sum_{i=1}^{N}X_i)^2} =$$

$$\frac{7 \cdot 586.605 - 40.9 \cdot 100{,}35}{7 \cdot 239.37 - (40.9)^2} = \frac{4106.235 - 4104.315}{1675.59 - 1672.81} = \frac{1.92}{2.78} = 0.69$$

$$a_{yx} = \frac{\sum_{i=1}^{N}Y_i - b_{yx} \cdot \sum_{i=1}^{N}X_i}{N} =$$

$$\frac{100.35 - 0.69 \cdot 40.9}{7} = \frac{100.35 - 28.221}{7} = \frac{72.129}{7} = 10.3$$

Συνεπώς η εξίσωση παλινδρόμησης θα είναι

$$Y' = b_{yx}X + a_{yx} = 0.69 \cdot X + 10.3$$

$$\Upsilon' = b_{yx}X + a_{yx} = 0.69 \cdot X + 10.3$$

Άρα αν κάποιος αθλητής, με τα ίδια χαρακτηριστικά των αθλητών του δείγματος, έχει επίδοση στο άλμα σε μήκος 6m, τότε η αναμενόμενη επίδοσή του στο άλμα τριπλούν θα είναι

$$\Upsilon' = b_{yx}X + a_{yx} = 0.69 \cdot X + 10.3 = 4.14 + 10.3 = 14.44$$

Συντελεστής προσδιορισμού

Η συνολική μεταβολή του Y, δηλαδή η συνολική διακύμανση των επιδόσεων στην μεταβλητή Y, εκφράζεται από

το άθροισμα των τετραγώνων

<u>της διαφοράς της μέσης τιμής όλων των ατόμων από την επίδοση του κάθε ατόμου</u>

$$\sum (Y - \bar{Y})^2$$

και οφείλεται σε δύο παράγοντες

1. Στη μεταβολή της μεταβλητής Y, η οποία οφείλεται στην μεταβολή της μεταβλητής X:

$$\sum (Y' - \bar{Y})^2$$

2. Σε άλλους παράγοντες που δεν μπορούν να εκτιμηθούν

$$\sum \left(Y - Y^{!}\right)^2$$

Το ποσό της διακύμανσης, δηλαδή της μεταβολής του Υ, το οποίο μπορεί να εκτιμηθεί βάσει της εξίσωσης παλινδρόμησης, σε σχέση με την συνολική διακύμανση του Υ, αποτελεί έναν δείκτη που εκφράζει το βαθμό της σχέσης μεταξύ των δύο μεταβλητών Χ και Υ.

Συντελεστής προσδιορισμού

$$r^2 = \sqrt{\frac{\Sigma(Y' - \bar{Y})^2}{\Sigma(Y - \bar{Y})^2}}$$

$$\frac{\text{προβλεπόμενη μεταβολή της μεταβλητής Υ που οφείλεται στην μεταβλητή Χ}}{\text{συνολική μεταβολή της μεταβλητής Υ,}}$$

παίρνει τιμές από 0 έως 1

μέτρο εκτίμησης της σχέσης μεταξύ Χ και Υ

Συντελεστής συσχέτισης

η διεύθυνση της σχέσης μεταξύ X και Y ορίζεται από την τετραγωνική ρίζα του συντελεστή προσδιορισμού

Συντελεστής συσχέτισης $$r = \pm \sqrt{\frac{\sum (Y' - \bar{Y})^2}{\sum (Y - \bar{Y})^2}}$$

Ο συντελεστής συσχέτισης δεν είναι λοιπόν μια αναλογία, αλλά η τετραγωνική ρίζα μιας αναλογίας.

Αυτό σημαίνει ότι, ο συντελεστής συσχέτισης που έχει την τιμή **0.80 δεν εκφράζει μια σχέση δύο φορές μεγαλύτερη**, από τη σχέση που εκφράζεται από έναν συντελεστή συσχέτισης με τιμή **0.40**.

Αν απαιτείται τέτοιας μορφής πληροφορία, τότε χρησιμοποιείται ο «συντελεστής προσδιορισμού».

