

## Περιγραφική στατιστική των 2 μεταβλητών

Μαρκαρουμε τις 2 μεταβλητες(income, food\_exp)→Enter→View→Descriptive Stats→common sample.

	FOOD_EXP	INCOME
Mean	283.5735	19.60475
Median	264.4800	20.03000
Maximum	587.6600	33.40000
Minimum	109.7100	3.690000
Std. Dev.	112.6752	6.847773
Skewness	0.492083	-0.626507
Kurtosis	2.851522	3.279728
Jarque-Bera	1.651045	2.747156
Probability	0.438006	0.253199
Sum	11342.94	784.1900
Sum Sq. Dev.	495132.2	1828.788
Observations	40	40

## Αποτελέσματα εκτίμησης γραμμικής παλινδρομησης

Is food\_exp c income

Dependent Variable: FOOD\_EXP

Method: Least Squares

Date: 10/19/22 Time: 19:16

Sample: 1 40

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	83.41600	43.41016	1.921578	0.0622
INCOME	10.20964	2.093264	4.877381	0.0000
R-squared	0.385002	Mean dependent var		283.5735
Adjusted R-squared	0.368818	S.D. dependent var		112.6752
S.E. of regression	89.51700	Akaike info criterion		11.87544
Sum squared resid	304505.2	Schwarz criterion		11.95988
Log likelihood	-235.5088	Hannan-Quinn criter.		11.90597
F-statistic	23.78884	Durbin-Watson stat		1.893880
Prob(F-statistic)	0.000019			

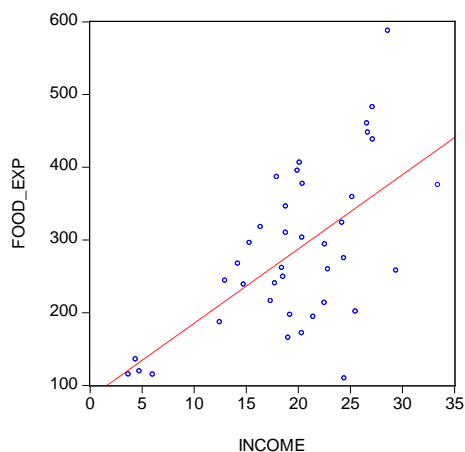
Για να αποθηκεύσουμε τα αποτελέσματα της εκτίμησης, κλικάρουμε “name” → Δίνουμε ένα όνομα, πχ. eq01 → OK

Αποθηκεύουμε τα καταλοιπα

series ehat=resid

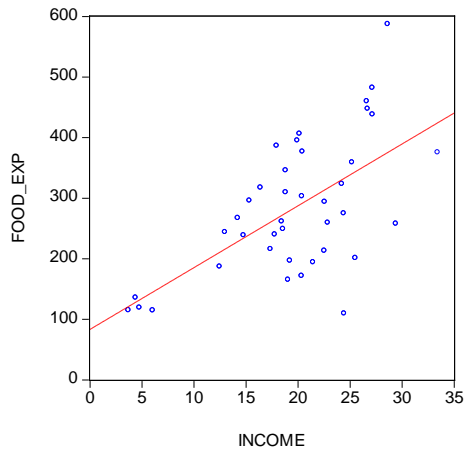
Απεικόνιση διαγραμματική της δειγματικής γραμμής παλινδρομησης μαζί με το νεφος των παρατηρησεων (διαγραμμα διασπορας)

Quick → graph → income food\_exp → OK → Επιλεγουμε “scatter” και στην επιλογή “fit lines” επιλεγουμε “Regression line” → OK



Καποιες επιπλεον βελτιωσεις στο ανωτερω διαγραμμα ωστε ο χγ-αξονας να αρχιζει απο το μηδεν

