ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Επικ. Καθ. Στέλιος Ζήμερας

Τμήμα Μαθηματικών
Κατεύθυνση Στατιστικής και Αναλογιστικά – Χρηματοοικονομικά Μαθηματικά

2015

Μοντελοποίηση

- Έστω για την σύγκριση k δειγμάτων συλλέγουμε n_i παρατηρήσεις Y_{ij} j=1,2,…, n_i από το δείγμα i με ι=1,2,…,k

$$1^0 \text{ δείγμα} \Rightarrow Y_{11}, Y_{12}, \ldots, Y_{1n_1} \Rightarrow \mathrm{N}\left(\mu_1, \sigma^2\right)$$

$$2^0 \text{ δείγμα} \Rightarrow Y_{21}, Y_{22}, \ldots, Y_{2n_2} \Rightarrow \mathrm{N}\left(\mu_2, \sigma^2\right)$$

$$\vdots$$

$$\vdots$$

$$k^{\tau\acute{\alpha}\xi\varepsilon\omega\varsigma} \text{ δείγμα} \Rightarrow Y_{k1}, Y_{k2}, \ldots, Y_{kn_k} \Rightarrow \mathrm{N}\left(\mu_k, \sigma^2\right)$$

Πειραματικό σχέδιο

$$H_0: \mu_1 = \mu_2 = \ldots\ldots = \mu_k$$

Μοντελοποίηση

Δηλ: οι μέσες τιμές των πληθυσμών από τους οποίους λαμβάνονται τα εν λόγω δείγματα είναι ίσες μεταξύ τους

Θεωρούμε ότι $n = n_1 + n_2 + \ldots\ldots + n_k$ όπου n: είναι το μέγεθος του δείγματος.

Θα έχουμε ότι για την παρατήρηση Y_{ij} που γράφεται ως

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \mathrm{i} = 1,2,\ldots,\mathrm{k};\ \mathrm{j} = 1,2,\ldots,\mathrm{n_i}$$

όπου οι τ.μ. ε_{ij} είναι ανεξάρτητες και $\varepsilon_{ij} \sim N(0,\sigma^2)$ με $\sigma^2>0$

Μοντελοποίηση

Αν θέσουμε:

$$Y' = \left(Y_{11}, Y_{12}, \ldots, Y_{1n_1}, Y_{21}, Y_{22}, \ldots, Y_{2n_2}, \ldots, Y_{k1}, Y_{k2}, \ldots, Y_{kn_k}\right)$$

$$\beta' = \left(\beta_1, \beta_2, \ldots, \beta_k\right) \Rightarrow \mu' = \left(\mu_1, \mu_2, \ldots, \mu_k\right)$$

$$\varepsilon' = \left(\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{1n_1}, \varepsilon_{21}, \varepsilon_{22}, \ldots, \varepsilon_{2n_2}, \ldots, \varepsilon_{k1}, \varepsilon_{k2}, \ldots, \varepsilon_{kn_k}\right)$$

$$X = \begin{bmatrix}
\left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\}n_1 & \begin{matrix}0\\ \vdots\\ 0\end{matrix} & \begin{matrix}\cdots\\ \cdots\\ \cdots\end{matrix} & \begin{matrix}0\\ \vdots\\ 0\end{matrix} \\
\begin{matrix}0\\ \vdots\\ 0\end{matrix} & \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\}n_2 & \begin{matrix}\cdots\\ \cdots\\ \cdots\end{matrix} & \begin{matrix}0\\ \vdots\\ 0\end{matrix} \\
\begin{matrix}0\\ \vdots\\ 0\end{matrix} & \begin{matrix}0\\ \vdots\\ 0\end{matrix} & \begin{matrix}\cdots\\ \cdots\\ \cdots\end{matrix} & \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\}n_k
\end{bmatrix}$$

Μοντελοποίηση

- Αν θέσουμε Y΄ για τα διανύσματα όλων των παρατηρήσεων

$$Y' = \left(Y_{11}, Y_{12}, \ldots, Y_{1n_1}, Y_{21}, Y_{22}, \ldots, Y_{2n_2}, \ldots, Y_{k1}, Y_{k2}, \ldots, Y_{kn_k}\right)$$

Για κάθε i=1,2,…,k έστω μ_i η κοινή μέση τιμή των Y_{ij} j=1,2,…, n_i τα οποία είναι ισόνομα. Έτσι μπορούμε να γράψουμε το μοντέλο μας ως:

$$Y_{ij} = E\left(Y_{ij}\right) + \varepsilon_{ij} = \mu_i + \varepsilon_{ij}$$

$$\Downarrow$$

$$\mathbf{Y} = E\left(\mathbf{Y}\right) + \boldsymbol{\varepsilon} = X_{nxk}\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix} + \boldsymbol{\varepsilon}$$

Μοντελοποίηση

Η τάξη του πίνακα X είναι k – k ανεξάρτητοι γραμμικοί συνδυασμοί

$$E(\mathbf{Y}) = X_{nxk}\begin{pmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_k\end{pmatrix}$$

Επομένως μπορούμε να παρατηρήσουμε μέσω του E(**Y**) τα **Y** ως:

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$$

$$X = [1_n, X] \qquad \mathbf{b} = (\mu, \alpha_1, \alpha_2, \cdots, \alpha_k)'$$

$$Y_{ij} = \mu + a_i + \varepsilon_{ij}$$

Μοντελοποίηση

Επειδή ο πίνακας **X** είναι ελλιπούς τάξης (k+1 στήλες είναι γραμμικά εξαρτημένες), οι παράμετροι μ, α_i, i=1,2,…,k δεν καθορίζονται μονοσήμαντα από την σχέση E(**Y**)

Για τον λόγο αυτό θα πρέπει να υπάρχουν περιορισμοί όπως

$$\sum_{i=1}^{k} a_{ij} = 0$$

Άρα ισχύει:

$$\mu = \bar{\mu}$$

$$\alpha_i = \mu_i - \bar{\mu}$$

$$\bar{\mu} = \frac{1}{k}\sum_{i=1}^{k}\mu_i$$

$\Rightarrow$

Το μ έχει την ερμηνεία του συνολικού μέσου όρου, ενώ τα α_i δείχνουν πόσο κάθε μ_i διαφέρει από το συνολικό μέσο όρο.

Τα α_i ονομάζονται επίδραση του δείγματος i.

Μοντελοποίηση

Τότε το παραπάνω μοντέλο μπορεί να πάρει την μορφή

$$\boxed{Y = X\beta + \varepsilon} \quad \longrightarrow \quad \boxed{Y = \mu + \varepsilon}$$

$$\mu = \mathrm{X}\beta = \mathrm{E(Y)}$$

όπου η μήτρα X είναι τάξεως k, αφού καμία στήλη τη; Δεν μπορεί να είναι γραμμικός συνδυασμός άλλων στηλών τ(X)=k

Για τα ε_{ij} παρατηρούμε ότι $\varepsilon \sim N_n\,(0,\sigma^2 I_n)$

$$\sigma^2 I = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

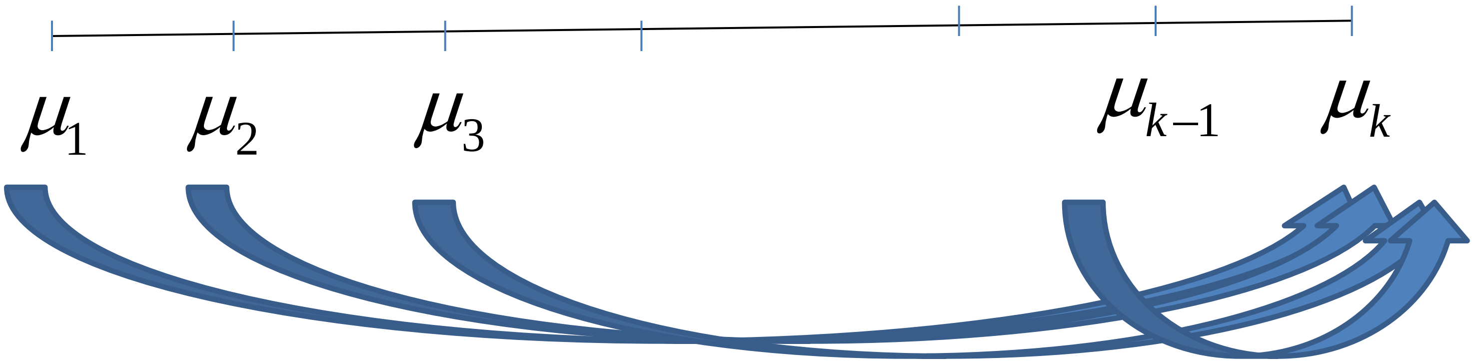
Μήτρα διακυμάνσεων-συνδιακυμάνσεων

Από τα παραπάνω έχουμε το κανονικό κλασσικό γραμμικό υπόδειγμα με τ(X)=k όπου

$$\boxed{\mathrm{Y} \sim N_n\,(\mathrm{X}\beta,\sigma^2 I_n)}$$

Μοντελοποίηση

Η υπόθεση H_0 μπορεί να γραφεί:



$$\mu_1 - \mu_k = 0, \mu_2 - \mu_k = 0, \ldots, \mu_{k-1} - \mu_k = 0$$

Μοντελοποίηση

Γραμμικοί συνδυασμοί:

$$\mu_1 - \mu_k = 1*\mu_1 + 0*\mu_2 + 0*\mu_3 + \ldots + 0*\mu_{k-1} + (-1)*\mu_k = l_1'\mu$$

$$\Rightarrow l_1' = (100\ldots0-1)$$

$$\mu_2 - \mu_k = 0*\mu_1 + 1*\mu_2 + 0*\mu_3 + \ldots + 0*\mu_{k-1} + (-1)*\mu_k = l_2'\mu$$

$$\Rightarrow l_2' = (010\ldots0-1)$$

....

$$\mu_{k-1} - \mu_k = 0*\mu_1 + 0*\mu_2 + 0*\mu_3 + \ldots 1*\mu_{k-1} + (-1)*\mu_k = l_{k-1}'\mu$$

$$\Rightarrow l_{k-1}' = (000\ldots1-1)$$

Μοντελοποίηση

Τα διανύσματα $l_1', l_2', \ldots, l_{k-1}'$ είναι ανεξάρτητα.

Άρα τα διανύσματα

$$l_1'\mu, l_2'\mu, \ldots, l_{k-1}'\mu$$

είναι ανεξάρτητες εκτιμήσιμες συναρτήσεις.

Γραμμικές εκτιμήσιμες συναρτήσεις

Πρόταση:

Δίνεται το υπόδειγμα $Y = X\beta + \varepsilon$. Τότε οι παρακάτω προτάσεις είναι ισοδύναμες:

- Η γραμμική συνάρτηση $l'\mu$ είναι εκτιμήσιμη
- Το διάνυσμα l_{kx1} είναι γραμμικός συνδυασμός των στηλών της X΄
- Ο γραμμικός συνδυασμός δίνεται από: $l = X_{kxn}\lambda_{nx1}, \quad \lambda \in R^n$

Μοντελοποίηση

Θεώρημα:

Δίνεται το σύστημα κανονικών εξισώσεων

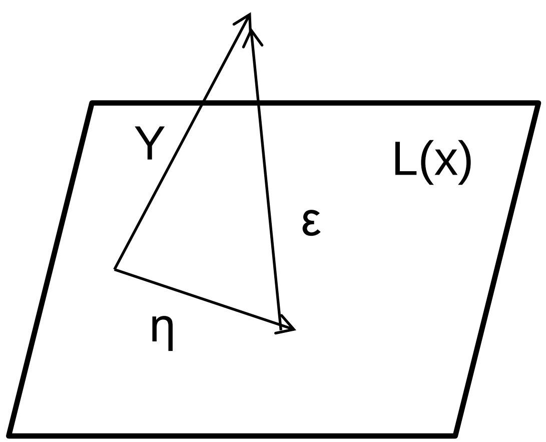
$Y=X\beta+\varepsilon$ με συνάρτηση $S(\beta)=(Y-X\beta)'(Y-X\beta)$. Οι παρακάτω προτάσεις είναι ισοδύναμες

1. Το διάνυσμα **β** είναι λύση του συστήματος των κανονικών εξισώσεων
2. Το διάνυσμα **β** ελαχιστοποιεί την συνάρτηση S

Μοντελοποίηση

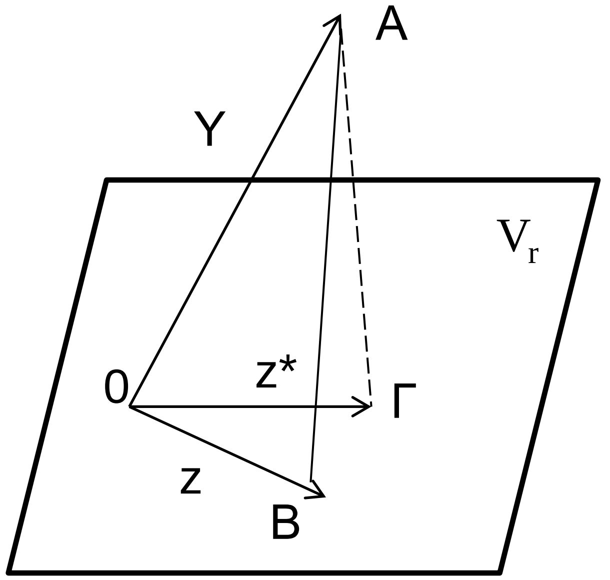
Γραφική παρουσίαση του ΓΓΜ **Υ=Χβ+ε**

Το Χβ είναι γραμμικός συνδυασμός των στηλών της μήτρας Χ άρα Χβ $\in$L(X) (πεδίο ορισμού της συνάρτησης). Αν συμβολίσουμε με η το διάνυσμα Χβ παρατηρούμε ότι η $\in$L(X) με Υ=η+ε



Μοντελοποίηση

Πρόταση: Ας θεωρήσουμε ένα διάνυσμα $Y \in R^n$ και έστω V_r ένας υπόχωρος του R^n. Σχηματίζουμε την συνάρτηση f με τύπο $f(z)=(V-z)'(V-z)$ με $z \in V_r$. Τότε η f παρουσιάζει ελάχιστο για z=z*, μόνο αν z* είναι η ορθή προβολή του Y πάνω στο V_r.



Παρατηρούμε ότι

$f(z)=(Y-z)'(Y-z)=||Y-z||^2$

(μήκος του $\overrightarrow{BA}$)^2

Το μήκος του $\overrightarrow{BA}$ γίνεται ελάχιστο αν αυτό συμπέσει με το $\overrightarrow{\Gamma A}$, οπότε το z=z*

Μοντελοποίηση

Αν θέσουμε

$$C_{(k-1)xk} = \begin{bmatrix} l_1' \\ l_2' \\ \vdots \\ l_{k-1}' \end{bmatrix} = \begin{bmatrix} 1*\mu_1 + 0*\mu_2 + \ldots + 0*\mu_{k-1} + (-1)\mu_k \\ 0*\mu_1 + 1*\mu_2 + \ldots + 0*\mu_{k-1} + (-1)\mu_k \\ \vdots \\ 0*\mu_1 + 0*\mu_2 + \ldots 1*\mu_{k-1} + (-1)\mu_k \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} l_1' \\ l_2' \\ \vdots \\ l_{k-1}' \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow \boxed{\mathbf{C}\boldsymbol{\mu} = \mathbf{0}}$$

Μοντελοποίηση

Έλεγχος υποθέσεων Cμ=0

- ✓ Επιλέγουμε επίπεδο σημαντικότητας α
- ✓ Επιλύουμε ως προς x_0 την εξίσωση

$$P\left\{F^* > x_0\right\} = a$$

όπου $F^* \sim F_{k-1,\, n-k}$

- ✓ Υπολογίζουμε την ποσότητα

$$F^* = \frac{\dfrac{\left(R_1^2 - R_0^2\right)}{k-1}}{\dfrac{R_0^2}{n-k}}$$

$$F^* = \frac{MST_r}{MSE} \sim F_{(k-1),(n-k)}$$

$$F^* = \frac{s_b^2}{s_w^2} = \frac{\text{μέσο άθροισμα των τετραγώνων των επεμβάσεων } (MST_r)}{\text{Μέσο άθροισμα τετραγώνων σφάλματος } (MSE)}$$

Μοντελοποίηση

Η ποσότητα

$$F^* = \frac{\dfrac{\left(R_1^2 - R_0^2\right)}{k-1}}{\dfrac{R_0^2}{n-k}}$$

$$R_0^2 = \min\left(Y - X\beta\right)'\left(Y - X\beta\right)$$

$$R_1^2 = \min\left(Y - X\beta\right)'\left(Y - X\beta\right)\Big|_{C\mu=0}$$

- ✓ Αν $F^*>x_0$ τότε απορρίπτουμε την υπόθεση H_0
- ✓ Αν $F^*<x_0$ τότε δεν απορρίπτουμε την υπόθεση H_0

όπου Cμ=0 ερμηνεύεται ως β=0 άρα οι παράμετροι β μέσα στο γραμμικό μοντέλο εξηγούνται από γραμμικούς συνδυασμούς των μέσων κάθε δείγματος.

