ΤΕΙ ΠΕΙΡΑΙΑ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΕΠΙΣΤΗΜΗ ΤΩΝ ΑΠΟΦΑΣΕΩΝ ΜΕ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΜΑΘΗΜΑ:

Ανάλυση Πολυδιάστατων (Πολυμεταβλητών) Δεδομένων και Συστήματα Εξόρυξης Δεδομένων (Multivariate Data Analysis and Data Mining Systems)

Θέμα εργασίας: **Ατομική Εργασία 1η**

Ονοματεπώνυμο φοιτητή: **ΧΟΥΡΔΑΚΗΣ ΕΥΣΤΡΑΤΙΟΣ** (Α.Μ.1317)

Επιβλέπουσα καθηγήτρια: Μοσχονά θ.

Μάιος 2014

Μέρος Α.

1. Μέτρα Θέσης (Δείκτες Κεντρικής Τάσης) της μεταβλητής "Bonus"

Η μεταβλητή "Bonus" αφορά στην ετήσια πρόσθετη οικονομική ανταμοιβή ενός δείγματος τριάντα (30) εργαζόμενων μιας επιχείρησης. Η εν λόγω μεταβλητή είναι ποσοτική και συνεχής. Έτσι, τα κυριότερα μέτρα θέσεως είναι:
α) Αριθμητικός μέσος (Mean)
β) Διάμεσος (Median)
γ) Επικρατούσα τιμή (Mode)
δ) Τεταρτημόρια

Με τη βοήθεια του SPSS (Analyze/Descriptive Statistics/Frequencies) αντλώ τους πίνακες 1 και 2.

Πίνακας 1

Statistics

BONUS		
N	Valid	30
	Missing	0
Mean		**739,3333**
Median		**750,00**
Mode		**800,00**
Percentiles	**25**	**695,00**
	50	750,00
	75	**800,00**

Πίνακας 2

BONUS

	Frequency	Percent	Valid Percent	Cumulative Percent
450,00	1	3,3	3,3	3,3
540,00	1	3,3	3,3	6,7
630,00	1	3,3	3,3	10,0
660,00	2	6,7	6,7	16,7
680,00	2	6,7	6,7	23,3
700,00	1	3,3	3,3	26,7
730,00	1	3,3	3,3	30,0
740,00	2	6,7	6,7	36,7
750,00	5	16,7	16,7	53,3
760,00	2	6,7	6,7	60,0
780,00	1	3,3	3,3	63,3
800,00	**6**	20,0	20,0	**83,3**
820,00	3	10,0	10,0	93,3
830,00	2	6,7	6,7	100,0
Total	30	100,0	100,0	

Από τον **Πίνακα 1** παρατηρούμε ότι το μέγεθος του δείγματος είναι N=30 και δεν υπάρχουν ελλείπουσες τιμές (missing values=0). Από τον ίδιο Πίνακα προσδιορίζω τα ζητούμενα μέτρα κεντρικής θέσης. Ειδικότερα:

α) Η **μέση τιμή** (δειγματικός μέσος) της μεταβλητής "Bonus" είναι: $\boldsymbol{\mu = 739,33}$ €

β) Η **διάμεσος** είναι: **M=750,00 €** . Ομοίως, από τον **Πίνακα 2** διαπιστώνουμε ότι διατάσσοντας τις τιμές που λαμβάνει η μεταβλητή Bonus κατά αύξουσα σειρά, η τιμή της 15^{ης} παρατήρησης (όπως αυτή μπορεί να προσδιοριστεί από τη στήλη frequency) είναι η *διάμεσος* η οποία ισούται με 750€.

γ) Η **επικρατούσα τιμή** είναι: **M_o=800,00 €.** Ομοίως, από τον Πίνακα 2 επαληθεύεται η ανωτέρω διαπίστωση καθόσον η τιμή αυτή (800,00€), που λαμβάνει η μεταβλητή bonus, εμφανίζεται έξι (6) φορές, δηλ. περισσότερες από κάθε άλλη τιμή.

δ) Το πρώτο τεταρτημόριο (Q_1) έχει τιμή **695,00€** (Πίνακας 1, Percentile 25)**.** Διατάσσοντας τις τιμές της μεταβλητής Bonus κατά αύξουσα σειρά, η **θέση**[1] του Q_1 είναι η $\frac{30+1}{4}$ = 7,75 διατεταγμένη παρατήρηση και η **τιμή** του είναι: $\frac{1}{4}\,680 + \frac{3}{4}\,700 = \mathbf{695}€$ (όπου, 680 είναι η τιμή της $7^{ης}$ διατεταγμένης παρατήρησης και 700 είναι η τιμή της $8^{ης}$ διατεταγμένης παρατήρησης, Πίνακας 2). Από τον ανωτέρω δείκτη συμπεραίνουμε ότι το 25% των εργαζομένων του δείγματος που λαμβάνουν χαμηλό ετήσιο bonus εισπράττουν (ως bonus) λιγότερα από 695,00€/έτος.

Το τρίτο τεταρτημόριο (Q_3) έχει τιμή **800,00€** (Πίνακας 1, Percentile 75)**.** Ομοίως με ανωτέρω, διατάσσοντας τις τιμές της μεταβλητής Bonus κατά αύξουσα σειρά, η **θέση** του Q_3 είναι $\frac{3\times 30+1}{4} = 22{,}75$ και η **τιμή** του είναι **800€** (η $22^{η}$ και η $23^{η}$ διατεταγμένη παρατήρηση έχουν την τιμή 800, και συνεπώς δεν γεννάται θέμα στάθμισης της τιμής του Q_3). Από τον δείκτη αυτό, στη συγκεκριμένη άσκηση, δεν μπορούμε να βγάλουμε κάποιο ασφαλές ποσοτικό συμπέρασμα (όπως, π.χ. ότι το 75% των εργαζομένων λαμβάνει ετήσιο bonus έως 800€) καθόσον η τιμή του ταυτίζεται με την επικρατούσα τιμή (M_o) την οποία έχουν οι $19^{η}$,$20^{η}$,$21^{η}$,$22^{η}$,$23^{η}$ και $24^{η}$ διατεταγμένες παρατηρήσεις. Σημειώνεται όμως, όπως προκύπτει από τον Πίνακα 2, ότι το **83,3%** των εργαζομένων λαμβάνει ετήσιο bonus έως και 800€.

Από τη σύγκριση των τιμών των Δεικτών Κεντρικής Τάσης διαπιστώνεται ότι:

$$\boldsymbol{\mu < M < M_o}$$

Γνωρίζουμε ότι στον υπολογισμό της *μέσης τιμής* λαμβάνονται υπόψη και οι τριάντα παρατηρήσεις και ως εκ τούτου οι ακραίες παρατηρούμενες τιμές επηρεάζουν την τιμή της *μέσης τιμής*. Αντίθετα, η *διάμεσος* δεν επηρεάζεται από την ύπαρξη ακραίων τιμών. Έτσι, το ότι η *μέση τιμή* είναι μικρότερη της *διαμέσου* οφείλεται στην ύπαρξη κάποιων παρατηρούμενων χαμηλών τιμών που απομακρύνουν τη *μέση τιμή* από την τιμή της *διαμέσου*. Επιπλέον, επειδή η *επικρατούσα τιμή* (M_o) είναι μεγαλύτερη της *διαμέσου* (M), η απόκλιση της *μέσης τιμής* (*μ*) από την *διάμεσο* είναι μικρή.

Η ύπαρξη ακραίων παρατηρούμενων τιμών διαπιστώνεται τόσο από το σχετικό φυλλογράφημα[2] (Πίνακας 3), όσο και από το θηκόγραμμα[3] (Box Plot, Γράφημα 1):

Πίνακας 3

BONUS Stem-and-Leaf Plot

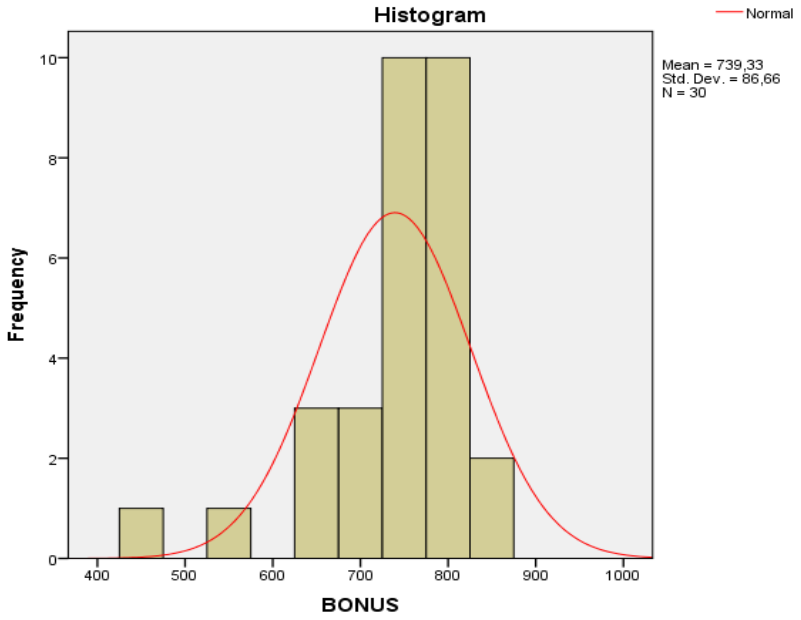
Frequency	Stem &	Leaf
2,00	**Extremes**	**(=<540)**
1,00	6 .	3
4,00	6 .	6688
4,00	7 .	0344
8,00	7 .	55555668
11,00	8 .	00000022233

Stem width: 100,00
Each leaf: 1 case(s)

Γράφημα 1

Το ιστόγραμμα των παρατηρούμενων τιμών και η καμπύλη κανονικής κατανομής απεικονίζονται στο Σχήμα 1.

Σχήμα 1



Κατηγοριοποιώντας[4] τις τιμές που λαμβάνει η μεταβλητή Bonus έχουμε τον πίνακα κατανομής συχνοτήτων της νέας μεταβλητής Bonus_Class (Πίνακας 4), από τον οποίο βλέπουμε ότι το 56,7% των εργαζομένων του δείγματος (ή 17 στους 30 εργαζόμενους) λαμβάνουν ετήσιο bonus μεταξύ 701 και 800 €, ενώ ένα σημαντικό ποσοστό των εργαζομένων (16,7%) του δείγματος λαμβάνει ετήσιο bonus πάνω από 801€.

Πίνακας 4

BONUS_CLASS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	400-500	1	3,3	3,3	3,3
	501-600	1	3,3	3,3	6,7
	601-700	6	20,0	20,0	26,7
	701-800	**17**	**56,7**	**56,7**	83,3
	801-++	5	16,7	16,7	100,0
	Total	30	100,0	100,0	

2. Μέτρα διασποράς της μεταβλητής Bonus.

Τα κυριότερα μέτρα διασποράς της μεταβλητής Bonus είναι:

α) Η διακύμανση (Variance)

β) Η τυπική απόκλιση (Standard Deviation)

γ) Το τυπικό σφάλμα του μέσου (Standard error of mean)

δ) Το εύρος (Range)

ε) Ο συντελεστής μεταβλητότητας (Coefficient of Variation)

στ) Το ενδοτεταρτημοριακό εύρος (Interquartile Range)

Με τη βοήθεια του SPSS (Analyze/Descriptive Statistics/Explore) αντλούμε τον **Πίνακα 5**, από τον οποίο λαμβάνουμε τα κυριότερα μέτρα διασποράς.

Πίνακας 5

Descriptives

			Statistic	Std. Error
BONUS	Mean		739,3333	**15,82180**
	95% Confidence Interval for Mean	Lower Bound	706,9741	
		Upper Bound	771,6926	
	5% Trimmed Mean		748,7037	
	Median		750,0000	
	Variance		**7509,885**	
	Std. Deviation		**86,65959**	
	Minimum		**450,00**	
	Maximum		**830,00**	
	Range		**380,00**	
	Interquartile Range		**105,00**	
	Skewness		-1,728	,427
	Kurtosis		3,569	,833

Ειδικότερα:

Η τιμή της ***διακύμανσης*** (*Variance*) είναι s^2=**7509,88** και η τετραγωνική ρίζα της διακύμανσης, δηλ. η ***τυπική απόκλιση*** *(Std Deviation)*, είναι s =**86,66 €**. Η *τυπική απόκλιση* (ή *διασπορά*) μάς πληροφορεί για το βαθμό διασκόρπισης των τιμών (του δείγματος) της μεταβλητής Bonus εκατέρωθεν της *μέσης τιμής* (Mean).

Το ***τυπικό σφάλμα του μέσου*** (Standard Error of mean), είναι μια καλή εκτίμηση του τυπικού σφάλματος της μέσης τιμής στον πληθυσμό[5]. Επομένως, έχει αξία όταν προσπαθούμε να εκτιμήσουμε τη *μέση τιμή* της ελεγχόμενης μεταβλητής (Bonus) στον πληθυσμό (στο σύνολο των εργαζομένων της επιχείρησης) γνωρίζοντας την τυπική απόκλιση ενός δείγματος (από τον πληθυσμό). Συνεπώς, το *τυπικό σφάλμα του μέσου* για τη μεταβλητή Bonus στον πληθυσμό (δηλ. στο σύνολο των εργαζομένων της επιχείρησης) είναι ίσο με **15,82€** ($\sigma_{\bar{x}}=\frac{86,659}{\sqrt{30}}$=15,82)[6].

Το ***Εύρος τιμών*** που παίρνει η μεταβλητή Bonus, δηλ. η διαφορά μεταξύ της μεγαλύτερης και της μικρότερης τιμής, είναι: Range= **380,00€** (Range = $X_{max} - X_{min}$= 830,00€ – 450,00€=380,00€).

