# Το Κλασσικό Γραμμικό Υπόδειγμα Πολλαπλής Παλινδρόμησης

# Γενική μορφή

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_{k-1} X_{k-1,i} + u_i$$

*i=1,2,…,n* ο αριθμός των παρατηρήσεων
*k* ο αριθμός των ανεξάρτητων μεταβλητών

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \ldots & X_{k-1,1} \\ 1 & X_{1,2} & X_{2,2} & \ldots & X_{k-1,2} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & X_{1,n} & X_{2,n} & \ldots & X_{k-1,n} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_{k-1} \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{bmatrix}$$

$$\beta_i = \frac{\partial Y}{\partial X_i} \qquad \text{ή} \qquad \beta_i = \frac{\Delta Y}{\Delta X_i}$$

Το $\beta_i$ αντιπροσωπεύει την μεταβολή στο $Y$ που προέρχεται από μια μεταβολή στο $X_i$ κατά μια μονάδα όταν ***όλες οι άλλες ανεξάρτητες μεταβλητές παραμένουν σταθερές***

# Γενική μορφή

## Παραδείγματα

$$FC = 2,2 + 0,009INC + 0,11AGE + 2,53FSIZE + 0,33EDUC + \hat{u}$$

*FC* = Μηνιαία οικογενειακή κατανάλωση φρούτων σε κιλά
*INC* = Οικογενειακό εισόδημα σε ευρώ
*AGE* = Ηλικία υπεύθυνου
*FSIZE* = Μέγεθος οικογένειας
*EDUC* = Εκπαίδευση υπεύθυνου (έτη σπουδών)

$$HP = 158,2 + 2,01SM + 17,42BDRMS + 9,54BTHRMS + \hat{u}$$

*HP* = Τιμή διαμερίσματος σε χιλ. ευρώ
*SM* = Επιφάνεια σε τμ
*BDRMS* = Υπνοδωμάτια
*BTHRMS* = Μπάνια

# Οι υποθέσεις

Όλες οι βασικές υποθέσεις που αναφέρθηκαν στο υπόδειγμα απλής παλινδρόμησης ισχύουν και στην περίπτωση της πολλαπλής παλινδρόμησης

$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ Το υπόδειγμα είναι γραμμικό στις παραμέτρους

Το σφάλμα $u$ ακολουθεί την κανονική κατανομή με μέσο ίσο με το μηδέν και σταθερή διακύμανση

$$E[\mathbf{u}] = 0$$

$$E[\mathbf{u}\mathbf{u}'] = \sigma^2 \mathbf{I}_{\mathrm{n}} = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ . & . & \cdots & . \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ . & . & \cdots & . \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

Ομοσκεδαστικότητα Δεν υπάρχει αυτοσυσχέτιση

Στην περίπτωση της πολλαπλής παλινδρόμησης προσθέτουμε τις υποθέσεις

Καμία από τις ανεξάρτητες μεταβλητές δεν μπορεί να εκφραστεί σαν γραμμικός μετασχηματισμός μιας ή περισσοτέρων από τις υπόλοιπες. (αποκλείει τελεια πολυσυγγραμικότητα)

Ο αριθμός των ανεξάρτητων μεταβλητών ($k$) θα πρέπει να είναι μικρότερος του αριθμού των παρατηρήσεων ($n$).

# Η μέθοδος των ελαχίστων τετραγώνων

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$ Πληθυσμός

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$ Δείγμα

$$\min \quad \sum \hat{u}_i^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}$$

$$\begin{aligned}
\hat{\mathbf{u}}'\hat{\mathbf{u}} &= \left[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right]' \left[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right] \\
&= \left[\mathbf{Y}' - \hat{\boldsymbol{\beta}}'\mathbf{X}'\right] \left[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right] \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\
&= \mathbf{Y}'\mathbf{Y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}
\end{aligned}$$

$$\frac{\partial\left(\hat{\mathbf{u}}'\hat{\mathbf{u}}\right)}{\partial\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$$ ***Κανονικές εξισώσεις***

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{Y}$$

Αποδεικνύεται ότι και στην περίπτωση της πολλαπλής παλινδρόμησης οι συντελεστές που προκύπτουν με την μέθοδο των ελαχίστων τετραγώνων είναι Άριστοι, Γραμμικοί, Αμερόληπτοι Εκτιμητές των συντελεστών του πληθυσμού.

Έτσι $\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \boldsymbol{\sigma}^2_{\hat{\boldsymbol{\beta}}}\right)$

Όπου $\boldsymbol{\sigma}^2_{\hat{\boldsymbol{\beta}}} = \sigma^2 \left[\mathbf{X}'\mathbf{X}\right]^{-1}$

Όταν το $\sigma^2$ δεν είναι γνωστό χρησιμοποιείται ο εκτιμητής του

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k} = \frac{\sum \hat{u}_i^2}{n-k}$$

Το $n$-$k$ εκφράζει τους βαθμούς ελευθερίας

Στην περίπτωση αυτή $\mathbf{S}^2_{\hat{\boldsymbol{\beta}}} = \hat{\sigma}^2 \left[\mathbf{X}'\mathbf{X}\right]^{-1}$

$$\mathbf{S}^2_{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} Var\left(\hat{\beta}_0\right) & Cov\left(\hat{\beta}_0, \hat{\beta}_1\right) & Cov\left(\hat{\beta}_0, \hat{\beta}_2\right) & . & . & . & Cov\left(\hat{\beta}_0, \hat{\beta}_{k-1}\right) \\ Cov\left(\hat{\beta}_1, \hat{\beta}_0\right) & Var\left(\hat{\beta}_1\right) & Cov\left(\hat{\beta}_2, \hat{\beta}_1\right) & . & . & . & Cov\left(\hat{\beta}_1, \hat{\beta}_{k-1}\right) \\ Cov\left(\hat{\beta}_2, \hat{\beta}_0\right) & Cov\left(\hat{\beta}_2, \hat{\beta}_1\right) & Var\left(\hat{\beta}_2\right) & . & . & . & Cov\left(\hat{\beta}_2, \hat{\beta}_{k-2}\right) \\ . & . & . & . & . & . & . \\ Cov\left(\hat{\beta}_{k-1}, \hat{\beta}_0\right) & Cov\left(\hat{\beta}_{k-1}, \hat{\beta}_1\right) & Cov\left(\hat{\beta}_{k-1}, \hat{\beta}_2\right) & . & . & . & Var\left(\hat{\beta}_{k-1}\right) \end{bmatrix}$$

# Παράδειγμα

| Y | $X_1$ | $X_2$ |
|---|---|---|
| 9 | 2 | 3 |
| 12 | 5 | 7 |
| 14 | 6 | 8 |
| 18 | 8 | 10 |
| 25 | 11 | 16 |
| 29 | 14 | 19 |
| 36 | 17 | 20 |

$$\mathbf{Y} = \begin{bmatrix} 9 \\ 12 \\ 14 \\ 18 \\ 25 \\ 29 \\ 36 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 5 & 7 \\ 1 & 6 & 8 \\ 1 & 8 & 10 \\ 1 & 11 & 16 \\ 1 & 14 & 19 \\ 1 & 17 & 20 \end{bmatrix}$$

$$\mathbf{X'X} = \begin{bmatrix} 7 & 63 & 83 \\ 63 & 735 & 951 \\ 83 & 951 & 1239 \end{bmatrix}$$

$$\mathbf{X'Y} = \begin{bmatrix} 143 \\ 1599 \\ 2074 \end{bmatrix}$$

# Παράδειγμα

$$\Rightarrow \hat{\mathbf{b}} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 7 & 63 & 83 \\ 63 & 735 & 951 \\ 83 & 951 & 1239 \end{bmatrix}^{-1} \begin{bmatrix} 143 \\ 1599 \\ 2074 \end{bmatrix} = \begin{bmatrix} 3,770 \\ 1,930 \\ -0,060 \end{bmatrix}$$

$$\hat{\mathbf{u}} = \begin{bmatrix} 1,55 \\ -1,00 \\ -0,87 \\ -0,61 \\ 0,96 \\ -0,65 \\ 0,62 \end{bmatrix} \Rightarrow \hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k} = \frac{6,26}{7-3} = 1,565$$

# Παράδειγμα

$$\mathbf{S}^{2}_{\hat{\mathbf{b}}} = \hat{\sigma}^{2}\left[\mathbf{X}'\mathbf{X}\right]^{-1} = 1,565\begin{bmatrix} 7 & 63 & 83 \\ 63 & 735 & 951 \\ 83 & 951 & 1239 \end{bmatrix}^{-1}$$

$$= 1,565\begin{bmatrix} 0,746 & 0,104 & -0.13 \\ 0,104 & 0.212 & -0.17 \\ -0.13 & -0.17 & 0,14 \end{bmatrix} = \begin{bmatrix} 1,167 & 0,163 & -0,203 \\ 0,163 & 0,332 & -0,266 \\ -0,203 & -0,266 & 0,219 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} S_{b_0} \\ S_{b_1} \\ S_{b_2} \end{bmatrix} = \begin{bmatrix} \sqrt{1,167} \\ \sqrt{0,332} \\ \sqrt{0,219} \end{bmatrix} \quad \Rightarrow \begin{bmatrix} S_{b_0} \\ S_{b_1} \\ S_{b_2} \end{bmatrix} = \begin{bmatrix} 1,080 \\ 0,576 \\ 0,468 \end{bmatrix}$$

# Ο συντελεστής πολλαπλού προσδιορισμού

Μέτρο του βαθμού προσαρμογής της ευθείας παλινδρόμησης στις παρατηρήσεις του δείγματος. Όπως και στην περίπτωση της απλής παλινδρόμησης συμβολίζεται με $R^2$.

---

$\sum \left(Y_i - \overline{Y}\right)^2$ Συνολικό Άθροισμα Τετραγώνων (Total Sum of Squares) - $TSS$ .

$\sum \left(\hat{Y}_i - \overline{Y}\right)^2$ Ερμηνευόμενο Άθροισμα Τετραγώνων (Regression Sum of Squares) - $RSS$ .

$\sum \left(Y_i - \hat{Y}_i\right)^2 = \sum \hat{u}_i^2$ Το Άθροισμα Τετραγώνων των σφαλμάτων (Error Sum of Squares) - $ESS$ .

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

$$R^2 = \frac{\sum\left(\hat{Y}_i - \overline{Y}\right)^2}{\sum\left(Y_i - \overline{Y}\right)^2} = 1 - \frac{\sum\left(Y_i - \hat{Y}_i\right)^2}{\sum\left(Y_i - \overline{Y}\right)^2}$$

Μετράει την ερμηνευτική ικανότητα της εξίσωσης παλινδρόμησης.

Το σημαντικότερο πρόβλημα του συντελεστή πολλαπλού προσδιορισμού που υπολογίζεται με αυτό τον τρόπο είναι ότι η τιμή του αυξάνει πάντα όταν αυξάνει ο αριθμός των ανεξάρτητων μεταβλητών.

# Ο διορθωμένος συντελεστής πολλαπλού προσδιορισμού

Όταν προστίθεται μια ανεξάρτητη μεταβλητή υπάρχει όφελος αλλά και κόστος.

- Αυξάνει η τιμή του συντελεστή πολλαπλού προσδιορισμού
- Χάνεται ένας βαθμός ελευθερίας (n-k)

Ο διορθωμένος συντελεστής πολλαπλού προσδιορισμού λαμβάνει υπόψη τόσο το κόστος όσο και το όφελος

$$\overline{R}^2 = 1 - \frac{\sum\left(Y_i - \hat{Y}_i\right)^2 \Big/ (n-k)}{\sum\left(Y_i - \overline{Y}\right)^2 \Big/ (n-1)} = 1 - \frac{n-1}{n-k}\left(1 - R^2\right)$$

Ο διορθωμένος συντελεστής είναι μικρότερος από τον $R^2$, θα είναι ίσοι μόνο όταν $R^2=1$ και είναι δυνατό να πάρει ***αρνητικές τιμές***

# Γραμμικό Υπόδειγμα Πολλαπλής παλινδρόμησης - Εκτίμηση

**Παράδειγμα:**

Προσδιοριστικοί παράγοντες της ζήτησης υπηρεσιών αστικών συγκοινωνιών

Στατιστικά στοιχεία από 40 περιοχές μιας χώρας για τις μεταβλητές

BTR = Λειτουργία αστικών συγκοινωνιών σε χιλ. ώρες
FR = Τιμή εισιτηρίου σε €
GP = Τιμή βενζίνης σε €
INC = Κατά κεφαλή μέσο ετήσιο εισόδημα σε €
POP = Πληθυσμός σε χιλ.
DENS = Πυκνότητα πληθυσμού (άτομα/τ. χιλ.)
AREA = Έκταση περιοχής σε τ. χιλ.

# Γραμμικό Υπόδειγμα Πολλαπλής παλινδρόμησης - Εκτίμηση

| A/A | BTR | FR | GP | INC | POP | DENS | AREA |
|---|---|---|---|---|---|---|---|
| 1 | 2073,0 | 0,85 | 0,88 | 17293 | 537,1 | 4099 | 131,0 |
| 2 | 2136,1 | 0,75 | 1,03 | 17768 | 787,0 | 9798 | 80,3 |
| 3 | 1878,8 | 0,60 | 0,91 | 17823 | 587,1 | 12438 | 47,2 |
| 4 | 937,5 | 1,00 | 0,91 | 15163 | 338,0 | 8070 | 41,8 |
| 5 | 7343,3 | 0,50 | 0,97 | 17480 | 3090,0 | 13547 | 228,1 |
| 6 | 837,9 | 0,85 | 0,88 | 15329 | 399,0 | 5110 | 78,1 |
| 7 | 1648,0 | 1,00 | 0,91 | 16141 | 561,8 | 7110 | 79,0 |
| 8 | 739,1 | 0,75 | 0,89 | 15326 | 585,1 | 3234 | 180,9 |
| 9 | 1070,7 | 1,50 | 0,89 | 17115 | 1142,4 | 3431 | 333,0 |
| 10 | 274,6 | 1,50 | 0,89 | 17117 | 486,5 | 2027 | 240,2 |
| 11 | 312,9 | 0,75 | 0,87 | 16127 | 198,7 | 4113 | 48,4 |
| 12 | 1879,1 | 1,00 | 0,94 | 17242 | 549,8 | 4975 | 110,6 |
| 13 | 1941,0 | 0,60 | 0,99 | 17340 | 1253,0 | 8913 | 135,6 |
| 14 | 2317,6 | 1,50 | 0,87 | 15108 | 1603,0 | 2885 | 556,4 |
| 15 | 471,4 | 1,05 | 0,93 | 15809 | 741,2 | 2105 | 352,0 |
| 16 | 594,3 | 0,70 | 0,79 | 16321 | 490,4 | 1551 | 316,3 |
| 17 | 7632,9 | 0,60 | 0,93 | 18027 | 3478,9 | 7486 | 464,7 |
| 18 | 510,1 | 0,60 | 0,93 | 18023 | 423,3 | 8508 | 49,8 |
| 19 | 630,6 | 0,60 | 0,93 | 12349 | 304,0 | 4997 | 60,0 |
| 20 | 1650,9 | 1,00 | 1,03 | 17886 | 377,2 | 10994 | 34,3 |

| A/A | BTR | FR | GP | INC | POP | DENS | AREA |
|---|---|---|---|---|---|---|---|
| 21 | 1618,3 | 0,50 | 0,86 | 16537 | 664,0 | 6702 | 95,8 |
| 22 | 2009,8 | 1,15 | 0,96 | 13019 | 368,0 | 6714 | 55,1 |
| 23 | 1562,4 | 1,15 | 0,96 | 13019 | 265,0 | 5144 | 52,4 |
| 24 | 1139,4 | 0,60 | 0,88 | 13130 | 572,0 | 2832 | 199,4 |
| 25 | 13130,0 | 1,00 | 1,00 | 20513 | 7323,3 | 24288 | 301,5 |
| 26 | 3739,6 | 1,35 | 0,92 | 17409 | 1760,2 | 12944 | 136,0 |
| 27 | 525,7 | 0,75 | 0,91 | 15944 | 991,6 | 3059 | 324,0 |
| 28 | 2385,8 | 1,00 | 0,89 | 15207 | 396,6 | 8147 | 55,4 |
| 29 | 1698,5 | 1,15 | 0,93 | 15409 | 387,0 | 3751 | 103,3 |
| 30 | 544,0 | 1,00 | 0,87 | 17743 | 167,0 | 8309 | 18,9 |
| 31 | 1769,1 | 0,85 | 0,81 | 16309 | 495,9 | 8077 | 61,4 |
| 32 | 1065,0 | 0,50 | 0,85 | 15092 | 794,0 | 2318 | 262,7 |
| 33 | 803,1 | 1,25 | 0,98 | 18014 | 1027,2 | 3208 | 320,0 |
| 34 | 1616,7 | 0,75 | 0,90 | 21886 | 753,6 | 16240 | 46,4 |
| 35 | 146,5 | 0,75 | 0,90 | 20744 | 376,0 | 6988 | 53,9 |
| 36 | 18,1 | 0,75 | 0,90 | 21313 | 698,1 | 4422 | 158,0 |
| 37 | 2056,1 | 1,00 | 0,88 | 17539 | 548,3 | 3790 | 144,6 |
| 38 | 470,1 | 0,75 | 0,92 | 17633 | 295,7 | 3497 | 84,4 |
| 39 | 242,5 | 0,75 | 0,92 | 17643 | 259,8 | 4675 | 55,5 |
| 40 | 3933,5 | 0,60 | 0,96 | 15522 | 693,6 | 11068 | 62,7 |

