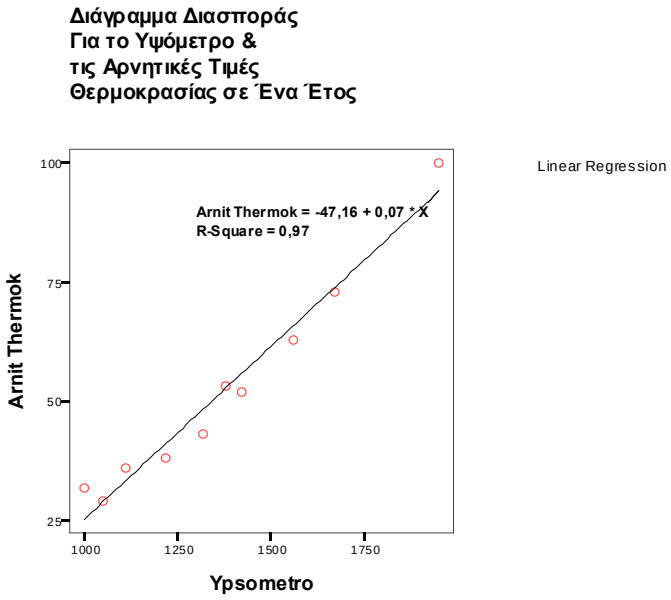
Ελένη Κανδηλώρου
Αναπλ. Καθηγήτρια

Αθήνα, 16-04-2017

Γραμμικά Μοντέλα

Λύσεις Ασκήσεων

1η Άσκηση:
(α) Είναι η σχέση μεταξύ των δύο μεταβλητών γραμμική;



Εντολές για το SPSS:

Graphs-Interactive-Scatterplot: (a) *Assign Variables*:

Arnit Themok

Ypsometro

(b)Fit-Regression-Include constant,
(*c*) Chart Title.....

Κάνοντας χρήση του Διαγράμματος Διασποράς, διαπιστώνουμε την ορθότητα της υπόθεσης που έχουμε κάνει, ότι, δηλαδή, η σχέση των δύο μεταβλητών είναι γραμμική και μάλιστα θετική, καθώς όσο αυξάνεται η X_i (*Ypsometro*) τόσο αυξάνεται και η Y_i (*Arnit. Thermok*). Μάλιστα φαίνεται ότι τα σημεία (X_i, Y_i) βρίσκονται «κοντά» σε μία ευθεία, π.χ. την $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i$, οι τιμές της X_i και της Y_i αυξάνονται πρός την ίδια κατεύθυνση.

(β) Σημειακές εκτιμήσεις των παραμέτρων β_0, β_1:

Υπολογισμός αθροισμάτων

X_i	Y_i	X_iY_i	X_i^2
1000	32	32000	1000000
1050	29	30450	1102500
1110	36	39960	1232100
1220	38	46360	1488400
1320	43	56760	1742400
1380	53	73140	1904400
1420	52	73840	2016400
1560	63	98280	2433600
1670	73	121910	2788900
1950	100	195000	3802500
$\Sigma X_i=13680$	$\Sigma Y_i=519$	$\Sigma X_iY_i=767700$	$\Sigma X_i^2= 19511200$

$X_{μέσος}= 13680/10 = 1368$ $\quad$ $Y_{μέσος}= 519/10 = 51,9$

(X_i-1368)	$(X_i-1368)^2$
-368	135424
-318	101124
-258	66564
-148	21904
-48	2304
12	144
52	2704
192	36864
302	91204
582	338724

796960 *όπου* $796960=\Sigma(Xi-1368)^2=S^2_{XX}$

και $S_{XX}=892,726$

Αντικατάσταση των αθροισμάτων στους επόμενους δύο τύπους υπολογισμού των εκτιμητριών β_1 και β_0.

$$\widehat{\beta}_1 = \frac{n\sum Y_i X_i - \sum Y_i \sum X_i}{n\sum X_i^2 - (\sum (X_i))^2} =$$

$$\frac{10(767700) - (13680)(519)}{10(19511200) - (13680)^2} = 0,07$$

$$\widehat{\beta}_0 = \overline{Y} - \beta_1 \overline{X} = \frac{519}{10} - 0,07(\frac{13680}{10}) = -47,157$$

Άρα, έχουμε: $\widehat{Y}_i = -47,157 + 0,07X_i$

Ερμηνεία των εκτιμητών:

- Ο $\widehat{\beta}_1 = 0,07$ δίνει την κλίση της γραμμής παλινδρόμησης και δηλώνει ότι: αν αυξηθεί το υψόμετρο κατά 100 μέτρα, ο αριθμός των ημερών με αρνητική θερμοκρασία θα αυξηθεί κατά επτά ημέρες!

