

600 ανεξάρτητες μεταβλητές, αυξάνεται η μέγεθος σταθερό το  $R^2$ .

Θέλουμε καλό μοντέλο, άρα υψώνει τη για  $R^2$ , αλλά μικρό αριθμό ανεξάρτητων μεταβλητών.

Το  $R^2$  δεν εξαρτάται από το  $k$ .

Adjusted  $R^2$ :

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

(Προσαρμοσμένος συντελεστής μεταβλητότητας)

- $0 \leq R^2 \leq 1$

- $R^2_{adj} \leq 1$

$R^2_{adj}$  μπορεί να πάρει και αρνητικές τιμές.

- Ανάμεσα σε δύο μοντέλα διαλέγω εκείνο με μεγαλύτερη τιμή  $R^2_{adj}$  (Σιγχαίμ μοντέλο)

- $k$  παρανοήσιμη: Δουλεύει σαν "ποινή".

Αυξάνοντας αρκεί  $k$  μειώνεται  $R^2_{adj}$

Μπορώ να διαλέγω ποιο μοντέλο είναι καλύτερο

Πχ Υποδομή χωραφών:

	$x_1$ ήλιος	$x_2$ εμβαλ.	$x_3$ εμβαλ.	$x_4$ θεωρία	$x_5$ σημεία	$x_6$ χιόνι	$x_7$ τρακτύ	$x_8$ δ-τάση	$x_9$ υποψήφιο	$x_{10}$ 114 βωί
$R^2$	0,4	0,6	0,7	0,78	0,9	0,91	0,91	0,91	0,92	max 0,93
$R^2_{adj}$	0,32	0,41	0,56	0,71	0,74	0,43	0,2	-0,5	-12	-13

$$Y = 150 + 30 \cdot x_1 + 100 \cdot x_2 + 0,2 \cdot x_3 - 2 \cdot x_4 + 5 \cdot x_6 + 0,07 \cdot x_7 + 0,08 \cdot x_8$$



# Έλεγχος του Durbin-Watson (d)

για αυτοσυσχέτιση

- Με αυτοσυσχέτιση πρώτου βαθμού έχουμε:

$$Y_t = b_0 + b_1 X_{t1} + \dots + b_k X_{tk} + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t \quad -1 \leq \rho \leq +1$$

- Υποθέσεις:  $H_0: \rho = 0$

$$H_1: \rho \neq 0 \rightarrow \text{Cov}(\varepsilon_t, \varepsilon_{t+s}) \neq 0$$

(δεν έχω αυτοσυσχέτιση)

(έχω αυτοσυσχέτιση)

- Υπολογίζουμε με την MET, τα κατάλοιπα του αρχικού μας μοντέλου και με βάση τα αποτελέσματα αυτά, υπολογίζουμε την στατιστική d:

$$d = \frac{\sum_{i=2}^T (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_{i=2}^T \hat{\varepsilon}_i^2}$$

Το SPSS υπολογίζει την στατιστική d!

