



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΠΕΡΙΓΡΑΦΙΚΗ και ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Επίλυση: Multiple Regression

Διδάσκων: Δαφέρμος Βασίλειος
ΤΜΗΜΑ ΠΟΛΙΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
ΣΧΟΛΗΣ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται στην άδεια χρήσης **Creative Commons** και ειδικότερα ***Αναφορά – Μη εμπορική Χρήση – Όχι Παράγωγο Έργο 3.0 Ελλάδα*** (***Attribution – Non Commercial – Non-derivatives 3.0 Greece***)



[ή επιλογή ενός άλλου από τους έξι συνδυασμούς]

[και αντικατάσταση λογότυπου άδειας όπου αυτό έχει μπει (σελ. 1, σελ. 2 και τελευταία)]

- Εξαιρείται από την ως άνω άδεια υλικό που περιλαμβάνεται στις διαφάνειες του μαθήματος, και υπόκειται σε άλλου τύπου άδεια χρήσης. Η άδεια χρήσης στην οποία υπόκειται το υλικό αυτό αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Τρίτη 20-5-14

ΠΕΡΙΓΡΑΦΙΚΗ ΚΑΙ ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Εισηγητής: Βασίλης Δαφέρμος, Αναπληρωτής Καθηγητής

MULTIPLE REGRESSION ANALYSIS



- Τι είναι και πότε κάνουμε **MULTIPLE REGRESSION ANALYSIS**
- **ΠΡΟΒΛΕΨΗ**. Πότε μπορούμε να κάνουμε πρόβλεψη.
- Η κατασκευή του Παλινδρομικού Μοντέλου.
- Από πού φαίνεται η προβλεπτική δύναμη του παλινδρομικού μας Μοντέλου.
- Πότε θα αποπλεύσουν τα Ελληνικά καράβια από την Αυλίδα της Βοιωτίας για την Τροία;
- Ή αλλιώς, πότε φυσάει ούριος άνεμος και οι οίωνοί είναι καλοί για την **MULTIPLE REGRESSION ANALYSIS**; Δυο σημαντικές ενδείξεις.
- Assumptions για την **MULTIPLE REGRESSION ANALYSIS**
- Αν δεν ικανοποιούνται οι παραδοχές τι κάνουμε;
- Μετασχηματισμός των Δεδομένων, Robust Regression, Προσομοίωση και άλλες Εναλλακτικές Λύσεις.
- Μέθοδοι Πολλαπλών Συγκρίσεων. Πότε χρησιμοποιούμε ποια.
- Μέθοδος Bonferroni
- Μέθοδος Tukey
- Μέθοδος Scheffe
- Κατασκευή και Ερμηνεία της εξίσωσης παλινδρόμησης.
- Outliers και Influential Points. Ποιες ακραίες τιμές είναι αθώες και ποιες υπονομεύουν το μοντέλο μας.
- Τελικά που γίνεται η ζημιά των παρατηρήσεων επίδρασης; Στους παλινδρομικούς συντελεστές, στην προβλεπτική δύναμη του μοντέλου ή και στα δυο;
- Τελικά τι κάνουμε με τους επικίνδυνους outliers ?? Τους πετάμε έξω και πως;
- Ή διαφορετικά: Πως τους εντοπίζουμε και ποιο πρόγραμμα μπορεί να κάνει αυτή τη δουλειά;
- Πρόβλημα- Παράδειγμα πάνω στην **MULTIPLE REGRESSION ANALYSIS**
- Λύση- Αποτελέσματα στο συγκεκριμένο παράδειγμα

Το πρόβλημα

Άσκηση_ *MULTIPLE_REGRESSION*

Ένας ερευνητής προκειμένου να εκτιμήσει το σημερινό μισθό που λαμβάνει ένας υπάλληλος που εργάζεται σε μια εταιρεία, θεώρησε πως σημαντικά στοιχεία γι' αυτήν την πρόβλεψη είναι ο αρχικός μισθός του υπαλλήλου, τα χρόνια υπηρεσίας του και τέλος το επίπεδο της μόρφωσής του. Για το σκοπό δε αυτό έλαβε τυχαίο δείγμα 20 υπαλλήλων και κατέγραψε τα στοιχεία τους (βλ. παρακάτω Πίνακα).