- Ο $\hat{\beta}_0 = -47,157$. Αν το υψόμετρο είναι μηδέν (βρισκόμαστε, δηλαδή, στο επίπεδο της θάλασσας), τότε δεν υπάρχουν μέρες με θερμοκρασία κάτω του μηδενός.

Εντολές για το SPSS:

Analyze → Regression →Linear:
Dependant: ***Arnit Themok***
Independent: ***Ypsometro***
Statistics: Estimates, Confidence Intervals, Model Fit, R-Squared
Change → **Continue**
Plots: →Y: AXIAY, X: ZRESID
x **Histogram**
x **Normal probability plot →** **Continue**
Save:

Predicted Values	Residuals
Unstandardized Standardized	Unstandardized Standardized

→ **Continue**

(γ) Το 95% Δ.Ε. της κλίσης της ευθείας δίνεται παρακάτω:

$$\hat{\beta}_1 + t_{n-2,1-a/2}(S_{\hat{\beta}_1})$$

όπου το τυπικό σφάλμα του $\hat{\beta}_1$ δίνεται παρακάτω, δεδομένου ότι:

$\sqrt{\frac{SSE}{n-2}} = S_e = \sqrt{\frac{150,2546}{8}} = \sqrt{18,782} = 4,33$:

$$S_{\hat{\beta}_1} = \frac{S_e}{\sqrt{\sum_i (X_i - \bar{X})^2}} = \frac{\sqrt{18,782}}{892,73} = \frac{4,33}{892,73} = 0,005$$

Y_i	$\hat{Y}_i = Y_i - (-47,157 + 0,07X_i)$	$Y_i - \hat{Y}_i = e_i$	$(Y_i - \hat{Y}_i)^2 = e_i^2$	
32	25,25306	6,74694	45,5212	
29	28,87357	0,12643	0,015985	
36	33,21818	2,78182	7,738523	
38	41,18330	-3,18330	10,1334	
43	48,42431	-5,42431	29,42314	
53	52,76892	0,23108	0,053398	
52	55,66533	-3,66533	13,43464	
63	65,80275	-2,80275	7,855408	
73	73,76787	-0,76787	0,589624	
100	94,04271	5,95729	35,4893	
519	**25,25306**	**6,74694**	***SSE***=150,2546	***Αθροίσματα***

95% Δ.Ε.:

$$\hat{\beta}_1 \pm t_{n-2,1-a/2}(S_{\hat{\beta}_1}) = 0,07 \pm 2,306 * 0,005 \Rightarrow [0,06 \qquad 0,08]$$

Το σχόλιο που μπορούμε να κάνουμε για τον αντίστοιχο στατιστικό έλεγχο είναι:

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 > 0$$

Δεδομένου ότι το μηδέν δεν περιλαμβάνεται στο 95% διάστημα εμπιστοσύνης, μπορούμε να συμπεράνουμε ότι η μηδενική υπόθεση απορρίπτεται, ή εναλλακτικά ότι η εκτίμηση της κλίσης της ευθείας θα είναι πάντα θετική!

(δ) Ο παρακάνω πίνακας μας πληροφορεί ότι η παλινδρόμησή μας υποδηλώνει την έντονη γραμμική σχέση που υπάρχει μεταξύ «*υψομέτρου*» και *«αριθμού ημερών με αρνητικές θερμοκρασίες, στη διάρκεια ενός έτους»,* η οποία είναι και στατικά σημαντικά (α=5%). Μάλιστα, η τιμή του Sig. μας πληροφορεί ότι η έντονη γραμμική σχέση που περιγράφεται από το υπόδειγμά μας, ισχύει για κάθε επίπεδο στατιστικής σημαντικότητας!