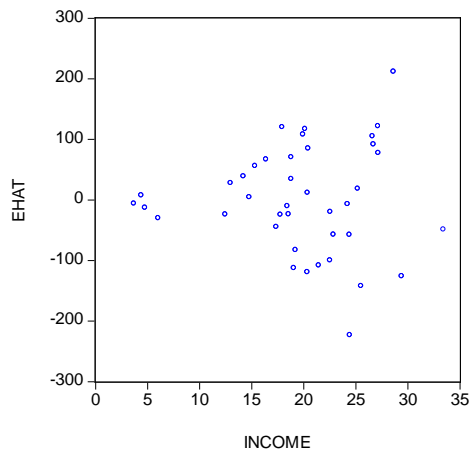
Κλικάρουμε Options → Axis/scale και στην επιλογή “Left axis scale endpoints” επιλεγουμε “user specified” και βαζουμε min=0, max=600 → OK



## ΕΛΕΓΧΟΙ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ

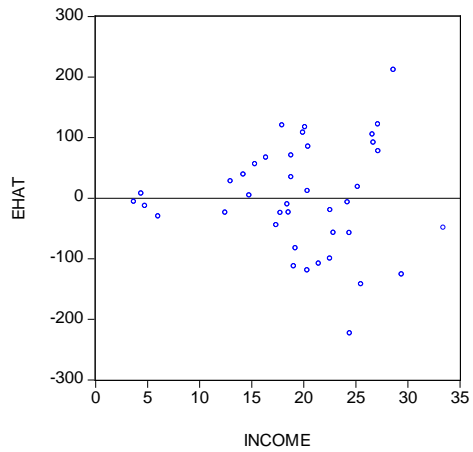
### 1) Διαγράμματα καταλοιπων

Quick → graph → income ehat → OK → Επιλεγουμε “scatter” → OK



Βαζουμε μια γραμμη στο υψος της τιμης  $ehat=0$ . Οποτε

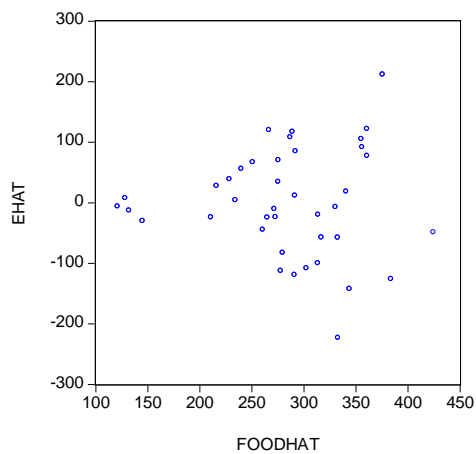
Κλικαρουμε “Line/Shade?” → Data value=0, και στο “Orientation” επιλεγουμε “Horizontal-Left Axis” → OK



Άλλα εναλλακτικά διαγράμματα διασποράς των καταλοίπων:

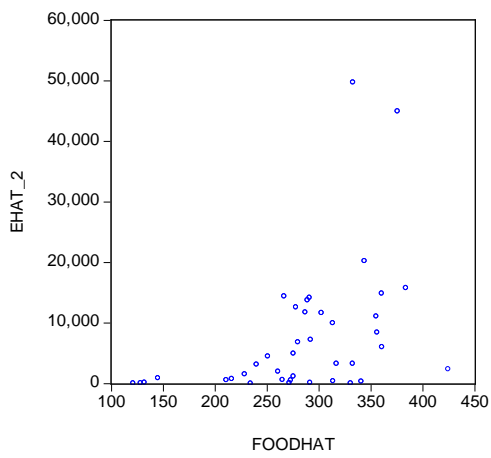
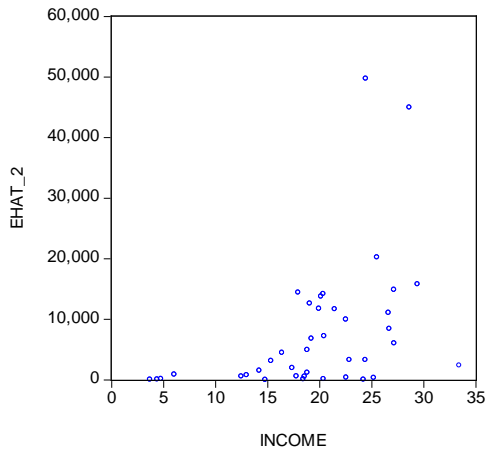
Οι εκτιμηθείσες τιμές της καταναλωτικής δαπάνης ( $\hat{y}$ ) αποκτώνται με την εντολή