Μοντελοποίηση

$$F^* = \frac{MST_r}{MSE} = \frac{\frac{SST_r}{k-1}}{\frac{SSE}{n-k}} = \frac{\frac{\sum_j n_j \left(\bar{y}_j - \bar{y}\right)^2}{k-1}}{\frac{\sum_{ij} \left(y_{ij} - \bar{y}_j\right)^2}{n-k}}$$

$$F^* = \frac{\frac{\left(R_1^2 - R_0^2\right)}{k-1}}{\frac{R_0^2}{n-k}}$$

$$R_1^2 - R_0^2 = \sum_j n_j \left(\bar{y}_j - \bar{y}\right)^2$$

$$R_0^2 = \sum_{ij} \left(y_{ij} - \bar{y}_j\right)^2$$

Μοντελοποίηση

Έχουμε την ποσότητα

$$R_0^2 = \left(Y - X\widehat{\beta}\right)'\left(Y - X\widehat{\beta}\right) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \widehat{\beta}_i\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \widehat{\mu}_i\right)^2$$

όπου $\widehat{\beta}$ είναι η λύση του συστήματος των κανονικών εξισώσεων ($\widehat{\beta} = \widehat{\mu}$)

Αν θέσουμε

$$S(\beta) = S(\beta_1, \beta_2, \cdots, \beta_k) = \left(Y - X\widehat{\beta}\right)'\left(Y - X\widehat{\beta}\right) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \beta_i\right)^2 =$$

$$= \sum_{j=1}^{n_1}\left(Y_{1j} - \beta_1\right)^2 + \sum_{j=1}^{n_2}\left(Y_{2j} - \beta_2\right)^2 + \cdots + \sum_{j=1}^{n_k}\left(Y_{kj} - \beta_k\right)^2$$

Μοντελοποίηση

Σύστημα κανονικών εξισώσεων:

$$\frac{\partial S}{\partial \beta_1} = -2\sum_{j=1}^{n_1}\left(Y_{1j} - \beta_1\right) = 0 \Rightarrow n_1\beta_1 = \sum_{j=1}^{n_1} Y_{1j}$$

$$\frac{\partial S}{\partial \beta_2} = -2\sum_{j=1}^{n_2}\left(Y_{2j} - \beta_2\right) = 0 \Rightarrow n_2\beta_2 = \sum_{j=1}^{n_2} Y_{2j}$$

$$\vdots$$

$$\frac{\partial S}{\partial \beta_k} = -2\sum_{j=1}^{n_k}\left(Y_{kj} - \beta_k\right) = 0 \Rightarrow n_k\beta_k = \sum_{j=1}^{n_k} Y_{kj}$$

Μοντελοποίηση

Εκτιμήσεις

$$\widehat{\beta}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{1j} = \bar{Y}_{1.} = \widehat{\mu}_1$$

$$\widehat{\beta}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{2j} = \bar{Y}_{2.} = \widehat{\mu}_2$$

$$\vdots$$

$$\widehat{\beta}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} Y_{kj} = \bar{Y}_{k.} = \widehat{\mu}_k$$

$$\Rightarrow \quad \widehat{\beta} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \vdots \\ \bar{Y}_{k.} \end{bmatrix}$$

Μοντελοποίηση

Επομένως η R_0^2 παίρνει την μορφή

$$R_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i.}\right)^2$$

Άθροισμα τετραγώνων των αποκλίσεων

Άθροισμα τετραγώνων μέσα στα δείγματα

Μοντελοποίηση

Ας θέσουμε μ^* την κοινή τιμή των μ_i, i=1,2,…,k κάτω από την H_0, δηλ: $\mu_1= \mu_2=\ldots= \mu_k= \mu^*$ ή $\beta_1= \beta_2=\ldots= \beta_k= \beta^*$

Έχουμε την ποσότητα

$$R_1^2 = \min_{C\mu=0}\left(Y - X\widehat{\beta}^*\right)'\left(Y - X\widehat{\beta}^*\right)$$

όπου $\widehat{\beta}^*$ είναι η ελάχιστων τετραγώνων εκτιμήτρια της παραμέτρου β κάτω από τους περιορισμούς Cμ=0 (Cβ=0)

Αν θέσουμε

$$S'\left(\beta^*\right) = S'\left(\beta^*\right)_{C\beta=0} = \left(Y - X\widehat{\beta}^*\right)'\left(Y - X\widehat{\beta}^*\right) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \beta^*\right)^2 =$$

$$= \sum_{j=1}^{n_1}\left(Y_{1j} - \beta^*\right)^2 + \sum_{j=1}^{n_2}\left(Y_{2j} - \beta^*\right)^2 + \cdots + \sum_{j=1}^{n_k}\left(Y_{kj} - \beta^*\right)^2$$

Μοντελοποίηση

Σύστημα κανονικών εξισώσεων:

$$\frac{\partial S'}{\partial \beta^*} = -2\sum_{j=1}^{n_1}\left(Y_{1j} - \beta^*\right) = 0 \Rightarrow n_1\beta^* = \sum_{j=1}^{n_1} Y_{1j}$$

$$\frac{\partial S'}{\partial \beta^*} = -2\sum_{j=1}^{n_1}\left(Y_{2j} - \beta^*\right) = 0 \Rightarrow n_2\beta^* = \sum_{j=1}^{n_2} Y_{2j}$$

$$\vdots$$

$$\frac{\partial S'}{\partial \beta^*} = -2\sum_{j=1}^{n_k}\left(Y_{kj} - \beta^*\right) = 0 \Rightarrow n_k\beta^* = \sum_{j=1}^{n_k} Y_{kj}$$

$$\Downarrow$$

$$\left(n_1 + n_2 + \cdots + n_k\right)\beta^* = \sum_{j=1}^{n_1} Y_{1j} + \sum_{j=1}^{n_2} Y_{2j} + \cdots + \sum_{j=1}^{n_k} Y_{kj} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}$$

Μοντελοποίηση

Εκτιμήσεις

$$\widehat{\beta}^* = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij} = \overline{Y} \quad \Rightarrow \quad \widehat{\beta}^* = \begin{bmatrix} \overline{Y} \\ \overline{Y} \\ \vdots \\ \overline{Y} \end{bmatrix}$$

$$\Downarrow$$

$$\widehat{\beta}_1^* = \overline{Y}, \widehat{\beta}_2^* = \overline{Y}, \cdots, \widehat{\beta}_k^* = \overline{Y}$$

Εκτιμήτριες ελαχίστων τετραγώνων

Μοντελοποίηση

Επομένως η R_1^2 παίρνει την μορφή

$$R_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}\right)^2$$

Ολικό άθροισμα τετραγώνων

$$R_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i.}\right)^2$$

$$R_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}\right)^2$$

$$R_1^2 - R_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}\right)^2 - \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i.}\right)^2$$

Άθροισμα τετραγώνων εκτός δειγμάτων

Μοντελοποίηση

Ως προς τις μορφές των μέσων έχουμε

$$\bar{Y} = \frac{1}{n}\sum_{i}^{k}\sum_{j}^{n_i} Y_{ij} \longrightarrow \text{Μεγάλος μέσος}$$

$$\bar{Y}_{i.} = \frac{1}{n_i}\sum_{j}^{n_i} Y_{ij} \longrightarrow \text{Μικρός μέσος}$$

Μοντελοποίηση

Μπορούμε να γράψουμε:

$$Y_{ij} - \overline{Y} = \left(Y_{ij} - \overline{Y}_{i.}\right) + \left(\overline{Y}_{i.} - \overline{Y}\right)$$

$$\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}\right)^2 = \sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i.}\right)^2 + \sum_{j=1}^{n_i}\left(\overline{Y}_{i.} - \overline{Y}\right)^2 + 2\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i.}\right)\left(\overline{Y}_{i.} - \overline{Y}\right)$$

$$\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i.}\right)\left(\overline{Y}_{i.} - \overline{Y}\right) = \left(\overline{Y}_{i.} - \overline{Y}\right)\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i.}\right) = \left(\overline{Y}_{i.} - \overline{Y}\right) * 0 = 0$$

