

ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Εργασία 1^η

- Στοιχεία Φοιτητή

ΠΑΝΤΕΛΕΗΜΩΝ ΜΑΘΙΟΥΔΑΚΗΣ
ΜΕΣ20022

- Διδάσκων
- Βοηθός Φροντιστηρίου

ΜΑΡΚΟΣ ΚΟΥΤΡΑΣ
ΔΗΜΗΤΡΙΟΣ ΑΥΜΠΕΡΟΠΟΥΛΟΣ

Πανεπιστήμιο Πειραιά
ΠΜΣ Εφαρμοσμένης Στατιστικής

Περιεχόμενα

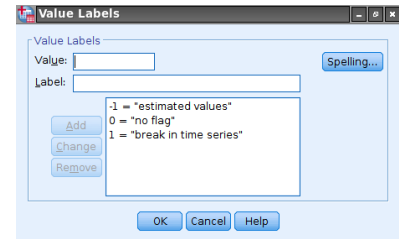
Άσκηση 1	3
Άσκηση 2	4
Άσκηση 3	4
Άσκηση 4	5
Υποερώτημα 1	5
Υποερώτημα 2	5
Υποερώτημα 3	6
Υποερώτημα 4	6
Άσκηση 5	7
Υποερώτημα 1	7
Άσκηση 6	7
Άσκηση 7	7
Άσκηση 8	8
Ερώτημα +	9

Κατάλογος Σχημάτων

1 Labels των παρατηρήσεων	3
2 Επιλογή της MIN Function	3
3 Δημιουργία του νέου Dataset	3
4 tourism2014valid	3
5 avgtripexp - accomarivals $R^2 = 0.129$	4
6 avgtripexp - avgaccomexp $R^2 = 0.687$	4
7 avgtripexp - avgdurexp $R^2 = 0.122$	4
8 avgtripexp - avgrestexp $R^2 = 0.120$	4
9 avgtripexp - avgtranspexp $R^2 = 0.619$	4
10 avgtripexp - accomarivals $R^2 = 0.082$	4
11 Στατιστικοί Έλεγχοι του βέλτιστου μοντέλου	5
12 Ιστόγραμμα Studentized residuals	5
13 P-P plot Studentized residuals	5
14 Έλεγχος KS-test	5
15 Scatter-Plot Studentized Residuals	6
16 Levene Test Studentized Residuals Sig = 0.232 > α	6
17 Διάγραμμα Ελέγχου Ανεξαρτησίας Καταλοίπων	6
18 Έλεγχος Ροών Ανεξαρτησίας Καταλοίπων P-value = 0.662 > $\alpha = 0.05$	6
19 Έλεγχος Μοντέλου avgtripexp-avgdurexp P-value = 0.102 > $\alpha = 0.05$	7
20 Έλεγχος των μετασχηματισμένων μεταβλητών (ανεξάρτητη μετα- βλητή avgaccomexp-transf). Διακρίνουμε Sig(β_0) = 0.001	8

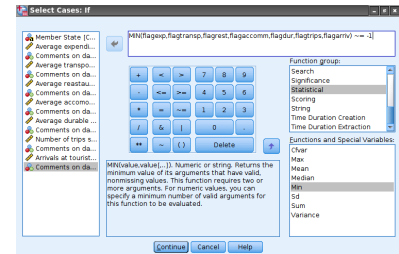
Άσκηση 1

Αρχικά, όπως ορίζει το πρώτο ερώτημα, επιλέγουμε με την εντολή Select Cases τις εκτιμημένες παρατηρήσεις, estimated values = -1, με βάση τον πίνακα των labels



Σχήμα 1: Labels των παρατηρήσεων

Στη συνέχεια, επιλέγουμε να εξαιρέσουμε τα δεδομένα με χαρακτηρισμό estimated values = -1 με την χρήση απο το Select Cases : $MIN(X1, X2, X3, X4, X5, X6) = -1$. Συνοπτικά, έστω κι αν μία παρατήρηση είναι Estimated Value, τότε το MIN Function θα επιστρέψει τιμή -1, συνεπώς θα αφαιρεθεί.