ANOVA

Υπόδειγμα	Αθρ. Τετραγ	β ε	Μέσα Τετραγ	F	Sig.
Παλινδρόμηση	SSR = 4178,645	k-1=1	SSR/(k-1) = 4178,645	222,484	,000
Κατάλοιπα	SSE = 150,255	n-2=8	SSE/(n-2) = 18,782		
Σύνολο	SST = 4328,900	n-1=9			

(δ) Τα 90% διάστηματα μέσης πρόβλεψης του *Y για* $E(Y)=\beta_0+\beta_1 X_0$*, όπου:*

i) X_0=1500

$$(\hat{\beta}_0+\hat{\beta}_1 X_0) \pm t_{n-2,1-a/2} * S\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{S_{XX}}} =$$

$$(-47,157+0,07*1500) \pm 1,860 * 4,33\sqrt{\frac{1}{10}+\frac{(1500-1368)^2}{892,73}} =$$

$$(57,843) \pm 35,672 \Rightarrow (22,171 \quad 93,515)$$

$$\text{όπου} \quad S=\sqrt{\frac{SSR}{n-2}}=4,33$$

ii) X_0=2000: Για υψόμετρο 2000 μέτρα δεν μπορούμε να κάνουμε πρόβλεψη γιατί η τιμή αυτή δεν είναι μέσα στο εύρος γνωστών υψομέτρων (1000-1950) για τα οποία ισχύει το γραμμικό υπόδειγμα.

2η Άσκηση:

(α) **Πίνακες**

	Συσχέτιση	Megethos	Kostos
Megethos	Pearson Correlation	1	,995(**)
	Sig. (2-tail.)		***,000***
	N	10	10
Kostos	Pearson Correlation	***,995(**)***	***1***
	Sig. (δικατλ. κριτήρ.)	,000	
	N	10	10

** *Correlation is significant at the 0.01 level (2-tailed).*

Ερμηνεία:Ο συντελεστής συσχέτισης μετρά το βαθμό της γραμμικής συσχέτισης των μεταβλητών *X & Y.* Υπάρχουν ενδείξεις έντονης γραμμικής σχέσης (r=0,995) μεταξύ των δύο μεταβλητών. Όταν αυξάνει το Megethos αυξάνει και το Kostos. Ο συντελεστής συσχέτισης είναι στατιστικά σημαντικός ακόμη και σε α=1%, εφ' όσον το Sig.=0,000.

ANOVA

Υπόδειγμα	Άθρ. Τετραγ=***SS***	β.ε	Μέσα Τετράγ=***MS***	F	Sig.
1 Regression	***SSR=292,152***	v_1***=k-1=1***	***MSR*** =292,152	F_0***= 760,633***	***,000***
Residual	***SSE***=3,073	v_2***=n-k=8***	***MSE= 0,384***		
Total	***SST***=299,225	***n-1***=9			
Σχόλια ⇒	$SSR=\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2$ $SSE=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$ $SST=\sum_{i=1}^{n}(Y_i-\bar{Y})^2$	***k =αριθ. ανεξαρ. μεταβ.***	$MSR=\frac{SSR}{k-1}$ $MSE=\frac{SSE}{n-k}$	$F\frac{SSR/\sigma^2}{SSE/\sigma^2}=\frac{\frac{SSR}{k-1}}{\frac{SSE}{n-k}}\sim F_{1,n-2}=F_{v_1,v_2}$	

Ερμηνεία: Αν ***Sig.*** $F_{v1, v2,\alpha}$ ***< 0.05***, τότε η Ηο απορρίπτεται, δηλ. οι μεταβλητές X_i και Y_i είναι γραμμικά συσχετισμένες. Άρα, μπορούμε να κατασκευάσουμε έναν έλεγχο για την υπόθεση H_0***:***β_1***=0***. Η H_0 θα απορρίπτεται όταν η παραπάνω στατιστική συνάρτηση λαμβάνει μεγάλες τιμές, δηλαδή, όταν:

$$F_0=\frac{SSR/v_1}{SSE/v_2}>F_{v_1,v_2,\alpha}$$

με αντίστοιχο p-value:

$$p-value = 1 - F_{F_{1,n-2}}\left(\frac{SSR}{SSE/(n-2)}\right)$$

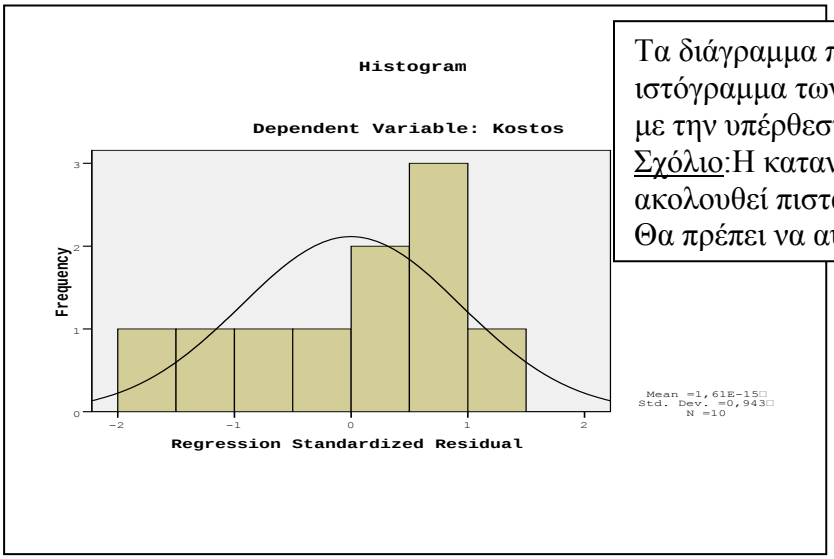
Πληροφορίες Υποδείγματος

Model					Change Statistics				Durbin-Watson
	R	R^2	Διορθωμένο R^2	Τυπ. Σφάλ. Εκτιμ.	F	df1	df2	Sig.	
1	,995	***,990***	,988	,61975	760,633	1	8	,000	1,377
	$R^2 = (R)^2$		$R^2_{adj} = 1 - [\frac{(1-R^2)/(n-1)}{n-k-1}]$						

Εκτιμήσεις Παραμέτρων

Mode	Εκτιμητές		Τυποποιημένοι Εκτιμητές			95% Confidence Interval for B	
	B	Τυπικό Σφάλμα	Beta	t	Sig.	Lower Bound	Upper Bound
1 (Constant)	1,100	,423		2,598	,032	,124	2,076
Megethos	,188	,007	***,995***	***27,580***	,000	***,172***	***,204***
		$beta = \hat{\beta}_1(S_X / S_Y)$					

(β) Να σχολιάσετε τα παρακάτω διαγράμματα

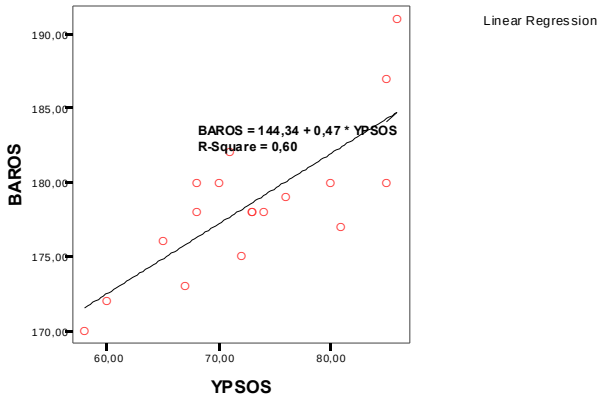


Τα διάγραμμα παραπλεύρως είναι ένα ιστόγραμμα των τυποποιημένων καταλοίπων με την υπέρθεση μιας κανονικής καμπύλης.
Σχόλιο:Η κατανομή δεν φαίνεται να ακολουθεί πιστά την κανονική κατανομή. Θα πρέπει να αυξήσουμε τα δεδομένα μας.

3η Άσκηση:

(α) *Να διερευνήσετε διαγραμματικά και να σχολιάσετε αν και πως συμμεταβάλλονται οι δυο μεταβλητές.*

Το παρακάτω διάγραμμα ελέγχει κατά πόσο ισχύει η ορθότητα της υπόθεσης, ότι η σχέση των δύο μεταβλητών μας είναι γραμμική.
Καθώς αυξάνεται η X (Ypsos), αυξάνεται και η Y (Baros).
Τα σημεία (X_i, Y_i) βρίσκονται «κοντά» σε μία ευθεία.



(β) *Αναζητούμε το* $\boldsymbol{\beta_1}$ *που ελαχιστοποιεί το σφάλμα και συγκεκριμένα ελαχιστοποιεί το άθροισμα των τετραγώνων της* $\boldsymbol{Q}$, που δίνεται παρακάτω:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \rightarrow \textit{εξίσωση παλινδρόμησης}$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \rightarrow \textit{εκτιμημένη εξίσωση παλινδρόμησης}$$

$$e_i = Y_i - \hat{Y}_i \rightarrow \textit{κατάλοιπα}$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2 = Q$$

$$MET \rightarrow \min Q \rightarrow \textit{ελάχιστο του} \sum_{i=1}^{n} e_i^2$$

$$\partial Q / \partial \hat{\beta}_1 = 0 \Rightarrow -2\sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)) = \sum_{i=1}^{n} Y_i X_i - \hat{\beta}_0 \sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$\hat{\beta}_1 = \frac{\sum Y_i X_i - \bar{Y} \sum X_i}{\sum X_i^2 - \bar{X} \sum X_i}$$