Μοντελοποίηση

Μπορούμε να γράψουμε:

$$\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}\right)^2=\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{i.}\right)^2+n_i\left(\bar{Y}_{i.}-\bar{Y}\right)^2\Rightarrow$$

$$\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}\right)^2}_{R_1^2}=\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{i.}\right)^2}_{R_0^2}+\sum_{i=1}^{k}n_i\left(\bar{Y}_{i.}-\bar{Y}\right)^2$$

$$R_1^2=R_0^2+\sum_{i=1}^{k}n_i\left(Y_{i.}-\bar{Y}\right)^2\Rightarrow R_1^2-R_0^2=\sum_{i=1}^{k}n_i\left(\bar{Y}_{i.}-\bar{Y}\right)^2$$

Μοντελοποίηση

Πίνακας Ανάλυσης Διακύμανσης

Πηγή μεταβλητότητας	**Άθροισμα τετραγώνων**	**Βαθμοί ελευθερίας**	**Μέσο άθροισμα τετραγώνων**	**F**
Απόκλιση από την H_0 (μεταξύ δειγμάτων)	$R_1^2 - R_0^2$	k-1	$\frac{R_1^2 - R_0^2}{k-1}$	$F^* = \frac{\frac{R_1^2 - R_0^2}{k-1}}{\frac{R_0^2}{n-k}}$
Κατάλοιπα (εντός δειγμάτων)	R_0^2	n-k	$\frac{R_0^2}{n-k}$	
Σύνολο	R_1^2	n-1		

Μοντελοποίηση

Απλοποιήσεις

$$R_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{1}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}\right)^2$$

Απόδειξη

$$R_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij}^2 - 2\overline{Y}Y_{ij} + \overline{Y}^2\right) = \sum_{i}\sum_{j} Y_{ij}^2 - 2\overline{Y}\left(\sum_{i}\sum_{j} Y_{ij}\right) + n\overline{Y}^2 =$$

$$= \sum_{i}\sum_{j} Y_{ij}^2 - 2n\overline{Y}^2 + n\overline{Y}^2 = \sum_{i}\sum_{j} Y_{ij}^2 - \frac{1}{n}\left(\sum_{i}\sum_{j} Y_{ij}\right)^2$$

Μοντελοποίηση

Απλοποιήσεις

$$R_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i.}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{n_i} Y_{ij}\right)^2$$

Απόδειξη

$$R_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i.}\right)^2 = \sum_{j}\left(Y_{1j} - \bar{Y}_{1.}\right)^2 + \cdots \sum_{j}\left(Y_{kj} - \bar{Y}_{k.}\right)^2 =$$

$$= \sum_{j} Y_{1j}^2 - \frac{1}{n_1}\left(\sum_{j} Y_{1j}\right)^2 + \cdots + \sum_{j} Y_{kj}^2 - \frac{1}{n_k}\left(\sum_{j} Y_{kj}\right)^2 =$$

$$= \sum_{i}\sum_{j} Y_{ij}^2 - \left\{\frac{1}{n_1}\left(\sum_{j} Y_{1j}\right)^2 + \cdots + \frac{1}{n_k}\left(\sum_{j} Y_{kj}\right)^2\right\} =$$

$$= \sum_{i}\sum_{j} Y_{ij}^2 - \sum_{i}\frac{1}{n_i}\left(\sum_{j} Y_{ij}\right)^2$$

Μοντελοποίηση

Απλοποιήσεις

$$R_1^2 - R_0^2 = \sum_{i=1}^{k} \frac{1}{n_i} \left(\sum_{j=1}^{n_i} Y_{ij} \right)^2 - \frac{1}{n} \left(\sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij} \right)^2$$

Μοντελοποίηση

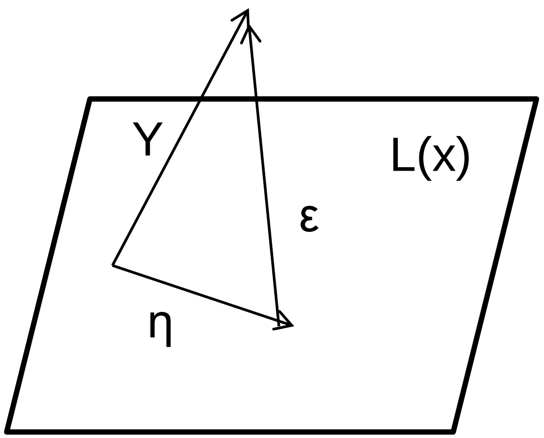
Πίνακας Υπολογισμού Διακύμανσης

# Δείγμα	Δείγμα	παρατηρήσεις	Άθροισμα παρατηρήσεων	Μέσο τετραγωνικό άθροισμα	Άθροισμα τετραγώνων	Μέση τιμή
1	$Y_{11}, Y_{12},\ldots, Y_{1n1}$	n_1	$\sum_j Y_{1j}$	$\frac{1}{n_1}\left(\sum_j Y_{1j}\right)^2$	$\sum_j Y_{1j}^2$	$\bar{Y}_{1.}$
2	$Y_{21}, Y_{22},\ldots, Y_{2n2}$	n_2	$\sum_j Y_{2j}$	$\frac{1}{n_2}\left(\sum_j Y_{2j}\right)^2$	$\sum_j Y_{2j}^2$	$\bar{Y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$Y_{k1}, Y_{k2},\ldots, Y_{knk}$	n_k	$\sum_j Y_{kj}$	$\frac{1}{n_k}\left(\sum_j Y_{kj}\right)^2$	$\sum_j Y_{kj}^2$	$\bar{Y}_{k.}$
Σύνολο		n	$\sum_i \sum_j Y_{ij}$	$\sum_i \frac{1}{n_i}\left(\sum_j Y_{ij}\right)^2$	$\sum_i \sum_j Y_{ij}^2$	$\bar{Y}$

Γεωμετρική ερμηνεία του F-test

Γραφική παρουσίαση του ΓΓΜ **Y=Xβ+ε**

Το Xβ είναι γραμμικός συνδυασμός των στηλών της μήτρας X άρα Xβ $\in$L(X) (πεδίο ορισμού της συνάρτησης). Αν συμβολίσουμε με η το διάνυσμα Xβ παρατηρούμε ότι η $\in$L(X) με Y=η+ε

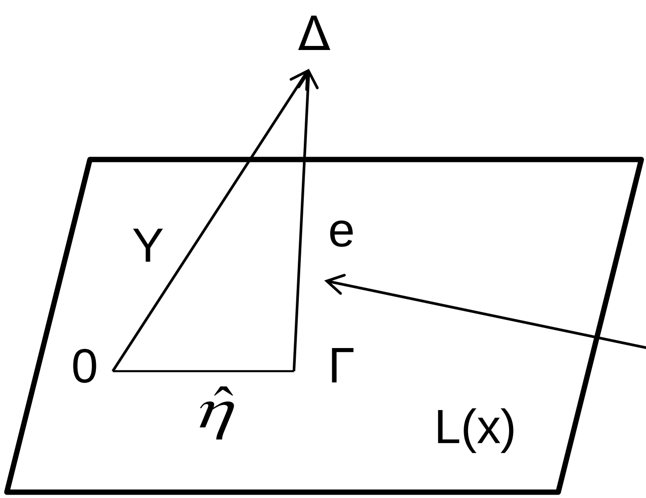


Γεωμετρική ερμηνεία του F-test

Αν $\hat{\eta}$ είναι η ορθή προβολή του Y πάνω στο χώρο L(X) τότε θα έχουμε

$$Y = \hat{\eta} + e$$

όπου $e = Y - \hat{\eta} = Y - X\hat{\beta}$ είναι το διάνυσμα των καταλοίπων (residuals), και $\hat{\beta}$ είναι μια εκτιμήτρια ελαχίστων τετραγώνων της παραμέτρου β.



$$R_0^2 = e'e$$

Τότε $\hat{\eta} \perp e$ και

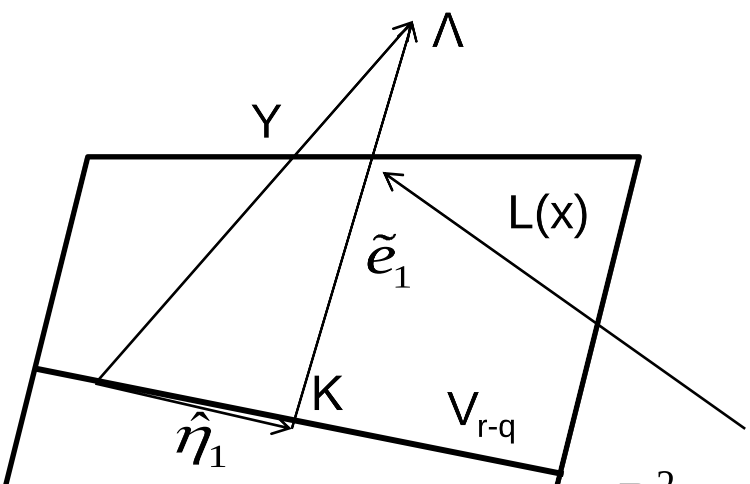
$$R_0^2 = e'e = (\mu\acute{\eta}\kappa o\varsigma \text{ του } \overrightarrow{\Gamma\Delta})^2 = \left\| \overrightarrow{\Gamma\Delta} \right\|^2$$