Εναλλακτικά από το SPSS μπορούμε να προσδιορίσουμε το *Εύρος τιμών* ακολουθώντας τη διαδρομή εντολών : Analyze/Descriptive Statistics/Descriptive απ' όπου λαμβάνουμε τον Πίνακα 6.

Πίνακας 6

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean
BONUS	30	380,00	450,00	830,00	739,3333
Valid N (listwise)	30				

Ο ***συντελεστής μεταβλητότητας*** (Coefficient of Variation) είναι ένας δείκτης **ομοιογένειας** των τιμών της μεταβλητής Bonus, και είναι ο λόγος της *τυπικής απόκλισης* (σ) προς τη *μέση τιμή* (μ) της εν λόγω μεταβλητής:

$$CV = \frac{\sigma}{\mu} = \frac{86,659}{739,333} = 0,117 \text{ ή } 11,7\%$$

Σύμφωνα με τη θεωρία[7] οι τιμές που λαμβάνει μια μεταβλητή χαρακτηρίζονται ομοιογενείς όταν ο συντελεστής μεταβλητότητας δεν ξεπερνά το 10%. Συνεπώς, μπορούμε να υποστηρίξουμε ότι τα bonus των εργαζομένων της συγκεκριμένης επιχείρησης δεν φαίνεται να απέχουν αισθητά από το να χαρακτηριστούν ως ομοιογενή.

Τέλος, όπως φαίνεται από τον Πίνακα 5, το ***ενδοτεταρτημοριακό εύρος*** (Interquartile Range), δηλ η διαφορά μεταξύ της τιμής του τρίτου και του πρώτου τεταρτημορίου ($Q_3 - Q_1$ = 800,00€ – 695,00€), είναι **105,00€**. Αυτό καταδεικνύει ότι το 50% των ετήσιων πρόσθετων αμοιβών (bonus), που βρίσκονται μεταξύ του πρώτου και του τρίτου τεταρτημορίου, δεν αποκλίνουν μεταξύ τους περισσότερο από 105€.

3. Μέτρα μορφής της μεταβλητής Bonus.

Τα κυριότερα μέτρα μορφής είναι:

α) η ασυμμετρία (skewness) , και

β) η κύρτωση (kurtosis)

Όπως φαίνεται στον **Πίνακα 5** η τιμή της **ασυμμετρίας** (S_K) είναι: **S_K= -1,728.** Αρνητική τιμή ασυμμετρίας (S_K<0) σημαίνει ότι η καμπύλη κατανομής (της μεταβλητής Bonus) παρουσιάζει αριστερή ασυμμετρία δηλ. η καμπύλη έχει την ουρά της στα αριστερά. Αυτό συνεπάγεται ότι το μεγαλύτερο πλήθος των παρατηρήσεων (δηλ. των ετήσιων πρόσθετων αμοιβών) βρίσκονται στο αριστερό μέρος της κορυφής (M_o) της καμπύλης κατανομής[8]. **Συνεπώς, η επιχείρηση ανταμείβει το μεγαλύτερο πλήθος των εργαζομένων του δείγματος με bonus μικρότερα της *επικρατούσας τιμής* (M_o)**. Το συμπέρασμα αυτό συνάγεται αβίαστα και από τον **Πίνακα 2** όπου παρατηρούμε ότι δεκαεννέα (19) τιμές της μεταβλητής Bonus είναι μικρότερες της *επικρατούσας τιμής* (Mo=800€), ενώ μόλις πέντε (5) τιμές είναι μεγαλύτερες από αυτή.

Προκειμένου να αξιολογήσουμε την ένταση της ασυμμετρίας της καμπύλης κατανομής της μεταβλητής Bonus, θα εξετάσουμε το πηλίκο: $\frac{Sk}{St\ Error\ of\ Sk} = \frac{-1,728}{0,427} = -4,047$. Επειδή ο βαθμός ασυμμετρίας είναι μικρότερος του -2, η ασυμμετρία αξιολογείται ως **έντονη**.

Η αριστερή ασυμμετρία της καμπύλης κατανομής της μεταβλητής Bonus αναδεικνύεται και από τη σύγκριση των τιμών της *μέσης τιμής (μ)*, της *διαμέσου (M)* και της *επικρατούσας τιμής (M_o)*, που όπως είδαμε παρουσιάζουν τη σχέση: $\mu < M < M_o$

Από τον **Πίνακα 5** προσδιορίζεται η τιμή της **κύρτωσης** (Ku), η οποία είναι: **Ku= 3,569**. Επειδή ο συντελεστής κύρτωσης έχει θετική τιμή (Ku>0) συμπεραίνουμε ότι η καμπύλη κατανομής της μεταβλητής Bonus είναι **λεπτόκυρτη**[9], δηλ. αρκετές παρατηρήσεις συγκεντρώνονται γύρω από την επικρατούσα τιμή (M_o).

Προκειμένου να αξιολογήσουμε την ένταση της κυρτότητας της καμπύλης κατανομή της μεταβλητής Bonus, θα εξετάσουμε το πηλίκο: $\frac{Ku}{St\ Error\ of\ Ku} = \frac{3,569}{0,833} = 4,284$. Επειδή ο βαθμός κύρτωσης είναι μεγαλύτερος του 2, η κύρτωση αξιολογείται ως **έντονη**.

Ένας δεύτερος τρόπος αξιολόγησης της έντασης της κύρτωσης είναι ο έλεγχος του διαστήματος (kurtosis -2*St. error of kurtosis , kurtosis + 2*St error of kurtosis) που είναι (3,569-2*0,833 , 3,569+2*0,833) ή (1,903 , 5,235), **το οποίο δεν περιλαμβάνει το μηδέν** και συνεπώς αξιολογήσουμε τον βαθμό κύρτωσης ως σημαντικό (έντονο).

4. Μέτρα θέσεως της μεταβλητής Bonus για κάθε φύλο ξεχωριστά.

Με τη βοήθεια του SPSS (Analyze/Descriptive Statistics/Explore) και θέτοντας ως Factor List το Φύλο (Gender) αντλούμε τον **Πίνακα 7**, από τον οποίο βλέπουμε ότι από τις τριάντα (30) παρατηρήσεις της μεταβλητής Bonus, οι δέκα οκτώ (18) αφορούν στις γυναίκες και οι δώδεκα (12) στους άνδρες, και τον **Πίνακα 8**, από τον οποίο λαμβάνουμε τα κυριότερα μέτρα θέσης της μεταβλητής Bonus για κάθε φύλο ξεχωριστά:

Πίνακας 7

Case Processing Summary

	ΦΥΛΟ	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
BONUS	ΓΥΝΑΙΚΑ	18	100,0%	0	0,0%	18	100,0%
	ΑΝΤΡΑΣ	12	100,0%	0	0,0%	12	100,0%

Πίνακας 8

Descriptives

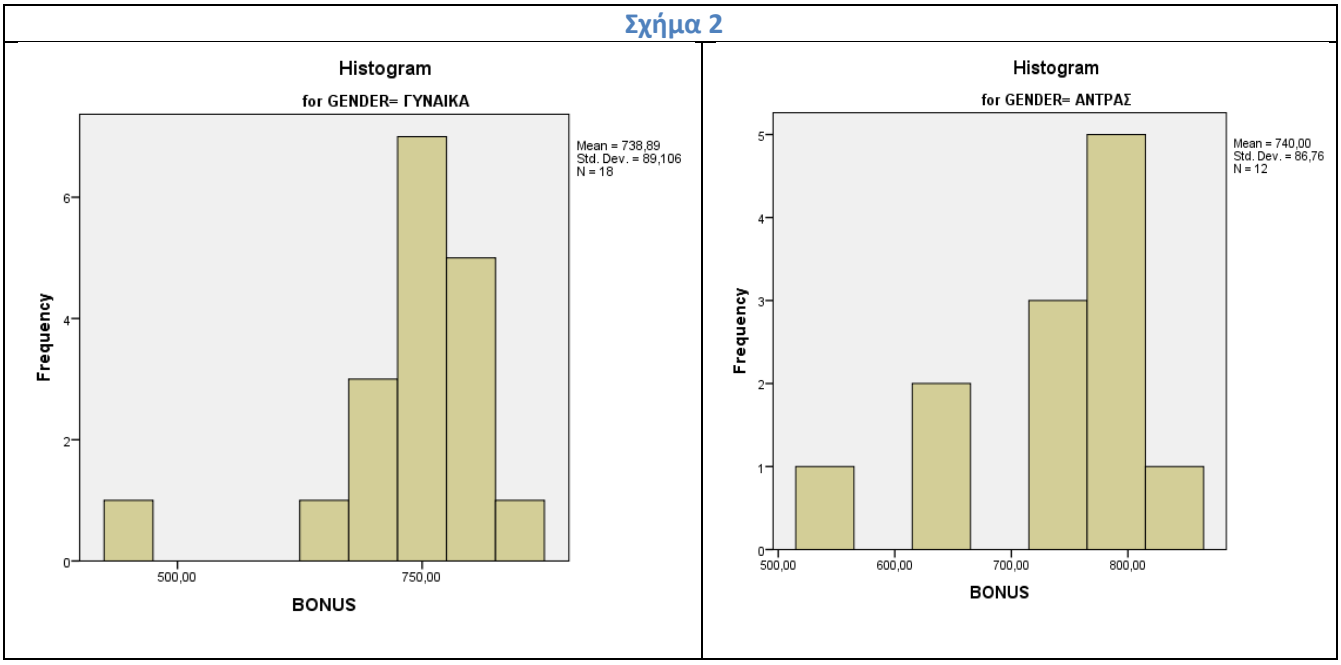
	ΦΥΛΟ			Statistic	Std. Error
BONUS	**ΓΥΝΑΙΚΑ**	**Mean**		**738,8889**	21,00247
		95% Confidence Interval for Mean	Lower Bound	694,5775	
			Upper Bound	783,2002	
		5% Trimmed Mean		749,8765	
		Median		**750,0000**	
		Variance		7939,869	
		Std. Deviation		89,10594	
		Minimum		450,00	
		Maximum		830,00	
		Range		380,00	
		Interquartile Range		110,00	
		Skewness		-2,080	,536
		Kurtosis		6,037	1,038

ΑΝΤΡΑΣ	**Mean**		**740,0000**	25,04541
	95% Confidence Interval for Mean	Lower Bound	684,8754	
		Upper Bound	795,1246	
	5% Trimmed Mean		746,1111	
	Median		**765,0000**	
	Variance		7527,273	
	Std. Deviation		86,75986	
	Minimum		540,00	
	Maximum		830,00	
	Range		290,00	
	Interquartile Range		120,00	
	Skewness		-1,367	,637
	Kurtosis		1,271	1,232

Ειδικότερα, συγκρίνοντας τα αντίστοιχα μέτρα κεντρικής τάσης διαπιστώνουμε ότι η διαφορά των μέσων ετήσιων πρόσθετων αμοιβών (bonus) μεταξύ των ανδρών και των γυναικών του δείγματος είναι περίπου **1 €** (=740,00-738,89), ενώ υπάρχει μια μεγαλύτερη διαφορά στις διαμέσους (υπέρ των ανδρών), ύψους **15€** (=765€-750€). Μια σημαντική διαφορά στο εύρος τιμών (Range) που υπάρχει μεταξύ ανδρών (Range=290€) και γυναικών (Range=380€) οφείλεται στο ότι μια έκτροπη χαμηλή τιμή στις πρόσθετες αμοιβές (bonus) των γυναικών είναι κατά 90€ μικρότερη της έκτροπης χαμηλής τιμής στις πρόσθετες αμοιβές των ανδρών, όπως φαίνεται τόσο από τον **Πίνακα 8**, όσο και από τα σχετικά φυλλογραφήματα (**Πίνακας 9**), αλλά και από το συγκεντρωτικό θηκόγραμμα (**Γράφημα 2**).

Στο **Σχήμα 2** εμφανίζονται το ιστόγραμμα που αφορά στις γυναίκες του δείγματος και το αντίστοιχο των ανδρών.