1. Να εκτελέσετε ανάλυση παλινδρόμησης με τη μέθοδο STEPWISE, με εξαρτημένη μεταβλητή το σημερινό μισθό (**SALARY**) και ανεξάρτητες τον αρχικό μισθό (**SAL1**), τα χρόνια υπηρεσίας (**XRONIA**), και το επίπεδο μόρφωσης (**EDLEVEL**). Ποιο είναι το παλινδρομικό σας μοντέλο σύμφωνα με τη μέθοδο αυτή, πόσο καλά προσαρμόζεται στα δεδομένα σας και σε ποια εξίσωση υπακούει ;
2. Να αναφέρετε τις παραδοχές της πολλαπλής παλινδρόμησης και στη συνέχεια να τις ελέγξετε σχολαστικά.
3. Ποια είναι η 'άριστη' μεταβλητή πρόβλεψης ('BEST PREDICTOR') αν εργαστούμε με τη μέθοδο STEPWISE ;
4. Ήταν ορθή η απόφασή μας να λάβουμε **3 ανεξάρτητες μεταβλητές** στην οικοδόμηση του παλινδρομικού μας μοντέλου; Μήπως έπρεπε να λάβουμε λιγότερες ανεξάρτητες μεταβλητές από όσες έχουμε, από όσες δηλ. διαθέτει ο παρακάτω Πίνακας, ή μήπως η έρευνα θα έπρεπε να αναζητήσει, από τη συγκεκριμένη Εταιρεία, περισσότερες ανεξάρτητες μεταβλητές; Τελικά ποιο είναι το κατάλληλο πλήθος ανεξάρτητων μεταβλητών που έχει ανάγκη η παλινδρομική μας ανάλυση για να είναι η 'βέλτιστη'; Ποιος δείκτης είναι σε θέση να μας παράσχει την αναγκαία πληροφορία;

5. Ποιο ακριβώς είναι το υποσύνολο των ανεξάρτητων μεταβλητών, δηλ. ποιες ακριβώς ανεξάρτητες μεταβλητές θα πρέπει να συμμετάσχουν στο παλινδρομικό μας μοντέλο, ώστε αυτό να μπορέσει να κάνει την καλύτερη δυνατή πρόβλεψη; Ποιες είναι οι πολιτικές που μας βοηθούν να πετύχουμε ένα 'άριστο' παλινδρομικό μοντέλο;

Πίνακας. Τα στοιχεία 20 υπαλλήλων μιας εταιρείας τυχαία επιλεγμένων

Αύξων αριθμός εγγραφής ή σειρά εισαγωγής δεδομένων (SEQUENCE)	Σημερινός Μισθός σε ευρώ (SALARY)	Αρχικός Μισθός σε ευρώ (SAL1)	Χρόνια Υπηρεσίας (XRONIA)	Επίπεδο Εκπαίδευσης (1=Χαμηλό, 2=Μέσο, 3=Ανώτερο) (EDLEVEL)
1	1400	150	17	3
2	890	90	12	1
3	1200	120	11	2
4	1250	200	13	3
5	1150	130	9	2
6	990	500	7	1
7	800	400	5	1
8	900	400	5	1
9	1700	520	19	3
10	1500	250	14	3
11	1250	300	6	3
12	1500	700	14	3
13	950	240	11	2
14	890	300	7	1
15	750	80	3	1

16	1000	300	4	1
17	980	120	10	1
18	870	300	4	1
19	1150	350	7	2
20	900	250	6	1

ΛΥΣΗ- ΑΠΟΤΕΛΕΣΜΑΤΑ

Descriptive Statistics			
	Mean	Std. Deviation	N
SALARY	1101,00	263,836	20
SAL1	285,00	161,196	20
XRONIA	9,20	4,561	20
EDLEVEL	1,80	,894	20

Correlations

		SALARY	SAL1	XRONIA	EDLEVEL
Pearson Correlation	SALARY	1,000	,344	,813	,900
	SAL1	,344	1,000	,092	,190
	XRONIA	,813	,092	1,000	,733
	EDLEVEL	,900	,190	,733	1,000
Sig. (1-tailed)	SALARY	.	,069	,000	,000
	SAL1	,069	.	,349	,211
	XRONIA	,000	,349	.	,000
	EDLEVEL	,000	,211	,000	.
N	SALARY	20	20	20	20
	SAL1	20	20	20	20
	XRONIA	20	20	20	20
	EDLEVEL	20	20	20	20

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	EDLEVEL		. Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
2	XRONIA		. Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
3	SAL1		. Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).