Για παράδειγμα,

όταν ο συντελεστής συσχέτισης μεταξύ δύο ποσοτικών μεταβλητών X και Y είναι **r = 0.80**, τότε ο συντελεστής προσδιορισμού θα είναι r^2= **0.64**

και αυτό σημαίνει ότι

το 64% της μεταβολής, δηλαδή της διακύμανσης, **της μεταβλητής Y μπορεί να προβλεφθεί από την μεταβολή της μεταβλητής X.**

Η αλγεβρική μορφή του συντελεστή συσχέτισης εκφράζεται από τη σχέση

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X})\cdot(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{N}(Y_i - \overline{Y})^2}}$$

Παράδειγμα:

Ένας προπονητής ενδιαφέρεται για το **βαθμό συσχέτισης** μεταξύ των **εβδομαδιαίων ωρών προπόνησης** των αθλητών του (**μεταβλητή Χ**) και της **επίδοσης τους στην σκοποβολή** (**μεταβλητή Υ**).

Για το σκοπό αυτό καταγράφει τις εβδομαδιαίες ώρες προπόνησης και τις επιδόσεις των 12 αθλητών του (Ν = 12).

α/α	(X_i)	(Y_i)
1	20	920
2	21	940
3	18	880
4	17	890
5	24	900
6	26	940
7	28	950
8	29	960
9	22	940
10	21	920
11	28	950
12	27	970

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X}) \cdot (Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{N}(Y_i - \overline{Y})^2}}$$

θα πρέπει αρχικά να υπολογιστούν

• η μέση τιμή της μεταβλητής X,

δηλαδή η $\overline{X}$

• η μέση τιμή της μεταβλητής Y,

δηλαδή η $\overline{Y}$

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

$$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$$

α/α	(X_i)	(Y_i)
1	20	920
2	21	940
3	18	880
4	17	890
5	24	900
6	26	940
7	28	950
8	29	960
9	22	940
10	21	920
11	28	950
12	27	970

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X}) \cdot (Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{N}(Y_i - \overline{Y})^2}}$$

θα πρέπει αρχικά να υπολογιστούν

- η μέση τιμή της μεταβλητής X,

δηλαδή η $\overline{X}$

- η μέση τιμή της μεταβλητής Y,

δηλαδή η $\overline{Y}$

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{281}{12} = 23.42$$

α/α	(X_i)	(Y_i)
1	20	920
2	21	940
3	18	880
4	17	890
5	24	900
6	26	940
7	28	950
8	29	960
9	22	940
10	21	920
11	28	950
12	27	970
	$= \sum_{i=1}^{N} X_i = 281$	

$$r = \frac{\sum_{i=1}^{N}(X_i-\overline{X})\cdot(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i-\overline{X})^2\cdot\sum_{i=1}^{N}(Y_i-\overline{Y})^2}}$$

θα πρέπει αρχικά να υπολογιστούν

- η μέση τιμή της μεταβλητής X,

δηλαδή η $\overline{X}$

- η μέση τιμή της μεταβλητής Y,

δηλαδή η $\overline{Y}$

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{281}{12} = 23.42$$

$$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{11160}{12} = 930$$

α/α	(X_i)	(Y_i)
1	20	920
2	21	940
3	18	880
4	17	890
5	24	900
6	26	940
7	28	950
8	29	960
9	22	940
10	21	920
11	28	950
12	27	970
	$= \sum_{i=1}^{N} X_i = 281$	$= \sum_{i=1}^{N} Y_i = 11160$

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X}) \cdot (Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{N}(Y_i - \overline{Y})^2}}$$

θα πρέπει να υπολογιστούν

$(X_i - \overline{X})$

$(X_i - \overline{X})^2$

$\sum_{i=1}^{N}(X_i - \overline{X})^2$

$(Y_i - \overline{Y})$

$(Y_i - \overline{Y})^2$

$\sum_{i=1}^{N}(Y_i - \overline{Y})^2$

$(X_i - \overline{X}) \cdot (Y_i - \overline{Y})$

$\sum_{i=1}^{N}(X_i - \overline{X}) \cdot (Y_i - \overline{Y})$

α/α	(X_i)	(Y_i)
1	20	920
2	21	940
3	18	880
4	17	890
5	24	900
6	26	940
7	28	950
8	29	960
9	22	940
10	21	920
11	28	950
12	27	970
	$= \sum_{i=1}^{N} X_i = 281$	$= \sum_{i=1}^{N} Y_i = 11160$

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X})\cdot(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{N}(Y_i - \overline{Y})^2}}$$