ή

$$\hat{\beta}_1 = \frac{\sum Y_i X_i - \bar{Y} \sum X_i}{\sum X_i^2 - \bar{X} \sum X_i} = \frac{234799\text{-}72{,}89*(178{,}56)}{574294\text{-}178{,}56*(3214)} = 0{,}19$$

(γ) **Ο συντελεστής *(1-R^2)* εκφράζει**: Το ποσοστό (39,8%) της μεταβλητότητας της εξαρτημένης μεταβλητής *Y* που δεν ερμηνεύεται από την ερμηνευτική μεταβλητή *X*.

$$1 - R^2 = 1 - \frac{SSR}{SST} = 1 - R^2 = 1 - 0{,}602 = 0{,}398$$

X	***Y***	***X^2***	***Y^2***	***X*Y***	***(Xi-178,56)***	***$(Xi-178{,}56)^2$***
170	58	28900	3364	9860	-8,56	73,2736
172	60	29584	3600	10320	-6,56	43,0336
173	67	29929	4489	11591	-5,56	30,9136
175	72	30625	5184	12600	-3,56	12,6736
176	65	30976	4225	11440	-2,56	6,5536
177	81	31329	6561	14337	-1,56	2,4336
178	73	31684	5329	12994	-0,56	0,3136
178	74	31684	5476	13172	-0,56	0,3136
178	73	31684	5329	12994	-0,56	0,3136
178	68	31684	4624	12104	-0,56	0,3136
179	76	32041	5776	13604	0,44	0,1936
180	68	32400	4624	12240	1,44	2,0736
180	80	32400	6400	14400	1,44	2,0736
180	70	32400	4900	12600	1,44	2,0736
180	85	32400	7225	15300	1,44	2,0736
182	71	33124	5041	12922	3,44	11,8336
187	85	34969	7225	15895	8,44	71,2336
191	86	36481	7396	16426	12,44	154,7536
3214	***1312***	***574294***	***96768***	***234799***	***-0,08***	***416,4448***
178,556	***72,8889***	**=Μέσοι**				

(δ) Να εκτιμήσετε τη Διακύμανση του σφάλματος.

Η εκτίμηση του $\boldsymbol{\sigma_\varepsilon^2}$ δίνεται από το $\boldsymbol{S_e^2}$:

$$S_e^2 = \frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{n-2} = \frac{SSE}{n-2} = \boxed{\text{[SSE/(18-2)] = 28,297 = MSE}}$$

e	***e²***
-3,91596	15,33474272
-4,48106	20,07989872
1,23639	1,528660232
3,67129	13,47837026
-4,61126	21,26371879
10,10619	102,1350763
0,82364	0,67838285
1,82364	3,32566285
0,82364	0,67838285
-4,17636	17,44198285
2,54109	6,457138388
-6,74146	45,44728293
5,25854	27,65224293
-4,74146	22,48144293
10,25854	105,2376429
-6,30656	39,77269903
1,28068	1,640141262
-2,84952	8,11976423
	452,7532331=Σ e²
	=SSE

(ε) *Να ελέγξετε τη στστιστική σημαντικότητα του συντελεστή παλινδρόμησης, σε α=1%.*

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 > 0$$

$$T_0 = \frac{\widehat{\beta}_1 - \beta_1}{S_{\widehat{\beta}_1}} = \frac{1,283 - 0}{0,261} = 4,92 > t_{16,0.005} = 2,921$$

Άρα, η μδενική υπόθεση απορρίπτεται. Δηλαδή, το β_1 δεν είναι μηδέν. Αντίθετα, το β_1 παίρνει θετικές τιμές.

Σχόλια για το αντίστοιχο Δ.Ε.*:* Δεδομένου ότι δεν γίνεται δεκτή η μηδενική υπόθεση, δηλ. απορρίπτεται, συμπεραίνουμε ότι το 99% Δ.Ε. δεν περιλαμβάνει το μηδέν. Επιπλέον, δεν θα υπάρχει αρνητική τιμή στο διαστημα αυτό!