Σχήμα 2



Πίνακας 9

BONUS Stem-and-Leaf Plot for
GENDER= ΓΥΝΑΙΚΑ

Frequency	Stem &	Leaf
1,00	**Extremes**	**(=<450)**
3,00	6 .	688
3,00	7 .	034
5,00	7 .	55566
6,00	8 .	002223

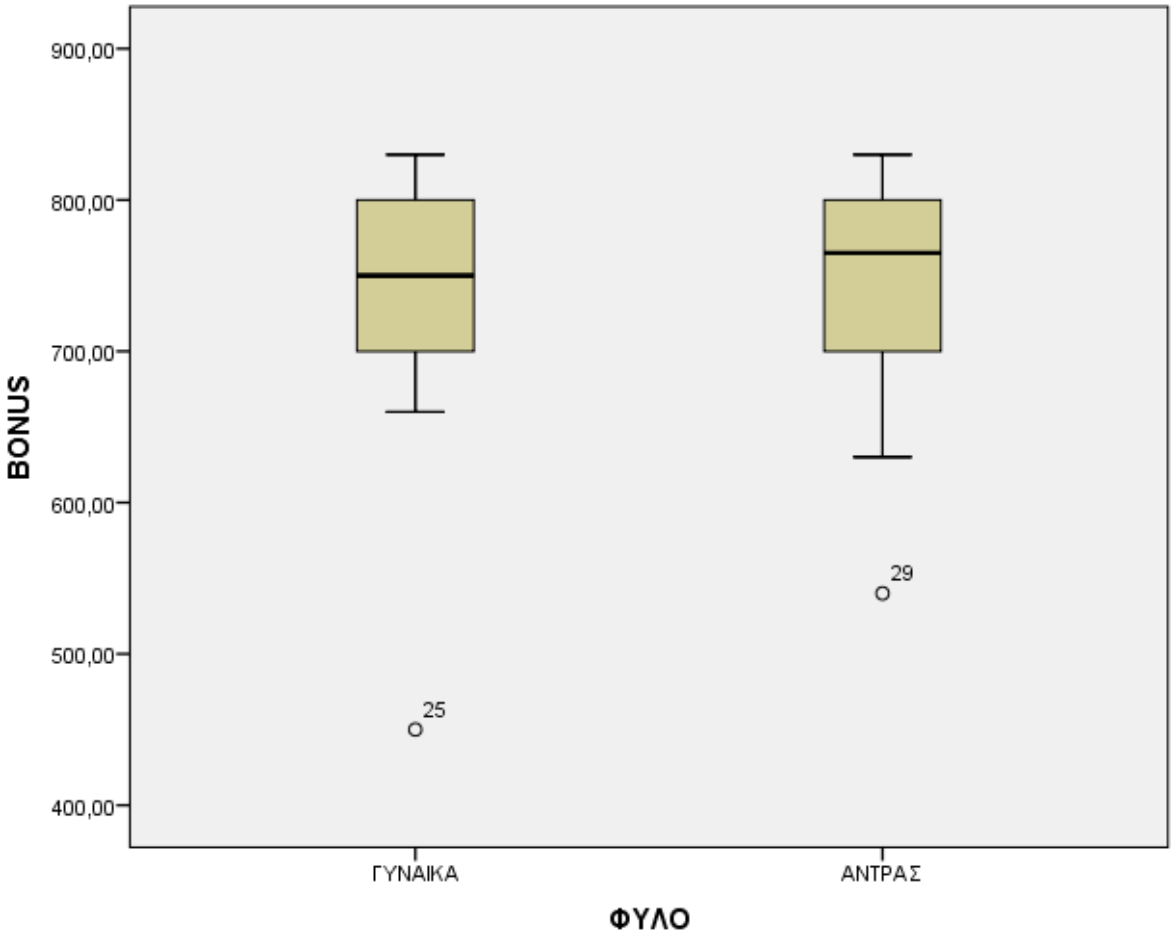
Stem width: 100,00
Each leaf: 1 case(s)

BONUS Stem-and-Leaf Plot for
GENDER= ΑΝΤΡΑΣ

Frequency	Stem &	Leaf
1,00	**Extremes**	**(=<540)**
1,00	6 .	3
1,00	6 .	6
1,00	7 .	4
3,00	7 .	558
5,00	8 .	00003

Stem width: 100,00
Each leaf: 1 case(s)

Γράφημα 2



Από το **Γράφημα 2** διαπιστώνεται μια διαφορά στο **κάτω φράγμα** μεταξύ των γυναικών και των ανδρών του δείγματος. Ο προσδιορισμός της διαφοράς γίνεται με τον υπολογισμό του κάτω φράγματος για καθένα από τα δύο φύλα:

Για τους άνδρες: $\mathbf{C_{1\alpha}=Q_{1\alpha} - 1,5\ (Q_{3\alpha} - Q_{1\alpha})}$ και για τις γυναίκες: $\mathbf{C_{1\gamma}=Q_{1\gamma} - 1,5\ (Q_{3\gamma} - Q_{1\gamma})}$

Από τον Πίνακα 10[10], λαμβάνουμε τις τιμές $Q_{1\alpha}$=680, $Q_{3\alpha}$=800 και $Q_{1\gamma}$=695, $Q_{3\gamma}$=805 και συνεπώς:

$\mathbf{C_{1\alpha}=680-1,5(800-680) = 500€}$ **και** $\mathbf{C_{1\gamma}=695-1,5(805-695)=530€}$

Πίνακας 10

Statistics

BONUS (ΓΥΝΑΙΚΕΣ)

N	Valid	18
	Missing	0
Percentiles	**25**	**695,0000**
	50	750,0000
	75	**805,0000**

Statistics

BONUS (ΑΝΔΡΕΣ)

N	Valid	12
	Missing	0
Percentiles	**25**	**680,0000**
	50	765,0000
	75	**800,0000**

Πίνακας 11

BONUS (ΑΝΔΡΕΣ)

	Frequency	Percent	Cumulative Percent
540,00	1	8,3	8,3
630,00	1	8,3	16,7
660,00	1	8,3	25,0
740,00	1	8,3	33,3
750,00	2	16,7	50,0
780,00	1	8,3	58,3
800,00	**4**	33,3	91,7
830,00	1	8,3	100,0
Total	12	100,0	

BONUS (ΓΥΝΑΙΚΕΣ)

	Frequency	Percent	Cumulative Percent
450,00	1	5,6	5,6
660,00	1	5,6	11,1
680,00	2	11,1	22,2
700,00	1	5,6	27,8
730,00	1	5,6	33,3
740,00	1	5,6	38,9
750,00	**3**	**16,7**	**55,6**
760,00	2	11,1	66,7
800,00	2	11,1	77,8
820,00	**3**	**16,7**	**94,4**
830,00	1	5,6	100,0
Total	18	100,0	

Από τον **Πίνακα 11** διαπιστώνουμε ότι η **επικρατούσα τιμή** της μεταβλητής Bonus για τους άνδρες του δείγματος είναι 800€ ενώ για στις γυναίκες εμφανίζονται οι τιμές 750€ και 820€ από τρεις φορές.

Ένας δεύτερος τρόπος υπολογισμού των κυριότερων μέτρων θέσης, ξεχωριστά για άνδρες και γυναίκες, είναι με τις εντολές: Data/Select Cases (Άνδρες =1 /Γυναίκες =0) και κατόπιν

Analyze/Descriptive Statistics/Frequencies, από τις οποίες αντλούμε διακριτούς πίνακες των ζητούμενων στατιστικών μέτρων (**Πίνακας 12** για γυναίκες και **Πίνακας 13** για άνδρες).

Πίνακας 12

Statistics

BONUS (ΓΥΝΑΙΚΕΣ)

N	Valid	18
	Missing	0
Mean		**738,8889**
Median		**750,0000**
Mode		**750,00[a]**
Percentiles	25	695,0000
	50	750,0000
	75	805,0000

a. Multiple modes exist. The smallest value is shown

Πίνακας 13

Statistics

BONUS (ΑΝΔΡΕΣ)

N	Valid	12
	Missing	0
Mean		**740,0000**
Median		**765,0000**
Mode		**800,00**
Percentiles	25	680,0000
	50	765,0000
	75	800,0000

5. Μέτρα θέσεως της μεταβλητής Bonus για κάθε επίπεδο σπουδών ξεχωριστά.

Τα κυριότερα μέτρα θέσης για κάθε επίπεδο σπουδών θα τα υπολογίσουμε βάσει των εντολών: Data/Select Cases (Λύκειο =0 /ΑΕΙ =1) και κατόπιν Analyze/Descriptive Statistics/Frequencies, από τις οποίες αντλούμε διακριτούς πίνακες των ζητούμενων στατιστικών μέτρων (**Πίνακας 14** για αποφοίτους Λυκείου και **Πίνακας 15** για αποφοίτους ΑΕΙ).

Πίνακας 14

Statistics

BONUS (**απόφοιτοι Λυκείου**)

N	Valid	16
	Missing	0
Mean		**720,0000**
Median		**750,0000**
Mode		**750,00[a]**
Percentiles	25	685,0000
	50	750,0000
	75	795,0000

a. Multiple modes exist. The smallest value is shown

Πίνακας 15

Statistics

BONUS (**απόφοιτοι ΑΕΙ**)

N	Valid	14
	Missing	0
Mean		**761,4286**
Median		**780,0000**
Mode		**800,00[a]**
Percentiles	25	725,0000
	50	780,0000
	75	820,0000

a. Multiple modes exist. The smallest value is shown

Από τη σύγκριση των ανωτέρω στοιχείων διαπιστώνεται ότι η *μέση* (Mean) και η *διάμεση (*Median*)* πρόσθετη ετήσια αμοιβή του δείγματος των εργαζομένων της επιχείρησης που έχουν αποφοιτήσει

από ΑΕΙ είναι υψηλότερες κατά 41,43€ (=761,43€-720,00€) και 30€ (=780,00€-750,00€) αντίστοιχα, σε σύγκριση με τους εργαζόμενους που έχουν μόνο απολυτήριο Λυκείου.

Από τους **Πίνακες 16** και **17** διαπιστώνεται ότι το 50% των αποφοίτων ΑΕΙ λαμβάνουν ετήσιο bonus από 800€ και άνω, ενώ το 75% των εργαζομένων που έχουν μόνο απολυτήριο Λυκείου λαμβάνουν ετήσιο bonus μικρότερο των 800€

Πίνακας 16

BONUS (απόφοιτοι Λυκείου)

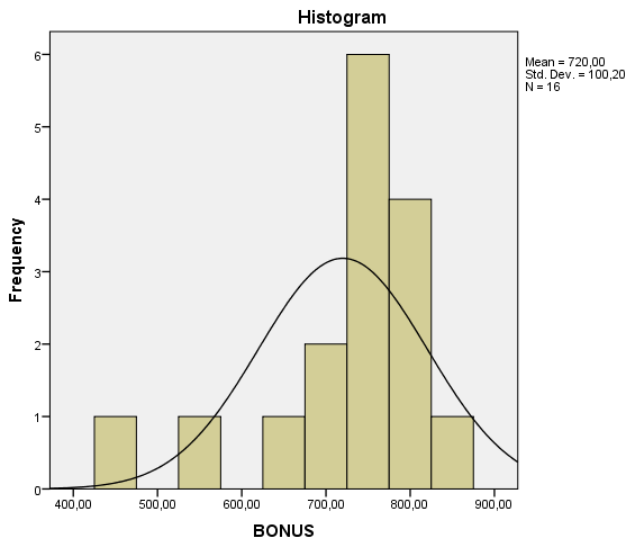
	Frequency	Valid Percent	Cumulative Percent
450,00	1	6,3	6,3
540,00	1	6,3	12,5
660,00	1	6,3	18,8
680,00	1	6,3	25,0
700,00	1	6,3	31,3
730,00	1	6,3	37,5
740,00	1	6,3	43,8
750,00	3	18,8	62,5
760,00	1	6,3	68,8
780,00	1	6,3	75,0
800,00	3	18,8	93,8
830,00	1	6,3	100,0
Total	16	100,0	

Πίνακας 17

BONUS (απόφοιτοι ΑΕΙ)

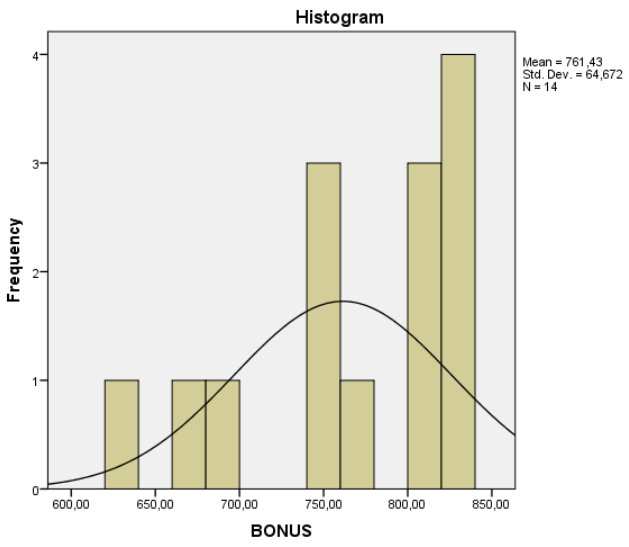
	Frequency	Valid Percent	Cumulative Percent
630,00	1	7,1	7,1
660,00	1	7,1	14,3
680,00	1	7,1	21,4
740,00	1	7,1	28,6
750,00	2	14,3	42,9
760,00	1	7,1	50,0
800,00	3	21,4	71,4
820,00	3	21,4	92,9
830,00	1	7,1	100,0
Total	14	100,0	

Γράφημα 3



Ιστόγραμμα Bonus, αποφοίτων Λυκείου

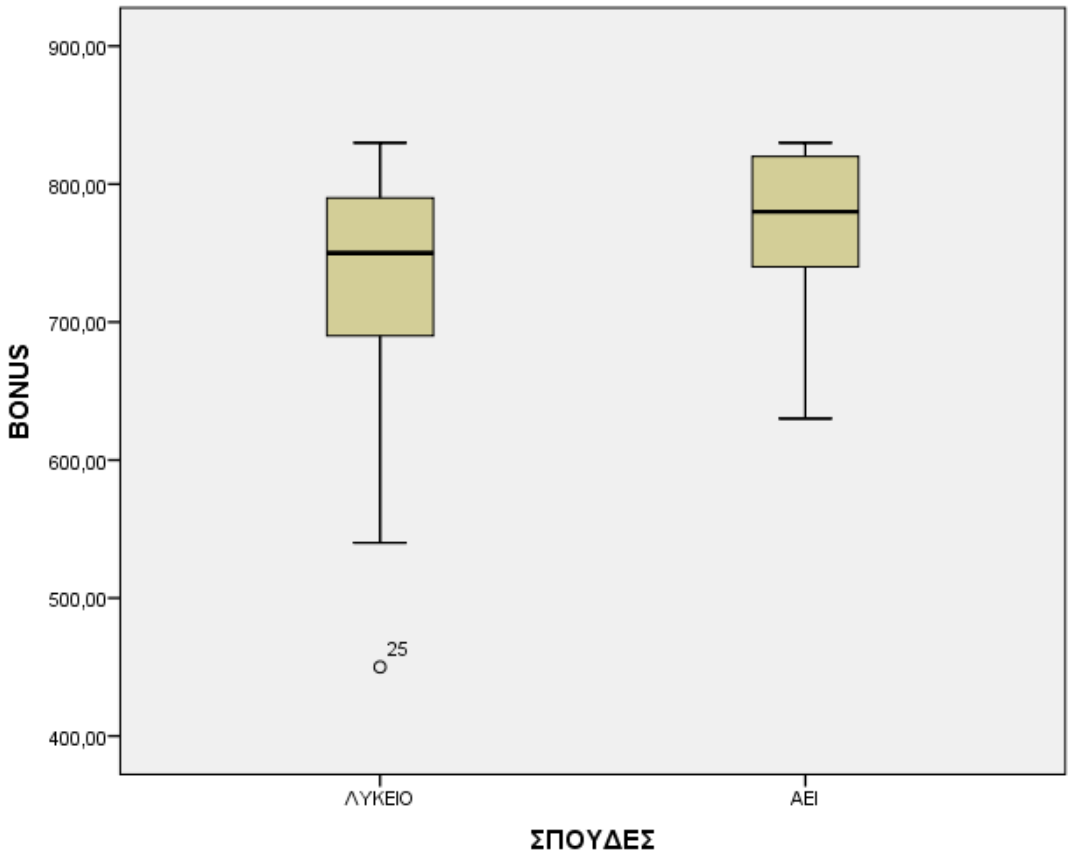
Γράφημα 4



Ιστόγραμμα Bonus, αποφοίτων ΑΕΙ

Στο θηκόγραμμα (**Γράφημα 5**) απεικονίζεται η διαφορά στις ***διαμέσους*** καθώς και οι υψηλότερες τιμές αναφοράς των τεταρτημορίων, που προσδιορίζουν την ποσοστιαία κατανομή των ετήσιων πρόσθετων αμοιβών εντός της κάθε κατηγορίας σπουδών.