Γεωμετρική ερμηνεία του F-test

Κάτω από την υπόθεση $C\mu=\omega_0$ το Y προβάλλεται σε ένα υπόχωρο του L(X) r-q διαστάσεων V_{r-q}

Αν $\hat{\eta}_1$ η ορθή προβολή του Y στον υπόχωρο V_{r-q} τότε τα αντίστοιχα κατάλοιπα (residuals) θα είναι

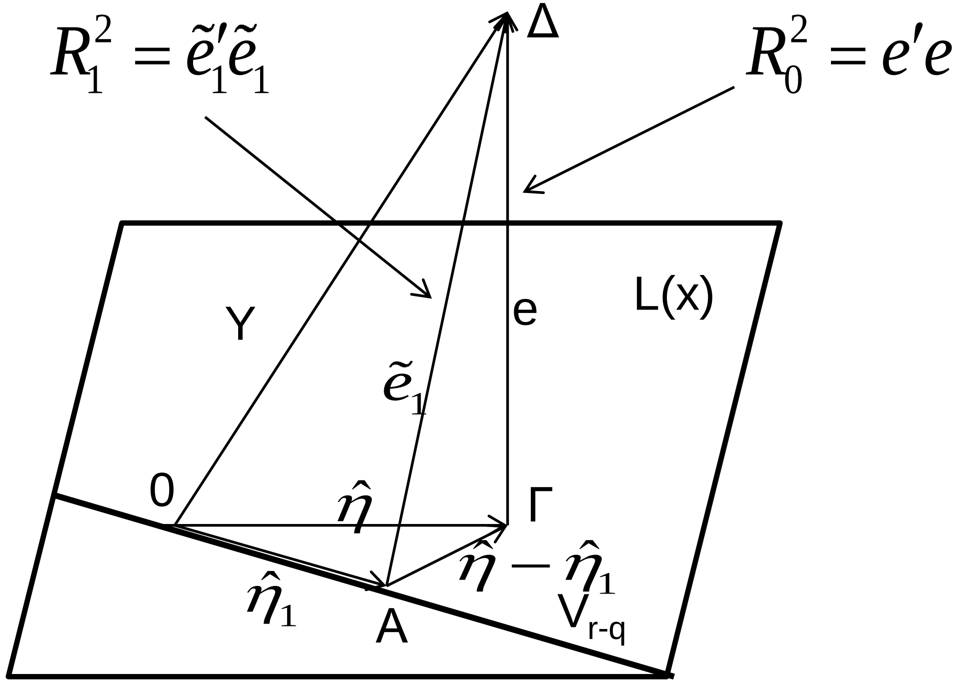
$$\tilde{e}_1 = Y - X\tilde{\beta}$$



όπου $\tilde{\beta}$ θέση όπου ελαχιστοποιείται η (Y-Xβ)'(Y-Xβ)

$$R_1^2 = \tilde{e}_1'\tilde{e}_1 = (\mu\acute{\eta}\kappa o\varsigma \text{ του } \overrightarrow{\mathrm{K\Lambda}})^2 = \left\|\overrightarrow{\mathrm{K\Lambda}}\right\|^2$$

Γεωμετρική ερμηνεία του F-test



$$\overrightarrow{\mathrm{O\Delta}} = \mathrm{Y}$$

$$\overrightarrow{\mathrm{O\Gamma}} = \hat{\eta}$$

$$\overrightarrow{\mathrm{OA}} = \hat{\eta}_1$$

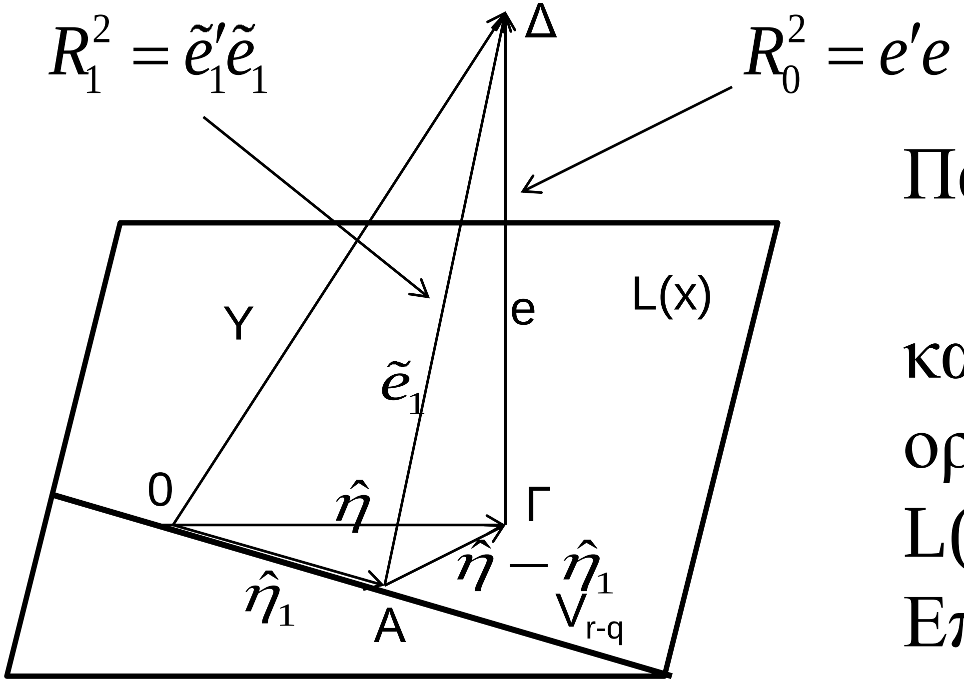
$$\overrightarrow{\mathrm{A\Delta}} = \tilde{e}_1 = \mathrm{Y} - \hat{\eta}_1$$

$$\overrightarrow{\mathrm{\Gamma\Delta}} = \tilde{e} = \mathrm{Y} - \hat{\eta}$$

$$\overrightarrow{\mathrm{A\Gamma}} = \hat{\eta} - \hat{\eta}_1$$

$$R_1^2 - R_0^2 = \left\| \hat{\eta} - \hat{\eta}_1 \right\|^2$$

Γεωμετρική ερμηνεία του F-test



Παρατηρούμε ότι

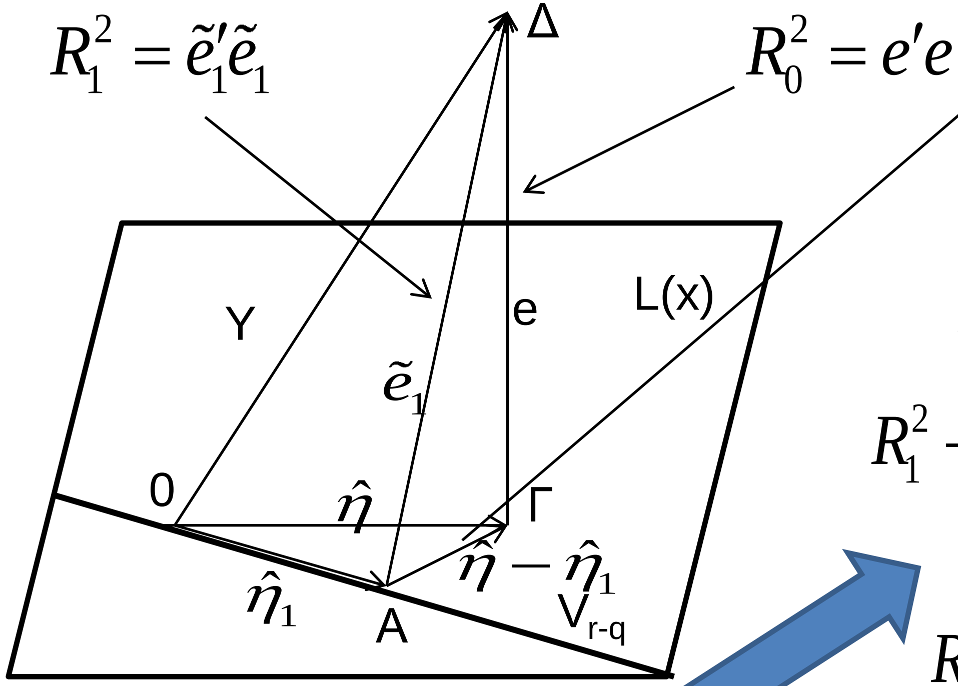
$$e \perp \hat{\eta} - \hat{\eta}_1 \in L(X)$$

και το e ανήκει στο ορθογώνιο σύμπλεγμα του L(X).