a. Dependent Variable: SALARY

Model Summary^d

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,900 ^a	,809	,799	118,316	
2	,928 ^b	,860	,844	104,199	
3	,947 ^c	,898	,878	92,008	1,780

a. Predictors: (Constant), EDLEVEL

b. Predictors: (Constant), EDLEVEL, XRONIA

c. Predictors: (Constant), EDLEVEL, XRONIA, SAL1

d. Dependent Variable: SALARY

Coefficients ^a													
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	623,289	60,695		10,269	,000	495,774	750,805					
	EDLEVEL	265,395	30,348	,900	8,745	,000	201,637	329,152	,900	,900	,900	1,000	1,000
2	(Constant)	575,808	56,748		10,147	,000	456,080	695,536					
	EDLEVEL	193,668	39,281	,657	4,930	,000	110,792	276,544	,900	,767	,447	,463	2,160
	XRONIA	19,194	7,704	,332	2,492	,023	2,941	35,448	,813	,517	,226	,463	2,160

3	(Constant)	501,004	58,950		8,499	,000	376,035	625,972				
	EDLEVEL	178,352	35,263	,605	5,058	,000	103,597	253,107	,900	,784	,405	,448
	XRONIA	20,344	6,819	,352	2,983	,009	5,889	34,800	,813	,598	,239	,461
	SAL1	,322	,134	,197	2,409	,028	,039	,606	,344	,516	,193	,959
a. Dependent Variable: SALARY												

Collinearity Diagnostics ^a							
Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	EDLEVEL	XRONIA	SAL1
1	1	1,900	1,000	,05	,05		
	2	,100	4,359	,95	,95		
2	1	2,833	1,000	,02	,01	,01	
	2	,116	4,943	,98	,12	,12	
	3	,051	7,479	,00	,86	,87	
3	1	3,632	1,000	,01	,01	,01	,01
	2	,227	4,002	,01	,05	,09	,61
	3	,092	6,287	,96	,11	,02	,32
	4	,049	8,594	,03	,83	,89	,05

a. Dependent Variable: SALARY

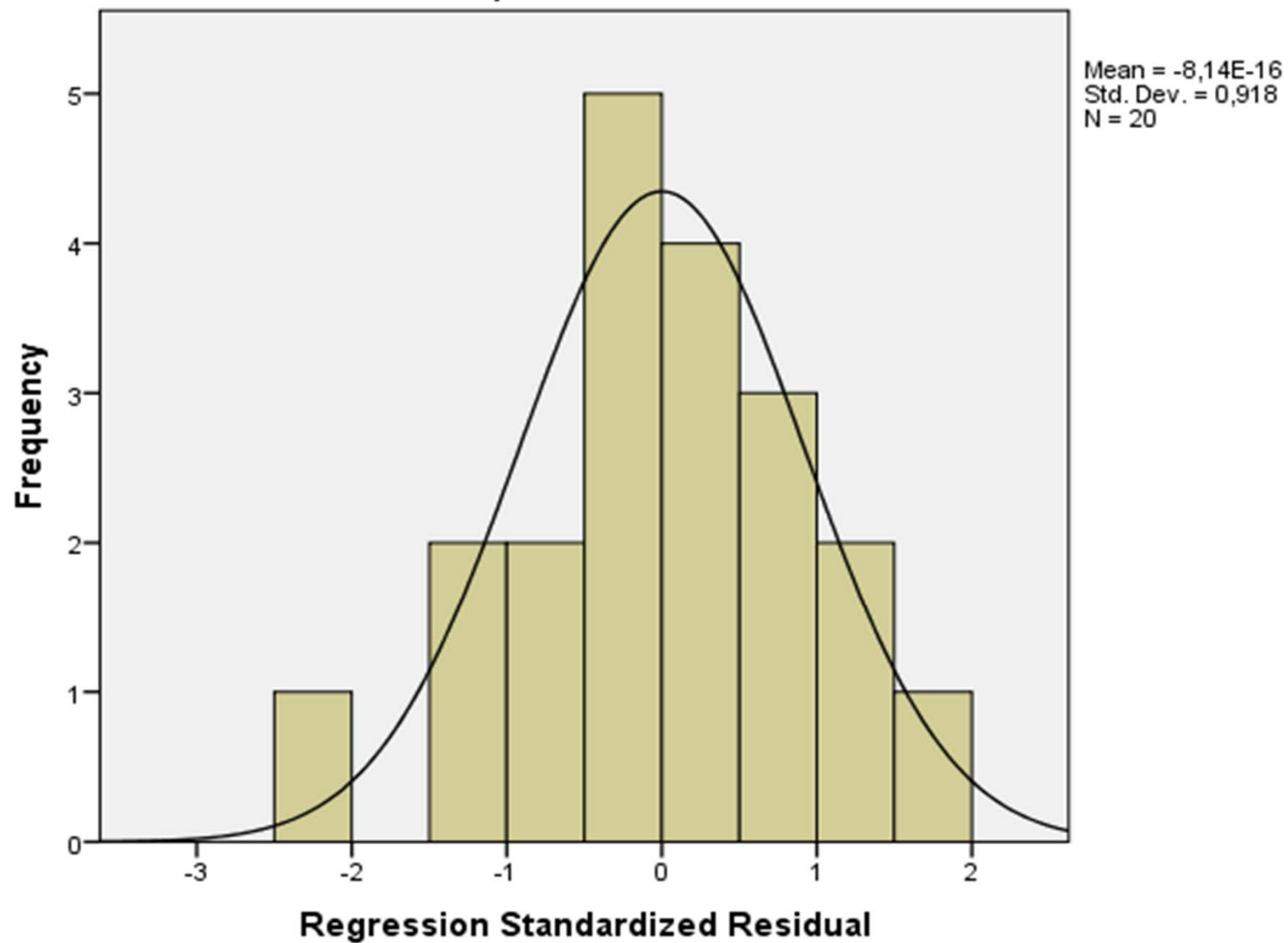
Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	766,16	1590,09	1101,00	249,961	20
Std. Predicted Value	-1,340	1,957	,000	1,000	20
Standard Error of Predicted Value	23,198	63,314	39,582	11,534	20
Adjusted Predicted Value	770,82	1583,10	1100,71	247,987	20
Residual	-208,797	142,641	,000	84,432	20
Std. Residual	-2,269	1,550	,000	,918	20
Stud. Residual	-2,345	1,654	,002	1,005	20
Deleted Residual	-222,971	181,652	,290	102,429	20
Stud. Deleted Residual	-2,803	1,759	-,015	1,083	20
Mahal. Distance	,258	8,047	2,850	2,289	20
Cook's Distance	,001	,385	,054	,087	20
Centered Leverage Value	,014	,424	,150	,120	20