- Η στατιστική d του Durbin-Watson:  $d = 2 - 2\rho$  όπου  $0 \leq d \leq 4$

➤ Όταν δεν υπάρχει αυτοσυσχέτιση:  $\rho = 0 \rightarrow d = 2$

➤ Με πλήρη θετική αυτοσυσχέτιση:  $\rho = +1 \rightarrow d = 0$

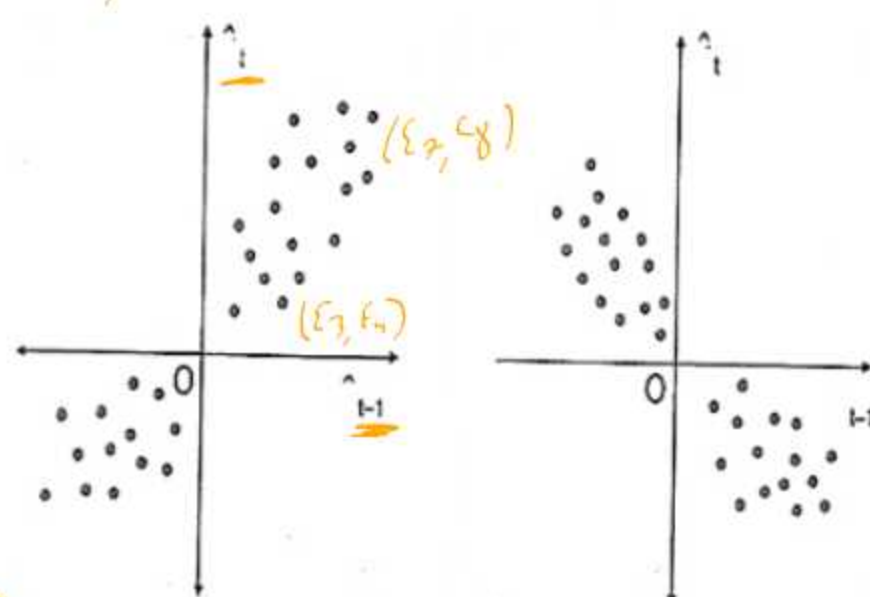
➤ Με πλήρη αρνητική αυτοσυσχέτιση:  $\rho = -1 \rightarrow d = 4$

(κοντά στο 2) (καλό μοντέλο)  
αλλιώς πρέπει να αλλάξω το μοντέλο μου

$$d = \frac{\sum_{n=2}^N (\varepsilon_n - \varepsilon_{n-1})^2}{\sum_{n=2}^N \varepsilon_n^2} = \frac{(\varepsilon_2 - \varepsilon_1)^2 + (\varepsilon_3 - \varepsilon_2)^2 + (\varepsilon_4 - \varepsilon_3)^2 + \dots + (\varepsilon_N - \varepsilon_{N-1})^2}{\varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_4^2 + \dots + \varepsilon_N^2}$$

## Διάγραμμα της Διασποράς των

σε σχέση με Καταλοίπων (1/4)



Εικόνα 1 (α): Διάγραμμα της Διασποράς των Καταλοίπων.



# 1. Έλεγχος του Durbin-Watson (d)

3

- Η κατανομή της στατιστικής  $d$  **δεν είναι ακριβής**, εξαρτάται από την ακολουθία τόσο των καταλοίπων όσο και των ερμηνευτικών μεταβλητών. Όμως οι Durbin-Watson έδειξαν ότι η κατανομή της στατιστικής  $d$  κυμαίνεται μεταξύ δύο άλλων τιμών:  $d_L$  (κατώτερο όρο) και  $d_U$  (ανώτερο όρο).
- Οι τιμές αυτές δίνονται από το Πίνακα του Durbin-Watson (σε έντυπη μορφή!)