# Γραμμικό Υπόδειγμα Πολλαπλής παλινδρόμησης - Εκτίμηση

$\hat{\beta}_0 = 2751,3$

$\hat{\beta}_1 = -236,1$

$\hat{\beta}_2 = 513,5$

$\hat{\beta}_3 = -0,195$

$\hat{\beta}_4 = 1,716$

$\hat{\beta}_5 = 0,116$

$\hat{\beta}_6 = -1,174$

$$S_{\hat{\boldsymbol{\beta}}}^{\ 2} = \begin{bmatrix} 6958971{,}0 & 13276{,}7 & -6359031{,}1 & -73{,}5 & 114{,}5 & 18{,}5 & 768{,}1 \\ -13276{,}7 & 203489{,}3 & -175380{,}1 & 1{,}3 & 14{,}5 & -1{,}4 & -200{,}9 \\ -6359031{,}0 & -175380{,}1 & 7046472{,}0 & 13{,}1 & -72{,}4 & -24{,}\ 3 & 540{,}7 \\ -73{,}5 & 1{,}3 & 13{,}1 & {,}004 & 0{,}001 & -0{,}001 & -0{,}016 \\ 114{,}5 & 14{,}5 & -72{,}4 & 0{,}001 & 0{,}053 & -0{,}011 & -0{,}349 \\ 18{,}5 & -1{,}{,}4 & -24{,}3 & 0{,}001 & -0{,}011 & 0{,}003 & 0{,}080 \\ -768{,}1 & -200{,}9 & 540{,}7 & -0{,}016 & -0{,}349 & 0{,}080 & 3{,}2 \end{bmatrix}$$

$$BTR = \underset{(2637,9)}{2751,3} - \underset{(451,0)}{236,1}FR + \underset{(2654,5)}{513,5}GP - \underset{(0,065)}{0,195}INC + \underset{(0,231)}{1,716}POP + \underset{(0,059)}{0,116}DENS - \underset{(1,800)}{1,174}AREA$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k} = \frac{18162484}{40-7} = 550378,3 \quad \hat{\sigma} = 741,9$$

$$R^2 = 1 - \frac{\sum \hat{u}^2}{\sum\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{18162484}{231220782} = 0,921 \quad \bar{R}^2 = 1 - \frac{n-1}{n-k}\left(1 - R^2\right) = 0,907$$

# Έλεγχοι Υποθέσεων

# Στατιστικός έλεγχος συγκεκριμένου συντελεστή

## ***<u>Μονόπλευρος έλεγχος</u>***

$H_0: \beta_i = \gamma \qquad H_1: \beta_i < \gamma$ Διατύπωση της υπόθεσης

$$t = \frac{\hat{\beta}_i - \gamma}{S_{\hat{\beta}_i}}$$ Υπολογισμός της στατιστικής $t$

$t_{\alpha, n-k}$ η τιμή της στατιστικής $t$ από τους πίνακες για επίπεδο σημαντικότητας $\alpha$ και $n$-$k$ βαθμούς ελευθερίας

Η $H_0$ απορρίπτεται αν $|t| > t_{\alpha, n-k}$

## Δίπλευρος έλεγχος

$H_0 : \beta_i = \gamma \qquad H_1 : \beta_i \neq \gamma$ Διατύπωση της υπόθεσης

$$t = \frac{\hat{\beta}_i - \gamma}{S_{\hat{\beta}_i}}$$ Υπολογισμός της στατιστικής $t$

$t_{\alpha/2,n-k}$ η τιμή της στατιστικής $t$ από τους πίνακες για επίπεδο σημαντικότητας $\alpha$ και $n$-$k$ βαθμούς ελευθερίας

Η $H_0$ απορρίπτεται αν $|t| > t_{\alpha/2,n-k}$

## Παράδειγμα:

$$BTR = 2751{,}3 - 236{,}1FR + 513{,}5GP - 0{,}195INC + 1{,}716POP + 0{,}116DENS - 1{,}174AREA$$

$$(2637{,}9) \quad (451{,}0) \quad (2654{,}5) \quad (0{,}065) \quad (0{,}231) \quad (0{,}059) \quad (1{,}800)$$

$$H_0 : \beta_2 = 1000 \qquad H_1 : \beta_2 \neq 1000$$

$$t = \frac{\hat{\beta}_i - \gamma}{S_{\hat{\beta}_i}} = \frac{513{,}5 - 1000}{2654{,}5} = -0{,}183 \qquad t_{\alpha/2, n-k} = t_{0{,}025{,}33} = 2{,}038$$

$\left|-0{,}183\right| < 2{,}038$ Η $H_0$ δεν απορρίπτεται

## Παράδειγμα:

$$BTR = 2751,3 - 236,1FR + 513,5GP - 0,195INC + 1,716POP + 0,116DENS - 1,174AREA$$
$$(2637,9) \quad (451,0) \quad (2654,5) \quad (0,065) \quad (0,231) \quad (0,059) \quad (1,800)$$

$$H_0: \beta_4 = 0 \qquad H_1: \beta_4 > 0$$

$$t = \frac{\hat{\beta}_i - \gamma}{S_{\hat{\beta}_i}} = \frac{1,716 - 0}{0,231} = 7,428 \qquad t_{\alpha,n-k} = t_{0,05,33} = 1,645$$

$7,428 > 1,645$ Η $H_0$ απορρίπτεται

# Στατιστικός έλεγχος δύο ή περισσοτέρων συντελεστών συγχρόνως

Ο έλεγχος αυτός είναι γνωστός σαν ***Wald Test***
Χρησιμοποιείται όταν θέλουμε να ελέξουμε αν ένα ***υποσύνολο*** των ανεξάρτητων μεταβλητών έχει ***στατιστικά σημαντική επίδραση*** στην ερμηνευόμενη μεταβλητή.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{k-1} X_{k-1i} + u_i$$

Το αρχικό υπόδειγμα στο οποίο δεν υπάρχει κανένας περιορισμός

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{m-1} X_{m-1i} + v_i$$

Το υπόδειγμα αυτό περιέχει $m$ από τους $k$ συντελεστές του προηγούμενου υποδείγματος. Στο υπόδειγμα αυτό υπάρχει ο περιορισμός ότι $k$-$m$ συντελεστές είναι ίσοι με το μηδέν έτσι $H_0: \beta_m = \beta_{m+1} = \beta_{m+2} =, \ldots, = \beta_{k-1} = 0$

$H_1$: Ένας τουλάχιστον από τους συντελεστές είναι διάφορος του μηδενός

Κριτήριο $\longrightarrow$ ***ESS***

$ESS_U$ $\sum \hat{u}_i^2$ Υπόδειγμα χωρίς περιορισμούς

$ESS_R$ $\sum \hat{v}_i^2$ Υπόδειγμα με περιορισμούς

Αν η διαφορά $ESS_R$ - $ESS_U$ είναι μικρή τότε θεωρούμε ότι οι ***k-m*** συντελεστές δεν συμβάλλουν στην ερμηνεία της $Y$ και άρα μπορούν να μην συμπεριληφθούν στο υπόδειγμα. Δηλαδή, η υπόθεση $H_0$ γίνεται δεκτή.

*πότε όμως η διαφορά αυτή είναι μικρή?*

$$F=\frac{\dfrac{ESS_R-ESS_U}{d.f.(R)-d.f.(U)}}{\dfrac{ESS_U}{d.f.(U)}}$$

***d.f.(U)*** οι βαθμοί ελευθερίας στο υπόδειγμα χωρίς περιορισμούς

***d.f.(R)*** οι βαθμοί ελευθερίας στο υπόδειγμα με περιορισμούς

$$F=\frac{\dfrac{ESS_R-ESS_U}{(n-m)-(n-k)}}{\dfrac{ESS_U}{n-k}}=\frac{\dfrac{ESS_R-ESS_U}{k-m}}{\dfrac{ESS_U}{n-k}}$$

$F_{\alpha,k-m,n-k}$ Η τιμή της στατιστικής $F$ από τους πίνακες, σε επίπεδο σημαντικότητας $\alpha$ και βαθμούς ελευθερίας $k$-$m$ και $n$-$k$.

Η $H_0$ απορρίπτεται αν $F>F_{\alpha,k-m,n-k}$

## Παράδειγμα:

$$BTR = 2751,3 - 236,1FR + 513,5GP - 0,195INC + 1,716POP + 0,116DENS - 1,174AREA$$

$$(2637,9) \quad (451,0) \quad (2654,5) \quad (0,065) \qquad (0,231) \qquad (0,059) \qquad (1,800)$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k} = \frac{18162484}{40-7} = 550378,3$$

---

$H_0: \ \beta_1 = \beta_2 = 0 \qquad H_1: \ \beta_1 \neq 0$ ή/και $\beta_2 \neq 0$

$$BTR = 3022,3 - 0,194INC + 1,735POP + 0,115DENS - 1,423AREA$$

$$(1010,0) \quad (0,063) \qquad (0,222) \qquad (0,057) \qquad (1,695)$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k} = \frac{18320949}{40-5} = 523455,7$$

$$F=\frac{\dfrac{ESS_R-ESS_U}{k-m}}{\dfrac{ESS_U}{n-k}}=\frac{\dfrac{18320949-18162484}{7-5}}{\dfrac{18162484}{40-7}}=0,144$$

$$F_{\alpha,k-m,n-k}=F_{0,05,2,33}=3,29 \qquad 0,144<3,29$$

Η $H_0$ δεν απορρίπτεται

---

$$H_0: \beta_2=\beta_4=\beta_5=0 \quad H_1: \beta_2\neq 0 \text{ ή/και } \beta_4\neq 0 \text{ ή/και } \beta_5\neq 0$$

$$\underset{(3313,9)}{BTR=-2343,1}-\underset{(1387,3)}{1033,9FR}+\underset{(0,178)}{0,252INC}+\underset{(2,974)}{6,238AREA}$$

$$\hat{\sigma}^2=\frac{\sum\hat{u}_i^2}{n-k}=\frac{195066760}{40-4}=5418521,1$$

$$F = \frac{\dfrac{ESS_R - ESS_U}{k-m}}{\dfrac{ESS_U}{n-k}} = \frac{\dfrac{195066760 - 18162484}{7-4}}{\dfrac{18162484}{40-7}} = 107,1$$

$$F_{\alpha, k-m, n-k} = F_{0,05,3,33} = 2,89$$

$$107,1 > 2,89$$

Η $H_0$ απορρίπτεται

# Έλεγχος της ερμηνευτικής ικανότητας του υποδείγματος συνολικά

$$H_0: \quad Y_i = \beta_0 + v_i$$

$$H_1: \quad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{k-1} X_{k-1i} + u_i$$

Αποδεικνύεται ότι στην περίπτωση αυτή η στατιστική $F$ έχει άμεση σχέση με τον συντελεστή προσδιορισμού $R^2$.

$$F = \frac{R^2/(k-1)}{\left(1-R^2\right)/(n-k)}$$

Η $H_0$ απορρίπτεται αν $F > F_{\alpha,k-1,n-k}$

## Παράδειγμα:

$$BTR = 2751,3 - 236,1FR + 513,5GP - 0,195INC + 1,716POP + 0,116DENS - 1,174AREA$$
$$(2637,9) \quad (451,0) \quad (2654,5) \quad (0,065) \quad (0,231) \quad (0,059) \quad (1,800)$$

$$R^2 = 0,921$$

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0,921/(7-1)}{(1-0,921)/(40-7)} = 64,94$$

$$F_{\alpha,k-m,n-k} = F_{0,05,6,33} = 2,39 \qquad 64,94 > 2,39$$

Η $H_0$ απορρίπτεται

# Στατιστικός έλεγχος γραμμικού συνδυασμού συντελεστών

***Περίπτωση 1η:*** Γραμμικός συνδυασμός συντελεστών = 0

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_3 + \beta_4 X_4 + u_i$$

$$H_0 : \beta_2 = \beta_3$$

***1ος τρόπος:***

Αν ισχύει η $H_0$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 \left( X_{2i} + X_3 \right) + \beta_4 X_4 + u_i$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 Z_i + \beta_4 X_4 + u_i$$

Όπου $Z_i = X_{2i} + X_{3i}$

Το υπόδειγμα αυτό ***(με περιορισμό)*** συγκρίνεται με το αρχικό ***(χωρίς περιορισμό)*** με βάση το κριτήριο $F$

**2ος τρόπος:**

Ορίζουμε νέο συντελεστή $d$ έτσι ώστε όταν ισχύει η $H_0$ το $d=0$

Δηλαδή, $d = \beta_2 - \beta_3$

Με βάση τον ορισμό αυτό

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + (\beta_2 - d) X_{3i} + \beta_4 X_4 + u_i$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 (X_{2i} + X_{3i}) - dX_{3i} + \beta_4 X_4 + u_i$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 Z_i - dX_{3i} + \beta_4 X_4 + u_i$$

Όπου $Z_i = X_{2i} + X_{3i}$

Εκτιμάται η τελευταία συνάρτηση και ελέγχεται η στατιστική σημαντικότητα του $d$ με βάση το κριτήριο $t$. Ελέγχεται δηλαδή η υπόθεση $d=0$. Αν αυτή γίνει αποδεκτή τότε $\beta_2=\beta_3$.