a. Dependent Variable: SALARY

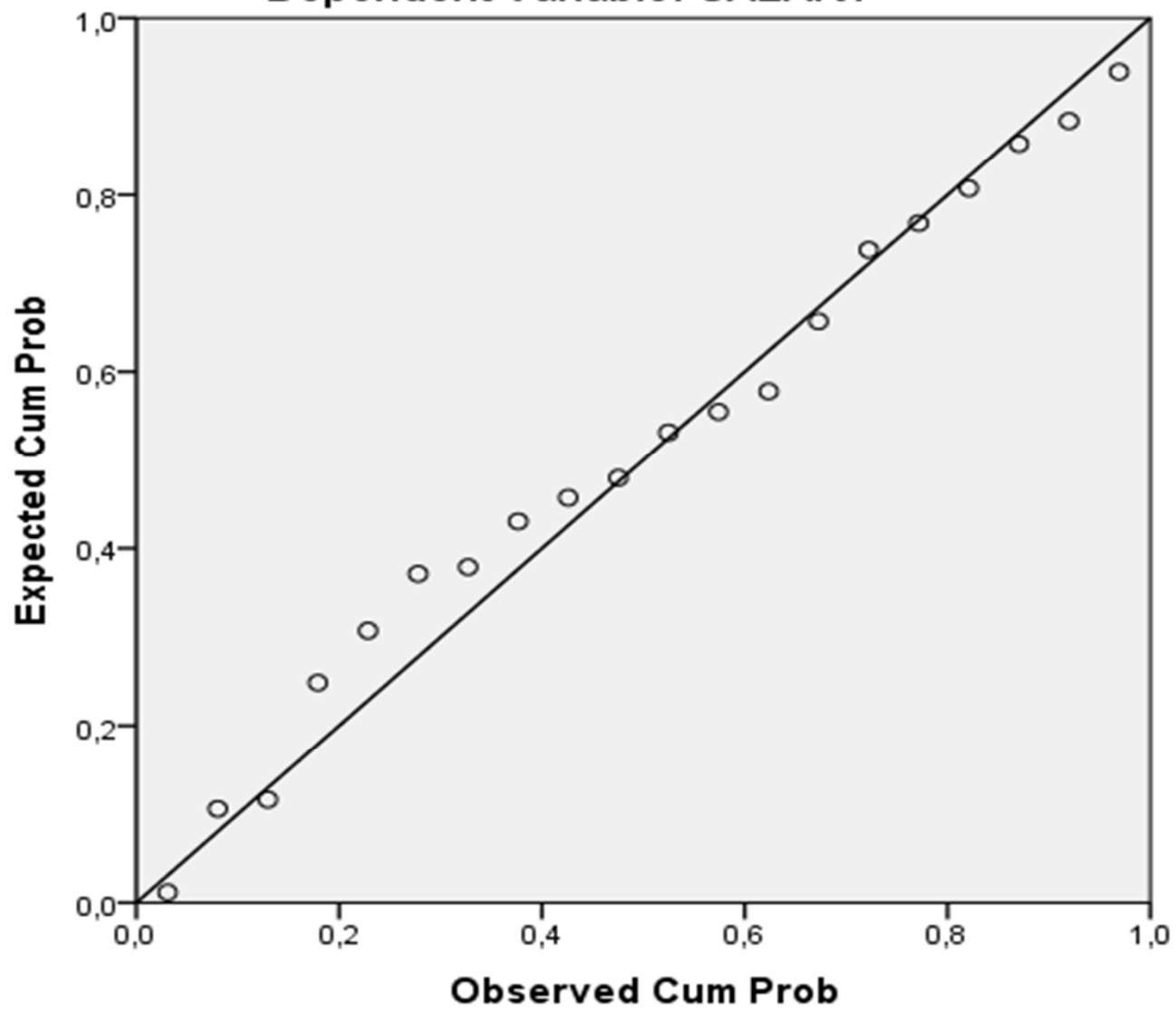
Histogram

Dependent Variable: SALARY