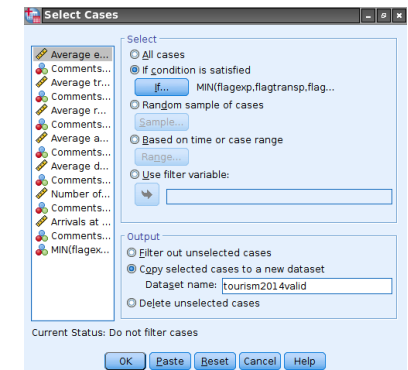


Σχήμα 2: Επιλογή της MIN Function

Τελικώς, αποθηκεύουμε το καινούριο Dataset με το όνομα tourism2014valid.

	COUNTRY	avgtripexp	flagexp	avgtranspexp	flagtransp	avgrestexp	flagrest	avgcommexp
1	Belgium	215.45	0	6.92	0	.00	0	24.82
2	Bulgaria	123.07	0	8.82	0	11.44	0	8.22
3	Czechia	51.86	0	3.26	0	3.06	0	2.85
4	Germany (until 1990 former territory of th...	271.74	0	24.54	0	.00	0	29.98
5	Estonia	65.62	0	11.05	0	7.97	0	7.34
6	Greece	254.28	0	6.01	0	8.34	0	.43
7	Spain	146.73	0	6.58	0	5.66	0	7.01
8	France	264.66	0	12.92	0	9.21	0	13.04
9	Croatia	126.35	0	7.52	0	6.36	0	5.12
10	Italy	235.85	1	12.16	1	.00	1	18.38
11	Cyprus	147.62	0	7.56	0	18.98	0	10.72
12	Latvia	38.73	0	5.14	0	1.82	0	2.72
13	Lithuania	51.94	0	.44	0	4.04	0	8.12
14	Luxembourg	473.33	0	78.33	0	26.22	0	68.13
15	Hungary	67.03	0	4.39	0	2.26	0	9.34
16	Netherlands	118.95	0	6.23	0	6.02	0	13.13
17	Austria	328.68	0	21.22	0	.00	0	42.87
18	Poland	105.47	1	4.55	1	6.98	1	6.05
19	Portugal	102.11	0	7.09	0	7.99	0	.58
20	Romania	9.14	0	7.86	0	6.37	0	5.05
21	Slovenia	102.38	0	6.75	0	7.55	0	16.83
22	Slovakia	133.16	0	8.29	0	7.26	0	13.19
23	Finland	189.69	0	21.13	0	13.07	0	15.05
24								

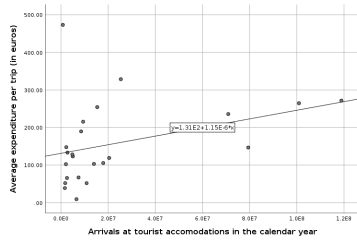
Σχήμα 4: tourism2014valid



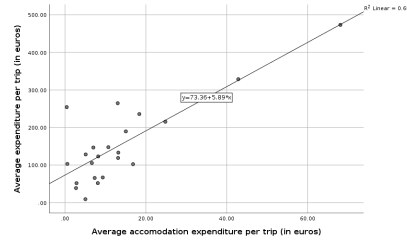
Σχήμα 3: Δημιουργία του νέου Dataset

Άσκηση 2

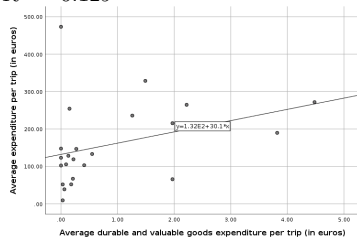
Παρακάτω διακρίνονται όλα τα διαγράμματα της μεταβλητής *avgtripexp* και όλες τις άλλες μεταβλητές