θα πρέπει να υπολογιστούν

$(X_i - \overline{X})$

$(X_i - \overline{X})^2$

$\sum_{i=1}^{N}(X_i - \overline{X})^2$

$(Y_i - \overline{Y})$

$(Y_i - \overline{Y})^2$

$\sum_{i=1}^{N}(Y_i - \overline{Y})^2$

$(X_i - \overline{X})\cdot(Y_i - \overline{Y})$

$\sum_{i=1}^{N}(X_i - \overline{X})\cdot(Y_i - \overline{Y})$

α/α	(X_i)	(Y_i)	$(X_i - \overline{X})$
1	20	920	-3.42
2	21	940	-2.42
3	18	880	-5.42
4	17	890	-6.42
5	24	900	0.58
6	26	940	2.58
7	28	950	4.58
8	29	960	5.58
9	22	940	-1.42
10	21	920	-2.42
11	28	950	4.58
12	27	970	3.58
	$= \sum_{i=1}^{N} X_i = 281$	$= \sum_{i=1}^{N} Y_i = 11160$	

$$r = \frac{\sum_{i=1}^{N}(X_i-\overline{X})\cdot(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i-\overline{X})^2\cdot\sum_{i=1}^{N}(Y_i-\overline{Y})^2}}$$

θα πρέπει να υπολογιστούν

$(X_i-\overline{X})$

$(X_i-\overline{X})^2$

$\sum_{i=1}^{N}(X_i-\overline{X})^2$

$(Y_i-\overline{Y})$

$(Y_i-\overline{Y})^2$

$\sum_{i=1}^{N}(Y_i-\overline{Y})^2$

$(X_i-\overline{X})\cdot(Y_i-\overline{Y})$

$\sum_{i=1}^{N}(X_i-\overline{X})\cdot(Y_i-\overline{Y})$

α/α	(X_i)	(Y_i)	$(X_i-\overline{X})$	$(X_i-\overline{X})^2$	$(Y_i-\overline{Y})$	$(Y_i-\overline{Y})^2$	$(X_i-\overline{X})(Y_i-\overline{Y})$
1	20	920	-3.42	11.69	-10	100	34.2
2	21	940	-2.42	5.85	10	100	-24.2
3	18	880	-5.42	29.37	-50	2500	271
4	17	890	-6.42	41.21	-40	1600	256.8
5	24	900	0.58	0.33	-30	900	-17.4
6	26	940	2.58	6.65	10	100	25.8
7	28	950	4.58	20.97	20	400	91.6
8	29	960	5.58	31.13	30	900	167.4
9	22	940	-1.42	2.01	10	100	-14.2
10	21	920	-2.42	5.85	-10	100	24.2
11	28	950	4.58	20.97	20	400	91.6
12	27	970	3.58	12.81	40	1600	143.2
	$=\sum_{i=1}^{N}X_i$ $= 281$	$=\sum_{i=1}^{N}Y_i$ $= 11160$		$=\sum_{i=1}^{N}(X_i-\overline{X})^2$ $= 188.84$		$=\sum_{i=1}^{N}(Y_i-\overline{Y})^2$ $= 8800$	$=\sum_{i=1}^{N}(X_i-\overline{X})(Y_i-\overline{Y})$ $= 1050$

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X}) \cdot (Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{N}(Y_i - \overline{Y})^2}} = \frac{1050}{\sqrt{188.84 \cdot 8800}} = \frac{1050}{1289.1} = 0.814$$

Συνεπώς ο **συντελεστής συσχέτισης** μεταξύ των εβδομαδιαίων ωρών προπόνησης και της επίδοσης στην σκοποβολή είναι

$r = 0.814$

Άρα, ο συντελεστής προσδιορισμού θα είναι

$r^2 = (0.814)^2 = 0.663$

Αυτό σημαίνει ότι το 66.3% της επίδοσης στη σκοποβολή οφείλεται στις εβδομαδιαίες ώρες προπόνησης,

ενώ

το υπόλοιπο 33.7% οφείλεται σε άλλους παράγοντες,

οι οποίοι δεν μπορούν να υπολογιστούν με αυτά τα δεδομένα.