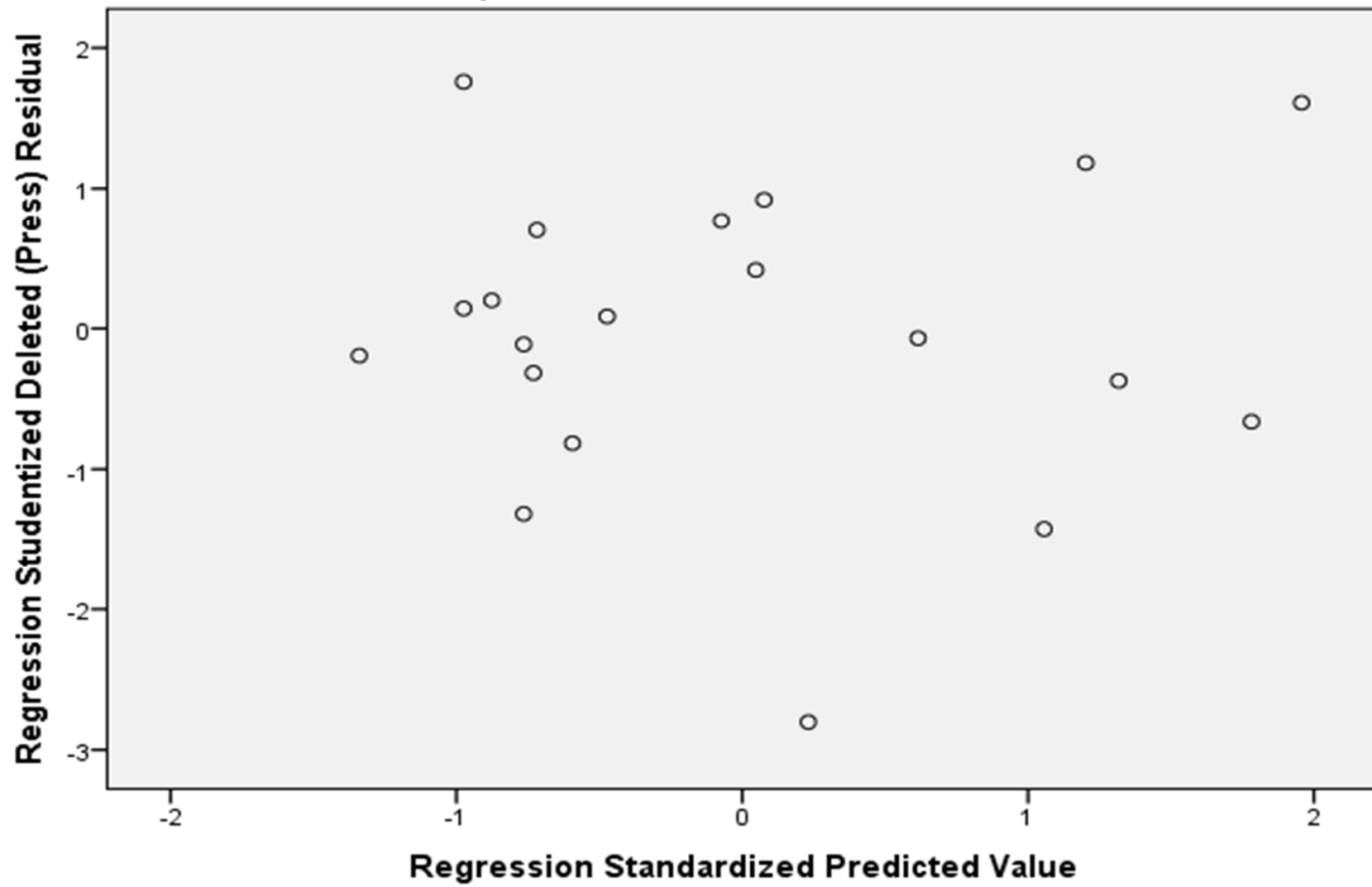
Normal P-P Plot of Regression Standardized Residual