Γράφημα 5



6. Πίνακας συχνοτήτων της μεταβλητής «Σπουδές».

Ο πίνακας συχνοτήτων της ποιοτικής διατάξιμης μεταβλητής «Σπουδές» αντλείται με τη βοήθεια του SPSS από τις εντολές Analyze/Descriptive Statistics/Frequencies και στα διαγράμματα (charts) επιλέγουμε Pie Charts ή Bar Charts.

Πίνακας 18

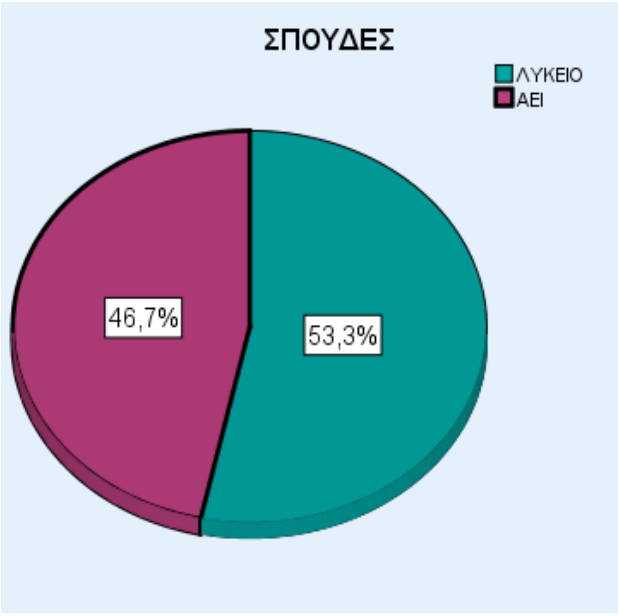
ΣΠΟΥΔΕΣ

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ΛΥΚΕΙΟ	16	53,3	53,3	53,3
	ΑΕΙ	14	46,7	46,7	100,0
	Total	30	100,0	100,0	

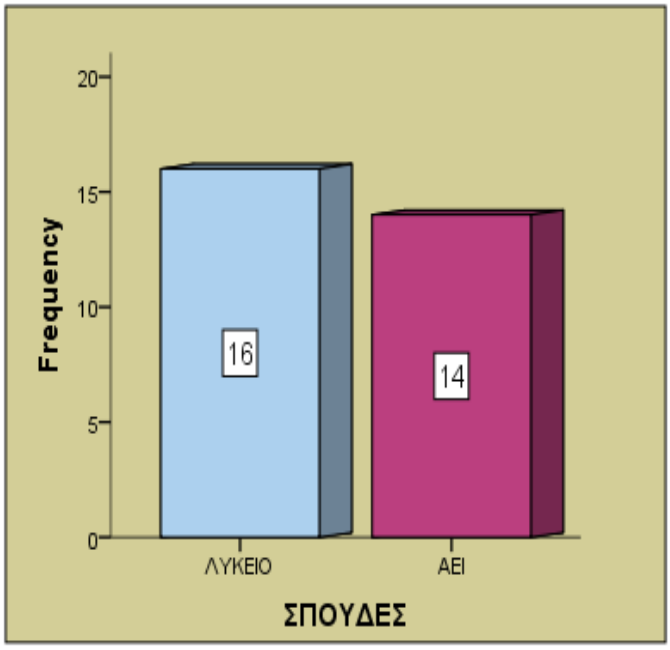
Από τον ανωτέρω πίνακα συχνοτήτων (**Πίνακας 18**) διαπιστώνεται ότι το 46,7% των εργαζομένων του δείγματος έχει αποφοιτήσει από κάποιο ΑΕΙ, ενώ το 53,3% έχει μόνο απολυτήριο Λυκείου.

Τα ανωτέρω στοιχεία απεικονίζονται στο **Γράφημα 6**, που ακολουθεί.

Γράφημα 6



Επίπεδο σπουδών (ποσοστιαία αναλογία)



Επίπεδο σπουδών (απόλυτη συχνότητα)

7. Πίνακας συνάφειας των μτβ «Σπουδές» και «Φύλο».

Η κατασκευή του πίνακα διπλής εισόδου (Πίνακας συνάφειας) γίνεται στο SPSS με τις εντολές Analyze/Descriptive Statistics/Crosstabs. Στο row(s) θέτω τη μτβ «φύλο» και στο column(s) θέτω τη μτβ «σπουδές», στα statistics επιλέγω chi-square και στα cells επιλέγω: counts (Observed και Expected) και percentages (row και column).

Στον **Πίνακα 19** παρουσιάζεται ο ζητούμενος πίνακας συνάφειας.

Πίνακας 19

ΦΥΛΟ * ΣΠΟΥΔΕΣ Crosstabulation

			ΣΠΟΥΔΕΣ		Total
			ΛΥΚΕΙΟ	ΑΕΙ	
ΦΥΛΟ	ΓΥΝΑΙΚΑ	Count	10	8	18
		Expected Count	9,6	8,4	18,0
		% within ΦΥΛΟ	55,6%	44,4%	100,0%
		% within ΣΠΟΥΔΕΣ	62,5%	57,1%	60,0%
	ΑΝΤΡΑΣ	Count	6	6	12
		Expected Count	6,4	5,6	12,0
		% within ΦΥΛΟ	50,0%	50,0%	100,0%
		% within ΣΠΟΥΔΕΣ	37,5%	42,9%	40,0%
Total		Count	16	14	30
		Expected Count	16,0	14,0	30,0
		% within ΦΥΛΟ	53,3%	46,7%	100,0%
		% within ΣΠΟΥΔΕΣ	100,0%	100,0%	100,0%

Από τον **Πίνακα 19** διαπιστώνουμε ότι το 55,6% των γυναικών του δείγματος που εργάζονται στην επιχείρηση έχουν απολυτήριο Λυκείου (μόνο) έναντι του 50% των ανδρών, ενώ το 44,4% των γυναικών έχουν πτυχίο ΑΕΙ έναντι του 50% των ανδρών. Επιπλέον, από τους εργαζόμενους της επιχείρησης που έχουν πτυχίο ΑΕΙ, το 57,1% είναι γυναίκες και το 42,9% είναι άνδρες.

Ο έλεγχος της ανεξαρτησίας των δύο μεταβλητών («φύλο» και «σπουδές) διεξάγεται με τον **έλεγχο X^2** (Pearson chi-square).
Προϋποθέσεις της εφαρμογής του ελέγχου **X^2** είναι[11]:

- Οι μεταβλητές για τις οποίες διεξάγουμε έλεγχο ανεξαρτησίας θα πρέπει να είναι ποιοτικές.
- Το ποσοστό των κελιών του πίνακα συνάφειας που έχουν αναμενόμενη συχνότητα μικρότερη του πέντε (5) δεν πρέπει να υπερβαίνει το 20%.
- Το μέγεθος του δείγματος θα πρέπει να κυμαίνεται από 25 έως 250 (παρατηρήσεις).

Στον έλεγχο ανεξαρτησίας που διεξάγουμε για τις ποιοτικές μεταβλητές «φύλο» (ονομαστική μτβ) και «σπουδές» (διατάξιμη μτβ), πληρούνται οι ανωτέρω προϋποθέσεις καθόσον -όπως φαίνεται από τον ανωτέρω Πίνακα 19- οι τέσσερις αναμενόμενες τιμές (Expected Counts) είναι μεγαλύτερες του πέντε (5) και το μέγεθος των παρατηρήσεων υπερβαίνει το ελάχιστο όριο των 25 παρατηρήσεων.

Συνεπώς μπορούμε να προχωρήσουμε στον έλεγχο ανεξαρτησίας εφαρμόζοντας τον **X^2** έλεγχο, διατυπώνοντας τη *μηδενική* (H_o) και την *εναλλακτική* υπόθεση (H_1):

H_o: Οι δύο μεταβλητές είναι ανεξάρτητες (Δεν υπάρχει σχέση ανάμεσα στο φύλο και το επίπεδο σπουδών)

H_1: Οι δύο μεταβλητές σχετίζονται (Υπάρχει σχέση μεταξύ του φύλου και του επιπέδου σπουδών)

Παράλληλα με την κατασκευή του πίνακα συνάφειας (Πίνακας 19) με την προαναφερόμενη διαδικασία στο SPSS, κατασκευάζεται και ο κατωτέρω **Πίνακας 20**, που περιλαμβάνει τον δείκτη Chi-Square.

Πίνακας 20

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	**,089[a]**	**1**	**,765**		
Continuity Correction[b]	,000	1	1,000		
Likelihood Ratio	,089	1	,765		
Fisher's Exact Test				1,000	,529
Linear-by-Linear Association	,086	1	,769		
N of Valid Cases	30				

a. **0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,60.**

b. Computed only for a 2x2 table

Από τον Πίνακα 20 προκύπτει ότι σε επίπεδο στατιστικής σημαντικότητας α=5%,

η **p-value=0,765> 0,05** και ως εκ τούτου αποδεχόμαστε την Ho ότι: **ΔΕΝ υπάρχει συσχέτιση μεταξύ του φύλου των εργαζομένων της επιχείρησης και του επιπέδου σπουδών τους**. Ομοίως, δεν εξαρτάται το επίπεδο σπουδών των εργαζομένων της επιχείρησης από το φύλο τους.

Στο ίδιο συμπέρασμα καταλήγουμε αν συγκρίνουμε το X^2=0,089 (που βρήκαμε στον Πίνακα 20) με τους πίνακες της X^2 κατανομής για df=1 και α=0,05 όπου η τιμή που αντιστοιχεί είναι 3,84. Επειδή, 0,089<3,84 αποδεχόμαστε την Ho.

8. Έλεγχος κανονικότητας της μεταβλητής Bonus

Ο έλεγχος κανονικότητας στο πλαίσιο του SPSS γίνεται με τρεις τρόπους:

α) Με τη βοήθεια στατιστικών κριτηρίων (κριτήριο Kolmogorov-Smirnov και Shapiro-Wilk),

β) Με τη βοήθεια γραφικών αναπαραστάσεων (Normal Q-Q Plots, Detrended Q-Q Plot και Box Plot), και

γ) Με τη βοήθεια του λόγου $\lambda = \frac{statistic}{standard\ error\ of\ statistic}$ (π.χ. $\frac{Skewness}{standard\ error\ of\ Skewness}$, $\frac{Kurtosis}{standard\ error\ of\ Kurtosis}$).

Ακολουθώντας τη διαδρομή Analyze/Descriptive Statistic/Explore κατασκευάζουμε τον **Πίνακα 21** (Test of Normality), που περιλαμβάνει τα κριτήρια Kolmogorov-Smirnov και Shapiro-Wilk.

Πίνακας 21

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
BONUS	,203	30	,003	,833	30	**,000**

a. Lilliefors Significance Correction

Διατυπώνουμε τη μηδενική (H_o) και την εναλλακτική υπόθεση (H_1), ως εξής:

H_o: Οι τιμές που λαμβάνει η μτβ Bonus ακολουθούν την κανονική κατανομή

H_1: Οι τιμές που λαμβάνει η μτβ Bonus ΔΕΝ ακολουθούν την κανονική κατανομή,

και ελέγχουμε τη μηδενική υπόθεση σε επίπεδο στατιστικής σημαντικότητας α=5%.

Σημειώνεται, ότι επειδή το μέγεθος του πληθυσμού μας είναι μικρότερο από 50 (n<50), ο έλεγχος κανονικότητας θα γίνει με το κριτήριο Shapiro-Wilk[12]:

Συνεπώς, επειδή **p-value** (Shapiro-Wilk)=0,0001 < 0,05 απορρίπτουμε την Ho και δεχόμαστε την H_1 δηλ. **δεχόμαστε ότι η μτβ bonus ΔΕΝ ακολουθεί κανονική κατανομή**.