## ***Περίπτωση 2η*: Γραμμικός συνδυασμός συντελεστών = σταθερά**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_3 + \beta_4 X_4 + u_i$$

$$H_0 : \beta_1 + \beta_2 + \beta_3 = 1$$

***1ος τρόπος:***

Αν ισχύει η $H_0$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + (1 - \beta_1 - \beta_2) X_{3i} + \beta_4 X_4 + u_i$$

$$Y_i - X_{3i} = \beta_0 + \beta_1 (X_{1i} - X_{3i}) + \beta_2 (X_{2i} - X_{3i}) + \beta_4 X_4 + u_i$$

$$W_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_4 X_4 + u_i$$ ***(Υπόδειγμα με περιορισμό)***

Το υπόδειγμα αυτό (με περιορισμό) συγκρίνεται με το αρχικό (χωρίς περιορισμό) με βάση το κριτήριο $F$

**2ος τρόπος:**

Ορίζουμε νέο συντελεστή $d$ έτσι ώστε όταν ισχύει η $H_0$ το $d=0$

Δηλαδή, $d = 1 - \beta_1 - \beta_2 - \beta_3$

Με βάση τον ορισμό αυτό

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + (1 - \beta_1 - \beta_2 - d)X_{3i} + \beta_4 X_4 + u_i$$

$$Y_i - X_{3i} = \beta_0 + \beta_1 (X_{1i} - X_{3i}) + \beta_2 (X_{2i} - X_{3i}) - dX_{3i} + \beta_4 X_4 + u_i$$

$$W_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} - dX_{3i} + \beta_4 X_4 + u_i$$

Εκτιμάται η τελευταία συνάρτηση και ελέγχεται η στατιστική σημαντικότητα του $d$ με βάση το κριτήριο $t$. Ελέγχεται δηλαδή η υπόθεση $d=0$. Αν αυτή γίνει αποδεκτή τότε $1 - \beta_1 - \beta_2 - \beta_3 = 0$ .

## Παράδειγμα:

$$BTR = 2751,3 - 236,1FR + 513,5GP - 0,195INC + 1,716POP + 0,116DENS - 1,174AREA$$

$$(2637,9) \quad (451,0) \quad (2654,5) \quad (0,065) \quad (0,231) \quad (0,059) \quad (1,800)$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k} = \frac{18162484}{40-7} = 550378,3$$

Αρχικό υπόδειγμα (χωρίς περιορισμούς)

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_3 INC + \beta_4 POP + \beta_5 DENS + \beta_6 AREA$$

## Υπόθεση 1 – 1ος τρόπος

$$H_0: \ \beta_3 + \beta_4 = 0 \quad H_1: \ \beta_3 + \beta_4 \neq 0$$

Αν ισχύει η $H_0$ $\qquad \beta_3 = -\beta_4$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP - \beta_4 INC + \beta_4 POP + \beta_5 DENS + \beta_6 AREA$$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_4 \left(POP - INC\right) + \beta_5 DENS + \beta_6 AREA$$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_4 Z + \beta_5 DENS + \beta_6 AREA$$

Εκτίμηση της παραπάνω συνάρτησης

$$B\hat{T}R = 1711{,}4 - 635{,}6FR + 2017{,}2GP + 0{,}332Z + 0{,}434DENS + 8{,}099AREA$$

$$ESS_R = \sum \hat{u}_R^2 = 39392398$$

$$F = \frac{\dfrac{ESS_R - ESS_U}{k-m}}{\dfrac{ESS_U}{n-k}} = \frac{\dfrac{39392398 - 18162484}{7-6}}{\dfrac{18162484}{40-7}} = 38,57$$

$$F_{\alpha, k-m, n-k} = F_{0,05,1,33} = 4,14$$

$$38,57 > 4,14$$

Η $H_0$ απορρίπτεται

## Υπόθεση 1 – 2ος τρόπος:

$$H_0: \ \beta_3 + \beta_4 = 0 \quad H_1: \ \beta_3 + \beta_4 \neq 0$$

Ορίζουμε $\beta_3 + \beta_4 = d \ \Rightarrow \ \beta_3 = d - \beta_4$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + \left(d - \beta_4\right) INC + \beta_4 POP + \beta_5 DENS + \beta_6 AREA$$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + dINC + \beta_4 \left(POP - INC\right) + \beta_5 DENS + \beta_6 AREA$$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + dINC + \beta_4 Z + \beta_5 DENS + \beta_6 AREA$$

Εκτίμηση της παραπάνω συνάρτησης

$$\underset{}{B\hat{T}R} = \underset{(2637,9)}{2751,3} - \underset{(451,1)}{236,1} FR + \underset{(2654,5)}{513,4} GP + \underset{(0,245)}{1,521} INC + \underset{(1,716)}{1,715} Z + \underset{(0,059)}{0,116} DENS - \underset{(1,800)}{1,174} AREA$$

$$t = \frac{\hat{d} - 0}{S_{\hat{d}}} = \frac{1,521 - 0}{0,245} = 6,2$$

$$t_{\alpha/2, n-k} = t_{0,025,33} = 2,03$$

$$6,2 > 2,03$$

Η $H_0$ απορρίπτεται

## Υπόθεση 2 – 1ος τρόπος:

$$H_0: \ \beta_1 + \beta_2 + \beta_5 = 100 \qquad H_1: \ \beta_1 + \beta_2 + \beta_5 \neq 100$$

Αν ισχύει η $H_0$ $\quad \beta_1 = 100 - \beta_2 - \beta_5$

$$BTR = \beta_0 + (100 - \beta_2 - \beta_5)FR + \beta_2 GP + \beta_3 INC + \beta_4 POP + \beta_5 DENS + \beta_6 AREA$$

$$BTR = \beta_0 + 100FR + \beta_2(GP - FR) + \beta_3 INC + \beta_4 POP + \beta_5(DENS - FR) + \beta_6 AREA$$

$$BTR - 100FR = \beta_0 + \beta_2(GP - FR) + \beta_3 INC + \beta_4 POP + \beta_5(DENS - FR) + \beta_6 AREA$$

$$W = \beta_0 + \beta_2 Z_1 + \beta_3 INC + \beta_4 POP + \beta_5 Z_2 + \beta_6 AREA$$

Εκτίμηση της παραπάνω συνάρτησης

$$\hat{W} = 2915{,}2 + 336{,}7Z_1 - 0{,}195INC + 1{,}717POP + 0{,}116Z_2 - 1{,}182AREA$$

$$ESS_R = \sum \hat{u}_R^2 = 18164997$$

$$F = \frac{\dfrac{ESS_R - ESS_U}{k-m}}{\dfrac{ESS_U}{n-k}} = \frac{\dfrac{18164997 - 18162484}{7-6}}{\dfrac{18162484}{40-7}} = 0,004$$

$$F_{\alpha,k-m,n-k} = F_{0,05,1,33} = 4,14$$

$$0,004 < 4,14$$

Η $H_0$ δεν απορρίπτεται

## Υπόθεση 2 – 2ος τρόπος:

$$H_0: \ \beta_1 + \beta_2 + \beta_5 = 100 \qquad H_1: \ \beta_1 + \beta_2 + \beta_5 \neq 100$$

Ορίζουμε $d = 100 - \beta_1 - \beta_2 - \beta_5 \ \Rightarrow \ \beta_1 = 100 - d - \beta_2 - \beta_5$

$$BTR = \beta_0 + \left(100 - d - \beta_2 - \beta_5\right) FR + \beta_2 GP + \beta_3 INC + \beta_4 POP + \beta_5 DENS + \beta_6 AREA$$

$$BTR = \beta_0 + 100FR - dFR + \beta_2 \left(GP - FR\right) + \beta_3 INC + \beta_4 POP + \beta_5 \left(DENS - FR\right) + \beta_6 AREA$$

$$BTR - 100FR = \beta_0 - dFR + \beta_2 \left(GP - FR\right) + \beta_3 INC + \beta_4 POP + \beta_5 \left(DENS - FR\right) + \beta_6 AREA$$

$$W = \beta_0 - dFR + \beta_2 Z_1 + \beta_3 INC + \beta_4 POP + \beta_5 Z_2 + \beta_6 AREA$$

Εκτίμηση της παραπάνω συνάρτησης

$$\hat{W} = 2751{,}3 + 177{,}4FR + 513{,}5Z_1 - 0{,}195INC + 1{,}716POP + 0{,}116Z_2 - 1{,}174AREA$$
$$(2637{,}9) \quad (2626{,}6) \quad (2654{,}5) \quad (0{,}065) \quad (0{,}231) \quad (0{,}059) \quad (1{,}800)$$

$$t = \frac{\hat{d} - 0}{S_{\hat{d}}} = \frac{177,4 - 0}{2626,6} = 0,067$$

$$t_{\alpha/2,n-k} = t_{0,025,33} = 2,03$$

$$0,067 < 2,03$$

Η $H_0$ δεν απορρίπτεται

## Υπόθεση 3 – 1ος τρόπος:

$$H_0: \ 2\beta_4 + 3\beta_6 = 5 \qquad H_0: \ 2\beta_4 + 3\beta_6 \neq 5$$

Αν ισχύει η $H_0$ $\beta_4 = 2,5 - 1,5\beta_6$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_3 INC + \left(2,5 - 1,5\beta_6\right) POP + \beta_5 DENS + \beta_6 AREA$$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_3 INC + 2,5POP + \beta_5 DENS + \beta_6 \left(AREA - 1,5POP\right)$$

$$BTR - 2,5POP = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_3 INC + \beta_5 DENS + \beta_6 \left(AREA - 1,5POP\right)$$

$$W = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_3 INC + \beta_5 DENS + \beta_6 Z$$

Εκτίμηση της παραπάνω συνάρτησης

$$\hat{W} = 2321,8 - 352,1FR + 812,1GP - 0,204INC + 0,160DEN + 0,649Z$$

$$ESS_R = \sum \hat{u}_R^2 = 18728834$$

$$F=\frac{\dfrac{ESS_R-ESS_U}{k-m}}{\dfrac{ESS_U}{n-k}}=\frac{\dfrac{18728834-18162484}{7-6}}{\dfrac{18162484}{40-7}}=1,029$$

$$F_{\alpha,k-m,n-k}=F_{0,05,1,33}=4,14$$

$$1,029<4,14$$

Η $H_0$ δεν απορρίπτεται

## Υπόθεση 3 – 2ος τρόπος:

$$H_0: \; 2\beta_4 + 3\beta_6 = 5 \qquad H_0: \; 2\beta_4 + 3\beta_6 \neq 5$$

Ορίζουμε $\quad d = 5 - 2\beta_4 - 3\beta_6 \;\Rightarrow\; \beta_4 = 2,5 - d - 1,5\beta_6$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_3 INC + \left(2,5 - d - 1,5\beta_6\right) POP + \beta_5 DENS + \beta_6 AREA$$

$$BTR = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_3 INC + 2,5POP - dPOP + \beta_5 DENS + \beta_6 \left(AREA - 1,5POP\right)$$

$$BTR - 2,5POP = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_3 INC - dPOP + \beta_5 DENS + \beta_6 \left(AREA - 1,5POP\right)$$

$$W = \beta_0 + \beta_1 FR + \beta_2 GP + \beta_3 INC - dPOP + \beta_5 DENS + \beta_6 Z$$

Εκτίμηση της παραπάνω συνάρτησης

$$\hat{W} = 2751,3 - 236,1FR + 513,4GP - 0,194INC - 2,545POP + 0,116DEN - 1,174Z$$

$$(2637,9) \quad (451,1) \quad (2654,5) \quad (0,065) \qquad (2,509) \qquad (0,059) \qquad (1,800)$$

# Αλλαγή στις μονάδες μέτρησης

Έστω ότι $Y^*, X_1^*, X_2^*$ είναι οι μεταβλητές με αλλαγμένες μονάδες μέτρησης και ότι: $Y^* = \lambda_y Y$

$$X_j^* = \lambda_j \mathrm{X}_j \quad \text{για} \quad j = 1,2$$

Αποδεικνύεται πως: $\hat{\beta}_0^* = \lambda_y \hat{\beta}_0, \quad \hat{\beta}_1^* = \dfrac{\lambda_y}{\lambda_1}\hat{\beta}_1, \quad \hat{\beta}_2^* = \dfrac{\lambda_y}{\lambda_2}\hat{\beta}_2$

Επομένως οι τιμές των εκτιμητών θα αλλάξουν εκτός και αν: $\lambda_y = \lambda_1 = \lambda_2$. Σε αυτήν την περίπτωση θα αλλάξει μόνο το $\hat{\beta}_0$.

Ομοίως και για τα τυπικά σφάλματα:

$$s_{\hat{\beta}_0^*} = \lambda_y s_{\hat{\beta}_0}, \quad \mathrm{s}_{\hat{\beta}_1^*} = \frac{\lambda_y}{\lambda_1} s_{\hat{\beta}_1}, \quad \mathrm{s}_{\hat{\beta}_2^*} = \frac{\lambda_y}{\lambda_2} s_{\hat{\beta}_2}$$

Όμως οι τιμές της στατιστικής $t$ για τον έλεγχο της υπόθεσης $\beta_j = 0$ δεν επηρεάζεται:

$$t_j^* = \frac{\beta_j^*}{s_{\beta_j^*}} = \frac{\left(\lambda_y / \lambda_j\right)\beta_j}{\left(\lambda_y / \lambda_j\right) s_{\beta_j}} = \frac{\beta_j}{s_{\beta_j}} = t_j$$

# Κριτήρια επιλογής παλινδρομήσεων

Εκτός από το διορθωμένο συντελεστή προσδιορισμού υπάρχουν και άλλα δύο κριτήρια που χρησιμοποιούνται:

Κριτήριο πληροφοριών Akaike $\mathrm{AIC} = n \cdot \ln \sum \hat{u}^2 + 2k$

Μπεϋσιανό κριτήριο Schwarz $\mathrm{SBC} = n \cdot \ln \sum \hat{u}^2 + k \cdot \ln n$

Επιλέγουμε το υπόδειγμα με την μικρότερη τιμή για τα κριτήρια αυτά.

# Άσκηση 1

Έστω ότι η προσφερόμενη ποσότητα ($Y$) ενός αγαθού, είναι γραμμική συνάρτηση της τιμής του ($X_1$) και τους ύψους των ημερομισθίων ($X_2$) τα οποία είναι απαραίτητα για την παραγωγή του. Δηλαδή, η συνάρτηση προσφοράς είναι: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$

| Y | 20 | 35 | 30 | 47 | 60 | 68 | 76 | 90 | 100 | 105 | 130 | 140 | 125 | 120 | 135 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 10 | 15 | 21 | 26 | 40 | 37 | 42 | 33 | 30 | 38 | 60 | 65 | 50 | 35 | 42 |
| $X_2$ | 12 | 10 | 9 | 8 | 5 | 7 | 4 | 5 | 7 | 5 | 3 | 4 | 3 | 1 | 2 |

Από την εκτίμηση της εξίσωσης βρέθηκε:

$$\hat{Y} = 89,534 + 1,053 X_1 - 7,470 X_2$$

**Α)** Να ελεγχθεί αν η συνδυασμένη επίδραση των μεταβλητών $X_1$, $X_2$ πάνω στην $Y$ είναι σημαντική. Δίνεται $R^2 = 0,827$.

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_1 : \beta_1 \neq 0 \ \acute{\eta}/\kappa\alpha\iota \ \beta_2 \neq 0$$

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0,827/(3-1)}{(1-0,827)/(15-3)} = 28,65$$

$$F = 28,65 > F_{(0.05),2,12} = 3,89$$

# Άσκηση 1

**Β)** Να υπολογισθεί η ελαστικότητα προσφοράς ως προς την τιμή στο σημείο των μέσων και να επεξηγηθεί.

$$\varepsilon_{\mathrm{YX}_1} = \frac{\partial \hat{Y}}{\partial X_1}\frac{\overline{X}_1}{\overline{Y}} = \hat{\beta}_1 \frac{\overline{X}_1}{\overline{Y}} = 1,053\frac{32,267}{85,4} = 0,398$$

Αυτό σημαίνει ότι αν η τιμή του αγαθού αυξηθεί κατά 1%, **δεδομένου ότι τα ημερομίσθια παραμένουν σταθερά**, τότε η προσφορά του αγαθού θα αυξηθεί κατά 0,398%.