Dependent Variable: SALARY

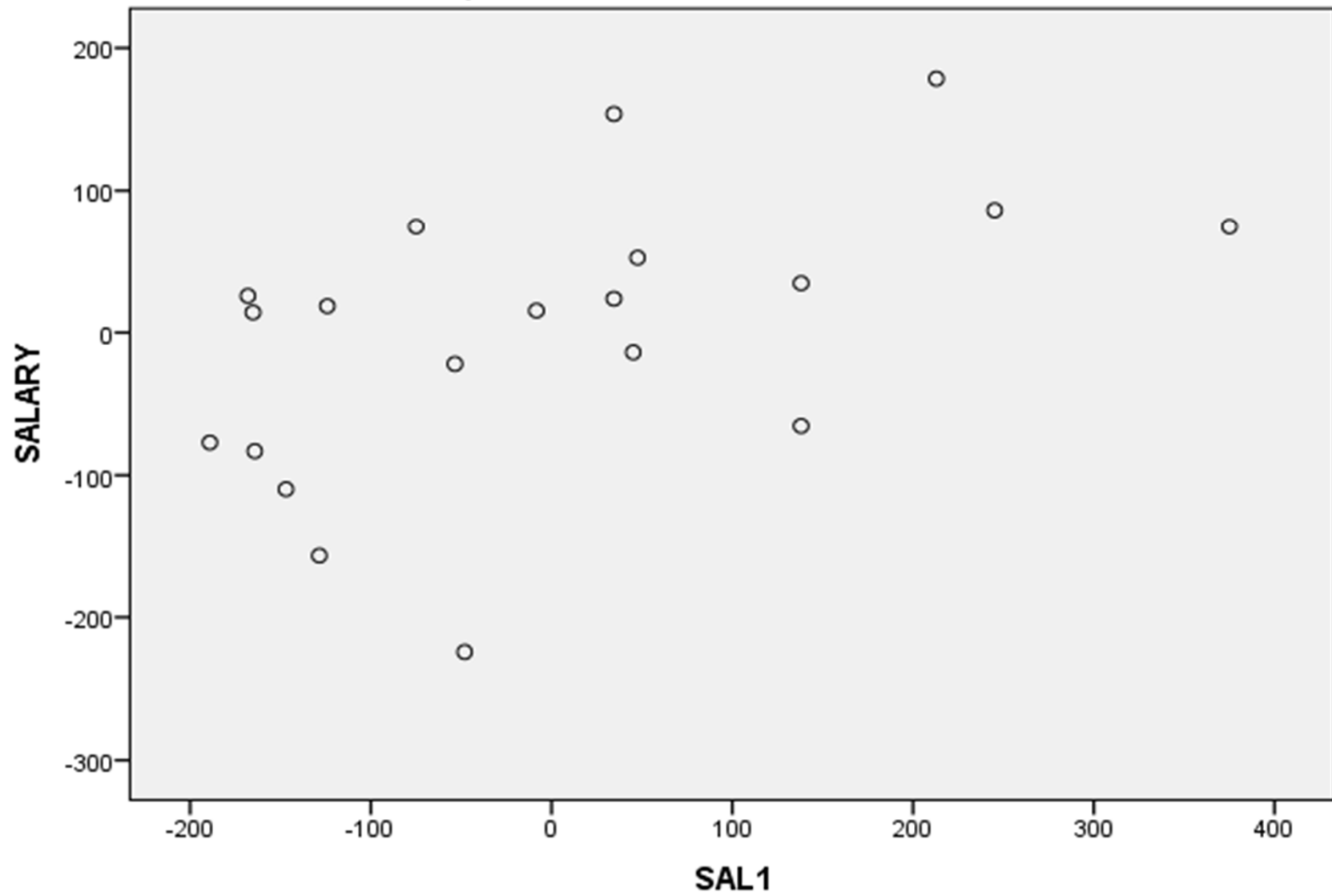


Scatterplot

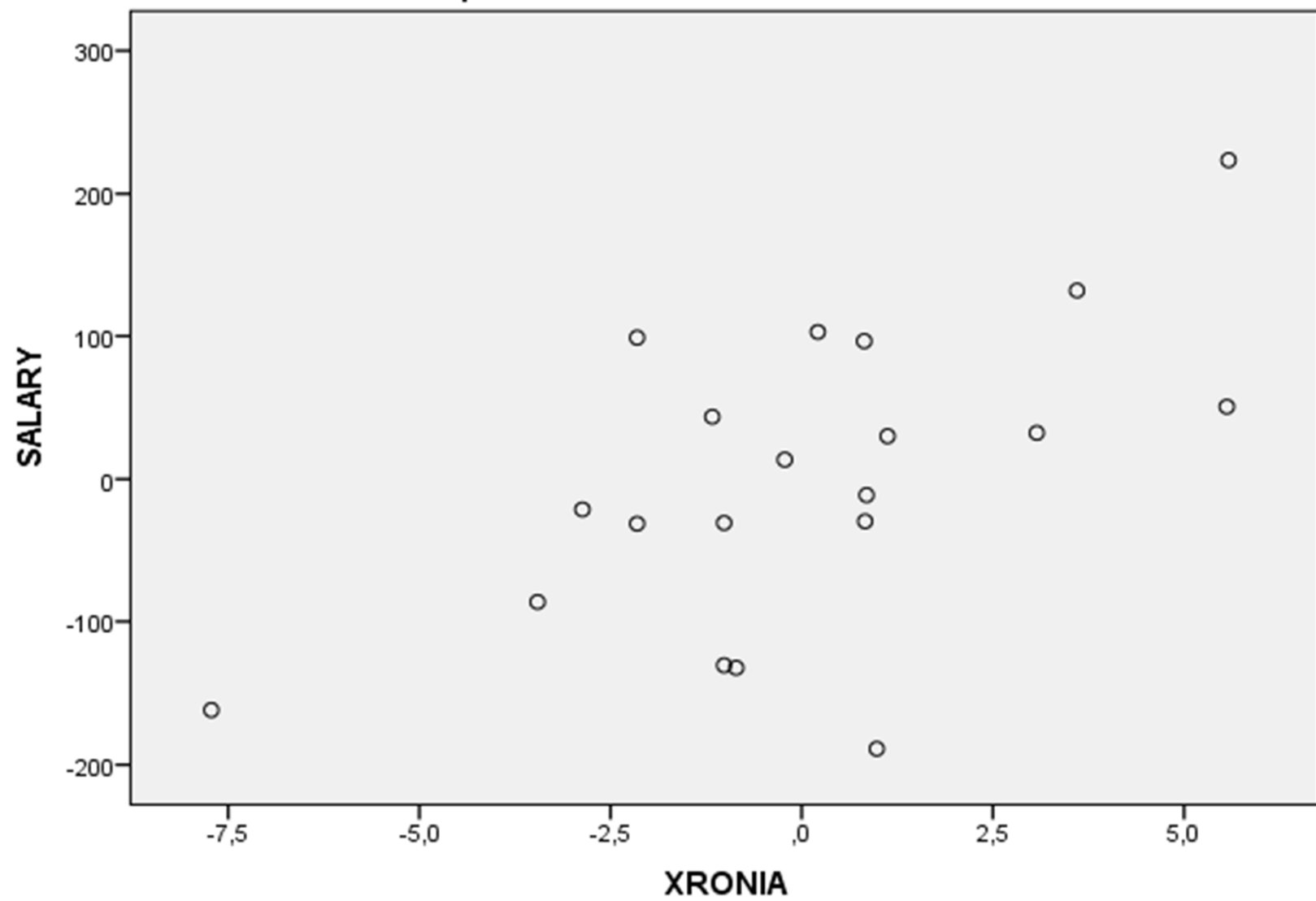
Dependent Variable: SALARY