Επισημαίνεται ότι και με το κριτήριο των Kolmogorov-Smirnov θα καταλήγαμε στο ίδιο συμπέρασμα, καθόσον **p-value** (K-S)=0,003 < 0,05.

Μαζί με τον Πίνακα 21, ο οποίος δημιουργήθηκε από την διαδρομή Analyze/Descriptive Statistic/Explore, κατασκευάζεται και ο **Πίνακας 22** που περιλαμβάνει τα στατιστικά κριτήρια για τον έλεγχο του λόγου **λ**.

Πίνακας 22

Descriptives

			Statistic	Std. Error
BONUS	Mean		739,3333	15,82180
	95% Confidence Interval for Mean	Lower Bound	706,9741	
		Upper Bound	771,6926	
	5% Trimmed Mean		748,7037	
	Median		750,0000	
	Variance		7509,885	
	Std. Deviation		86,65959	
	Minimum		450,00	
	Maximum		830,00	
	Range		380,00	
	Interquartile Range		105,00	
	Skewness		**-1,728**	**,427**
	Kurtosis		**3,569**	**,833**

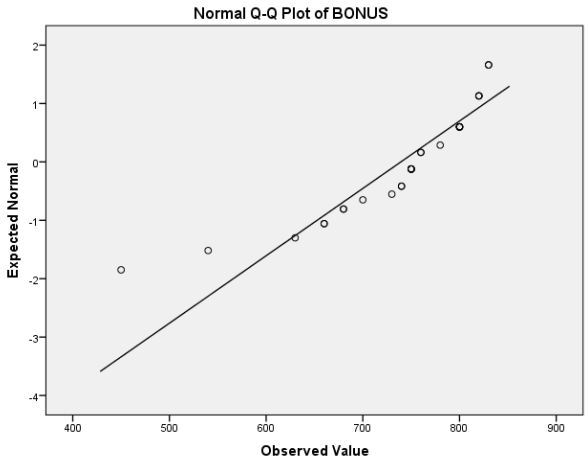
$$\boldsymbol{\lambda sk} = \frac{Skewness}{standard\ error\ of\ Skewness} = \frac{-1{,}728}{0{,}427} = \text{-4,047 και}$$

$$\boldsymbol{\lambda ku} = \frac{Kurtosis}{standard\ error\ of\ Kurtosis} = \frac{3{,}569}{0{,}833} = 4{,}284$$

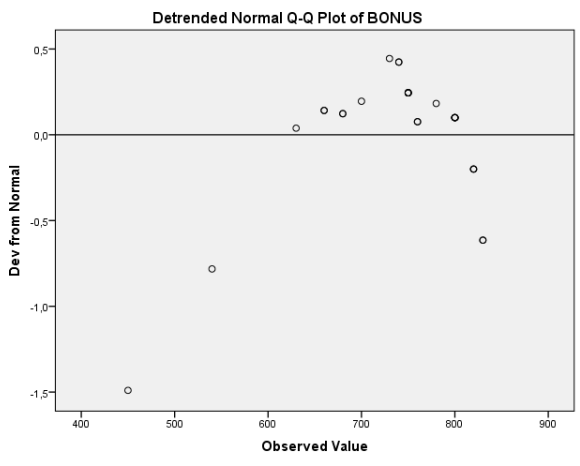
Επειδή, $\boldsymbol{\lambda sk}$**< -2** και $\boldsymbol{\lambda ku}$ **> 2** δηλ. οι δείκτες λ είναι εκτός του διαστήματος (-2,2), **συμπεραίνουμε ότι η μτβ Bonus δεν ακολουθεί την κανονική κατανομή**.

Ο τρίτος τρόπος ελέγχου της κανονικότητας της μεταβλητής Bonus είναι μέσω των γραφικών αναπαραστάσεων που κατασκευάζονται στο SPSS με την ίδια προαναφερόμενη διαδικασία (Analyze/Descriptive Statistic/Explore).

Γράφημα 7



Γράφημα 8



Στο **Γράφημα 7** (Normal Q-Q Plot of Bonus) κάθε σημείο αναπαριστά ένα ζεύγος (α,β), όπου α είναι η παρατηρούμενη τιμή από τα δεδομένα μας, και β είναι η αναμενόμενη τιμή (expected normal) της μεταβλητής Bonus αν αυτή ακολουθεί την κανονική κατανομή. Παρότι τα σημεία δεν «πέφτουν» ακριβώς πάνω ή πολύ κοντά στην ευθεία κανονικής κατανομής, ώστε να συμπεράνουμε με βεβαιότητα ότι τα δεδομένα μας ακολουθούν την κανονική κατανομή, εντούτοις η απομάκρυνσή τους από αυτή δεν μας επιτρέπει να είμαστε σίγουροι για το αντίθετο (ότι δηλ. η μτβ Bonus δεν ακολουθεί κανονική κατανομή).

Στο **Γράφημα 8** (Detrended Normal Q-Q Plot of Bonus) ο άξονας Χ (observed value) αφορά στις παρατηρούμενες τιμές της μτβ bonus, ενώ στον κατακόρυφο άξονα είναι οι αντίστοιχες Z-τιμές που αναμένονται δεδομένης της κανονικότητας. Από το εν λόγω γράφημα μπορούμε να διαγνώσουμε απομακρύνσεις των δεδομένων μας από την κανονικότητα αν τα σημεία (κουκίδες) δεν είναι τυχαία κατανεμημένα πάνω και κάτω από την οριζόντια γραμμή δηλ. αν τα σημεία ακολουθούν κάποιο **πρότυπο** (όπως π.χ. αν οι κουκίδες σχηματίζουν ευθεία γραμμή ή παραβολή) ή **συσσωρεύσεις** (π.χ. κατά τόπους διαφορετική πυκνότητα) και επομένως δεν είναι τυχαία κατανεμημένα[13]. Στο εν λόγω γράφημα, μπορεί κάποιος να εντοπίσει μια οιονεί παραβολή με τα κοίλα στραμμένα προς τα κάτω ή τουλάχιστον μια συσσώρευση σημείων που μας προϊδεάζει για το ότι η μτβ Bonus δεν ακολουθεί την κανονική κατανομή.

Από το θηκόγραμμα της μτβ Bonus (Γράφημα 1) δεν διαπιστώνεται κάποια σημαντική ασυμμετρία που θα μας επέτρεπε να υποθέσουμε ότι η μεταβλητή μας δεν ακολουθεί την κανονική κατανομή. Σημειώνεται ότι η συμμετρία του θηκογράμματος είναι «ένας προάγγελος της κανονικότητας[14]» καθόσον κάθε κανονική κατανομή «δεν είναι δυνατόν να μην είναι συμμετρική».

Τέλος, μια από τις ιδιότητες της κανονικής κατανομής είναι ότι οι τιμές των δεικτών κεντρικής τάσης (κεντρικής θέσης) γειτνιάζουν. Κάτι, που όπως διαπιστώνεται από τον **Πίνακα 23**, δεν επαληθεύεται από τις παρατηρήσεις μας, καθόσον η τιμή της *μέση τιμής (μ)*, της *διαμέσου (Μ)* και της *επικρατούσας τιμής (Μο)* δεν γειτνιάζουν (δηλ. μ≠ Μ ≠Μο)

Πίνακας 23

Statistics

BONUS

N	Valid	30
	Missing	0
Mean		739,3333
Median		750,0000
Mode		800,00

9. Έλεγχος της υπόθεσης ότι ο μ.ο. ηλικίας των εργαζομένων της επιχείρησης είναι 40 έτη

Ο έλεγχος της υπόθεσης ότι η μέση τιμή της ηλικίας των εργαζομένων της επιχείρησης είναι ίσος με 40 έτη, θα γίνει με τη βοήθεια του SPSS (Analyze/Compare Means/One Sample T-Test).

Προϋποθέσεις για την ασφαλή χρήση του στατιστικού κριτηρίου T-Test είναι[15]:

α) Η ελεγχόμενη μτβ να είναι ποσοτική

β) Το δείγμα να έχει επιλεγεί τυχαία από τον πληθυσμό ενδιαφέροντος

γ) Η μτβ θα πρέπει να ακολουθεί κανονική κατανομή.

Η μτβ Age της άσκησης είναι ποσοτική και μάλιστα συνεχής και θεωρώντας ότι το δείγμα μας έχει επιλεγεί τυχαία, θα ελέγξουμε την προϋπόθεση της κανονικότητας της μτβ Age.

Με τη βοήθεια του SPSS (Analyze/Descriptive Statistics/Explore) ελέγχω την υπόθεση ότι η μτβ Age ακολουθεί κανονική κατανομή.
Ειδικότερα, σε επίπεδο στατιστικής σημαντικότητας 5% (α=0,05) ελέγχω τη μηδενική (H_o) και την εναλλακτική υπόθεση (H_1), οι οποίες διατυπώνονται ως εξής:
H_o: Η μτβ Age ακολουθεί την κανονική κατανομή
H_1: Η μτβ Age Δεν ακολουθεί την κανονική κατανομή

Πίνακας 24

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ΗΛΙΚΙΑ	,166	30	,034	,952	30	,189

a. Lilliefors Significance Correction

Επειδή το μέγεθος του δείγματός μας είναι μικρότερο από 50 (n<50), ο έλεγχος κανονικότητας θα γίνει με το κριτήριο Shapiro-Wilk:

Συνεπώς, επειδή **p-value** (Shapiro-Wilk)=0,189 > 0,05 αποδεχόμαστε την Ho δηλ. **δεχόμαστε ότι η μτβ Age ακολουθεί κανονική κατανομή**.

Πέρα από τον ανωτέρω έλεγχο κανονικότητας των Shapiro-Wilk, θα εξετάσω τα λsk και λku, τις τιμές των οποίων παίρνω μέσω των εντολών (Analyze/Descriptive Statistics/Frequencies, Πίνακας 25)

Πίνακας 25

Statistics

ΗΛΙΚΙΑ

N	Valid	30
	Missing	0
Mean		34,2033
Median		34,2000
Mode		34,20
Skewness		,196
Std. Error of Skewness		,427
Kurtosis		-,478
Std. Error of Kurtosis		,833

Από τον παραπλεύρως πίνακα διαπιστώνω ότι:

Mean≈ Median ≈Mode = 34,20

και

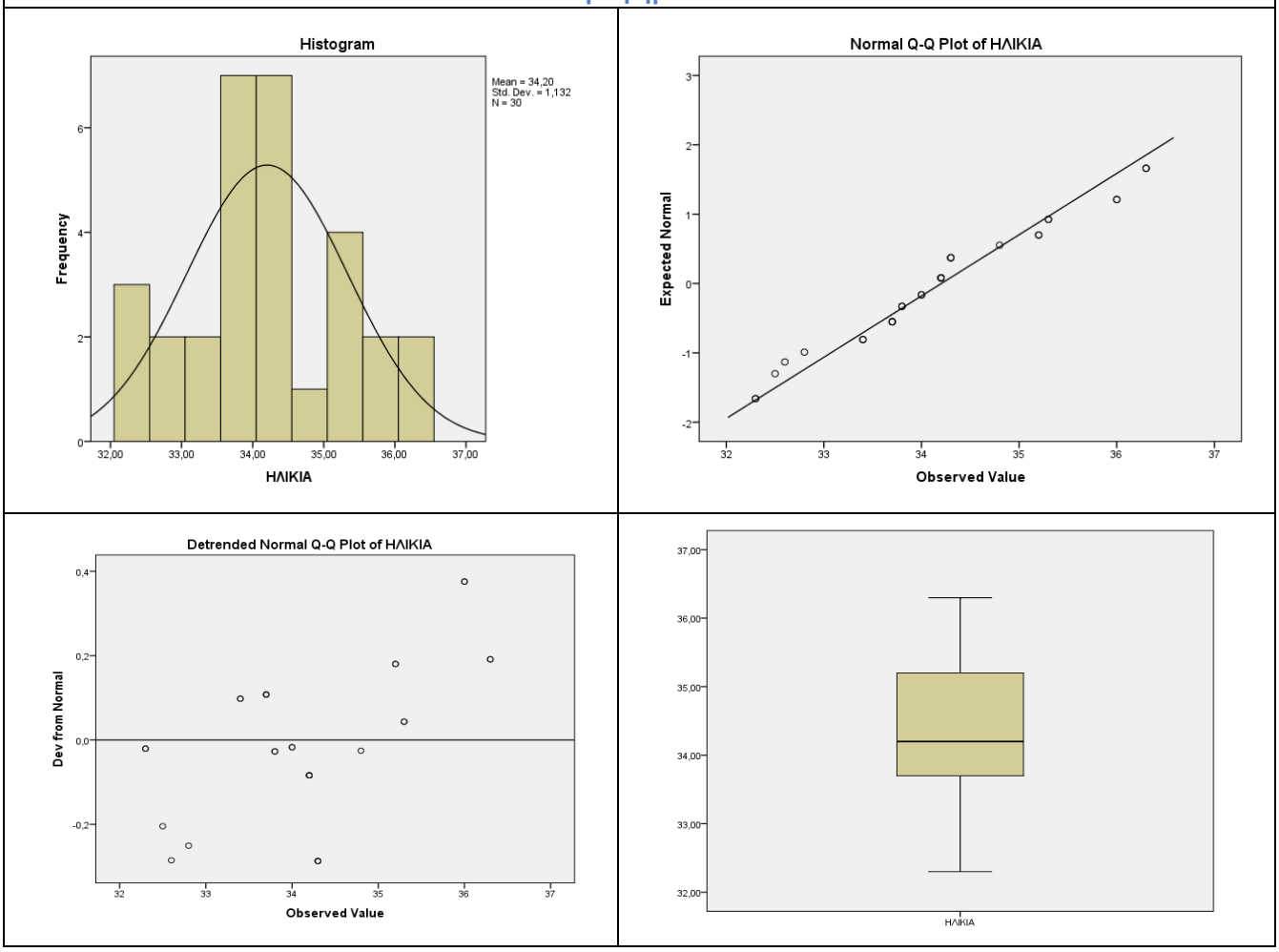
$$\lambda sk = \frac{Skewness}{standard\ error\ of\ Skewness} = \frac{0,196}{0,427} = 0,459 < 2$$

$$\lambda ku = \frac{Kurtosis}{standard\ error\ of\ Kurtosis} = \frac{-0,478}{0,833} = -0,573 > -2$$

Συνεπώς, επειδή τα λsk και λku λαμβάνουν τιμές εντός του διαστήματος (-2,2), **η μτβ Age ακολουθεί την κανονική κατανομή**.