**Γ)** Να υπολογιστεί ο διορθωμένος συντελεστής προσδιορισμού.

$$\overline{R}^2 = 1 - \frac{n-1}{n-k}\left(1-R^2\right) = 1 - \frac{15-1}{15-3}\left(1-0,827\right) = 1 - \frac{14}{12}0,173 = 0,798$$

# Άσκηση 2

Έστω ότι από ένα δείγμα 10 παρατηρήσεων εκτιμήθηκε η συνάρτηση:

$$\hat{Y} = 8.79 + 3.87X_1 - 3.45X_2 \qquad R^2 = 0.957$$

$$(2.57) \quad (1.22) \quad (0.56) \qquad ESS = 4.006$$

**Α)** Να γίνει στατιστικός έλεγχος του υποδείγματος ($\alpha$=0.05).

$$H_0: \beta_0 = 0 \qquad H_1: \beta_0 \neq 0$$

$$t = \frac{\hat{\beta}_0 - 0}{S_{\hat{\beta}_0}} = \frac{8.79}{2.57} = 3.42 \qquad t = 3.42 > t_{(0.025),7} = 2.365$$

$$H_0: \beta_1 = 0 \qquad H_1: \beta_1 \neq 0$$

$$t = \frac{\hat{\beta}_1 - 0}{S_{\hat{\beta}_1}} = \frac{3.87}{1.22} = 3.17 \qquad t = 3.17 > t_{(0.025),7} = 2.365$$

$$H_0: \beta_2 = 0 \qquad H_1: \beta_2 \neq 0$$

$$t = \frac{\hat{\beta}_2 - 0}{S_{\hat{\beta}_2}} = \frac{-3.45}{0.56} = -6.20 \qquad |t| = 6.20 > t_{(0.025),7} = 2.365$$

$$H_0: \beta_1 = \beta_2 = 0 \qquad H_1: \beta_1 \neq 0 \text{ ή/και } \beta_2 \neq 0$$

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0,957/(3-1)}{(1-0,957)/(10-3)} = 77.89$$

$$F = 77.89 > F_{(0.05),2,7} = 4.74$$

# Άσκηση 2

**B)** Να ελεγχθεί αν $\beta_1+\beta_2=2$ $(\alpha=0.05)$ με το κριτήριο t και F.

$$H_0 : \beta_1 + \beta_2 = 2$$
$$H_1 : \beta_1 + \beta_2 \neq 2$$

Ορίζουμε $d = 2 - \beta_1 - \beta_2 \Rightarrow \beta_1 = 2 - d - \beta_2$

$$Y = \beta_0 + \left(2 - d - \beta_2\right) X_1 + \beta_2 X_2$$

$$Y = \beta_0 + 2X_1 - dX_1 + \beta_2\left(X_2 - X_1\right)$$

$$Y - 2X_1 = \beta_0 - dX_1 + \beta_2\left(X_2 - X_1\right)$$

$$W = \beta_0 - dX_1 + \beta_2 Z$$

Εκτίμηση της παραπάνω συνάρτησης

$$\hat{W} = 8.79 - 1.58X_1 - 3.45Z$$
$$(2.57) \quad (1.66) \quad (0.56)$$

$$t = \frac{\hat{d} - 0}{S_{\hat{d}}} = \frac{-1.58}{1.66} = -0.95$$

$$|t| = 0.95 < t_{\alpha/2, n-k} = t_{0,025,7} = 2,365$$

Η $H_0$ δεν απορρίπτεται

# Άσκηση 2

**B)** Να ελεγχθεί αν $\beta_1+\beta_2=2$ ($\alpha=0.05$) με το κριτήριο t και F.

$$H_0 : \beta_1 + \beta_2 = 2$$
$$H_1 : \beta_1 + \beta_2 \neq 2$$

$$\beta_1 = 2 - \beta_2$$

$$Y = \beta_0 + \left(2 - \beta_2\right)X_1 + \beta_2 X_2$$

$$Y - 2X_1 = \beta_0 + \beta_2\left(X_2 - X_1\right)$$

$$W = \beta_0 + \beta_2 Z + u$$

Εκτίμηση της παραπάνω συνάρτησης $\hat{W} = 6.71 - 3.00Z$ $\quad R^2 = 0.93$

$(1.32) \quad (0.28) \quad ESS = 4.52$

$$F = \frac{\dfrac{ESS_R - ESS_U}{k-m}}{\dfrac{ESS_U}{n-k}} = \frac{\dfrac{4.52 - 4.006}{3-2}}{\dfrac{4.006}{10-3}} = 0.90 \quad < F_{(0,05),1,7} = 5,59$$

Η $H_0$ δεν απορρίπτεται

# Άσκηση 2

Γ) Να ελεγχθεί αν $\beta_1+\beta_2=5$ ($\alpha=0.05$) με το κριτήριο t και F.

$$H_0 : \beta_1 + \beta_2 = 5$$
$$H_1 : \beta_1 + \beta_2 \neq 5$$

Ορίζουμε $d = 5 - \beta_1 - \beta_2 \Rightarrow \beta_1 = 5 - d - \beta_2$

$$Y = \beta_0 + \left(5 - d - \beta_2\right) X_1 + \beta_2 X_2$$

$$Y = \beta_0 + 5X_1 - dX_1 + \beta_2\left(X_2 - X_1\right)$$

$$Y - 5X_1 = \beta_0 - dX_1 + \beta_2\left(X_2 - X_1\right)$$

$$W = \beta_0 - dX_1 + \beta_2 Z$$

Εκτίμηση της παραπάνω συνάρτησης

$$\hat{W} = 8.79 - 4.58X_1 - 3.45Z$$
$$(2.57) \quad (1.67) \quad (0.56)$$

$$t = \frac{\hat{d} - 0}{S_{\hat{d}}} = \frac{-4.58}{1.67} = -2.75$$

$$\left|t\right| = 2.75 > t_{\alpha/2,n-k} = t_{0,025,7} = 2,365$$

Η $H_0$ απορρίπτεται

# Άσκηση 2

Γ) Να ελεγχθεί αν $\beta_1+\beta_2=5$ ($\alpha=0.05$) με το κριτήριο t και F.

$$H_0: \beta_1+\beta_2=5$$
$$H_1: \beta_1+\beta_2 \neq 5$$

$$\beta_1 = 5-\beta_2$$

$$Y = \beta_0 + \left(5-\beta_2\right)X_1 + \beta_2 X_2$$

$$Y - 5X_1 = \beta_0 + \beta_2\left(X_2 - X_1\right)$$

$$W = \beta_0 + \beta_2 Z + u$$

Εκτίμηση της παραπάνω συνάρτησης $\hat{W} = 2.75 - 2.14Z$ $\quad R^2 = 0.79$

$\quad\quad (1.80) \quad (0.39) \quad\quad ESS = 8.33$

$$F = \frac{\dfrac{ESS_R - ESS_U}{k-m}}{\dfrac{ESS_U}{n-k}} = \frac{\dfrac{8.33-4.006}{3-2}}{\dfrac{4.006}{10-3}} = 7.56 \quad > F_{(0,05),1,7} = 5,59$$

Η $H_0$ απορρίπτεται

# Άσκηση 3

Έστω ότι η ζητούμενη ποσότητα ($Y$) ενός αγαθού, είναι γραμμική συνάρτηση της τιμής του ($X_1$) και της τιμής ενός υποκατάστατου αγαθού ($X_2$). Δηλαδή, η συνάρτηση ζήτησης είναι:

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

| Y | 100 | 82 | 86 | 90 | 60 | 70 |
|---|---|---|---|---|---|---|
| $X_1$ | 13.0 | 13.5 | 13.2 | 13.0 | 14.0 | 13.7 |
| $X_2$ | 6.5 | 6.2 | 6.3 | 6.4 | 6.4 | 6.5 |

Από την εκτίμηση της εξίσωσης βρέθηκε:

$$\hat{Y} = \underset{(143.21)}{558.87} - \underset{(5.43)}{34.183} X_1 - \underset{(18.84)}{3.051} X_2 \qquad R^2 = 0.929 \qquad ESS = 72.6$$

**Α)** Να ελεγχθεί η στατιστική σημαντικότητα των συντελεστών $\beta_1$, $\beta_2$ (α=0.05).

$$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0$$

$$t = \frac{\hat{\beta}_1 - 0}{S_{\hat{\beta}_1}} = \frac{-34.183}{5.43} = -6.28 \qquad |t| = 6.28 > t_{(0.025),3} = 3.182$$

$$H_0 : \beta_2 = 0 \qquad H_1 : \beta_2 \neq 0$$

$$t = \frac{\hat{\beta}_2 - 0}{S_{\hat{\beta}_2}} = \frac{-3.051}{18.84} = -0.162 \qquad |t| = 0.162 < t_{(0.025),3} = 3.182$$

# Άσκηση 3

**Β)** Να ελεγχθεί αν η συνδυασμένη επίδραση των συντελεστών $\beta_1$, $\beta_2$ είναι σημαντική (α=0.05).

$$H_0: \beta_1 = \beta_2 = 0$$
$$H_1: \beta_1 \neq 0 \text{ ή/και } \beta_2 \neq 0$$

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0{,}929/(3-1)}{(1-0{,}929)/(6-3)} = 20.19$$

$$F = 20{,}19 > F_{(0.05),2,3} = 9{,}55$$

**Γ)** Να ελεγχθεί αν $\beta_2=0$ και με το κριτήριο F.

Εκτιμάμε την συνάρτηση: $Y = \beta_0 + \beta_1 X_1 + v$

$$\hat{Y} = 538.89 - 34.14 X_1 \quad R^2 = 0.928$$
$$ESS = 73.24$$

$$F = \frac{\dfrac{ESS_R - ESS_U}{k-m}}{\dfrac{ESS_U}{n-k}} = \frac{\dfrac{73.24-72.6}{3-2}}{\dfrac{72.6}{6-3}} = 0.026 < F_{(0.05),1,3} = 10.1$$

# Άσκηση 3

**Δ)** Να εκτιμηθεί η σταυροειδής ελαστικότητα ζήτησης στο σημείο των μέσων.

$$\varepsilon_{\mathrm{YX}_1} = \frac{\partial \hat{Y}}{\partial X_2}\frac{\overline{X}_2}{\overline{Y}} = \hat{\beta}_2 \frac{\overline{X}_2}{\overline{Y}} = -3.051\frac{6.4}{81.3} = -0.24$$

Αν η τιμή του αγαθού 2 αυξηθεί κατά 1% **ενώ η τιμή του αγαθού 1 παραμένει σταθερή**, η ζητούμενη ποσότητα από το αγαθό 1 θα μειωθεί κατά 0.24%.

Το πρόσημο δεν είναι το αναμενόμενο για υποκατάστατα αγαθά.

Σφάλμα εξειδίκευσης;;

# Εξειδίκευση του Υποδείγματος

# Εξειδίκευση

**Επιλογή των κατάλληλων ανεξάρτητων μεταβλητών**

Παράλειψη σχετικής μεταβλητής

Άσχετη μεταβλητή στο υπόδειγμα

**Επιλογή της κατάλληλης μορφής της συνάρτησης**

# Παράλειψη σχετικής μεταβλητής

Το $\hat{\beta}_k$ αντιπροσωπεύει την μεταβολή της $Y$ ύστερα από μια μεταβολή της $X_k$ κατά μία μονάδα όταν ***όλες*** οι άλλες ***ανεξάρτητες μεταβλητές*** παραμένουν ***σταθερές*** σε κάποιο επίπεδο.

Παράλειψη μιας μεταβλητής → Δεν παραμένουν όλες οι ανεξάρτητες μεταβλητές σταθερές

↓

Πιθανή μεροληψία στους συντελεστές της συνάρτησης

Έστω $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

Αν παραλείψουμε την $X_2$ $Y_i = \beta_0 + \beta_1 X_{1i} + u_i^*$

Όπου $u_i^* = u_i + \beta_2 X_{2i}$

$$\mathrm{E}u_i^* = \mathrm{E}\left(u_i + \beta_2 X_{2i}\right) = \mathrm{E}\left(\beta_2 X_{2i}\right)$$

Γενικά $Cov\left(X_1, X_2\right) \neq 0 \Rightarrow Cov\left(u^*, X_1\right) \neq 0$

Παραβίαση βασικής υπόθεσης

$\Rightarrow$ $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{u}_i^*$

$$E\left(\hat{\beta}_1\right) \neq \beta_1$$

$$E\left(\hat{\beta}_0\right) \neq \beta_0$$

Όταν $Cov\left(X_1, X_2\right) \neq 0$ Μέρος της διακύμανσης του $Y$ που αποδίδεται στο $X_1$ οφείλεται στο $X_2$

$$E\left(\hat{\beta}_1\right) = \beta_1 + \beta_2 \hat{a}_1$$

$$X_{2i} = \hat{a}_0 + \hat{a}_1 X_{1i} + \hat{w}_i$$

Συνήθως τα $\beta_2$ ή/και $\hat{\alpha}_1$ είναι άγνωστα.

*Από τα* **αναμενόμενα** πρόσημα είναι δυνατό να προβλέψουμε το πρόσημο της μεροληψίας

Μεροληψία $> 0$ αν $\beta_2 > 0$ και $\hat{\alpha}_1 > 0$ ή $\beta_2 < 0$ και $\hat{\alpha}_1 < 0$

Μεροληψία $< 0$ αν $\beta_2 > 0$ και $\hat{\alpha}_1 < 0$ ή $\beta_2 < 0$ και $\hat{\alpha}_1 > 0$

Μεροληψία $= 0$ όταν $\beta_2 = 0$ ή/και $\hat{\alpha}_1 = 0$

Αποδεικνύεται ομοίως ότι οι διακυμάνσεις των συντελεστών είναι θετικά μεροληπτικές, δηλαδή οδηγούν σε υπερεκτιμήσεις ακόμα και αν οι $X_1, X_2$ είναι ασυσχέτιστες.

Άρα ο έλεγχος υποθέσεων και τα διαστήματα εμπιστοσύνης δεν είναι αξιόπιστα.

**Συμπέρασμα**: όταν παραλείπεται μια σχετική μεταβλητή επηρεάζεται η αξιοπιστία των αποτελεσμάτων αφού οι **εκτιμητές των συντελεστών και των διακυμάνσεων** είναι μεροληπτικοί.

**Παράδειγμα:**

Έστω $Q_D = 54 + 0.65P + 0.25I$ Η συνάρτηση ζήτησης ενός αγαθού.

Πιθανή ένδειξη προβλήματος εξειδίκευσης

Πιθανή περίπτωση η παράλειψη της τιμής ενός άλλου αγαθού.

Το ερώτημα είναι αν αυτό είναι υποκατάστατο ή συμπληρωματικό.