Επίσης έχουμε

$$\tilde{e}_1 = (\hat{\eta} - \hat{\eta}_1) - e$$

Γεωμετρική ερμηνεία του F-test



Όσο μεγαλύτερο το μήκος του $\overrightarrow{\text{ΑΓ}}$ τόσο απορρίπτουμε την $\mathbf{H_0}$

$$R_1^2 = R_0^2 + (\mu\acute{\eta}\kappa o\varsigma \text{ του } \overrightarrow{\text{ΑΓ}})^2$$

$$R_1^2 - R_0^2 = (\mu\acute{\eta}\kappa o\varsigma \text{ του } \overrightarrow{\text{ΑΓ}})^2 = \left\|\hat{\eta} - \hat{\eta}_1\right\|^2$$

$$\frac{R_1^2 - R_0^2}{q}$$ αριθμητής της F-κατανομής

Άρα

$$R_1^2 = \tilde{e}_1^{'} - \tilde{e}_1 = \left[e + (\hat{\eta} - \hat{\eta}_1)\right]^{'}\left[e + (\hat{\eta} - \hat{\eta}_1)\right] =$$

$$e'e + e'(\hat{\eta} - \hat{\eta}_1) + (\hat{\eta} - \hat{\eta}_1)'e + (\hat{\eta} - \hat{\eta}_1)'(\hat{\eta} - \hat{\eta}_1) =$$

$$e'e + (\hat{\eta} - \hat{\eta}_1)'(\hat{\eta} - \hat{\eta}_1) = R_0^2 + \left\|\hat{\eta} - \hat{\eta}_1\right\|^2$$

Παραδείγματα

Διαθέτουμε 3 είδη χοιριδίων. Μελετάμε αν το είδος της τροφής επιδρά στην αύξηση του βάρους.

Ένα σύνολο 12 χοιριδίων χωριστήκαν τυχαία σε 3 ισοπληθείς ομάδες και σε κάθε μια ομάδα δόθηκε ένα είδος τροφής για ένα χρονικό διάστημα στο τέλος του οποίου μετρήθηκε η αύξηση του βάρους καθενός από τα 12 χοιρίδια.

Παραδείγματα

Ομάδα χοιριδίων	Αύξηση βάρους			
1	1.42	1.28	1.37	1.37
2	1.4	1.33	1.41	1.42
3	1.17	1.39	1.4	1.37

Υποθέτουμε ότι οι παρατηρήσεις είναι κανονικές, ανεξάρτητες και προέρχονται από κανονικό πληθυσμό.
Θέλουμε να ελέγξουμε την υπόθεση της ισότητας των μέσων έναντι της εναλλακτικής

Παραδείγματα

Ομάδα χοιριδίων					n_i	$\sum_j Y_{ij}$	$\frac{1}{n_i}\left(\sum_j Y_{ij}\right)^2$	$\sum_j Y_{ij}^2$	$\bar{Y}_{i.}$
1	1.42	1.28	1.37	1.37	4	5.44	7.398	7.408	1.36
2	1.4	1.33	1.41	1.42	4	5.66	7.728	7.733	1.39
3	1.17	1.39	1.4	1.37	4	5.33	7.102	7.137	1.33
	Σύνολο				12	16.33	22.229	22.779	1.36

Παραδείγματα

$$R_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{1}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}\right)^2 = 22.279 - \frac{1}{12}(16.33)^2 = 0.0574$$

$$R_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{n_i} Y_{ij}\right)^2 = 22.279 - 22.229 = 0.0509$$

$$R_1^2 - R_0^2 = 0.0574 - 0.0509 = 0.000659$$

Παραδείγματα

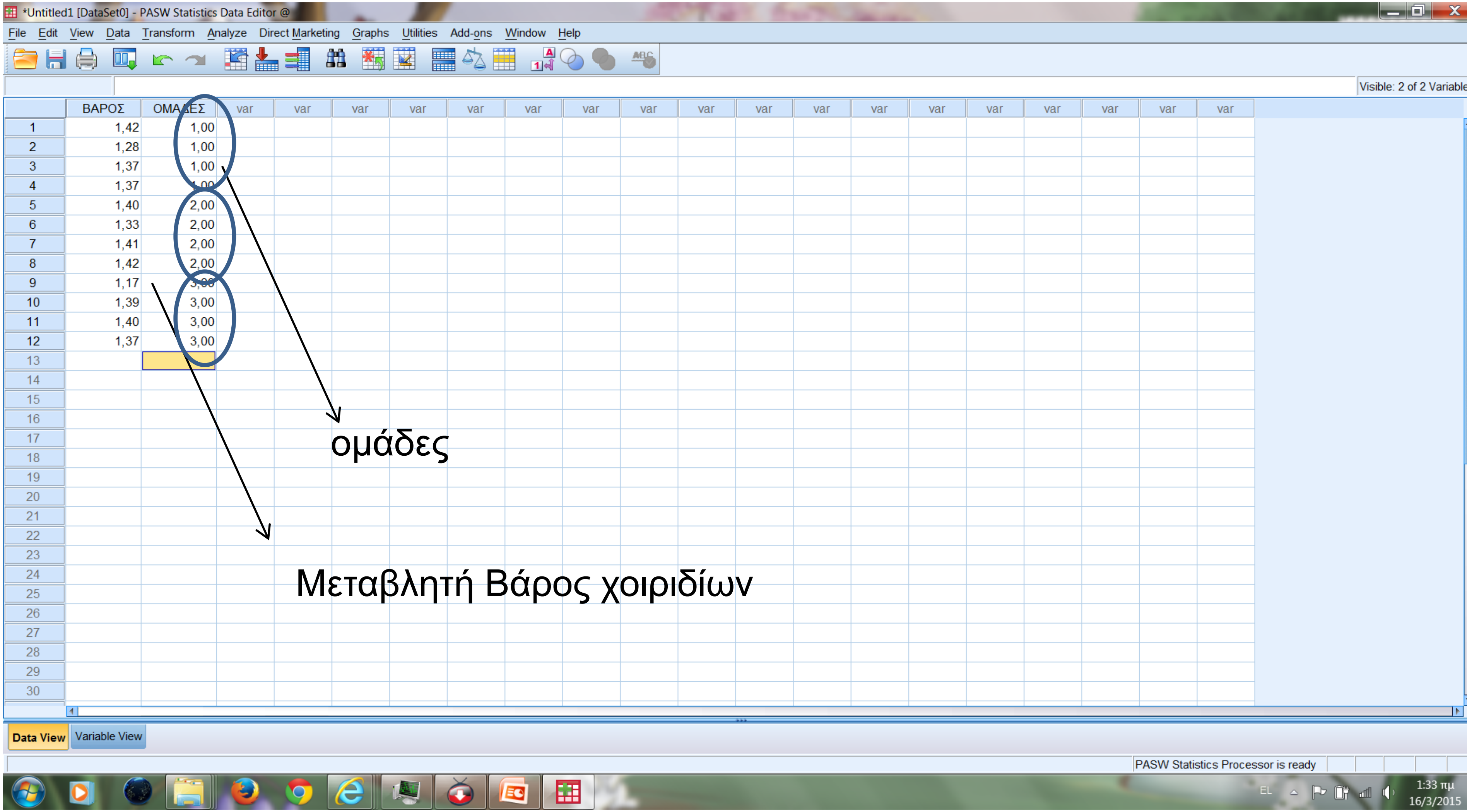
Πίνακας Ανάλυσης Διακύμανσης

Πηγή μεταβλητότητας	**Άθροισμα τετραγώνων**	**Βαθμοί ελευθερίας**	**Μέσο άθροισμα τετραγώνων**	**F**
Απόκλιση από την H_0 (μεταξύ δειγμάτων)	0.00659	2	0.00329	0.583
Κατάλοιπα (εντός δειγμάτων)	0.0509	9	0.00565	
Σύνολο	0.0574	11		

Με επίπεδο σημαντικότητας α=0.05,παρατηρούμε ότι η τιμή της συνάρτησης $F_{2,9,0.05}=4.26$