Ομοίως, και από το Γράφημα 9, **δεν** φαίνεται να υπάρχει πρόβλημα κανονικότητας για τη μτβ Age.

Γράφημα 9



Εφόσον δεν παραβιάζονται οι προϋποθέσεις εφαρμογής του T-Test, ακολουθώ τη διαδρομή εντολών:Analyze/Compare Means/One Sample T-Test και θέτω ως διάστημα εμπιστοσύνης (Confidence Interval Percentage) 95%. Από την εκτέλεση της εντολής λαμβάνω τους Πίν. 26 και 27.

Πίνακας 26

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
ΗΛΙΚΙΑ	30	34,2033	1,13213	,20670

Πίνακας 27

One-Sample Test

	Test Value = 40					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
ΗΛΙΚΙΑ	-28,044	29	,000	-5,79667	-6,2194	-5,3739

Από τον **Πίνακα 26** βλέπω ότι η μέση δειγματική ηλικία των εργαζομένων (δηλ. όπως αυτή προσδιορίζεται από το δείγμα μας) είναι 34,203 έτη και θέλω να ελέγξω αν η διαφορά της μέσης ηλικίας του συνόλου των εργαζομένων της επιχείρησης από την τιμή ελέγχου (40 έτη) είναι στατιστικώς σημαντική.

Διατυπώνω τη μηδενική και την εναλλακτική υπόθεση, την οποία θα ελέγξω σε επίπεδο σημαντικότητας α=5%:

H_o : Η μέση ηλικία των εργαζομένων της επιχείρησης είναι ίση με 40 έτη

H_1: Η μέση ηλικία των εργαζομένων της επιχείρησης ΔΕΝ είναι ίση με 40 έτη.

Από τον **Πίνακα 27** παρατηρώ ότι **p-value = 0,0001 < 0,05** και συνεπώς απορρίπτω τη μηδενική υπόθεση, σε επίπεδο στατιστικής σημαντικότητας 5%, και αποδέχομαι την εναλλακτική ότι **η μέση ηλικία των εργαζομένων της επιχείρησης δεν είναι ίση με 40 έτη.** Ομοίως, η διαφορά της μέσης ηλικίας των εργαζομένων από την τιμή ελέγχου (40έτη) είναι στατιστικώς σημαντική σε επίπεδο 5%

Μπορούμε να ελέγξουμε τη μηδενική υπόθεση συγκρίνοντας το t=-28,044 με df=29 (βαθμοί ελευθερίας), που λαμβάνουμε από τον Πίνακα 27, με την τιμή που προκύπτει από τους πίνακες της t κατανομής με 29 βαθμούς ελευθερίας και επίπεδο σημαντικότητας α=0,05. Η τιμή που λαμβάνουμε από τους πίνακες της t κατανομής είναι **2,045**. Επειδή[16] t=-28,044 < -2,045 ($t_{\alpha/2}$) απορρίπτω την H_o σε επίπεδο στατιστικής σημαντικότητας 5%.

Όπως φαίνεται στον Πίνακα 27 η μέση διαφορά (Mean Difference) της δειγματικής μέσης τιμής της μτβ Age από την τιμή ελέγχου (40) είναι -5,797 και **με βεβαιότητα 95% η διαφορά αυτή θα βρίσκεται στο διάστημα (-6,219 , -5,374)**.

Ένας πρόσθετος έλεγχος της αποδοχής ή της απόρριψης της μηδενικής υπόθεσης είναι να εξετάσουμε αν στο ανωτέρω διάστημα, που βρίσκεται (με 95% βεβαιότητα) η διαφορά της δειγματικής μέσης τιμής από την τιμή ελέγχου, περιλαμβάνεται το μηδέν. Επειδή, λοιπόν, στο διάστημα (-6,219 , -5,374) **δεν περιλαμβάνεται το μηδέν** μπορούμε να απορρίψουμε την H_o σε επίπεδο στατιστικής σημαντικότητας 5%.

Η τιμή της μτβ Age με βεβαιότητα 95% (95% διάστημα εμπιστοσύνης) βρίσκεται στο διάστημα:
(34,203 -2,045*0,207 , 34,203 + 2,045*0,207) = (33,779 , 34,626) όπως φαίνεται και από τον **Πίνακα 28**, ο οποίος εξάγεται με την εντολή Explore από το SPSS.

Πίνακας 28

Descriptives

			Statistic	Std. Error
ΗΛΙΚΙΑ	Mean		34,2033	,20670
	95% Confidence Interval for Mean	Lower Bound	**33,7806**	
		Upper Bound	**34,6261**	
	5% Trimmed Mean		34,1926	
	Median		34,2000	
	Variance		1,282	
	Std. Deviation		1,13213	
	Minimum		32,30	
	Maximum		36,30	
	Range		4,00	
	Interquartile Range		1,58	
	Skewness		,196	,427
	Kurtosis		-,478	,833

10. Έλεγχος της υπόθεσης ότι το μέσο Bonus των ανδρών ισούται με το μέσο Bonus των γυναικών.

Ο έλεγχος της υπόθεσης της ισότητας των μέσων της μτβ Bonus για δύο ανεξάρτητα δείγματα (άνδρες / γυναίκες εργαζόμενοι στην επιχείρηση) θα γίνει με τη βοήθεια του SPSS (Analyze/ Compare Means/ Independent Sample T-Test).
Για να εξάγουμε έγκυρα συμπεράσματα από τον εν λόγω έλεγχο θα πρέπει να ισχύουν δύο προϋποθέσεις:
α) τα ανεξάρτητα δείγματα, που ελέγχονται ως προς μια μτβ, θα πρέπει να ακολουθούν την κανονική κατανομή, και
β) οι διακυμάνσεις (variance) των δύο ανεξάρτητων δειγμάτων θα πρέπει να είναι ίσες.

Ελέγχω την υπόθεση της κανονικότητας της μτβ bonus ως προς τα δύο ανεξάρτητα δείγματα (άνδρες / γυναίκες) με την εντολή Analyze/Descriptive Statistics / Explore , Dependent → Bonus και Factor → Gender.
Από τον **Πίνακα 29** διαπιστώνω ότι το μέγεθος του δείγματος (άνδρες / γυναίκες) διαφέρει αισθητά. Οι γυναίκες είναι 18 ενώ οι άνδρες 12.

Πίνακας 29

Case Processing Summary

	ΦΥΛΟ	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
BONUS	ΓΥΝΑΙΚΑ	18	100,0%	0	0,0%	18	100,0%
	ΑΝΤΡΑΣ	12	100,0%	0	0,0%	12	100,0%

Η εξέταση της μτβ Bonus ξεχωριστά για τα δύο δείγματα έχει γίνει στην ενότητα 4 και αναλυτικά στοιχεία περιέχονται στον Πίνακα 8.

Τον έλεγχο κανονικότητας τον διεξάγω από τον **Πίνακα 30**.

Πίνακας 30

Tests of Normality

	ΦΥΛΟ	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
BONUS	ΓΥΝΑΙΚΑ	,182	18	,116	,798	18	,001
	ΑΝΤΡΑΣ	,250	12	,037	,837	12	,025

a. Lilliefors Significance Correction

Επειδή το μέγεθος του δείγματος είναι μικρότερο από 50 (n<50), ο έλεγχος κανονικότητας θα γίνει με τη μέθοδο των Shapiro-Wilk σε επίπεδο στατιστικής σημαντικότητας α=0,05.
Διατυπώνω τη μηδενική και την εναλλακτική υπόθεση για τις γυναίκες:
H_o : Η ετήσια πρόσθετη αμοιβή των γυναικών της επιχείρησης ακολουθεί κανονική κατανομή
H_1: Η ετήσια πρόσθετη αμοιβή των γυναικών της επιχείρησης ΔΕΝ ακολουθεί κανονική κατανομή
Επειδή p-value = 0,001 < 0,05 απορρίπτω την Ho και **αποδέχομαι ότι η ετήσια πρόσθετη αμοιβή (Bonus) των γυναικών δεν ακολουθεί κανονική κατανομή**.

Ομοίως για τους άνδρες, επειδή p-value = 0,025 < 0,05 απορρίπτω την Ho και **αποδέχομαι ότι η ετήσια πρόσθετη αμοιβή (Bonus) των ανδρών δεν ακολουθεί την κανονική κατανομή.**
Επειδή το μέγεθος του δείγματος είναι μεγάλο (n≥ 30), γνωρίζουμε ότι βάσει του Κεντρικού Οριακού Θεωρήματος **το t-test για δύο ανεξάρτητα δείγματα είναι ανθεκτικό έναντι της μη κανονικότητας.**
Όμως, επειδή το μέγεθος του δείγματός μας μπορεί οριακά να θεωρηθεί ως «μεγάλο», μιας και έχουμε ακριβώς τριάντα (30) παρατηρήσεις, και προκειμένου να πληρωθεί με βεβαιότητα η υπόθεση της κανονικότητας, δοκιμάζω να «απομακρύνω» την έκτροπη τιμή Bonus για τις γυναίκες (case 25) και την αντίστοιχη για τους άντρες (case 29). Η «απομάκρυνση» γίνεται με την εντολή Data /Select Cases/If Condition is Satisfied → Bonus>550.

Μετά από τον εν λόγω μετασχηματισμό λαμβάνω τους εξής Πίνακες:

Πίνακας 31

Case Processing Summary

	ΦΥΛΟ	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
BONUS	ΓΥΝΑΙΚΑ	17	100,0%	0	0,0%	17	100,0%
	ΑΝΤΡΑΣ	11	100,0%	0	0,0%	11	100,0%

Πίνακας 32

Descriptives

	ΦΥΛΟ			Statistic	Std. Error
BONUS	ΓΥΝΑΙΚΑ	Mean		755,8824	13,09071
		95% Confidence Interval for Mean	Lower Bound	728,1313	
			Upper Bound	783,6334	
		5% Trimmed Mean		757,0915	
		Median		750,0000	
		Variance		2913,235	
		Std. Deviation		53,97439	
		Minimum		660,00	
		Maximum		830,00	
		Range		170,00	
		Interquartile Range		95,00	
		Skewness		-,262	,550
		Kurtosis		-,979	1,063
	ΑΝΤΡΑΣ	Mean		758,1818	18,86884
		95% Confidence Interval for Mean	Lower Bound	716,1394	
			Upper Bound	800,2242	
		5% Trimmed Mean		761,3131	
		Median		780,0000	
		Variance		3916,364	

	Std. Deviation	62,58086	
	Minimum	630,00	
	Maximum	830,00	
	Range	200,00	
	Interquartile Range	60,00	
	Skewness	-1,190	,661
	Kurtosis	,639	1,279

Πίνακας 33

Tests of Normality

	ΦΥΛΟ	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
BONUS	ΓΥΝΑΙΚΑ	,146	17	,200*	,929	17	,207
	ΑΝΤΡΑΣ	,204	11	,200*	,852	11	,045

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Από τον **Πίνακα 33** ελέγχω τη μηδενική και την εναλλακτική υπόθεση που διατυπώσαμε ανωτέρω για τις γυναίκες σε επίπεδο στατιστικής σημαντικότητας **α=0,01**.

Επειδή (για τις γυναίκες) **p-vanue=0,207 > 0,01** αποδέχομαι την **Ho ότι οι ετήσιες πρόσθετες αποδοχές των γυναικών ακολουθούν την κανονική κατανομή**.