Αναμενόμενο πρόσημο $<0$ $\Rightarrow$ μεροληψία $> 0$

Αν, για λόγους ευκολίας, υποθέσουμε ότι $P$ και $I$ δεν συσχετίζονται $\Rightarrow$ $\beta_X \alpha_{PX} > 0$

Όπου $X$ η μεταβλητή που έχει παραληφθεί

$\beta_X$ ο συντελεστής του $X$ (αν είχε συμπεριληφθεί)

$a_{PX}$ ο συντελεστής που προκύπτει από

$$P = a_0 + a_{PX} X + w_i$$

Γενικά, αναμένουμε ότι οι τιμές κινούνται στην ίδια κατεύθυνση. Δηλαδή $\alpha_{PX}>0$

Αν $X$ η τιμή ενός συμπληρωματικού $\Rightarrow \quad \beta_X < 0 \qquad a_{PX} > 0$

Μεροληψία $<$ 0

Αν $X$ η τιμή ενός υποκατάστατου $\Rightarrow \quad \beta_X > 0 \qquad a_{PX} > 0$

Μεροληψία $>$ 0

Συμπέρασμα: Το πιθανότερο είναι ότι έχει παραληφθεί η τιμή ενός υποκατάστατου

Η διαπίστωση πόσες και ποιες μεταβλητές έχουν παραληφθεί είναι δύσκολη υπόθεση

Συνήθως το υπόδειγμα που εκτιμάται θεωρείται το σωστό
*(Για ποιο λόγο να εκτιμηθεί ένα υπόδειγμα που δεν θεωρείται σωστό)*

Γι' αυτό απαιτείται προσεκτική ανάλυση πριν την εκτίμηση

- Ποιες είναι οι μεταβλητές που προκύπτουν από τη θεωρία
- Ποια είναι τα αναμενόμενα πρόσημα
- Ποια είναι η εμπειρία από άλλες παρόμοιες εργασίες

# Προσθήκη άσχετης μεταβλητής

Έστω $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$ η πραγματική συνάρτηση

και $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i^*$ η συνάρτηση που πρόκειται να εκτιμηθεί

όπου $u_i^* = u_i - \beta_2 X_{2i}$

Αφού $\beta_2 = 0$ στον πληθυσμό η προσθήκη του $X_2$ **δεν δημιουργεί πρόβλημα μεροληψίας**

<u>Η προσθήκη άσχετης μεταβλητής</u>

- Αυξάνει τις τιμές των $S_{\hat{\beta}_i}$ και άρα μειώνει την στατιστική σημαντικότητα των $\hat{\beta}_i$
- Μειώνει την τιμή του $\overline{R}^2$

## Κριτήρια για την αφαίρεση άσχετης μεταβλητής

- Η θεωρία
- Η στατιστική σημαντικότητα του αντίστοιχου συντελεστή (έλεγχος $t$)
- Το $\bar{R}^2$
- Σημαντικές αλλαγές στους υπόλοιπους συντελεστές

Αν κανένα από τα παραπάνω δεν ισχύει η μεταβλητή αφαιρείται

Αν ισχύουν ορισμένα απαιτείται η κρίση του ερευνητή βασισμένη περισσότερο στην **θεωρητική επιχειρηματολογία** και λιγότερο στην **στατιστική**

Η επιλογή των μεταβλητών πρέπει να βασίζεται περισσότερο στην **θεωρία** και την **εμπειρία** και να ελαχιστοποιείται ο αριθμός των **«εναλλακτικών εξειδικεύσεων»**.

Θα πρέπει να αποφεύγονται τεχνικές που βασίζονται στην λογική επαναληπτικών εκτιμήσεων και επιλογή μιας συνάρτησης με κριτήριο την στατιστική σημαντικότητα μόνο

Stepwise παλινδρόμηση
Data mining

***Εξαίρεση:*** όταν οι μεθοδολογίες αυτές χρησιμοποιούνται για να διατυπωθεί μια νέα θεωρία ή όταν γίνεται προσπάθεια να αποδεχτεί η αντοχή των αποτελεσμάτων σε διαφορετικές εξειδικεύσεις

## *Έλεγχος εξειδίκευσης RESET (Regression Equation Specification Error Test)*

Έστω $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$ η συνάρτηση που εκτιμήθηκε

Εκτιμάται η συνάρτηση

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 \hat{Y}_i^2 + \beta_4 \hat{Y}_i^3 + \beta_5 \hat{Y}_i^4 + v_i$$

Ελέγχεται $H_0 : \hat{\beta}_3 = \hat{\beta}_4 = \hat{\beta}_5 = 0$

- Αν η υπόθεση $H_0$ απορριφθεί υπάρχει ένδειξη προβλήματος εξειδίκευσης
- Το *RESET* δεν βοηθά στην κατεύθυνση της λύσης του προβλήματος
- Χρησιμοποιείται κυρίως για να επιβεβαιώσει το πρόβλημα

## Παράδειγμα:

Έστω $\hat{Y}_i = 6,12 + 0,16X_{1i} + 1,07X_{2i}$ η συνάρτηση που εκτιμήθηκε

$$R^2 = 0,46$$

Όπου $Y$ η οικογενειακή δαπάνη για τρόφιμα, $X_1$ το οικογενειακό εισόδημα και $X_2$ ο αριθμός ατόμων στην οικογένεια.

Εκτιμάμε την συνάρτηση:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 \hat{Y}_i^2 + \beta_4 \hat{Y}_i^3 + \beta_5 \hat{Y}_i^4 + v_i$$

Και παίρνουμε:

$$Y_i = -133,3 - 17,4X_{1i} - 116,6X_{2i} + 8,3\hat{Y}_i^2 - 0,27\hat{Y}_i^3 + 0,003\hat{Y}_i^4 + v_i$$

$$R^2 = 0,48$$

$$F = \frac{\frac{R_U^2 - R_R^2}{k-m}}{\frac{1-R_U^2}{n-k}} = \frac{\frac{0,48-0,46}{6-3}}{\frac{1-0,48}{40-6}} = 0,67 < F_{(0.05),2,35} = 3,27$$

Δεχόμαστε ότι $\beta_3 = \beta_4 = \beta_5 = 0$, άρα δεν υπάρχει σφάλμα εξειδίκευσης.

## *Έλεγχος με τον πολλαπλασιαστή Lagrange για προσθήκη μεταβλητών*

Έστω $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$ η συνάρτηση που εκτιμήθηκε (1)

Μήπως θα έπρεπε να εκτιμήσουμε την (2) ;; :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + v_i$$

1ον: Εκτιμάται η περιορισμένη μορφή (1) και υπολογίζονται τα κατάλοιπα $\hat{u}$

2ον: Εκτιμάται η παλινδρόμηση

$$\hat{u} = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + b_4 X_{4i} + b_5 X_{5i} + e_i$$

Υπολογίζεται το $nR^2$

3ον: Εάν $nR^2 > \chi^2_{\alpha,k-m}$ η περιορισμένη μορφή (1) απορρίπτεται.

**Παράδειγμα:**

Έστω $\hat{Y}_i = 10,67 + 0,164 X_{1i}$ η συνάρτηση που εκτιμήθηκε

$R^2 = 0,34$

Όπου Y η οικογενειακή δαπάνη για τρόφιμα, $X_1$ το οικογενειακό εισόδημα.

Είναι σημαντική η προσθήκη της $X_2$ (αριθμός ατόμων στην οικογένεια);

Εκτιμάμε την συνάρτηση:

$$u_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + v_i$$

Και παίρνουμε:

$$\hat{u}_i = -4,553 - 0,004 X_{1i} + 1,071 X_{2i}$$

$$R^2 = 0,18$$

$$nR^2 = 7,2 > \chi^2_{(0.05),1} = 3,841$$

Απορρίπτουμε την περιορισμένη μορφή άρα η προσθήκη της $X_2$ είναι σημαντική.

# Η μορφή της συνάρτησης

Η επιλογή της μορφής της συνάρτησης θα πρέπει να στηρίζεται κυρίως στην ***θεωρία***, στην ***εμπειρία*** και στην ***κοινή λογική*** και ελάχιστα στον ***βαθμό προσαρμογής*** της στα στοιχεία του δείγματος.

<u>*Γραμμική και προσθετική μορφή*</u>

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{k-1} X_{k-1i} + u_i$$

$$\beta_i = \frac{\partial Y}{\partial X_i} = \frac{\Delta Y}{\Delta X}$$

Υποθέτει ότι η μεταβολή στο $Y$ ύστερα από μια μεταβολή στο $X$ κατά μία μονάδα είναι ***σταθερή*** (ανεξάρτητη από το $X$ και $Y$)

- Συνήθως η θεωρία προτείνει μόνο το πρόσημο
- 1η επιλογή εκτός αν υπάρχει σοβαρός λόγος για το αντίθετο

## *Διπλή λογαριθμική μορφή*

$$\ln Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \ldots + \beta_{k-1} \ln X_{k-1i} + u_i$$

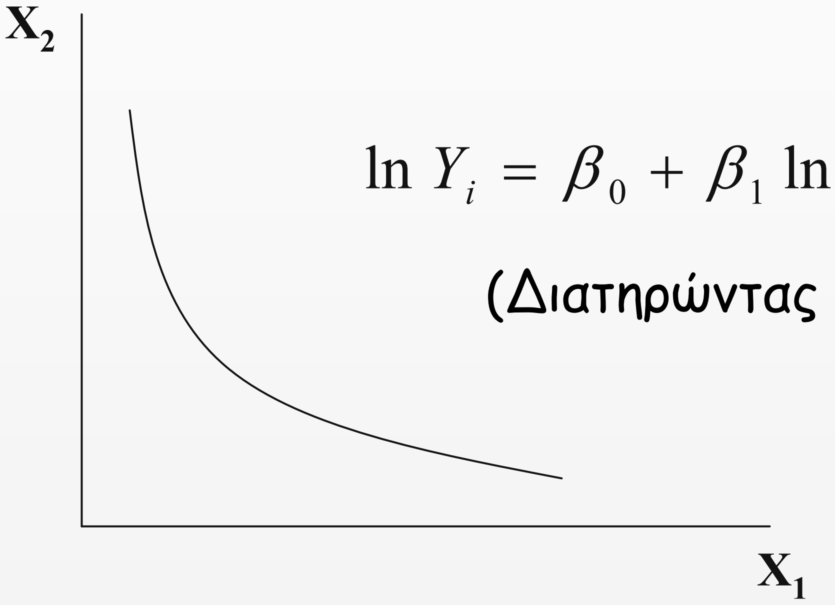
$$\beta_i = \frac{\partial \ln Y}{\partial \ln X_i} = \frac{\%\Delta Y}{\%\Delta X}$$ ***Ελαστικότητα***

Υποθέτει ότι η ελαστικότητα είναι ***σταθερή***.
Το % μεταβολής στο $Y$ ύστερα από μια μεταβολή στο $X$ κατά 1% είναι ***σταθερό***

Πραγματική μορφή

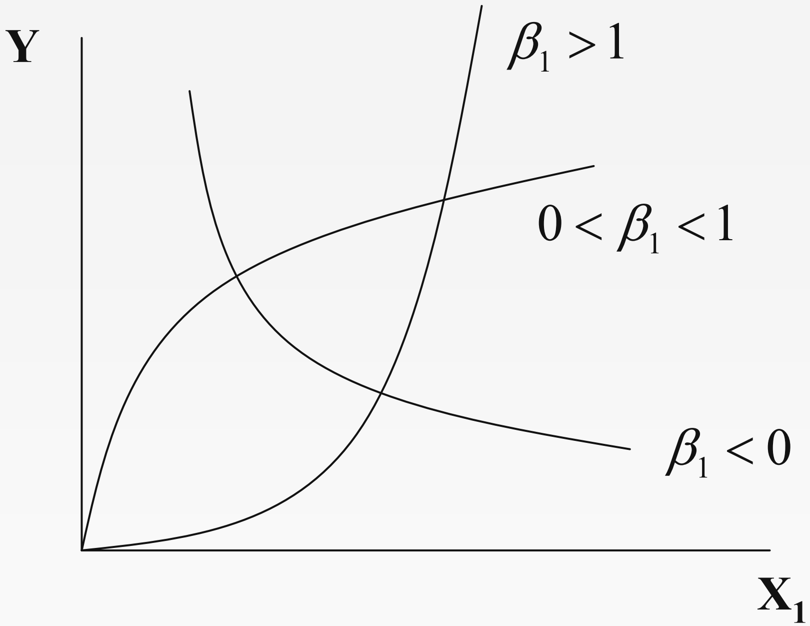
$$Y_i = e^{\beta_0} X_{1i}^{\beta_1} X_{2i}^{\beta_2} \cdots X_{k-1i}^{\beta_{k-1}} e^{u_i}$$

Προσοχή! Στις μηδενικές και αρνητικές τιμές των μεταβλητών




$$\ln Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i}$$

(Διατηρώντας το $Y$ σταθερό)



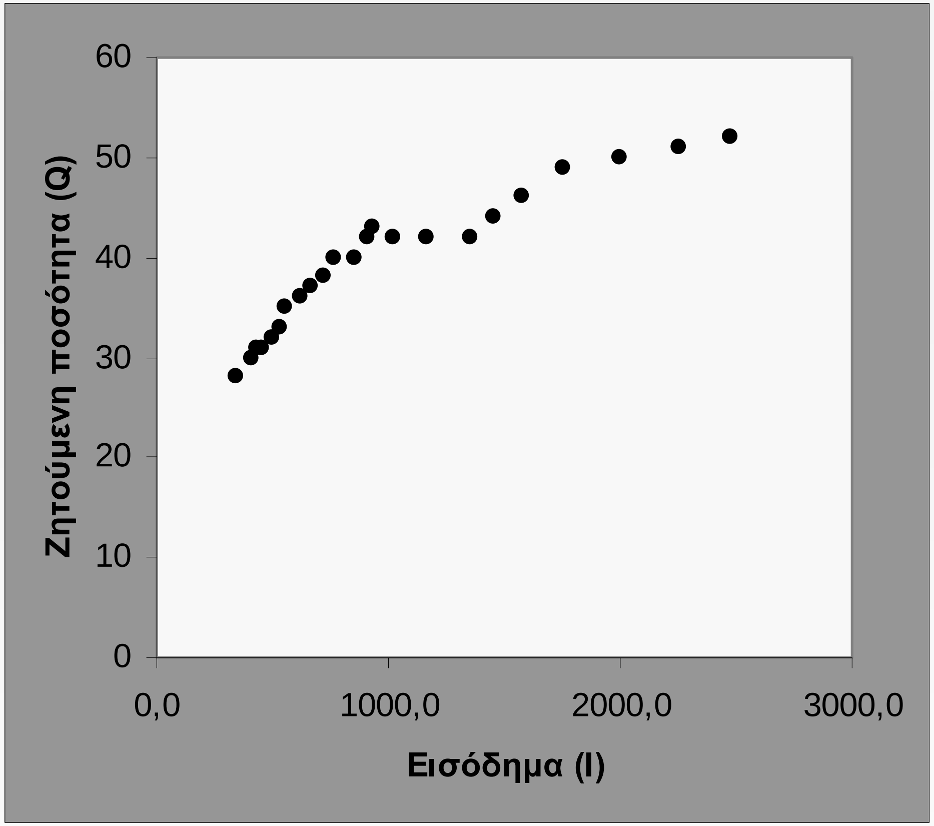

$$\ln Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i}$$

(Διατηρώντας το $X_2$ σταθερό)