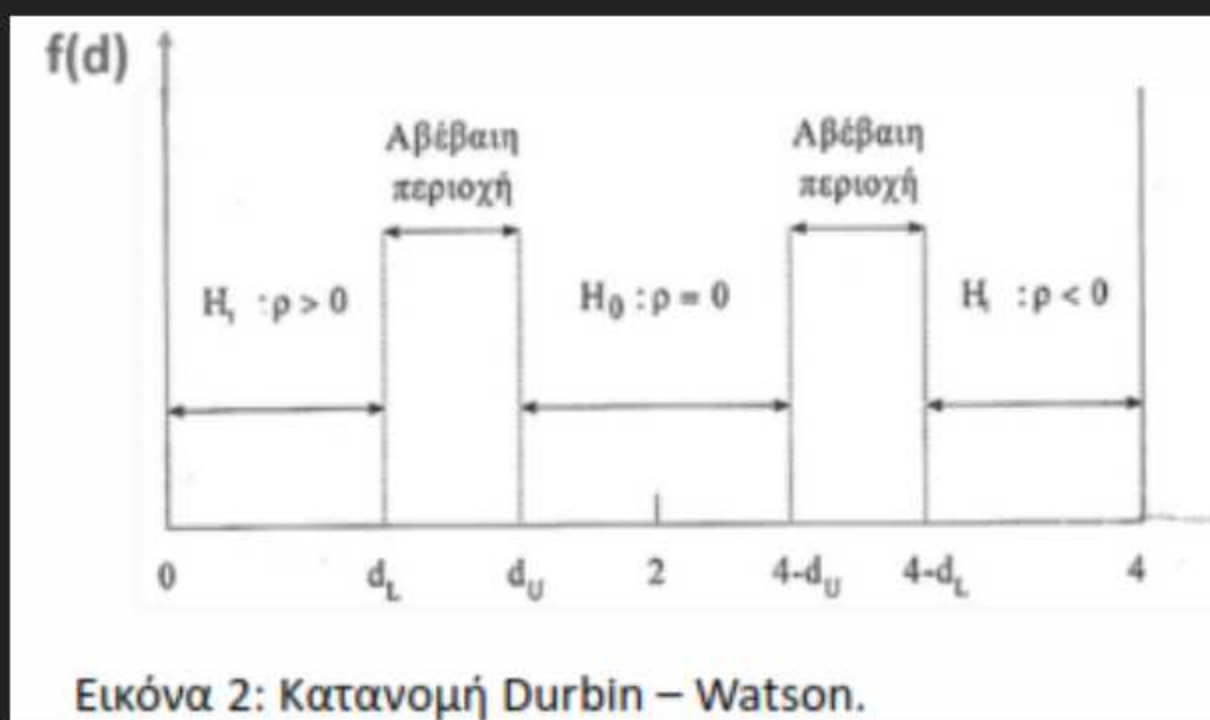
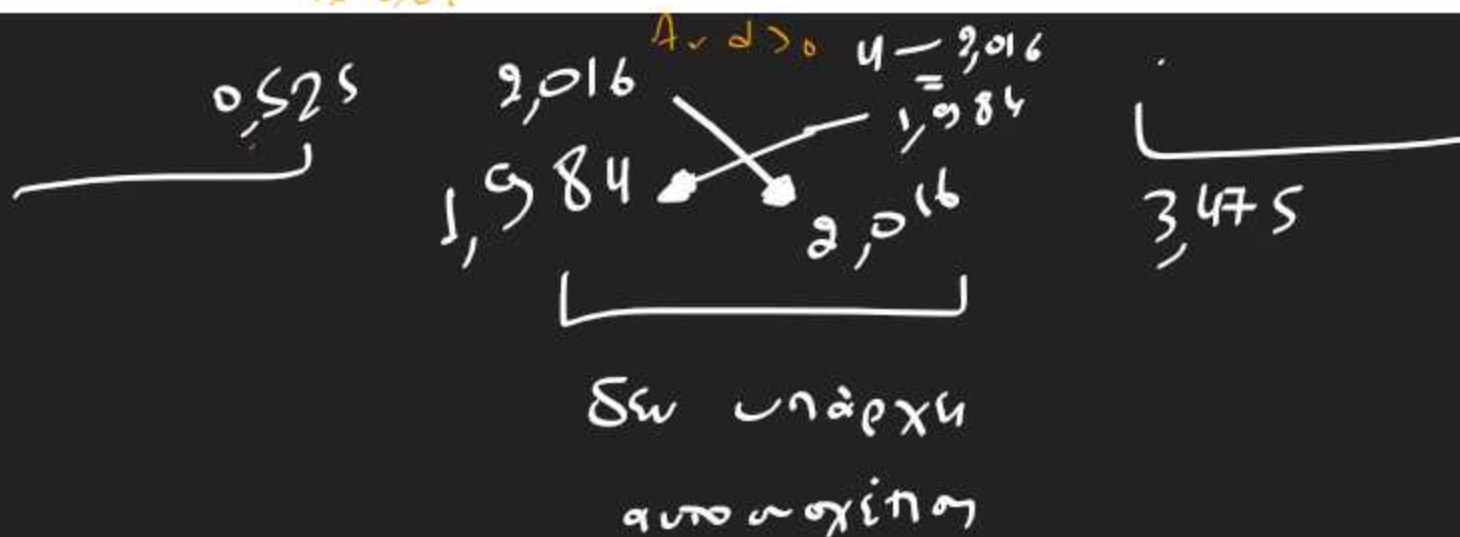
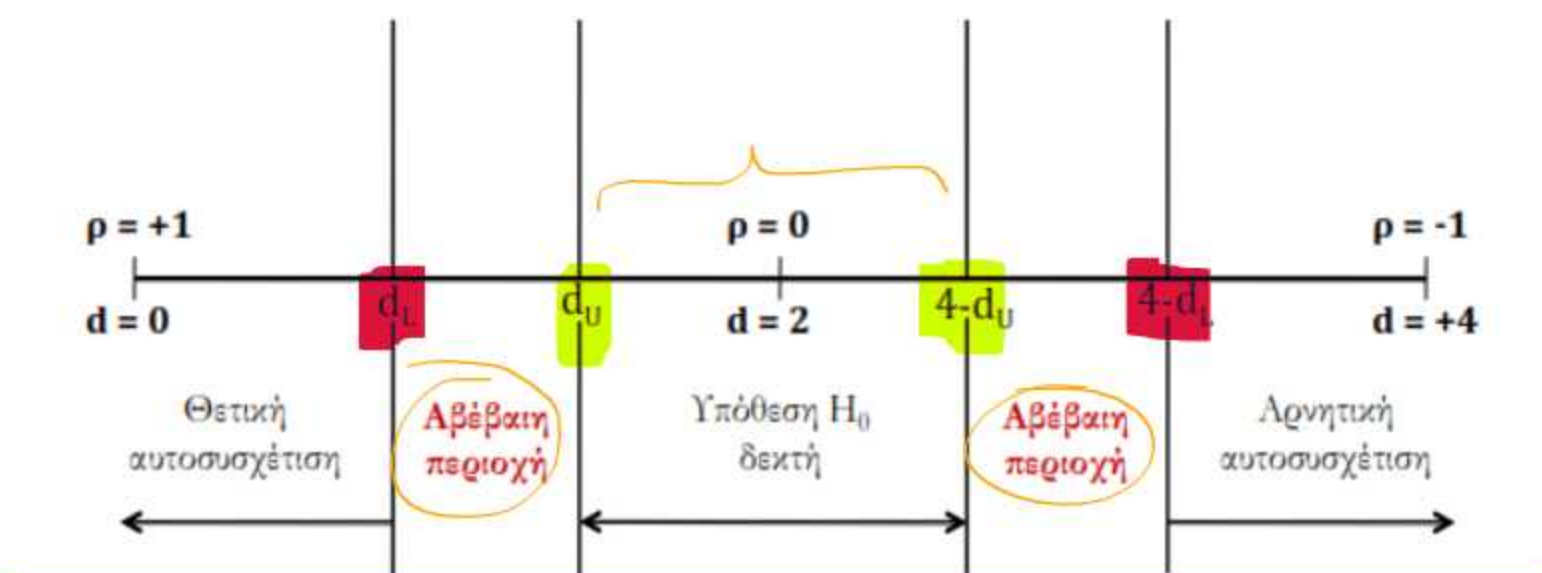
## Κριτική περιοχή του D-W: $d = 2 - 2\rho$

Υποθέσεις:

$H_0: \rho = 0 \rightarrow d = 2 \rightarrow$  δεν υπάρχει αυτοσυσχέτιση

$H_1: \rho \neq 0$

- $\rho > 0, d < 2 \rightarrow$  θετική αυτοσυσχέτιση
- $\rho < 0, d > 2 \rightarrow$  αρνητική αυτοσυσχέτιση





$Y$	$X_1$	$X_2$	$X_3$	$\hat{Y}$	$\varepsilon$
$y_1$	...	...	...	$\hat{y}_1$	$y_1 - \hat{y}_1 = \varepsilon_1$
$y_2$	...	...	...	$\hat{y}_2$	$y_2 - \hat{y}_2 = \varepsilon_2$
$y_3$	...	...	...	$\hat{y}_3$	$y_3 - \hat{y}_3 = \varepsilon_3$

$$y_{\text{pred}} = \hat{y}$$

Προβλεπτική:

(4)

$$d = \frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_n - \hat{y}_n)^2}{\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i$$

π.χ  $Y = 3 + 2,5 X_1 - 3 X_2 + \underline{\varepsilon_i}$