Ομοίως, (για τους άνδρες) επειδή **p-value= 0,045 > 0,01** αποδέχομαι την **Ho ότι οι ετήσιες πρόσθετες αποδοχές των ανδρών ακολουθούν την κανονική κατανομή.**

Ο λόγος των διακυμάνσεων των δύο πληθυσμών (όπως οι τελευταίοι μετασχηματίστηκαν) είναι:

$$\frac{Variance\ Ανδρών}{Variance\ Γυναικών} = \frac{3916,364}{2913,235} = 1,344$$

Εφόσον ισχύουν οι προϋποθέσεις εφαρμογής του Independent Sample T-Test, διατυπώνω τη μηδενική και την εναλλακτική υπόθεση:

H_o: **Η μέση πρόσθετη ετήσια αμοιβή των γυναικών δεν διαφέρει από τη μέση πρόσθετη ετήσια αμοιβή των ανδρών ($\mu_{bonus\ γυναικών} = \mu_{bonus\ ανδρών}$)**

H_1: **Η μέση πρόσθετη ετήσια αμοιβή των γυναικών διαφέρει από τη μέση πρόσθετη ετήσια αμοιβή των ανδρών ($\mu_{bonus\ γυναικών} \neq \mu_{bonus\ ανδρών}$)**

και εκτελώ τον σχετικό έλεγχο στο SPSS, απ' όπου λαμβάνω τους εξής Πίνακες:

Πίνακας 34

Group Statistics

	ΦΥΛΟ	N	Mean	Std. Deviation	Std. Error Mean
BONUS	ΓΥΝΑΙΚΑ	17	**755,8824**	53,97439	13,09071
	ΑΝΤΡΑΣ	11	**758,1818**	62,58086	18,86884

Από τον **Πίνακα 34** συμπεραίνω ότι η μέση πρόσθετη ετήσια αμοιβή των **γυναικών** στο δείγμα μας είναι **755,88€** ενώ η μέση πρόσθετη ετήσια αμοιβή των **ανδρών** είναι **758,18€.**

Αυτό που απομένει είναι να ελέγξουμε αν αυτή η διαφορά είναι στατιστικώς σημαντική στον πληθυσμό (δηλ. στο σύνολο των εργαζομένων της επιχείρησης) σε επίπεδο στατιστικής σημαντικότητας 1%. Ο σχετικός έλεγχος γίνεται από τον Πίνακα 35:

Πίνακας 35

Independent Samples Test

			BONUS	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		,130	
	Sig.		**,721**	
t-test for Equality of Means	t		-,103	-,100
	df		26	19,168
	Sig. (2-tailed)		**,918**	,921
	Mean Difference		-2,29947	-2,29947
	Std. Error Difference		22,22557	22,96519
	99% Confidence Interval of the Difference	Lower	**-64,05798**	-67,93842
		Upper	**59,45905**	63,33949

Με το **Test Levene's** ελέγχεται η υπόθεση της **ομοιογένειας** του πληθυσμού (δηλ. ελέγχεται αν οι δύο ανεξάρτητοι πληθυσμοί έχουν ίση *διακύμανση*, $\sigma^2_1 = \sigma^2_2$) Επειδή **p-value = 0,721 > 0,01**, η υπόθεση της ομοιογένειας ισχύει και συνεπώς αποδέχομαι ότι οι πληθυσμοί μας έχουν ίσες διακυμάνσεις σε επίπεδο στατιστικής σημαντικότητας 1%.

Με το t=-0,103 και df=26 το **p-value=0,918 > 0,01** και συνεπώς αποδέχομαι τη μηδενική υπόθεση ότι σε επίπεδο στατιστικής σημαντικότητας 1% **η μέση πρόσθετη ετήσια αμοιβή των γυναικών της επιχείρησης δεν διαφέρει από τη μέση πρόσθετη αμοιβή των ανδρών ($\mu_{\text{bonus γυναικών}} = \mu_{\text{bonus ανδρών}}$)**

Όπως φαίνεται στον **Πίνακα 35** η μέση διαφορά (Mean Difference) της μέσης πρόσθετης ετήσιας αμοιβής των γυναικών σε σχέση με τη μέση πρόσθετη ετήσια αμοιβή των ανδρών, **με βεβαιότητα 99%, βρίσκεται στο διάστημα (-64,06 € , 59,46 €)**.

Ένας πρόσθετος έλεγχος της αποδοχής ή της απόρριψης της μηδενικής υπόθεσης είναι να εξετάσουμε αν στο ανωτέρω διάστημα, που βρίσκεται (με 99% βεβαιότητα) η διαφορά της μέσης ετήσιας πρόσθετης αμοιβής των γυναικών σε σχέση με την αντίστοιχη των ανδρών, περιλαμβάνεται το μηδέν. Επειδή, λοιπόν, στο διάστημα **(-64,06 € , 59,46 €)** περιλαμβάνεται το μηδέν μπορούμε να αποδεχθούμε την H_o σε επίπεδο στατιστικής σημαντικότητας 1%.

Επειδή τα δύο ανεξάρτητα δείγματα έχουν μικρό μέγεθος και επειδή -προκειμένου να ικανοποιήσουμε την υπόθεση κανονικότητας του δείγματος- «απομακρύναμε» δύο έκτροπες παρατηρήσεις, κρίνεται σκόπιμο να εκτελέσουμε τον μη παραμετρικό έλεγχο των **Mann-Whitney** για τον συνολικό αριθμό των παρατηρήσεών μας (δηλ. χωρίς την «απομάκρυνση» των έκτροπων τιμών). Σημειώνεται ότι οι μη παραμετρικοί έλεγχοι δεν προϋποθέτουν έλεγχο κανονικότητας.

Εκτελώ τη διαδρομή: **Analyze/Nonparametric Tests/Legacy Dialogs/2 Independent Samples**, από την οποία αντλώ τους Πίνακες 36 και 37.

Πίνακας 36

Ranks

	ΦΥΛΟ	N	Mean Rank	Sum of Ranks
BONUS	ΓΥΝΑΙΚΑ	18	15,47	278,50
	ΑΝΤΡΑΣ	12	15,54	186,50
	Total	30		

Πίνακας 37

Test Statistics[a]

	BONUS
Mann-Whitney U	107,500
Wilcoxon W	278,500
Z	-,021
Asymp. Sig. (2-tailed)	**,983**
Exact Sig. [2*(1-tailed Sig.)]	,983[b]

a. Grouping Variable: ΦΥΛΟ

b. Not corrected for ties.

Από τον **Πίνακα 37** (για το σύνολο των εργαζόμενων ανδρών και γυναικών της επιχείρησης) το **p-value = 0,983 > 0,05**, και συνεπώς σε επίπεδο στατιστικής σημαντικότητας 5% αποδεχόμαστε τη μηδενική υπόθεση (H_o) ότι η **μέση πρόσθετη ετήσια αμοιβή των γυναικών δεν διαφέρει από τη μέση πρόσθετη αμοιβή των ανδρών ($\mu_{\text{bonus γυναικών}} = \mu_{\text{bonus ανδρών}}$) ή ομοίως, δεν υπάρχει στατιστικώς σημαντική διαφορά στο μέσο ετήσιο Bonus των γυναικών σε σχέση με το αντίστοιχο των ανδρών.**

11. Έλεγχος της υπόθεσης ότι η μέση ηλικία των ανδρών είναι μικρότερη της μέσης ηλικίας των γυναικών

Ο έλεγχος της υπόθεσης ότι η μέση ηλικία των ανδρών της επιχείρησης είναι μικρότερη της μέσης ηλικίας των γυναικών αφορά στον έλεγχο της μτβ Age για δύο ανεξάρτητους πληθυσμούς (άνδρες / γυναίκες εργαζόμενοι στην επιχείρηση) και θα γίνει με τη βοήθεια του SPSS (Analyze/ Compare Means/ Independent Sample T-Test).

Για να εξάγουμε έγκυρα συμπεράσματα από τον εν λόγω έλεγχο θα πρέπει να ισχύουν δύο προϋποθέσεις:
α) τα ανεξάρτητα δείγματα, που ελέγχονται ως προς μια μτβ, θα πρέπει να ακολουθούν την κανονική κατανομή, και
β) οι διακυμάνσεις (variance) των δύο ανεξάρτητων δειγμάτων θα πρέπει να είναι ίσες.

Ελέγχω την υπόθεση της κανονικότητας της μτβ Age ως προς τους δύο πληθυσμούς (άνδρες / γυναίκες) με την εντολή Analyze/Descriptive Statistics / Explore , Dependent → Age και Factor → Gender.
Από τον **Πίνακα 38** διαπιστώνω ότι το μέγεθος των δύο δειγμάτων (άνδρες / γυναίκες) διαφέρει αισθητά. Οι γυναίκες είναι 18 ενώ οι άνδρες 12.

Πίνακας 38

Case Processing Summary

	ΦΥΛΟ	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
ΗΛΙΚΙΑ	ΓΥΝΑΙΚΑ	18	100,0%	0	0,0%	18	100,0%
	ΑΝΤΡΑΣ	12	100,0%	0	0,0%	12	100,0%

Πίνακας 39

Descriptives

	ΦΥΛΟ			Statistic	Std. Error
ΗΛΙΚΙΑ	ΓΥΝΑΙΚΑ	**Mean**		**34,2056**	,26500
		95% Confidence Interval for Mean	Lower Bound	33,6464	
			Upper Bound	34,7647	
		5% Trimmed Mean		34,1951	
		Median		34,1000	
		Variance		**1,264**	
		Std. Deviation		1,12432	
		Minimum		32,30	
		Maximum		36,30	
		Range		4,00	
		Interquartile Range		,90	
		Skewness		,585	,536
		Kurtosis		,098	1,038
	ΑΝΤΡΑΣ	**Mean**		**34,2000**	,34466
		95% Confidence Interval for Mean	Lower Bound	33,4414	
			Upper Bound	34,9586	
		5% Trimmed Mean		34,2056	
		Median		34,2000	
		Variance		**1,425**	
		Std. Deviation		1,19392	
		Minimum		32,30	
		Maximum		36,00	
		Range		3,70	
		Interquartile Range		2,23	
		Skewness		-,313	,637
		Kurtosis		-,946	1,232

Από τον **Πίνακα 39** διαπιστώνω ότι η μέση ηλικία των ανδρών του δείγματος είναι 34,2 έτη ενώ αντίστοιχη είναι και η μέση ηλικία των γυναικών.

Διεξάγω τον έλεγχο κανονικότητας του δείγματος ως προς την μτβ Age, χρησιμοποιώντας τον **Πίνακα 40**.

Πίνακας 40

Tests of Normality

	ΦΥΛΟ	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
ΗΛΙΚΙΑ	ΓΥΝΑΙΚΑ	,244	18	,006	,912	18	**,093**
	ΑΝΤΡΑΣ	,132	12	,200*	,940	12	**,500**

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Επειδή το μέγεθος του δείγματος είναι μικρότερο από 50 (n<50), ο έλεγχος κανονικότητας θα γίνει με τη μέθοδο των Shapiro-Wilk σε επίπεδο στατιστικής σημαντικότητας α=0,05.
Διατυπώνω τη μηδενική και την εναλλακτική υπόθεση για τους άνδρες:
H_o : Η ηλικία των ανδρών του δείγματος ακολουθεί κανονική κατανομή
H_1: Η ηλικία των ανδρών του δείγματος ΔΕΝ ακολουθεί κανονική κατανομή
Επειδή **p-value = 0,500 > 0,05** αποδέχομαι την Ηο **ότι η ηλικία των ανδρών του δείγματος ακολουθεί κανονική κατανομή**.
Ομοίως για τις γυναίκες, επειδή **p-value = 0,093 > 0,05** αποδέχομαι την Ηο **ότι η ηλικία των γυναικών του δείγματος ακολουθεί κανονική κατανομή.**

Για τις γυναίκες του δείγματος η αποδοχή της μηδενικής υπόθεσης περί κανονικότητας κρίνεται οριακή, και για το λόγο αυτό θεωρείται σκόπιμο να ελεγχθούν οι λόγοι $\boldsymbol{\lambda sk}$ και $\boldsymbol{\lambda ku}$:

Πίνακας 41

Statistics

ΗΛΙΚΙΑ (ΓΥΝΑΙΚΕΣ)

N	Valid	18
	Missing	0
Mean		34,2056
Median		34,1000
Mode		33,70[a]
Std. Deviation		1,12432
Variance		1,264
Skewness		,585
Std. Error of Skewness		,536
Kurtosis		,098
Std. Error of Kurtosis		1,038

a. Multiple modes exist. The smallest value is shown

Από τον παραπλεύρως πίνακα διαπιστώνω ότι:
Mean≈ Median ≈Mode = 33,70
και

$$\lambda sk = \frac{Skewness}{standard\ error\ of\ Skewness} = \frac{0{,}585}{0{,}536} = 1{,}09 < 2$$

$$\lambda ku = \frac{Kurtosis}{standard\ error\ of\ Kurtosis} = \frac{0{,}098}{1{,}038} = 0{,}09 < 2$$

Συνεπώς, επειδή τα λsk και λku λαμβάνουν τιμές εντός του διαστήματος (-2,2), **η ηλικία των γυναικών του δείγματος ακολουθεί την κανονική κατανομή**.

Ο λόγος των διακυμάνσεων των ηλικιών των δύο δειγμάτων είναι:

$$\frac{Variance\ Age\ Ανδρών}{Variance\ Age\ Γυναικών} = \frac{1{,}425}{1{,}264} = 1{,}127$$

Εφόσον ισχύουν οι προϋποθέσεις εφαρμογής του Independent Sample T-Test, διατυπώνω τη μηδενική και την εναλλακτική υπόθεση:

H_o: **Η μέση ηλικία των ανδρών της επιχείρησης δεν διαφέρει από τη μέση ηλικία των γυναικών της επιχείρησης (μ ηλικία ανδρών = μ ηλικία γυναικών)**

H_1: **Η μέση ηλικία των ανδρών της επιχείρησης είναι μικρότερη από τη μέση ηλικία των γυναικών της επιχείρησης (μ ηλικία ανδρών < μ ηλικία γυναικών)**

και εκτελώ τον σχετικό έλεγχο στο SPSS, απ' όπου λαμβάνω τους εξής Πίνακες:

Πίνακας 42

Group Statistics

	ΦΥΛΟ	N	Mean	Std. Deviation	Std. Error Mean
ΗΛΙΚΙΑ	ΓΥΝΑΙΚΑ	18	34,2056	1,12432	,26500
	ΑΝΤΡΑΣ	12	34,2000	1,19392	,34466

Από τον **Πίνακα 42** βλέπω τη μέση ηλικία των ανδρών και των γυναικών του δείγματος και θέλω να ελέγξω αν αυτή η διαφορά είναι στατιστικώς σημαντική στον πληθυσμό (δηλ. στο σύνολο των εργαζομένων της επιχείρησης), σε επίπεδο στατιστικής σημαντικότητας 5%.