# Παράδειγμα:

| $Q_A$ | I | $P_A$ | $P_B$ | $P_C$ |
|---|---|---|---|---|
| 28 | 340,0 | 42,2 | 50,7 | 78,3 |
| 30 | 415,0 | 38,1 | 52,0 | 79,2 |
| 31 | 435,0 | 40,3 | 54,0 | 79,2 |
| 31 | 460,0 | 39,5 | 55,3 | 79,2 |
| 32 | 495,0 | 37,3 | 54,7 | 77,4 |
| 33 | 530,0 | 38,1 | 63,7 | 80,2 |
| 35 | 560,0 | 39,3 | 69,8 | 80,4 |
| 36 | 625,0 | 37,8 | 65,9 | 83,9 |
| 37 | 665,0 | 38,4 | 64,5 | 85,5 |
| 38 | 720,0 | 40,1 | 70,0 | 93,7 |
| 40 | 770,0 | 38,6 | 73,2 | 106,1 |
| 40 | 850,0 | 39,8 | 67,8 | 104,8 |
| 42 | 915,0 | 39,7 | 79,1 | 114,0 |
| 43 | 930,0 | 52,1 | 95,4 | 124,1 |
| 42 | 1020,0 | 48,9 | 94,2 | 127,6 |
| 42 | 1170,0 | 58,3 | 123,5 | 142,9 |
| 42 | 1350,0 | 57,9 | 129,9 | 143,6 |
| 44 | 1450,0 | 56,5 | 117,6 | 139,2 |
| 46 | 1575,0 | 63,7 | 130,9 | 165,5 |
| 49 | 1760,0 | 61,6 | 129,8 | 203,3 |
| 50 | 2000,0 | 58,9 | 128,0 | 219,6 |
| 51 | 2260,0 | 66,4 | 141,0 | 221,6 |
| 52 | 2480,0 | 70,4 | 168,2 | 232,6 |

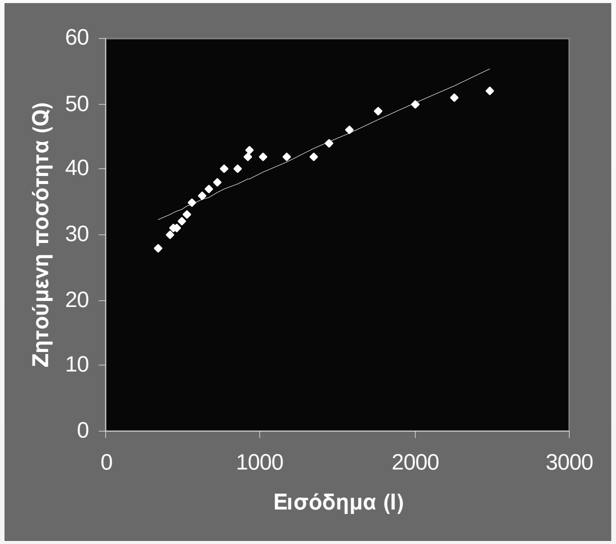
Υποθέτουμε όλες τις τιμές σταθερές



## Γραμμική

$$\hat{Q}_A = 28,67 + 0,011I$$

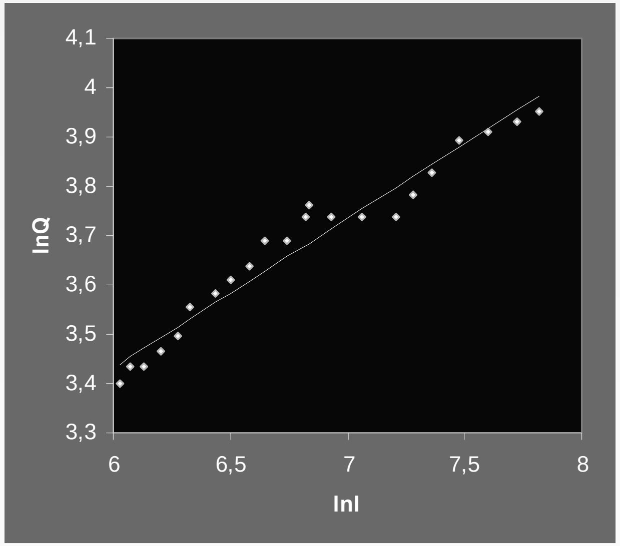
$$(0,987)\,(0,0008) \qquad R^2 = 0,889$$

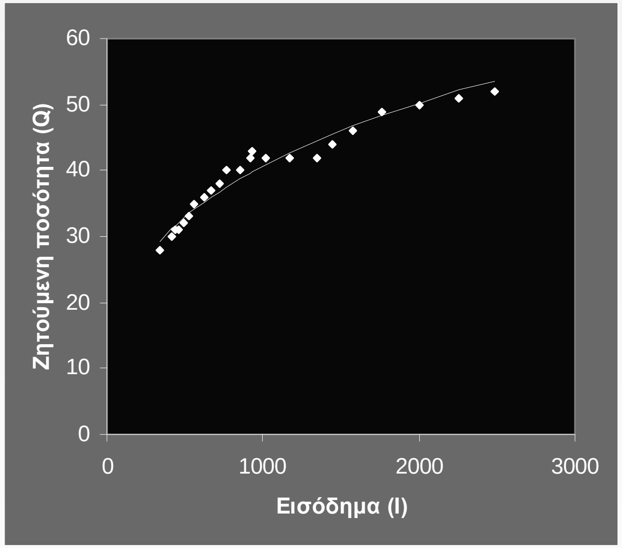



## Διπλή λογαριθμική

$$\ln \hat{Q}_A = 1,610 + 0,303 \ln I$$

$$(0,095)\,(0,014) \qquad R^2 = 0,957$$



## ***Ημιλογαριθμική μορφή*** **(1) (λογαριθμική-γραμμική)**

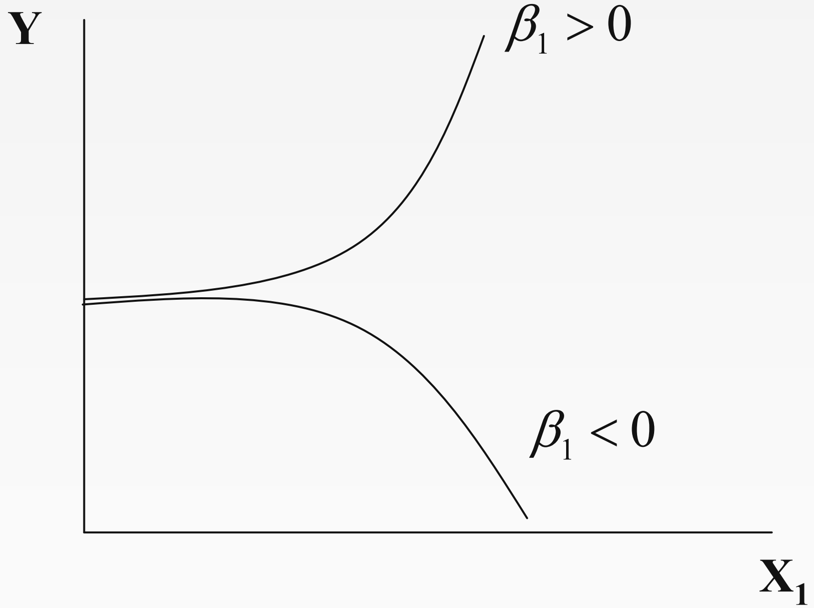
$$\ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{k-1} X_{k-1i} + u_i$$

$$\beta_i = \frac{\partial \ln Y}{\partial X_i} = \frac{\% \Delta Y}{\Delta X}$$

Πραγματική μορφή

$$Y_i = e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{\kappa-1} X_{k-1} + u_i}$$

Υποθέτει ότι η ***ποσοστιαία μεταβολή*** στο $Y$ ύστερα από μια μεταβολή στο $X$ κατά μία μονάδα είναι ***σταθερή***




$$\ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

(Διατηρώντας το $X_2$ σταθερό)

## Εφαρμογή: Μέσος ετήσιος ρυθμός μεταβολής

$$Y_t = Y_0(1 + r)^T$$

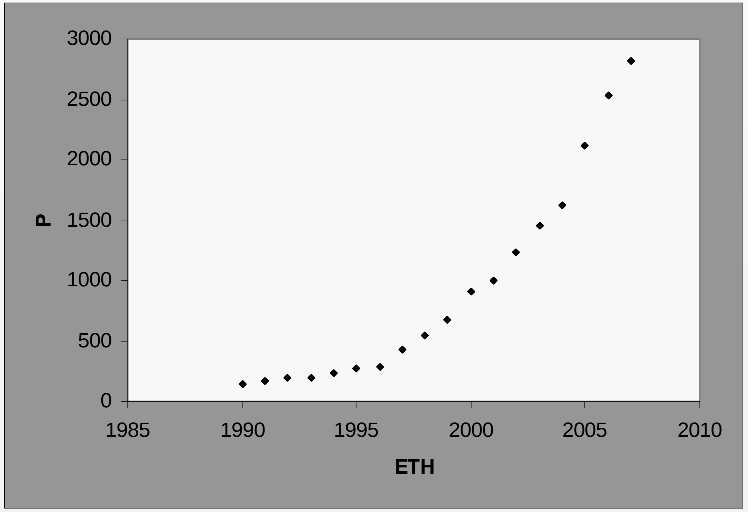
$$\ln Y_t = \ln Y_0 + T \ln(1 + r)$$

$$\ln Y_t = \beta_0 + \beta_1 T$$

$$\left.\begin{aligned} \beta_0 &= \ln Y_0 \\ \beta_1 &= \ln(1 + r) \end{aligned}\right\} \quad \begin{aligned} Y_0 &= e^{\beta_0} \\ r &= e^{\beta_1} - 1 \end{aligned}$$

# Παράδειγμα:

| T | X | P |
|---|---|---|
| 1990 | 1 | 143 |
| 1991 | 2 | 167 |
| 1992 | 3 | 195 |
| 1993 | 4 | 201 |
| 1994 | 5 | 234 |
| 1995 | 6 | 267 |
| 1996 | 7 | 287 |
| 1997 | 8 | 435 |
| 1998 | 9 | 543 |
| 1999 | 10 | 678 |
| 2000 | 11 | 908 |
| 2001 | 12 | 1000 |
| 2002 | 13 | 1234 |
| 2003 | 14 | 1456 |
| 2004 | 15 | 1623 |
| 2005 | 16 | 2123 |
| 2006 | 17 | 2534 |
| 2007 | 18 | 2819 |

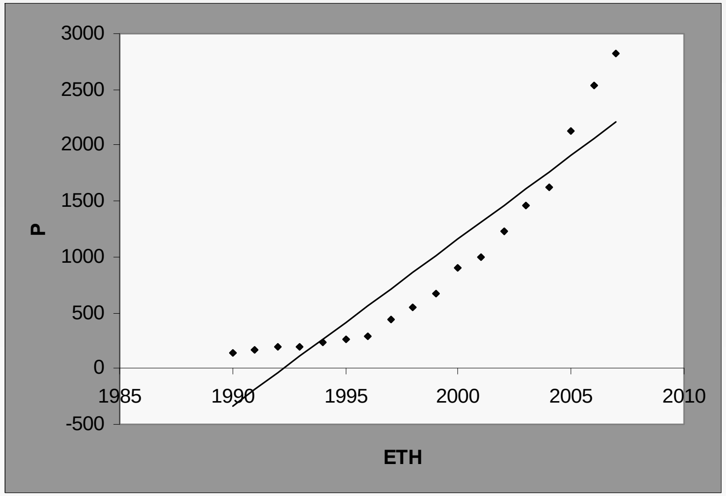



Γραμμική

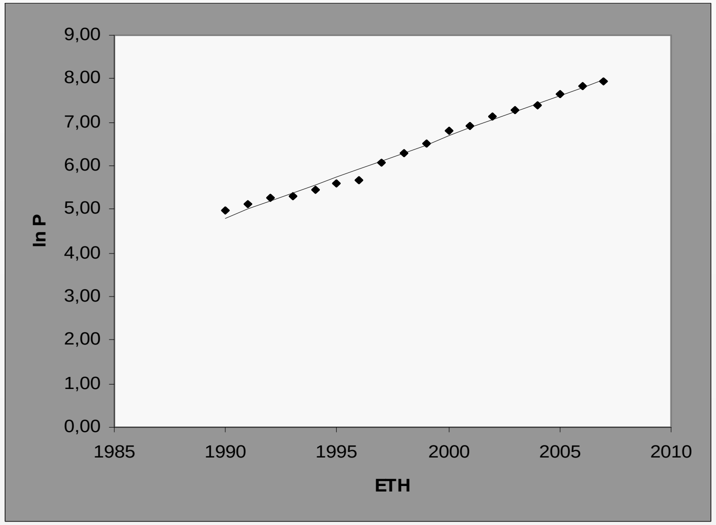
$$\hat{P} = -481,6 + 149,2X$$

$$(160,3) \quad (14,8) \qquad R^2 = 0,863$$

Μέση ετήσια μεταβολή



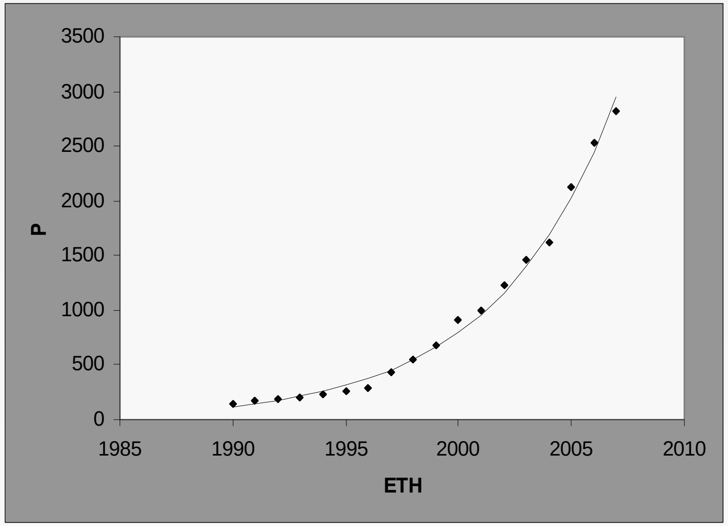
$$\ln \hat{P} = 4,622 + 0,187X$$
$$(0,054) \quad (0,005) \qquad R^2 = 0,988$$




$$P_0 = e^{\beta_0} = 101,755$$

$$r = e^{\beta_1} - 1 = 0,206$$

Μέση ετήσια ποσοστιαία αύξηση



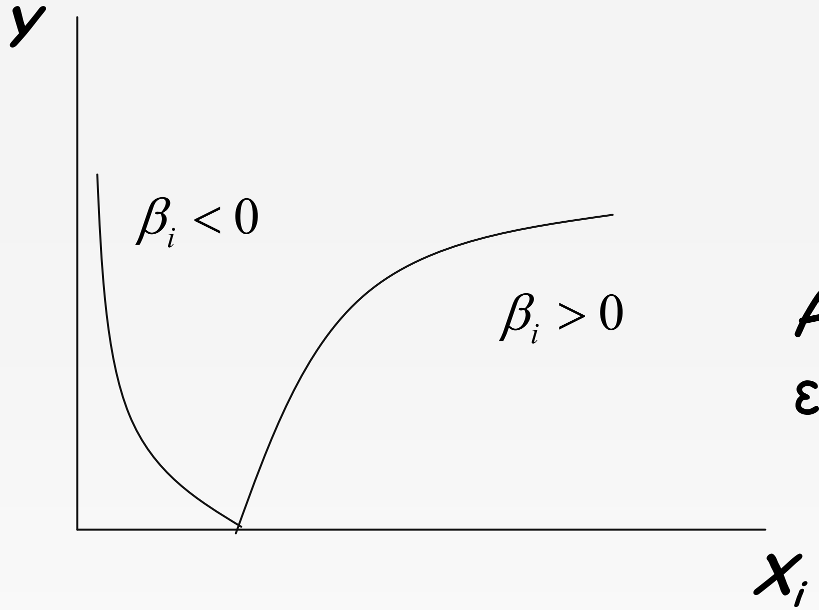

$$\hat{P}_t = 101,755(1 + 0,206)^T$$

## Ημιλογαριθμική μορφή (2) (γραμμική-λογαριθμική)

$$Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \ldots + \beta_{k-1} \ln X_{k-1i} + u_i$$

$$\beta_i = \frac{\partial Y}{\partial \ln X_i} = \frac{\Delta Y}{\% \Delta X}$$

Υποθέτει ότι η μεταβολή στο $Y$ ύστερα από μια μεταβολή στο $X$ κατά 1% είναι **σταθερή**