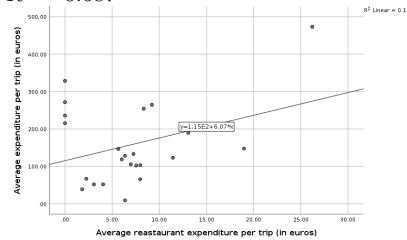
Σχήμα 5: *avgtripexp* - *accomarivals*
 $R^2 = 0.129$



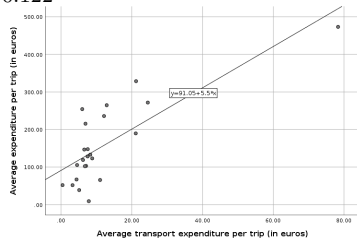
Σχήμα 6: *avgtripexp* - *avgaccomexp*
 $R^2 = 0.687$



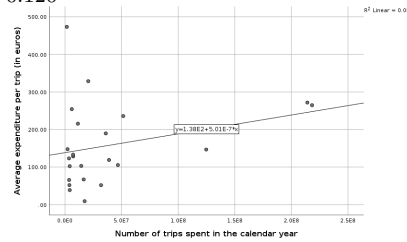
Σχήμα 7: *avgtripexp* - *avgdurexp* $R^2 = 0.122$



Σχήμα 8: *avgtripexp* - *avgrestexp* $R^2 = 0.120$



Σχήμα 9: *avgtripexp* - *avgtranspexp*
 $R^2 = 0.619$



Σχήμα 10: *avgtripexp* - *accomarivals*
 $R^2 = 0.082$

***Encoding: UTF-8.

***ERWTHMA II SPSS OUTPUT

GRAPH

/SCATTERPLOT(BIVAR)=

avgtranspexp WITH

avgtripexp

/MISSING=LISTWISE.

GRAPH

/SCATTERPLOT(BIVAR)=

avgrestexp WITH avgtripexp

/MISSING=LISTWISE.

GRAPH

/SCATTERPLOT(BIVAR)=

avgaccomexp WITH

avgtripexp

/MISSING=LISTWISE.

GRAPH

/SCATTERPLOT(BIVAR)=

avgdurexp WITH avgtripexp

/MISSING=LISTWISE.

GRAPH

/SCATTERPLOT(BIVAR)=trips

WITH avgtripexp

/MISSING=LISTWISE.

GRAPH

/SCATTERPLOT(BIVAR)=

accomarivals WITH

avgtripexp

/MISSING=LISTWISE.

Άσκηση 3

Παρατηρούμε απο τα διαγράμματα καθώς και τους συντελεστές R^2 η μεταβλητή που ερμηνεύει καλύτερα τα δεδομένα είναι η *avgaccomexp* με $R^2 = 0.687$. Συνεπώς, εάν ισχύουν και οι προϋποθέσεις των γραμμικών μοντέλων (Κανονικότητα καταλοίπων, ασυσχέτιστα και ομοσκεδαστικά κατάλοιπα που θα ελεγχθούν στα επόμενα ερωτήματα) τότε θα αποτελέσει ένα αρκετά καλό μοντέλο.

Συγκεκριμένα, το βέλτιστο εκτιμηθέν υπόδειγμα είναι :

$$y = 73.36 + 5.89 \cdot avgaccomexp \quad \text{βέλτιστο εκτιμηθέν υπόδειγμα} \quad (1)$$

$$y = \beta_0 + \beta_1 \cdot x + \epsilon \quad \text{αντίστοιχο θεωρητικό} \quad (2)$$

Άσκηση 4

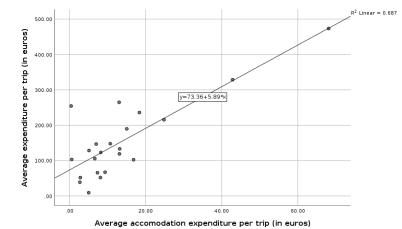
Υποερώτημα 1

Coefficients ^a							
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B
		B	Std. Error	Beta			Lower Bound Upper Bound
1	(Constant)	73.357	18.000		4.075	.001	35.923 110.791
	Average accommodation expenditure per trip (in euros)	5.885	.867	.829	6.786	.000	4.082 7.689

a. Dependent Variable: Average expenditure per trip (in euros)