`series foodhat=food_exp-ehat`



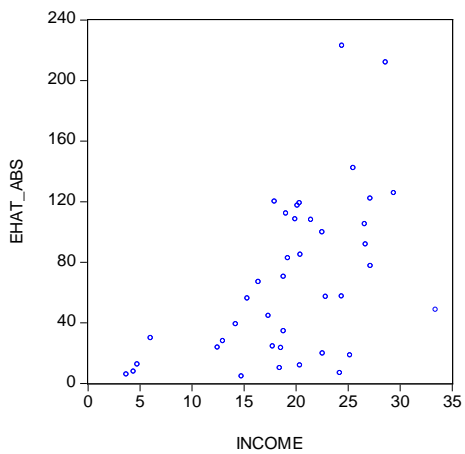
Τα τετράγωνα των καταλοίπων:

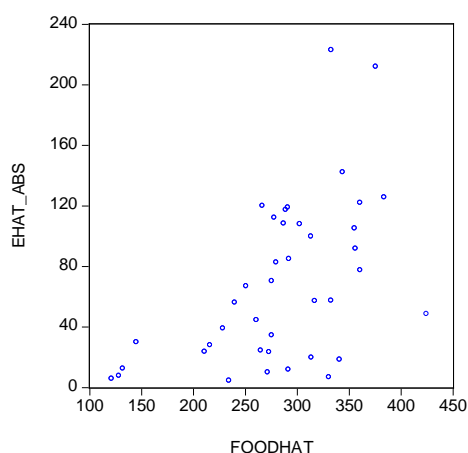
`series ehat_2=resid^2`



Οι απολυτες τιμες των καταλοιπων:

series ehat\_abs=@abs(ehat)





## 2) Goldfeld-Quandt ελεγχος

Sort income

Τρεχουμε το μοντελο με τις 20 πρωτες παρατηρησεις και αποθηκευουμε το τετραγωνο του τυπικου σφαλματος της παλινδρομησης με την ονομασια sig21  
 smpl 1 20

ls food\_exp c income  
 scalar sig21=@se^2

Dependent Variable: FOOD\_EXP  
 Method: Least Squares  
 Date: 10/19/22 Time: 20:23  
 Sample: 1 20  
 Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	72.96174	38.83435	1.878794	0.0766
INCOME	11.50038	2.507514	4.586367	0.0002
R-squared	0.538873	Mean dependent var		240.1830
Adjusted R-squared	0.513254	S.D. dependent var		85.69849
S.E. of regression	59.78939	Akaike info criterion		11.11417
Sum squared resid	64345.89	Schwarz criterion		11.21375
Log likelihood	-109.1417	Hannan-Quinn criter.		11.13361
F-statistic	21.03477	Durbin-Watson stat		2.081129
Prob(F-statistic)	0.000229			

Κανουμε το ιδιο για τις τελεταιες 20 παρατηρησεις και αποθηκεουμε το τετραγωνο του τυπικου σφαλματος της παλινδρομησης με την ονομασια sig22

```
smpl 21 40  
ls food_exp c income  
scalar sig22=@se^2
```

Dependent Variable: FOOD\_EXP  
Method: Least Squares  
Date: 10/19/22 Time: 20:50  
Sample: 21 40  
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-24.91465	184.9249	-0.134729	0.8943
INCOME	14.26400	7.425093	1.921054	0.0707