Αν διατηρήσουμε όλα τα $X$, εκτός του $X_i$, σταθερά

# Παράδειγμα:

Γραμμική

$$\hat{Q}_A = 28,67 + 0,011I$$

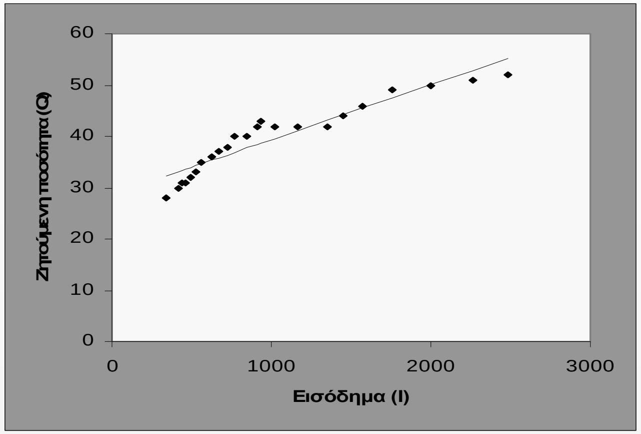
$$(0,987)\,(0,0008) \qquad R^2 = 0,889$$

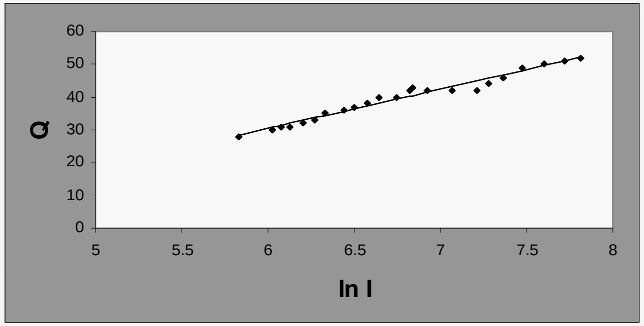
Ημιλογαριθμική

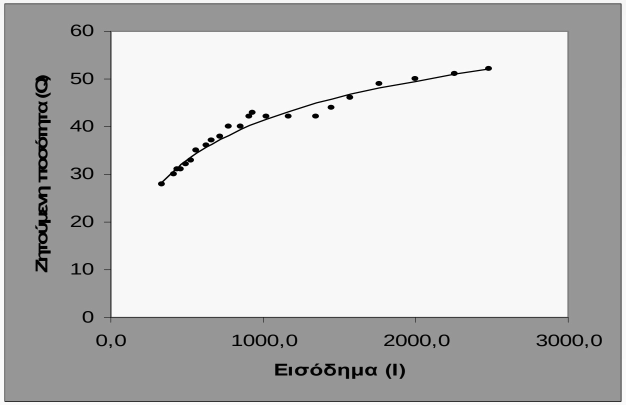
$$\hat{Q}_A = -41,28 + 11,95\ln I$$

$$(3,103) \quad (0,456) \qquad R^2 = 0,970$$

Απόλυτη μεταβολή στη ζητούμενη ποσότητα ύστερα από μια μεταβολή στο εισόδημα κατά 1%



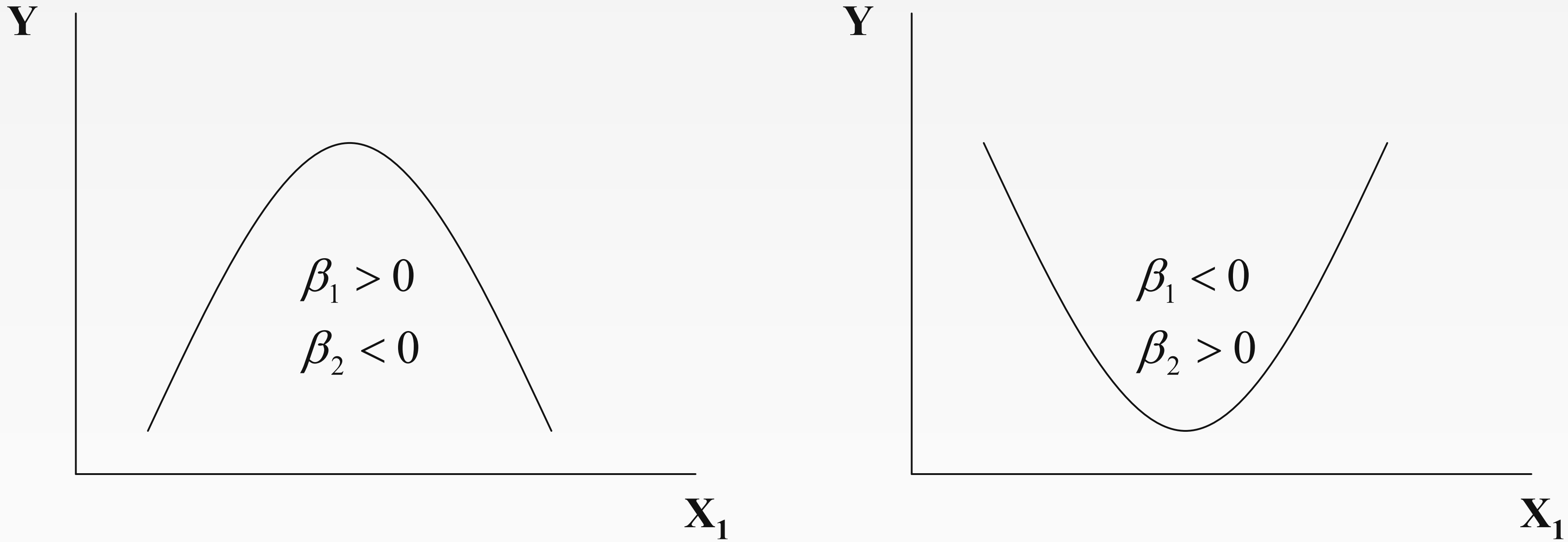



## *Πολυωνυμική μορφή*

Παράδειγμα $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \beta_3 X_{2i} + u_i$

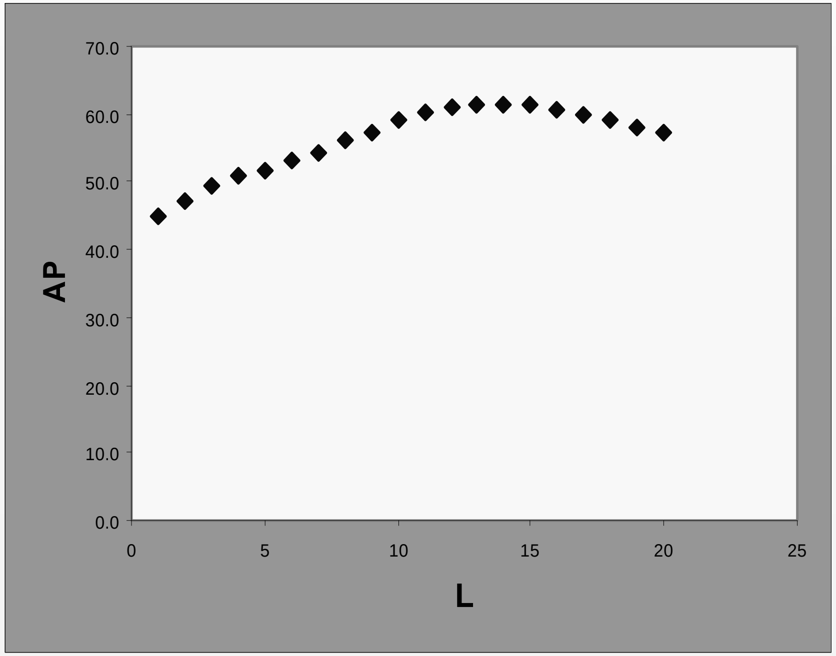
$$\frac{\partial Y}{\partial X_1} = \beta_1 + 2\beta_2 X_1$$

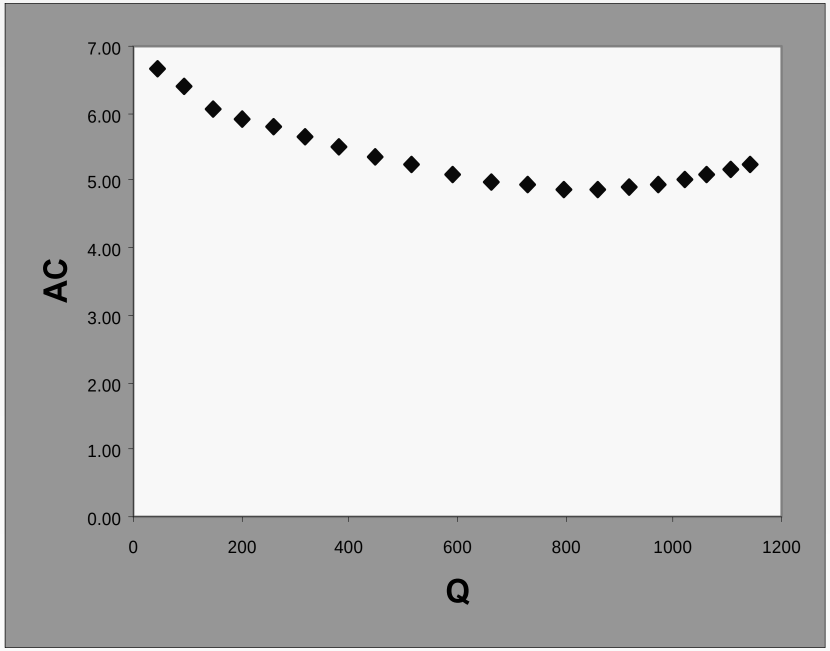
Υποθέτει ότι η μεταβολή στο $Y$ ύστερα από μια μεταβολή στο $X$ δεν είναι σταθερή αλλά ***εξαρτάται από την ίδια τη μεταβλητη.***



# Παράδειγμα:

| L | Q | MP | AP | TC | AC |
|---|---|---|---|---|---|
| 1 | 45 | 45 | 45.0 | 300 | 6.67 |
| 2 | 94 | 49 | 47.0 | 600 | 6.38 |
| 3 | 148 | 54 | 49.3 | 900 | 6.08 |
| 4 | 203 | 55 | 50.8 | 1200 | 5.91 |
| 5 | 259 | 56 | 51.8 | 1500 | 5.79 |
| 6 | 319 | 60 | 53.2 | 1800 | 5.64 |
| 7 | 381 | 62 | 54.4 | 2100 | 5.51 |
| 8 | 448 | 67 | 56.0 | 2400 | 5.36 |
| 9 | 517 | 69 | 57.4 | 2700 | 5.22 |
| 10 | 590 | 73 | 59.0 | 3000 | 5.08 |
| 11 | 662 | 72 | 60.2 | 3300 | 4.98 |
| 12 | 730 | 68 | 60.8 | 3600 | 4.93 |
| 13 | 799 | 69 | 61.5 | 3900 | 4.88 |
| 14 | 861 | 62 | 61.5 | 4200 | 4.88 |
| 15 | 919 | 58 | 61.3 | 4500 | 4.90 |
| 16 | 971 | 52 | 60.7 | 4800 | 4.94 |
| 17 | 1019 | 48 | 59.9 | 5100 | 5.00 |
| 18 | 1062 | 43 | 59.0 | 5400 | 5.08 |
| 19 | 1104 | 42 | 58.1 | 5700 | 5.16 |
| 20 | 1142 | 38 | 57.1 | 6000 | 5.25 |



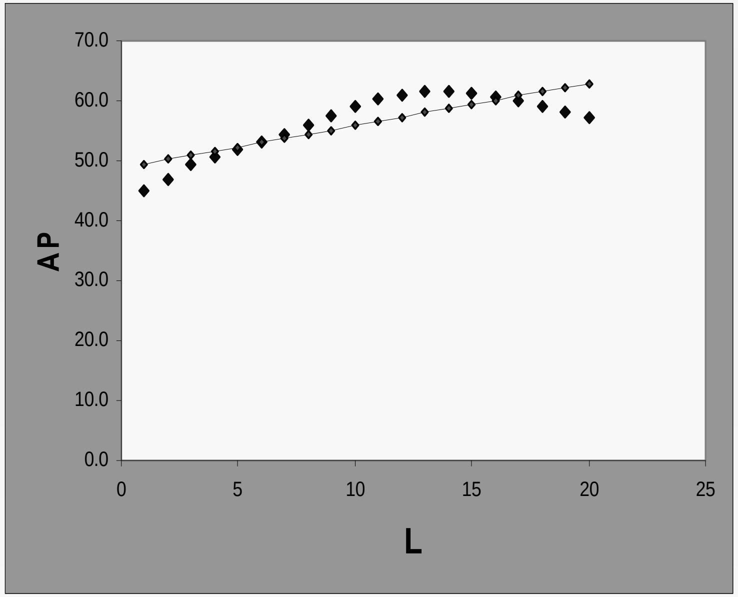



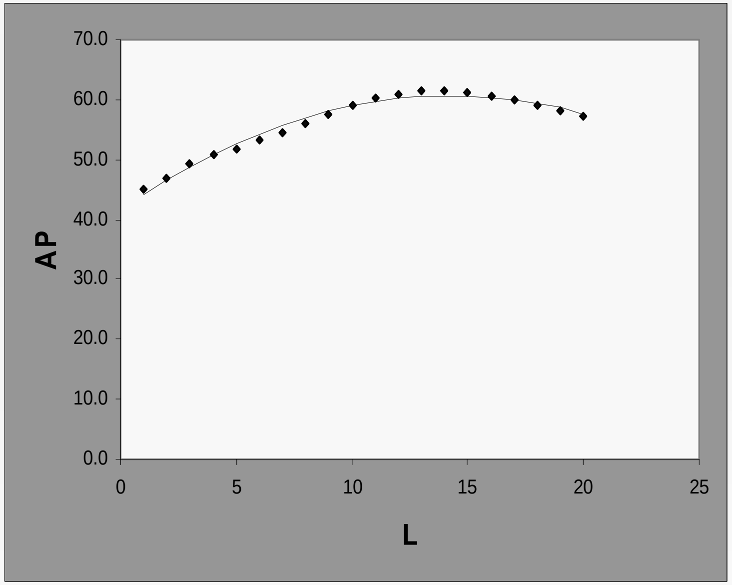
## Γραμμική Μέσου Προϊόντος

$$AP = 48,76 + 0,708L$$
$$(1,383) \quad (0,115) \quad R^2 = 0,676$$

## Πολυωνυμική Μέσου Προϊόντος

$$AP = 41,64 + 2,651L - 0,092L^2$$
$$(0,554) \quad (0,121) \quad (0,005) \quad R^2 = 0,981$$