Σχήμα 11: Στατιστικοί Έλεγχοι του βέλτιστου μοντέλου

Όπως διακρίνουμε απο το παραπάνω αποτέλεσμα του γραμμικού μοντέλου, Απορρίπτεται η υπόθεση ο συντελεστής β_1 της μεταβλητής avgaccomexp να είναι μηδέν με Sig = .000 σε επίπεδο σημαντικότητας $\alpha = 0.05$. Συνεπώς, η μεταβλητή είναι στατιστικά σημαντική. Αυτό βέβαια φαίνεται και απο το διπλανό γράφημα με εμφανή ένδειξη ύπαρξης τάσης.

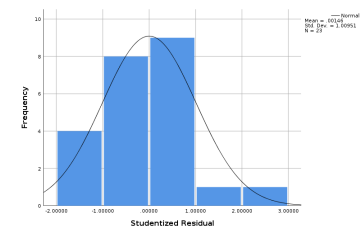


Υποερώτημα 2

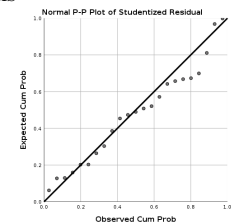
Για τον έλεγχο κανονικής κατανομής των καταλοίπων, μπορούμε να χρησιμοποιήσουμε τόσο γραφικούς ελέγχους, όσο και πιό αυστηρούς όπως τον έλεγχο KS-test. Δίπλα παρουσιάζονται οι γραφικές αναπαραστάσεις των τυποποιημένων Studentized residuals, ιστογράμματος και P-P Plot.

Απο το ιστόγραμμα δεν διακρίνεται κάποια ισχυρή ένδειξη κανονικότητας των καταλοίπων. Στο P-P Plot τα σημεία δεν βρίσκονται αρκετά κοντά στις αντίστοιχες θεωρητικές τιμές των ποσοστιαίων σημείων της κανονικής κατανομής.

Βέβαια, για να καταλήξουμε σε κάποιο ασφαλές συμπέρασμα, πρέπει να εφαρμόσουμε και τον έλεγχο KS-test.



Σχήμα 12: Ιστόγραμμα Studentized residuals



Σχήμα 13: P-P plot Studentized residuals

One-Sample Kolmogorov-Smirnov Test		
		Studentized Residual
N		23
Normal Parameters ^{a,b}	Mean	.0014612
	Std. Deviation	1.00951288
Most Extreme Differences	Absolute	.170
	Positive	.170
	Negative	-.084
Test Statistic		.170
Asymp. Sig. (2-tailed)		.082 ^c

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Σχήμα 14: Έλεγχος KS-test

Όπως διακρίνουμε απο τον έλεγχο KS-test, $p\text{-value} = 0.082 > \alpha = 0.05$, συνεπώς Δέν Απορρίπτονται την υπόθεση τα δεδομένα να ακολουθούν κανονική κατανομή.

Υποερώτημα 3

Για την υπόθεση ομοσκεδαστικότητας, πάλι θα ξεκινήσουμε με γραφικούς ελέγχους και καταλήγοντας με στατιστικό έλεγχο θα αποφανθούμε αν δέν παραβιάζεται η ομοσκεδαστικότητα. Δεξιά, παρουσιάζεται το Scatter-Plot Studentized Residuals όπου αν εξαιρέσουμε κάποιες απομονωμένες παρατηρήσεις, δέν φαίνεται να μην είναι ομοσκεδαστικά τα δεδομένα.