R-squared	0.170142	Mean dependent var	326.9640
Adjusted R-squared	0.124038	S.D. dependent var	121.4566
S.E. of regression	113.6747	Akaike info criterion	12.39920
Sum squared resid	232594.7	Schwarz criterion	12.49877
Log likelihood	-121.9920	Hannan-Quinn criter.	12.41864
F-statistic	3.690447	Durbin-Watson stat	1.887953
Prob(F-statistic)	0.070707		

Υπολογιζουμε την τιμη της F-στατιστικης  
scalar F\_GQ=sig22/sig21

και η κριτικη τιμη για 5% επιπεδο σημαντικοτητας ειναι  
scalar F\_c\_GQ=@qfdist(.95,18,18)

Ισοδυναμα, αντι για το @se^2 μπορουμε να χρησιμοποιησουμε παραπανω το αθροισμα των τετραγωνων των καταλοιπων με την εντολη  
scalar rss=@ssr

Επειδη σπασαμε το αρχικο δειγμα σε 2 κομματια λογω του GQ ελεγχου, πρεπει τωρα να ξαναχρησιμοποιησουμε ολο το δειγμα για να κανουμε ολους τους υπολοιπους ελεγχους που ακολουθουν

```
smpl 1 40
```

### 3) Breusch-Pagan-Godfrey ελεγχος

#### A) Breusch-Pagan εκδοχη

Το ehat\_2 το δημιουργησαμε πιο πανω

```
scalar sigma2_BP=@sum(ehat_2)/40  
series p_i=ehat_2/sigma2_BP  
ls p_i c income
```

Dependent Variable: P\_I  
Method: Least Squares  
Date: 10/20/22 Time: 18:21  
Sample: 1 40  
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.756949	0.633618	-1.194645	0.2396
INCOME	0.089619	0.030553	2.933172	0.0057
R-squared	0.184611	Mean dependent var		1.000000
Adjusted R-squared	0.163153	S.D. dependent var		1.428298
S.E. of regression	1.306597	Akaike info criterion		3.421437
Sum squared resid	64.87348	Schwarz criterion		3.505881
Log likelihood	-66.42873	Hannan-Quinn criter.		3.451969
F-statistic	8.603501	Durbin-Watson stat		2.344523
Prob(F-statistic)	0.005659			

$R^2=0.1846=ESS/TSS$ . Αλλα TSS?

#### Περιγραφικη στατιστικη για το p\_i

Κλικάρουμε την μεταβλητη p\_i → View → Descriptive Statistics and tests → Stats table

	P_I
Mean	1.000000
Median	0.431354
Maximum	6.533927
Minimum	0.002779
Std. Dev.	1.428298
Skewness	2.513502
Kurtosis	9.748365
Jarque-Bera	118.0186
Probability	0.000000
Sum	40.00000
Sum Sq. Dev.	79.56135



Observations 40

Αρα TSS=79.561 και ESS=14.686

scalar LM\_BP=14.686/2=7.343

scalar LM\_c\_BP=@qchisq(0.95,1)

## B) Godfrey εκδοχη

Κλικαρουμε το eq01 που περιεχει τα αποτελεσματα εκτιμησης του βασικου μοντελου → View→Residual tests→ Heteroscedasticity tests→ Breusch-Pagan-Godfrey→ OK

### Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	8.603501	Prob. F(1,38)	0.0057
Obs*R-squared	7.384424	Prob. Chi-Square(1)	0.0066
Scaled explained SS	6.627901	Prob. Chi-Square(1)	0.0100

### Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 10/19/22 Time: 21:11

Sample: 1 40

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5762.370	4823.501	-1.194645	0.2396
INCOME	682.2326	232.5920	2.933172	0.0057

R-squared	0.184611	Mean dependent var	7612.629
Adjusted R-squared	0.163153	S.D. dependent var	10873.10
S.E. of regression	9946.642	Akaike info criterion	21.29656
Sum squared resid	3.76E+09	Schwarz criterion	21.38101
Log likelihood	-423.9313	Hannan-Quinn criter.	21.32710
F-statistic	8.603501	Durbin-Watson stat	2.344523
Prob(F-statistic)	0.005659		

$$LM_G = N * R^2 = 40 * 0.1846 = 7.384$$

Η κριτική τιμή υπολογίστηκε στο Breusch-Pagan ελεγχό πιο πριν.