## Γραμμική Μέσου Κόστους

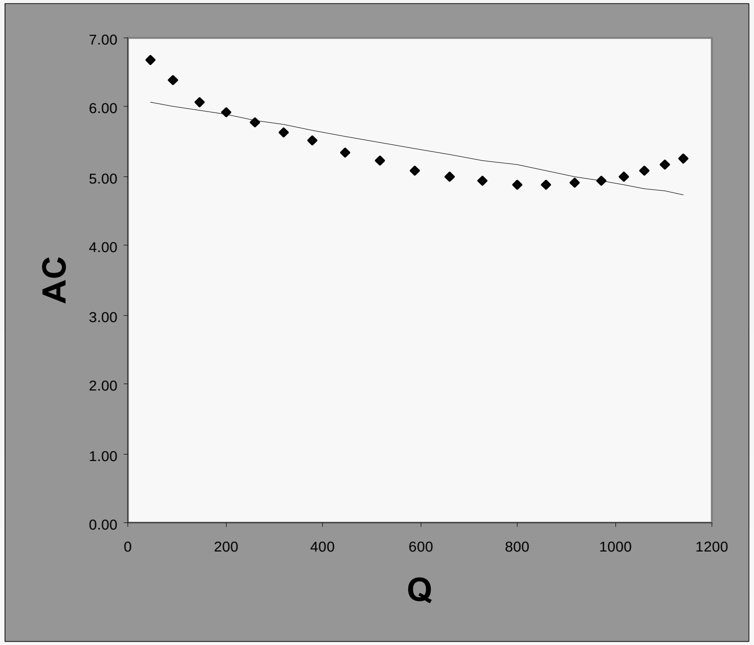
$$AC = 6,138 - 0,001Q$$

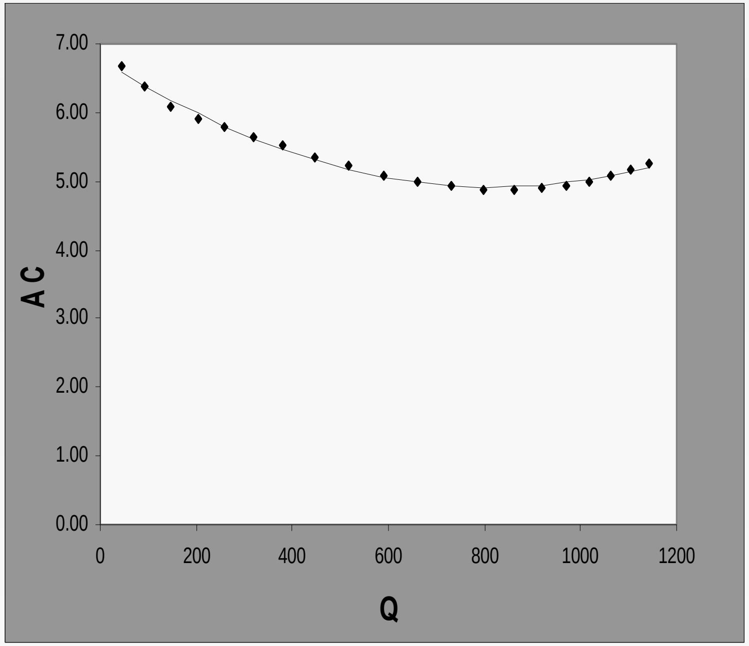
$$(0,133) \quad (0,0002) \qquad R^2 = 0,703$$

## Πολυωνυμική Μέσου Κόστους

$$AC = 6,798 - 0,005Q + 0,000003Q^2$$

$$(0,033) \quad (0,0001) \quad (0,0000001) \qquad R^2 = 0,993$$



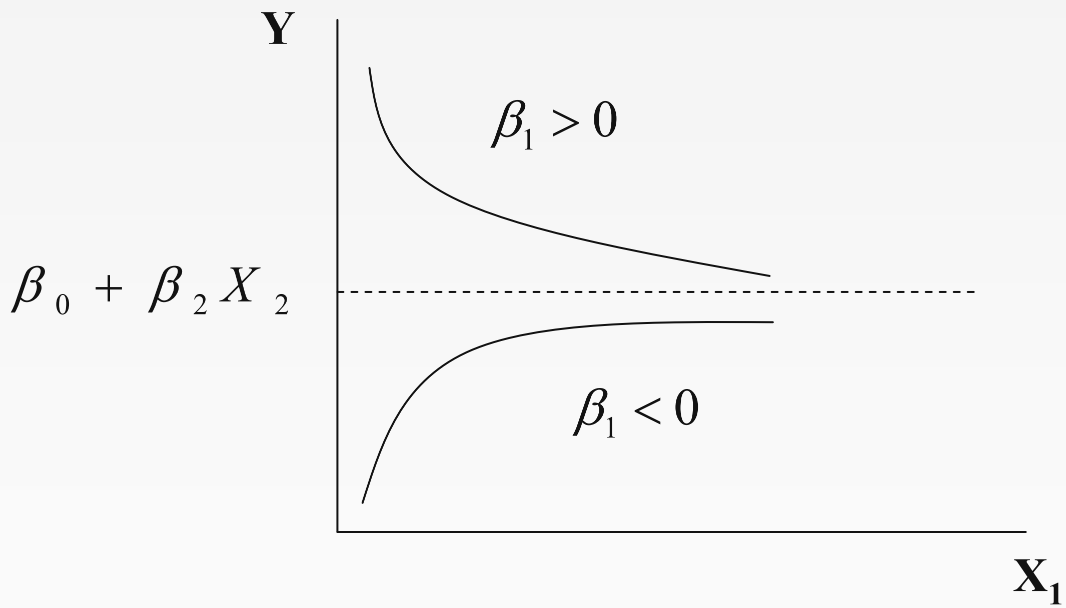
## *Αντίστροφη μορφή*

Παράδειγμα $$Y_i = \beta_0 + \beta_1 \frac{1}{X_1} + \beta_2 X_{2i} + u_i$$

$$\frac{\partial Y}{\partial X_1} = \frac{-\beta_1}{X_1^2}$$

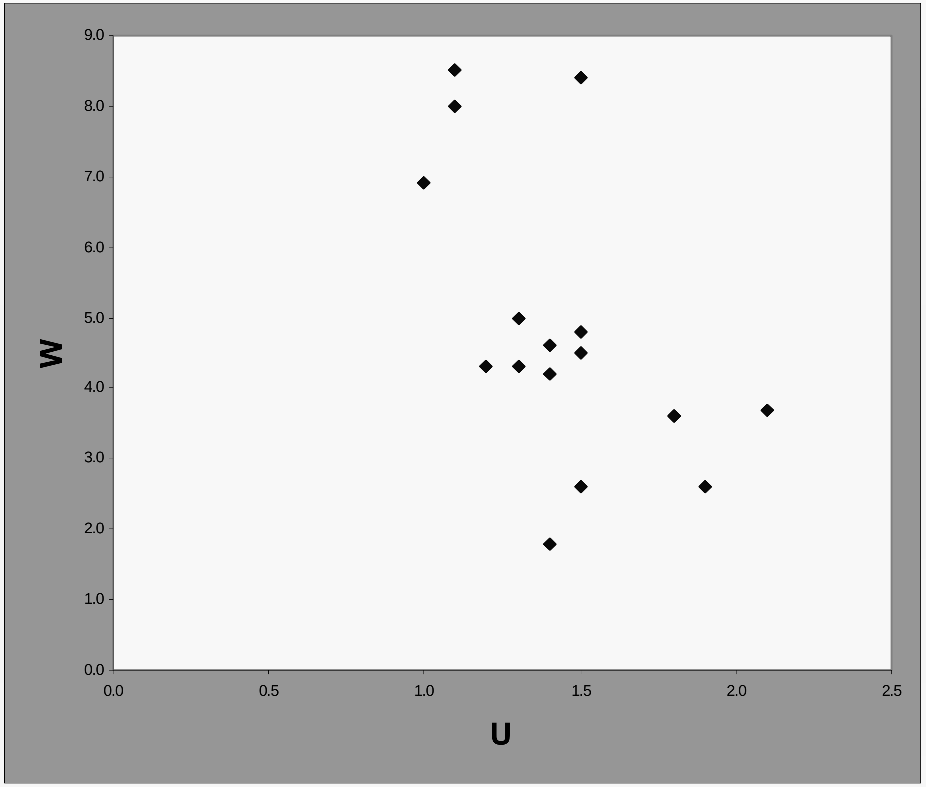
Υποθέτει ότι η μεταβολή στο $Y$ ύστερα από μια μεταβολή στο $X$ δεν είναι σταθερή αλλά ***εξαρτάται από την ίδια τη μεταβλητή.***

Αν διατηρήσουμε όλα τα *X*, εκτός του $X_1$, σταθερά



# Παράδειγμα:

| W | U |
|---|---|
| 1,8 | 1,4 |
| 8,5 | 1,1 |
| 8,4 | 1,5 |
| 4,5 | 1,5 |
| 4,3 | 1,2 |
| 6,9 | 1,0 |
| 8,0 | 1,1 |
| 5,0 | 1,3 |
| 3,6 | 1,8 |
| 2,6 | 1,9 |
| 2,6 | 1,5 |
| 4,2 | 1,4 |
| 3,6 | 1,8 |
| 3,7 | 2,1 |
| 4,8 | 1,5 |
| 4,3 | 1,3 |
| 4,6 | 1,4 |



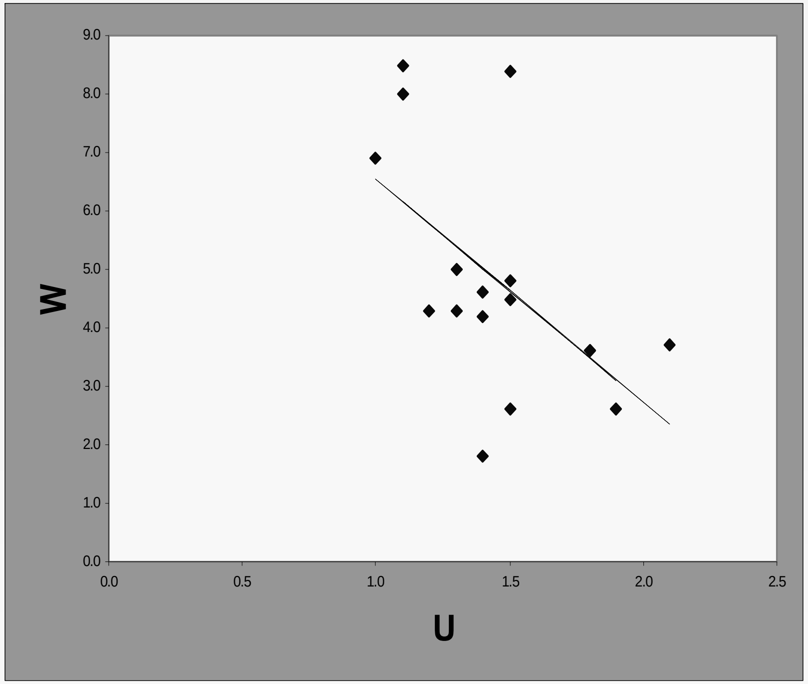

W = Ποσοστιαία μεταβολή στους μισθούς
U = Ποσοστό ανεργίας

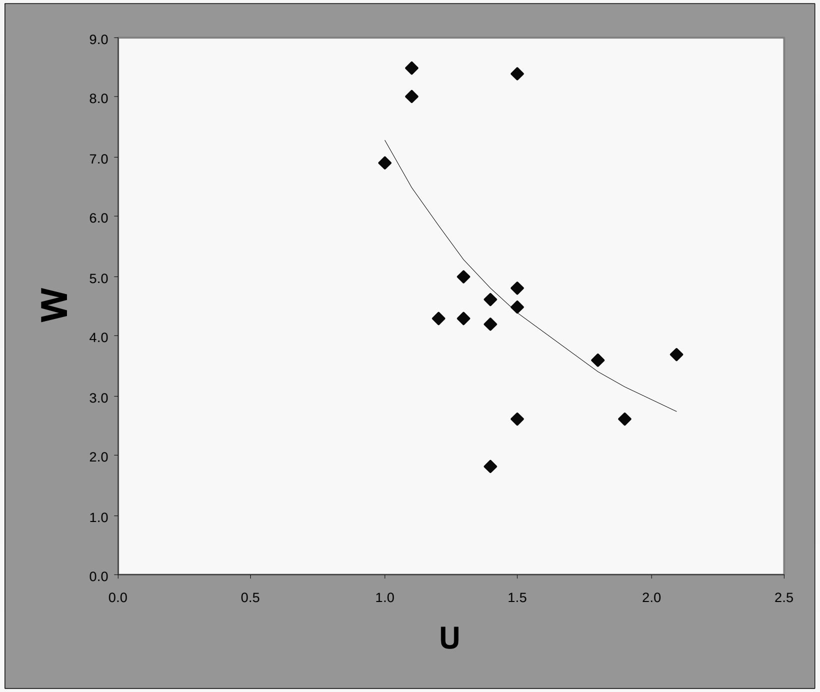
## Γραμμική

$$W = 10,343 - 3,808U$$
$$(2,127) \quad (1,430) \qquad R^2 = 0,321$$

## Αντίστροφη

$$W = -1,428 + 8,724\frac{1}{U}$$
$$(2,067) \quad (2,847) \qquad R^2 = 0,384$$



## *Αλληλεπίδραση μεταβλητών*

Παράδειγμα $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + u_i$

$$\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_3 X_2$$

Υποθέτει ότι η μεταβολή στο $Y$ ύστερα από μια μεταβολή στο $X$ δεν είναι σταθερή αλλά ***εξαρτάται από την δεύτερη μεταβλητή***

# Παράδειγμα:

| $Q_A$ | I | $P_A$ | $P_B$ | $P_C$ |
|---|---|---|---|---|
| 28 | 340,0 | 42,2 | 50,7 | 78,3 |
| 30 | 415,0 | 38,1 | 52,0 | 79,2 |
| 31 | 435,0 | 40,3 | 54,0 | 79,2 |
| 31 | 460,0 | 39,5 | 55,3 | 79,2 |
| 32 | 495,0 | 37,3 | 54,7 | 77,4 |
| 33 | 530,0 | 38,1 | 63,7 | 80,2 |
| 35 | 560,0 | 39,3 | 69,8 | 80,4 |
| 36 | 625,0 | 37,8 | 65,9 | 83,9 |
| 37 | 665,0 | 38,4 | 64,5 | 85,5 |
| 38 | 720,0 | 40,1 | 70,0 | 93,7 |
| 40 | 770,0 | 38,6 | 73,2 | 106,1 |
| 40 | 850,0 | 39,8 | 67,8 | 104,8 |
| 42 | 915,0 | 39,7 | 79,1 | 114,0 |
| 43 | 930,0 | 52,1 | 95,4 | 124,1 |
| 42 | 1020,0 | 48,9 | 94,2 | 127,6 |
| 42 | 1170,0 | 58,3 | 123,5 | 142,9 |
| 42 | 1350,0 | 57,9 | 129,9 | 143,6 |
| 44 | 1450,0 | 56,5 | 117,6 | 139,2 |
| 46 | 1575,0 | 63,7 | 130,9 | 165,5 |
| 49 | 1760,0 | 61,6 | 129,8 | 203,3 |
| 50 | 2000,0 | 58,9 | 128,0 | 219,6 |
| 51 | 2260,0 | 66,4 | 141,0 | 221,6 |
| 52 | 2480,0 | 70,4 | 168,2 | 232,6 |

$$\hat{Q}_A = 19,37 + 0,027I - 0,032P_A + 0,077P_B + 0,029P_C - 0,0003(IP_A)$$
$$(4,79)\quad(0,007)\quad(0,180)\quad(0,060)\quad(0,042)\quad(0,00007)$$
$$R^2 = 0,963$$

$$\frac{\partial \hat{Q}_A}{\partial I} = 0,027 - 0,0003P_A$$

$$\frac{\partial \hat{Q}_A}{\partial P_A} = -0,032 - 0,0004I$$