Προχωρώντας σε στατιστικό έλεγχο Levene, με κατηγοριοποίηση των δεδομένων ως προς την διάμεσο, δέν απορρίπτουμε τα τυποποιημένα κατάλοιπα να είναι ομοσκεδαστικά

Independent Samples Test									
Levene's Test for Equality of Variances					t-test for Equality of Means				
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Studentized Residual	1.515	.232	-4.783	21	.000	-1.4271198	.29840380	-2.0476845	-.80655511
Equal variances assumed									
Equal variances not assumed			-4.666	15.174	.000	-1.4271198	.30586250	-2.0784007	-.77583890

Σχήμα 16: Levene Test Studentized Residuals Sig = 0.232 > α

Υποερώτημα 4

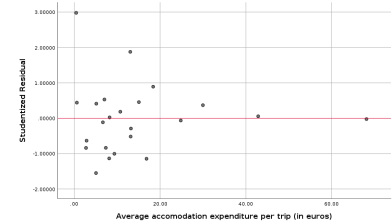
Καταλήγοντας σε έλεγχο ανεξαρτησίας, πάλι παρουσιάζονται γραφικοί και στατιστικοί έλεγχοι για τον έλεγχο ανεξαρτησίας των καταλοίπων. Απο Scatter-Plot Studentized Residuals ως προς Unstandardized Predicted Values, δέν φαίνεται κάποια σαφή ένδειξη συσχέτισης καταλοίπων. Εν κατακλείδη, ο έλεγχος ροών κι αυτός Δέν Απορρίπτει ανεξαρτησία καταλοίπων.

Runs Test

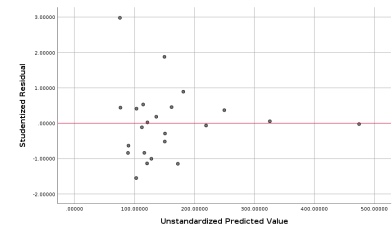
	Unstandardized Residual
Test Value ^a	-.99407
Cases < Test Value	11
Cases >= Test Value	12
Total Cases	23
Number of Runs	14
Z	.437
Asymp. Sig. (2-tailed)	.662

a. Median

Σχήμα 18: Έλεγχος Ροών Ανεξαρτησίας Καταλοίπων P-value = 0.662 > $\alpha = 0.05$



Σχήμα 15: Scatter-Plot Studentized Residuals



Σχήμα 17: Διάγραμμα Ελέγχου Ανεξαρτησίας Καταλοίπων

Άσκηση 5

Υποερώτημα 1

$$\text{AVGTRIPEXP} = 73.36 + 5.89 \cdot \text{AVGACCOMEXP}$$

Άυξηση κατά 50 ευρώ δαπάνης καταλυμάτων \Rightarrow

Άυξηση κατά 294.5 της ταξιδιωτικής δαπάνης (3)

$$\text{AVGTRIPEXP} = 132 + 30.1 \cdot \text{AVGDUREXP}$$

Μείωση κατά 103 ευρώ δαπάνης πολυτελών αγαθών \Rightarrow

Μείωση κατά 3100.3 της ταξιδιωτικής δαπάνης (4)

Βέβαια, για το μοντέλο (4) από έλεγχο

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Δέν Απορρίπτουμε την H_0 , άρα μια μεταβολή στο avgdurexp δεν πρόκειται να προκαλέσει μεταβολή στο avgtripexp.

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	132.113	26.438		.000
	Average durable and valuable goods expenditure per trip (in euros)	30.104	17.603	.350	.102

a. Dependent Variable: Average expenditure per trip (in euros)

Σχήμα 19: Έλεγχος Μοντέλου
avgtripexp-avgdurexp P-value =
0.102 > $\alpha = 0.05$

Άσκηση 6

Analyze \rightarrow Regression \rightarrow Linear \rightarrow avgtripexp with avgaccomexp \rightarrow

Save Mean, Individual, Conf.Int 93%

93%ΔΕ

1. μέση πρόβλεψη avgtripexp της Ιταλίας [155.77784 , 207.28088]

2. πρόβλεψη Βέλγικη avgtripexp [96.39631 , 342.46534]