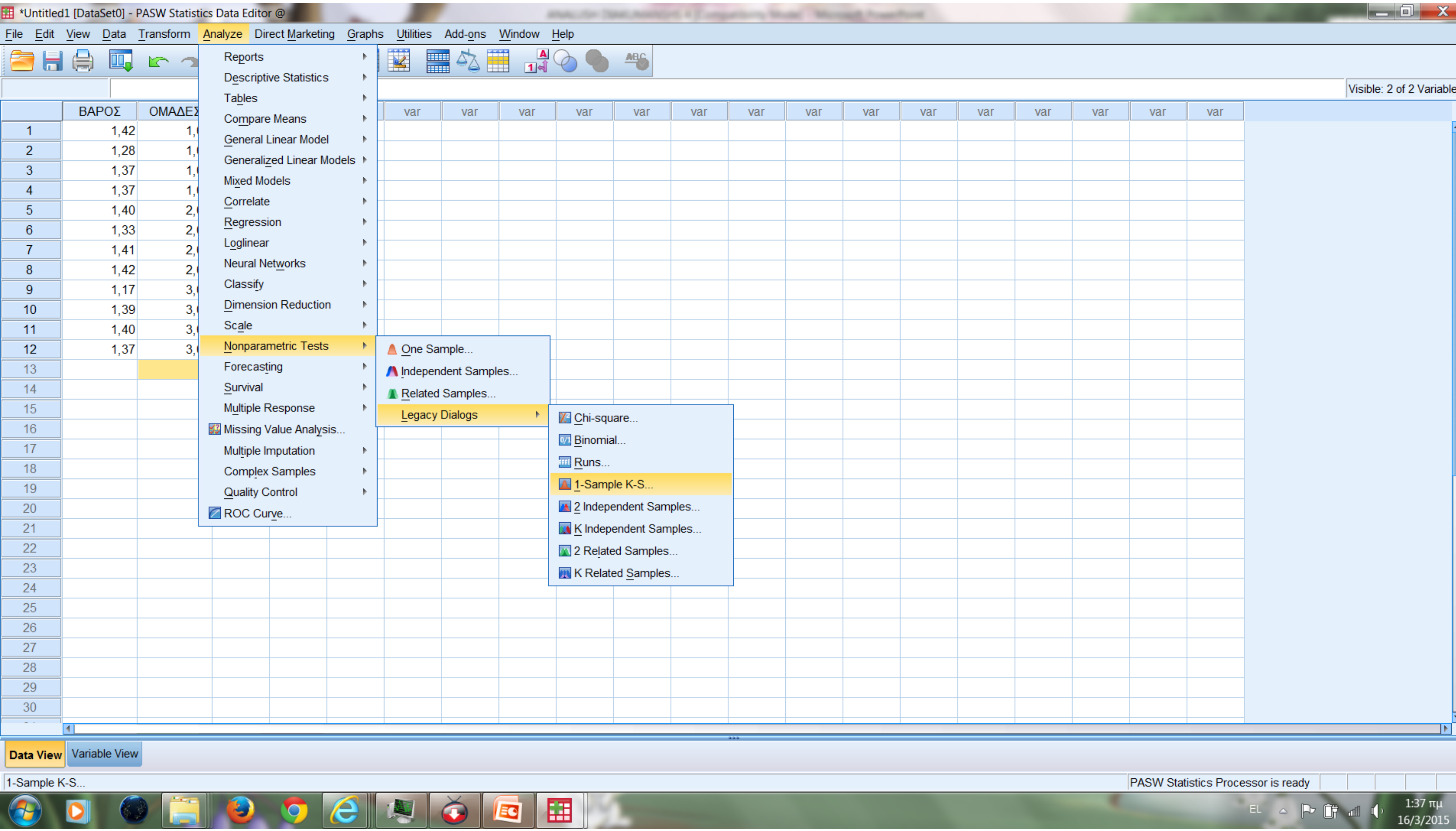
$4.26>0.583$ άρα αποδέχομαι την αρχική υπόθεση H_0

Παραδείγματα

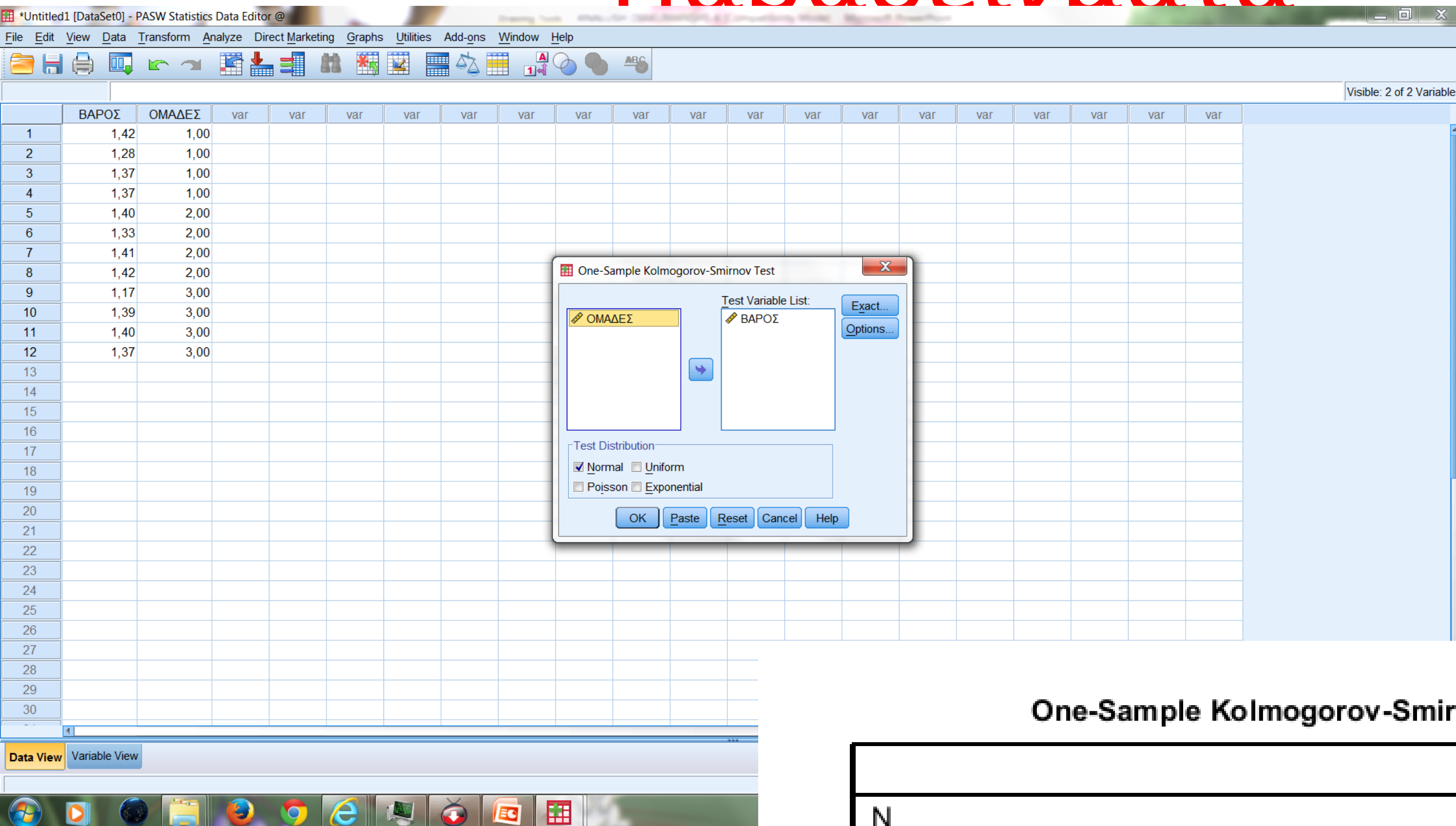


	ΒΑΡΟΣ	ΟΜΑΔΕΣ
1	1,42	1,00
2	1,28	1,00
3	1,37	1,00
4	1,37	1,00
5	1,40	2,00
6	1,33	2,00
7	1,41	2,00
8	1,42	2,00
9	1,17	3,00
10	1,39	3,00
11	1,40	3,00
12	1,37	3,00

Παραδείγματα



Παραδείγματα



One-Sample Kolmogorov-Smirnov Test

		ΒΑΡΟΣ
N		12
Normal Parameters [a,b]	Mean	1,3608
	Std. Deviation	,07229
Most Extreme Differences	Absolute	,300
	Positive	,207
	Negative	-,300
Kolmogorov-Smirnov Z		1,041
Asymp. Sig. (2-tailed)		,229