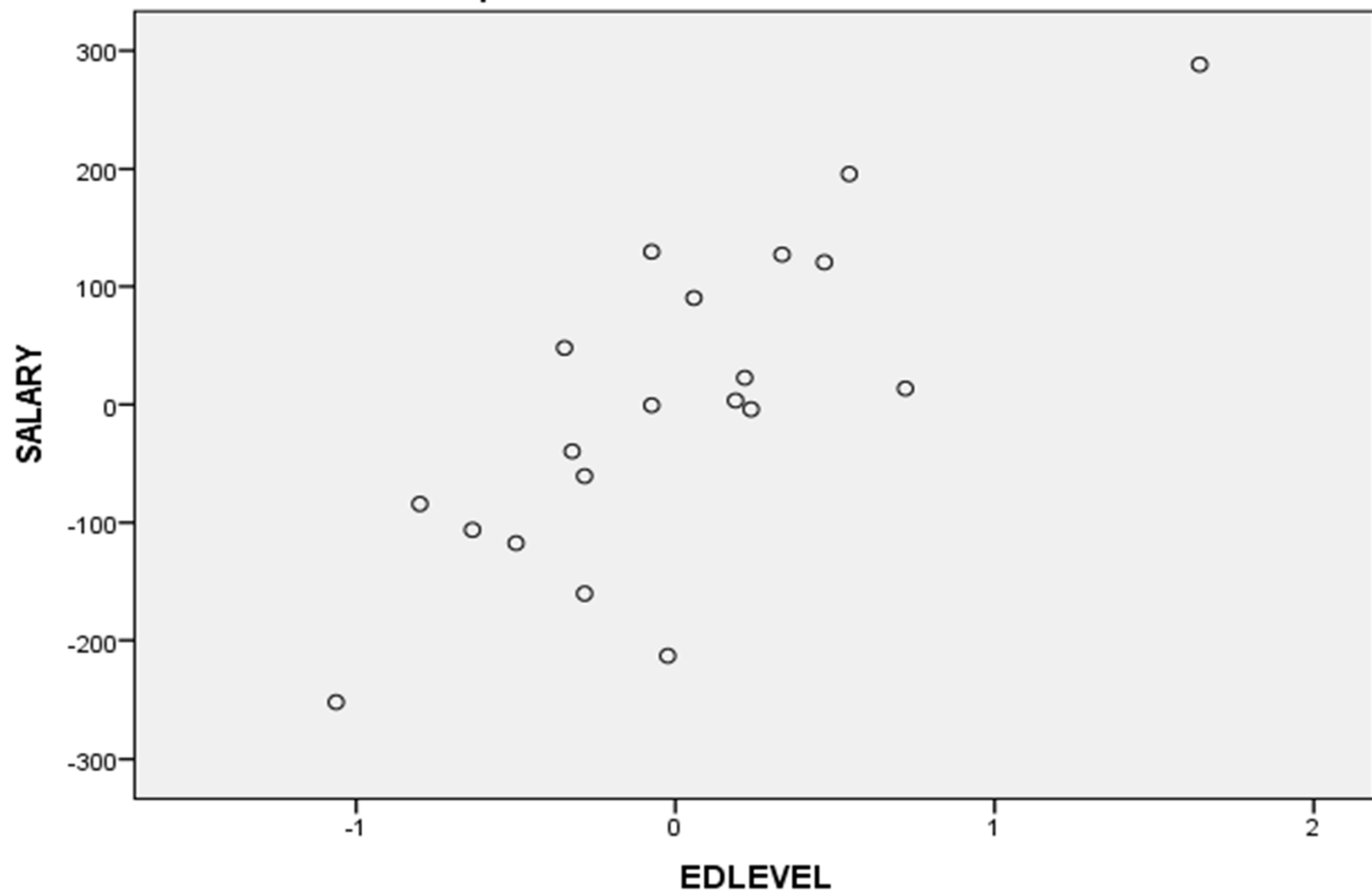
Partial Regression Plot
Dependent Variable: SALARY



Partial Regression Plot
Dependent Variable: SALARY



Partial Regression Plot
Dependent Variable: SALARY



Τέλος Ενότητας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