Άσκηση 7

Άρχικα, θα μετασχηματίσουμε κατάλληλα τον έλεγχο ώστε να γραφί συναρτήσει του συντελεστή β_0 και στη συνέχεια θα μετασχηματίσουμε και τις μεταβλητές καταλλήλως για να εφαρμόσουμε τον έλεγχο υπόθεσης χωρίς την σταθερά (αφού θα γραφεί συναρτήσει του β_1)

$$H_0 : 2 \cdot \beta_1 - 5 \cdot \beta_0 = 4 \Rightarrow$$

$$\beta_0 = 0.4 \cdot \beta_1 - 0.8$$

```
COMPUTE avgtripexp_transf=
    avgtripexp + 0.8.
EXECUTE.
```

```
COMPUTE avgaccommexp_transf=
    avgaccommexp +0.4.
EXECUTE.
```

$$Y = \overbrace{0.4 \cdot \beta_1 - 0.8}^{\beta_0} + \beta_1 \cdot X \Rightarrow$$

$$Y^* = Y + 0.8$$

$$X^* = X + 0.4$$

(5)

Αποτέλεσμα ελέγχου :

Sig = 0.001

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	98.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	71.803	18.242		3.936	.001	25.877	117.729
	avgaccommexp_transf	5.885	.867	.829	6.786	.000	3.702	8.069

a. Dependent Variable: avgtripexp_transf

Σχήμα 20: Έλεγχος των μετασχηματισμένων μεταβλητών (ανεξάρτητη μεταβλητή avgaccommexp-transf). Διακρίνουμε Sig(β_0) = 0.001

Συνεπώς, απορρίπτουμε την $H_0 : 2 \cdot \beta_1 - 5 \cdot \beta_0 = 4$

Άσκηση 8

Μετασχηματίζουμε τα δεδομένα ως εξής :

$$y = e^{\beta_0 + \beta_1 \cdot x} \Rightarrow$$

$$\log(y) = \beta_0 + \beta_1 \cdot x \Rightarrow$$

$$\ln(\text{avgtripexp}) = 4.454 + 0.028 \cdot \text{avgtranspexp}$$

$$R^2 = 0.356$$

(6)

Όταν τα μεταφορικά κόστη ανέλθουν σε 31.8 τότε η μέση ταξιδιωτική δαπάνη θα εκτιμηθεί ως :

$$\ln(\text{avgtripexp}) = 4.454 + 0.028 \cdot 31.8 = 5.344 \Rightarrow$$

$$\text{avgtripexp} = e^{5.344}$$

(7)

```
COMPUTE avgtripexp_transf=
  avgtripexp + 0.8.
EXECUTE.
```

```
COMPUTE avgaccommexp_transf=
  avgaccommexp + 0.4.
EXECUTE.
```

```
DATASET ACTIVATE DataSet1.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI
    (98) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT avgtripexp_transf
  /METHOD=ENTER
  avgaccommexp_transf.
```

```
COMPUTE log_avgtripexp=LN(
  avgtripexp).
EXECUTE.
```

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI
    (95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT log_avgtripexp
  /METHOD=ENTER
  avgaccommexp.
```

```
COMPUTE log_avgtripexp=LN(
  avgtripexp).
EXECUTE.
```

* Encoding: UTF-8.

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI
    (95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT log_avgtripexp
  /METHOD=ENTER avgtranspexp.
```


Ερώτημα +

1. Ο ορισμός του επισκέπτη με βάση τα metadata της ιστοσελίδας είναι κάθε άτομο (Ευρωπαίο) που ταξιδεύει απο ένα μέρος διαφορετικό της κατοικίας του, σε ένα άλλο για λιγότερο απο 12 μήνες, με σκοπό διαφορετικό απο αυτόν που ασκεί στο καθημερινό περιβάλλον του.
2. Τρόποι συλλογής δεδομένων
 - Ερωτηματολόγια
 - Τηλεφωνική επικοινωνία
 - Συνέντευξη πρόσωπο με πρόσωπο
3. Δεδομένης της 15.2.1 παραγράφου απο τα Πολωνικά αρχεία,λόγω καινούριας μεθοδολογίας υπολογισμού δεδομένων(έτος 2014),δεν έχουν υπολογιστεί αυτά τα οποία προέρχονται απο το 2014