Έλεγχος ανεξαρτησίας των δύο ποσοτικών μεταβλητών Χ και Υ

κατά πόσο οι δύο ποσοτικές μεταβλητές Χ και Υ είναι ανεξάρτητες ή όχι μεταξύ τους σε **επίπεδο πληθυσμού**.

r = δειγματικός συντελεστής συσχέτισης

ρ = συντελεστής συσχέτισης πληθυσμού

με βάση τον r να εκτιμήσουμε τον ρ

μηδενική υπόθεση: $\boldsymbol{H_o}$**: ρ=0**

(οι μεταβλητές είναι ανεξάρτητες μεταξύ τους)

εναλλακτική υπόθεση: $\boldsymbol{H_A}$**: ρ ≠ 0**

(οι μεταβλητές δεν είναι ανεξάρτητες μεταξύ τους, αλλά συσχετίζονται)

Έλεγχος μηδενικής υπόθεσης $\boldsymbol{H_o}$: ρ=0

(οι μεταβλητές είναι ανεξάρτητες μεταξύ τους)

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

Αν η υπολογιζόμενη t – τιμή είναι **μικρότερη από την κρίσιμη t – τιμή**, η οποία εντοπίζεται στον πίνακα της t – κατανομής για

- **N – 2 βαθμούς ελευθερίας**, όπου N = το μέγεθος του δείγματος, και
- προεπιλεγμένο **επίπεδο σημαντικότητας α**,

τότε γίνεται **αποδεκτή η μηδενική υπόθεση**.

Παράδειγμα:

Ο συντελεστής συσχέτισης μεταξύ των εβδομαδιαίων ωρών προπόνησης και της επίδοσης στην σκοποβολή ενός δείγματος N= 12 ατόμων είναι **r = 0.814**

Ερώτημα*:*

κατά πόσο αυτός ο βαθμός συσχέτισης μεταξύ των δύο μεταβλητών

μπορεί να επεκταθεί και σε επίπεδο πληθυσμού, με επίπεδο σημαντικότητας α= 0.05.

Αρχικά θα πρέπει να υπολογιστεί το στατιστικό t.

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

N= 12

r = 0.814

Αρχικά θα πρέπει να υπολογιστεί το στατιστικό t.

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} = \frac{0.814\sqrt{12-2}}{\sqrt{1-\ 0.814^2}}$$

N= 12

r = 0.814

Αρχικά θα πρέπει να υπολογιστεί το στατιστικό t.

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} = \frac{0.814\sqrt{12-2}}{\sqrt{1-0.814^2}} = \frac{2.57}{0.58} = 4.43$$

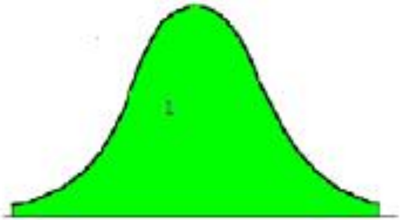
N= 12

r = 0.814

Στη συνέχεια πρέπει **να εντοπιστεί η κρίσιμη t – τιμή** μέσα από τον πίνακα της t-κατανομής.

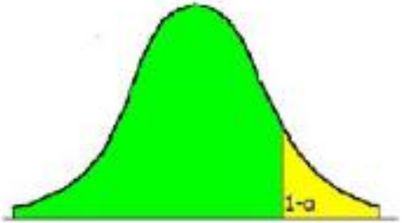
Εφόσον η εναλλακτική υπόθεση διατυπώνεται ως H_A: ρ ≠ 0, πρόκειται για **δίπλευρο έλεγχο.**

Το συνολικό εμβαδόν κάτω από την καμπύλη της t – κατανομής ισούται με 1.

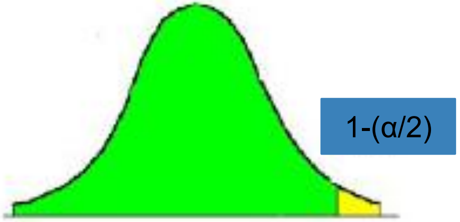


Κατά συνέπεια για να εντοπιστεί η t – τιμή που αντιστοιχεί στο επίπεδο σημαντικότητας α = 0.05 θα πρέπει να δημιουργηθεί η διαφορά:

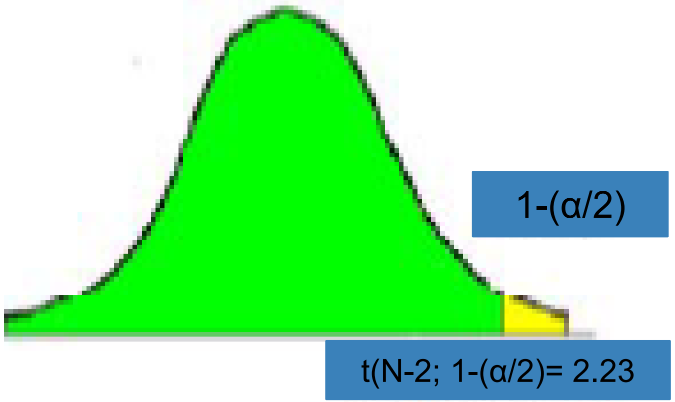
1 – α = 1- 0.05.



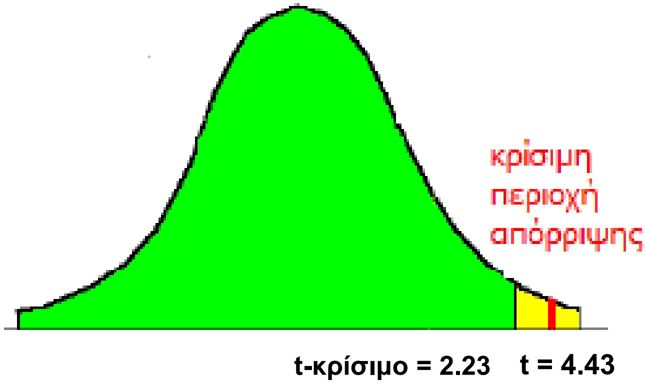
Επειδή όμως μας ενδιαφέρουν οι τιμές οι οποίες βρίσκονται και δεξιά και αριστερά από τον μέσο όρο, θα πρέπει να υπολογιστεί η διαφορά: 1 – α/2= 1- 0.05/2= 1- 0.025 = 0.975.



- 1 – (α/2)= 1- (0.05/2)=
  1- 0.025 = 0.975
- N – 2 = 12 – 2 = 10



df | p	.60	.70	.80	.90	.95	.975	.99	.995
1	.325	.727	1.367	3.08	6.31	12.71	31.82	63.66
2	.289	.617	1.061	1.89	2.92	4.30	6.96	9.92
3	.277	.584	.978	1.64	2.35	3.18	4.54	5.84
4	.271	.569	.941	1.53	2.13	2.78	3.75	4.60
5	.267	.559	.920	1.48	2.02	2.57	3.36	4.03
6	.265	.553	.906	1.44	1.94	2.45	3.14	3.71
7	.263	.549	.896	1.42	1.90	2.36	3.00	3.50
8	.262	.546	.889	1.40	1.86	2.31	2.90	3.36
9	.261	.543	.883	1.38	1.83	2.26	2.82	3.25
10	.260	.542	.879	1.37	1.81	2,23	2.76	3.17
11	.260	.540	.876	1.36	1.80	2.20	2.72	3.11
12	.259	.539	.873	1.36	1.78	2.18	2.68	3.06
13	.259	.538	.870	1.35	1.77	2.16	2.65	3.01
14	.258	.537	.868	1.34	1.76	2.14	2.62	2.98
15	.258	.536	.866	1.34	1.75	2.13	2.60	2.95
16	.258	.535	.865	1.34	1.75	2.12	2.58	2.92
17	.257	.534	.863	1.33	1.74	2.11	2.57	2.90
18	.257	.534	.862	1.33	1.73	2.10	2.55	2.88
19	.257	.533	.861	1.33	1.73	2.09	2.54	2.86
20	.257	.533	.860	1.32	1.72	2.09	2.53	2.84



Εφόσον η υπολογιζόμενη **t – τιμή (t= 4.43) είναι μεγαλύτερη από την $t_{κρίσιμη}$= 2.23**, βρίσκεται δηλαδή στην κρίσιμη περιοχή, απορρίπτεται η μηδενική υπόθεση.

Άρα **οι δύο ποσοτικές μεταβλητές**, εβδομαδιαίες ώρες προπόνησης και επίδοση στην σκοποβολή, δεν είναι ανεξάρτητες μεταξύ τους, αλλά **συσχετίζονται στατιστικά σημαντικά**.

Ερωτήσεις & Συζήτηση