#### 4) White ελεγχος

Κλικάρουμε το eq01 που περιεχει τα αποτελεσματα εκτιμησης του βασικου μοντελου →  
View→Residual tests→ Heteroscedasticity tests→ White-→ OK

#### Heteroskedasticity Test: White

F-statistic	4.307884	Prob. F(2,37)	0.0208
Obs*R-squared	7.555079	Prob. Chi-Square(2)	0.0229
Scaled explained SS	6.781072	Prob. Chi-Square(2)	0.0337

#### Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 10/19/22 Time: 22:34

Sample: 1 40

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2908.783	8100.109	-0.359104	0.7216
INCOME	291.7457	915.8462	0.318553	0.7519
INCOME^2	11.16529	25.30953	0.441150	0.6617

R-squared	0.188877	Mean dependent var	7612.629
Adjusted R-squared	0.145032	S.D. dependent var	10873.10
S.E. of regression	10053.75	Akaike info criterion	21.34132
Sum squared resid	3.74E+09	Schwarz criterion	21.46798
Log likelihood	-423.8264	Hannan-Quinn criter.	21.38712
F-statistic	4.307884	Durbin-Watson stat	2.365567
Prob(F-statistic)	0.020802		

$$LM\_W = N * R^2 = 40 * 0.188 = 7.555$$

$$\text{scalar LM\_c\_W} = @qchisq(0.95, 2)$$

## ΑΝΤΙΜΕΤΩΠΙΣΗ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ

### A) Τυπικά σφάλματα του White

ls(h) food\_exp c income

Dependent Variable: FOOD\_EXP

Method: Least Squares

Date: 10/19/22 Time: 23:35

Sample: 1 40

Included observations: 40

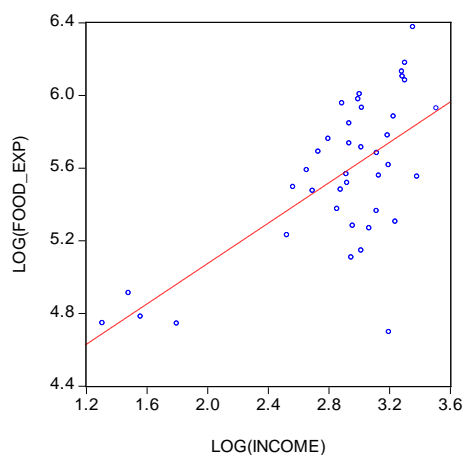
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	83.41600	27.46375	3.037313	0.0043
INCOME	10.20964	1.809077	5.643565	0.0000

R-squared	0.385002	Mean dependent var	283.5735
Adjusted R-squared	0.368818	S.D. dependent var	112.6752
S.E. of regression	89.51700	Akaike info criterion	11.87544
Sum squared resid	304505.2	Schwarz criterion	11.95988
Log likelihood	-235.5088	Hannan-Quinn criter.	11.90597
F-statistic	23.78884	Durbin-Watson stat	1.893880
Prob(F-statistic)	0.000019		

### B) Λογαριθμικός μετασχηματισμός του υποδειγματος



## Is log(food\_exp) c log(income)

Dependent Variable: LOG(FOOD\_EXP)