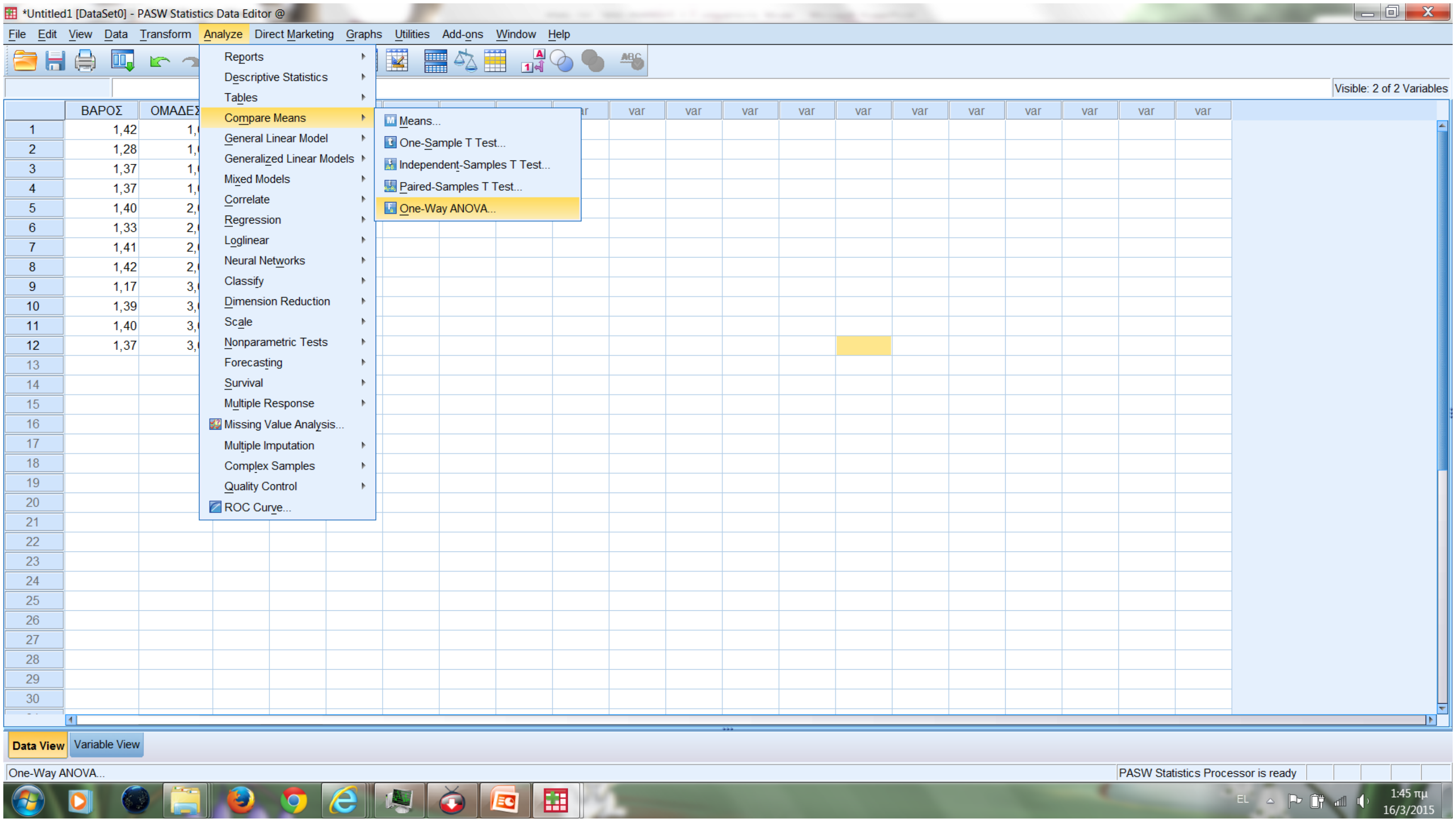
a. Test distribution is Normal.

b. Calculated from data.

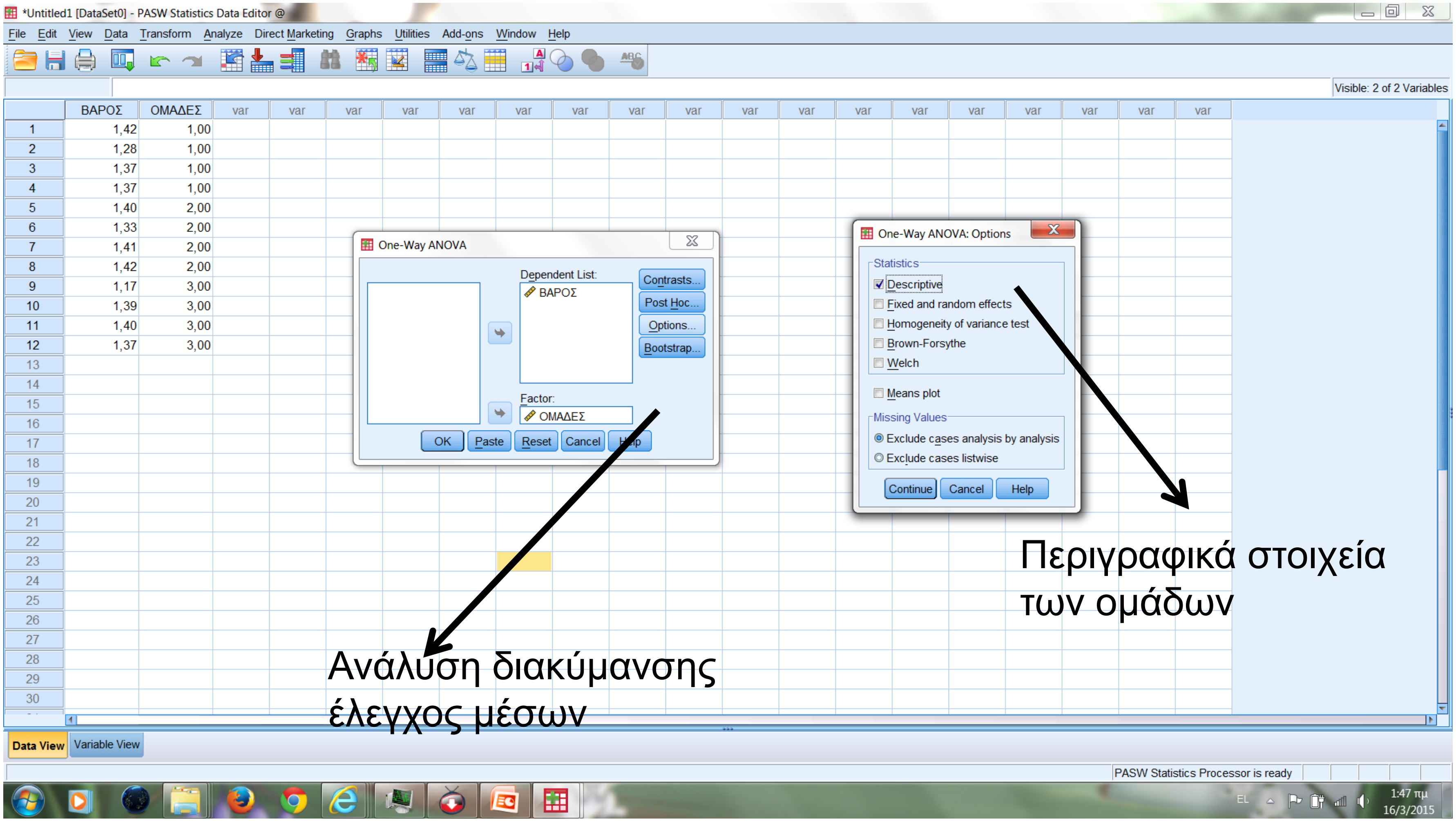
$p>0.05$, άρα δεν υπάρχει στατιστικά σημαντική διαφορά

Αποδέχομαι την H_0 κανονική κατανομή

Παραδείγματα



Παραδείγματα



Ανάλυση διακύμανσης
έλεγχος μέσων

Περιγραφικά στοιχεία
των ομάδων

Παραδείγματα

Descriptives

ΒΑΡΟΣ

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1,00	4	1,3600	,05831	,02915	1,2672	1,4528
2,00	4	1,3900	,04082	,02041	1,3250	1,4550
3,00	4	1,3325	,10905	,05452	1,1590	1,5060
Total	12	1,3608	,07229	,02087	1,3149	1,4068

$Y_{i.} \Rightarrow$ έλεγχος υποθέσεων H_0

ANOVA

ΒΑΡΟΣ

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,007	2	,003	,585	,577
Within Groups	,051	9	,006		
Total	,057	11			

p=0.577>0.05
Αποδέχομαι την H_0
Δεν διαφέρουν σημαντικά οι μέσοι