Method: Least Squares

Date: 10/21/22 Time: 02:20

Sample: 1 40

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.963567	0.294373	13.46444	0.0000
LOG(INCOME)	0.555881	0.100660	5.522391	0.0000
R-squared	0.445230	Mean dependent var		5.565019
Adjusted R-squared	0.430630	S.D. dependent var		0.424068
S.E. of regression	0.319987	Akaike info criterion		0.607634
Sum squared resid	3.890883	Schwarz criterion		0.692078
Log likelihood	-10.15268	Hannan-Quinn criter.		0.638166
F-statistic	30.49680	Durbin-Watson stat		1.982420
Prob(F-statistic)	0.000003			

## Έλεγχος ετεροσκεδαστικότητας των καταλοίπων της ανωτέρω παλινδρόμησης

### BPG ελεγχος:

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	2.317020	Prob. F(1,38)	0.1362
Obs*R-squared	2.298801	Prob. Chi-Square(1)	0.1295
Scaled explained SS	3.394959	Prob. Chi-Square(1)	0.0654

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 10/21/22 Time: 02:21

Sample: 1 40

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.144525	0.161249	-0.896284	0.3757
LOG(INCOME)	0.083930	0.055138	1.522176	0.1362
R-squared	0.057470	Mean dependent var		0.097272
Adjusted R-squared	0.032667	S.D. dependent var		0.178215
S.E. of regression	0.175280	Akaike info criterion		-0.596159
Sum squared resid	1.167475	Schwarz criterion		-0.511715
Log likelihood	13.92318	Hannan-Quinn criter.		-0.565627
F-statistic	2.317020	Durbin-Watson stat		2.274738
Prob(F-statistic)	0.136244			

White ελεγχος:

Heteroskedasticity Test: White

F-statistic	1.376771	Prob. F(2,37)	0.2650
Obs*R-squared	2.770612	Prob. Chi-Square(2)	0.2502
Scaled explained SS	4.091748	Prob. Chi-Square(2)	0.1293

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 10/21/22 Time: 02:22

Sample: 1 40

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.192347	0.518060	0.371283	0.7125
LOG(INCOME)	-0.218917	0.445737	-0.491136	0.6262
(LOG(INCOME))^2	0.062627	0.091458	0.684766	0.4978

R-squared	0.069265	Mean dependent var	0.097272
Adjusted R-squared	0.018955	S.D. dependent var	0.178215
S.E. of regression	0.176518	Akaike info criterion	-0.558753
Sum squared resid	1.152864	Schwarz criterion	-0.432087
Log likelihood	14.17505	Hannan-Quinn criter.	-0.512954
F-statistic	1.376771	Durbin-Watson stat	2.309990
Prob(F-statistic)	0.265020		

### 3) ΓΕΝΙΚΕΥΜΕΝΗ ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ (GLS)

Υποθετούμε ότι  $\sigma_i^2 = \sigma^2 \text{income}_i$

ls(W=1/sqr(income)) food\_exp c income

Dependent Variable: FOOD\_EXP

Method: Least Squares

Date: 10/21/22 Time: 01:16

Sample: 1 40

Included observations: 40

Weighting series: 1/SQR(INCOME)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	78.68408	23.78872	3.307621	0.0021
INCOME	10.45101	1.385891	7.541002	0.0000

Weighted Statistics

R-squared	0.599438	Mean dependent var	263.3689
Adjusted R-squared	0.588897	S.D. dependent var	76.43899

S.E. of regression	76.30741	Akaike info criterion	11.55612
Sum squared resid	221267.2	Schwarz criterion	11.64057
Log likelihood	-229.1225	Hannan-Quinn criter.	11.58666
F-statistic	56.86672	Durbin-Watson stat	1.905701
Prob(F-statistic)	0.000000		

---



---

Unweighted Statistics

---



---

R-squared	0.384787	Mean dependent var	283.5735
Adjusted R-squared	0.368597	S.D. dependent var	112.6752
S.E. of regression	89.53266	Sum squared resid	304611.7
Durbin-Watson stat	1.892377		

---



---