Πίνακας 43

Independent Samples Test

			ΗΛΙΚΙΑ	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		,215	
	Sig.		**,646**	
t-test for Equality of Means	t		,013	,013
	df		28	22,714
	Sig. (2-tailed)		**,990**	,990
	Mean Difference		,00556	,00556
	Std. Error Difference		,42939	,43476
	95% Confidence Interval of the Difference	Lower	-,87400	-,89444
		Upper	,88511	,90555

Στον **Πίνακα 43**, με το **Test Levene's** ελέγχεται η υπόθεση της **ομοιογένειας** του πληθυσμού (δηλ. ελέγχεται αν οι δύο ανεξάρτητοι πληθυσμοί έχουν ίση *διακύμανση*, $\sigma^2_1 = \sigma^2_2$). Επειδή **p-value = 0,646 > 0,05**, η υπόθεση της ομοιογένειας ισχύει και συνεπώς αποδέχομαι ότι οι δύο πληθυσμοί έχουν ίσες διακυμάνσεις σε επίπεδο στατιστικής σημαντικότητας 5%.

Με το t=0,013 και df=28 το **Sig. (2-tailed)=0,990.**

Επειδή ο έλεγχος είναι μονόπλευρος, η ανισότητα είναι **<** και το t>0, το **p-value=1-**$\frac{1}{2}$$\boldsymbol{Sig.}$ **= 0,505**

Έτσι, επειδή p-value=0,505 > 0,05 αποδέχομαι τη μηδενική υπόθεση ότι σε επίπεδο στατιστικής σημαντικότητας 5% **η μέση ηλικία των ανδρών της επιχείρησης δεν διαφέρει από τη μέση ηλικία**

των γυναικών της επιχείρησης (μ $_{\text{ηλικία ανδρών}}$ = μ $_{\text{ηλικία γυναικών}}$) ή ομοίως, η μέση ηλικία των ανδρών δεν είναι μικρότερη της μέσης ηλικίας των γυναικών της επιχείρησης.
Όπως φαίνεται στον **Πίνακα 43** η μέση διαφορά (Mean Difference) της μέσης ηλικίας των ανδρών σε σχέση με τη μέση ηλικία των γυναικών, **με βεβαιότητα 95%, βρίσκεται στο διάστημα (-0,874 , 0,885) έτη**.

Ένας πρόσθετος έλεγχος της αποδοχής ή της απόρριψης της μηδενικής υπόθεσης είναι να εξετάσουμε αν στο ανωτέρω διάστημα, που βρίσκεται (με 95% βεβαιότητα) η διαφορά της μέσης ηλικίας των ανδρών της επιχείρησης σε σχέση με την αντίστοιχη των γυναικών, περιλαμβάνεται το μηδέν. Επειδή, λοιπόν, στο διάστημα **(-0,874 , 0,885)** περιλαμβάνεται το μηδέν μπορούμε να αποδεχθούμε την H_o σε επίπεδο στατιστικής σημαντικότητας 5%.

12. Έλεγχος υπόθεσης ότι το μέσο Bonus αποφοίτων Λυκείου ισούται με το μέσο Bonus των αποφοίτων ΑΕΙ.

Ο έλεγχος της υπόθεσης της ισότητας των μέσων της μτβ Bonus για δύο ανεξάρτητα δείγματα (απόφοιτοι Λυκείου / απόφοιτοι ΑΕΙ στην επιχείρηση) αφορά στον έλεγχο της μτβ Bonus για δύο ανεξάρτητους πληθυσμούς (απόφοιτοι Λυκείου / απόφοιτοι ΑΕΙ στην επιχείρηση) και θα γίνει με τη βοήθεια του SPSS (Analyze/ Compare Means/ Independent Sample T-Test).
Για να εξάγουμε έγκυρα συμπεράσματα από τον εν λόγω έλεγχο θα πρέπει να ισχύουν δύο προϋποθέσεις:
α) τα ανεξάρτητα δείγματα, που ελέγχονται ως προς μια μτβ, θα πρέπει να ακολουθούν την κανονική κατανομή, και
β) οι διακυμάνσεις (variance) των δύο ανεξάρτητων δειγμάτων θα πρέπει να είναι ίσες.
Ελέγχω την υπόθεση της κανονικότητας της μτβ bonus ως προς τα δύο ανεξάρτητα δείγματα (απόφοιτοι Λυκείου / Απόφοιτοι ΑΕΙ) με την εντολή Analyze/Descriptive Statistics / Explore , Dependent → Bonus και Factor → Studies.

Από τον **Πίνακα 44** διαπιστώνω ότι το δείγμα μας περιλαμβάνει δεκαέξι (16) αποφοίτους Λυκείου και δεκατέσσερις (14) αποφοίτους ΑΕΙ.

Πίνακας 44

Case Processing Summary

	ΣΠΟΥΔΕΣ	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
BONUS	ΛΥΚΕΙΟ	16	100,0%	0	0,0%	16	100,0%
	ΑΕΙ	14	100,0%	0	0,0%	14	100,0%

Στον **Πίνακα 45** αναφέρονται τα περιγραφικά στατιστικά των δύο ανεξάρτητων δειγμάτων που ελέγχονται ως προς τη μεταβλητή Bonus. Ειδικότερα, η μέση πρόσθετη αμοιβή των εργαζομένων του δείγματος που είναι απόφοιτοι Λυκείου είναι 720,00€ ενώ η αντίστοιχη αμοιβή για τους αποφοίτους ΑΕΙ είναι 761,43€. Επιπλέον διαπιστώνεται ότι το εύρος της πρόσθετης αμοιβής των αποφοίτων Λυκείου (Range= 380) είναι αρκετά μεγαλύτερο του εύρους της πρόσθετης αμοιβής των αποφοίτων ΑΕΙ (Range=200), και ανέρχεται στα 180€ (380€-200€). Αυτή η παρατήρηση σε συνδυασμό με τη σύγκριση της διαμέσου της πρόσθετης αμοιβής, η οποία διαφέρει μόλις κατά

30€ (=780€-750€), μας προϊδεάζει για την ύπαρξη έκτροπων τιμών στις πρόσθετες αποδοχές των αποφοίτων Λυκείου.

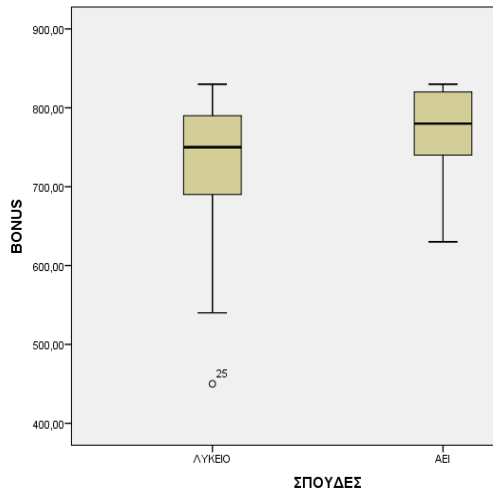
Πίνακας 45

Descriptives

	ΣΠΟΥΔΕΣ			Statistic	Std. Error
BONUS	ΛΥΚΕΙΟ	**Mean**		**720,0000**	25,04995
		95% Confidence Interval for Mean	Lower Bound	666,6073	
			Upper Bound	773,3927	
		5% Trimmed Mean		728,8889	
		Median		**750,0000**	
		Variance		10040,000	
		Std. Deviation		100,19980	
		Minimum		450,00	
		Maximum		830,00	
		Range		**380,00**	
		Interquartile Range		110,00	
		Skewness		-1,709	,564
		Kurtosis		2,881	1,091
	ΑΕΙ	**Mean**		**761,4286**	17,28422
		95% Confidence Interval for Mean	Lower Bound	724,0883	
			Upper Bound	798,7688	
		5% Trimmed Mean		764,9206	
		Median		**780,0000**	
		Variance		4182,418	
		Std. Deviation		64,67161	
		Minimum		630,00	
		Maximum		830,00	
		Range		**200,00**	
		Interquartile Range		95,00	
		Skewness		-,919	,597
		Kurtosis		-,256	1,154

Η ανωτέρω παρατήρηση περί έκτροπων δειγματικών τιμών στα bonus των αποφοίτων Λυκείου επιβεβαιώνεται τόσο από το σχετικό θηκόγραμμα (Γράφημα 10) όσο και από τον πίνακα με τις ακραίες τιμές ανά πληθυσμιακή ομάδα (Πίνακας 46).

Γράφημα 10



Πίνακας 46

Extreme Values

	ΣΠΟΥΔΕΣ			Case Number	Value
BONUS	ΛΥΚΕΙΟ	Highest	1	7	830,00
			2	3	800,00
		Lowest	**1**	**25**	**450,00**
			2	29	540,00
			3	15	660,00
	ΑΕΙ	Highest	1	24	830,00
			2	13	820,00
		Lowest	1	1	630,00
			2	22	660,00

Προκειμένου να προχωρήσω στον αναγκαίο έλεγχο κανονικότητας της μτβ Bonus ως προς το επίπεδο σπουδών, διατυπώνω τη μηδενική και την εναλλακτική υπόθεση για τους αποφοίτους Λυκείου και διεξάγω τον έλεγχο σε επίπεδο στατιστικής σημαντικότητας 5%:

H_o: Η πρόσθετη ετήσια αμοιβή των αποφοίτων Λυκείου του δείγματος ακολουθεί κανονική κατανομή.

H_1: Η πρόσθετη ετήσια αμοιβή των αποφοίτων Λυκείου του δείγματος ΔΕΝ ακολουθεί κανονική κατανομή.

Πίνακας 47

Tests of Normality

	ΣΠΟΥΔΕΣ	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
BONUS	ΛΥΚΕΙΟ	,227	16	,027	,820	16	,005
	ΑΕΙ	,225	14	,054	,872	14	,045

a. Lilliefors Significance Correction

Ο έλεγχος γίνεται με το κριτήριο Shapiro-Wilk επειδή το μέγεθος του δείγματός μας είναι μικρότερο από 50. Για τους αποφοίτους Λυκείου, το **p-value =0,005 < 0,05** και συνεπώς απορρίπτω την Ηο σε επίπεδο στατιστικής σημαντικότητας 5% και αποδέχομαι την εναλλακτική ότι η πρόσθετη ετήσια αμοιβή των αποφοίτων Λυκείου του δείγματος ΔΕΝ ακολουθεί την κανονική κατανομή.

Αντίστοιχα και για τους αποφοίτους ΑΕΙ, το **p-value =0,045 < 0,05** και συνεπώς απορρίπτω **οριακά** την Ηο σε επίπεδο στατιστικής σημαντικότητας 5% και αποδέχομαι την εναλλακτική, σύμφωνα με την οποία η πρόσθετη ετήσια αμοιβή των αποφοίτων ΑΕΙ του δείγματος ΔΕΝ ακολουθεί την κανονική κατανομή.

Επειδή το μέγεθος του δείγματος είναι μεγάλο ($n \geq 30$), γνωρίζουμε ότι βάσει του Κεντρικού Οριακού Θεωρήματος **το t-test για δύο ανεξάρτητα δείγματα είναι ανθεκτικό έναντι της μη κανονικότητας.**

Βασιζόμενοι στην ανθεκτικότητα του t-test ως προς την υπόθεση της κανονικότητας λόγω του Κεντρικού Οριακού Θεωρήματος, εφαρμόζουμε του Independent Sample T-Test, και διατυπώνουμε τη μηδενική και την εναλλακτική υπόθεση:

H_o: **Η μέση πρόσθετη ετήσια αμοιβή των αποφοίτων Λυκείου της επιχείρησης δεν διαφέρει από τη μέση ετήσια πρόσθετη αμοιβή των αποφοίτων ΑΕΙ ($\mu_{\text{bonus Λυκείου}} = \mu_{\text{bonus ΑΕΙ}}$)**

H_1: **Η μέση πρόσθετη ετήσια αμοιβή των αποφοίτων Λυκείου της επιχείρησης διαφέρει από τη μέση ετήσια πρόσθετη αμοιβή των αποφοίτων ΑΕΙ ($\mu_{\text{bonus Λυκείου}} \neq \mu_{\text{bonus ΑΕΙ}}$)**

Πίνακας 48

Group Statistics

	ΣΠΟΥΔΕΣ	N	Mean	Std. Deviation	Std. Error Mean
BONUS	ΛΥΚΕΙΟ	16	720,0000	100,19980	25,04995
	ΑΕΙ	14	761,4286	64,67161	17,28422

Από τον **Πίνακα 48** βλέπω το μέσο Bonus των αποφοίτων Λυκείου του δείγματος είναι 720,00€ και το αντίστοιχο των αποφοίτων ΑΕΙ είναι 761,43€ και θέλω να ελέγξω αν αυτή η διαφορά είναι στατιστικώς σημαντική στον πληθυσμό (δηλ. στο σύνολο των εργαζομένων της επιχείρησης), σε επίπεδο στατιστικής σημαντικότητας 5%.

Πίνακας 49

Independent Samples Test

			BONUS	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		,947	
	Sig.		,339	
t-test for Equality of Means	t		-1,323	-1,361
	df		28	25,907
	Sig. (2-tailed)		,197	,185
	Mean Difference		-41,42857	-41,42857
	Std. Error Difference		31,31153	30,43426
	95% Confidence Interval of the Difference	Lower	-105,56733	-103,99801
		Upper	22,71019	21,14087

Στον **Πίνακα 49**, με το **Test Levene's** ελέγχεται η υπόθεση της **ομοιογένειας** του πληθυσμού (δηλ. ελέγχεται αν οι δύο ανεξάρτητοι πληθυσμοί έχουν ίση *διακύμανση*, $\sigma^2_1 = \sigma^2_2$). Επειδή **p-value = 0,339 > 0,05**, η υπόθεση της ομοιογένειας ισχύει και συνεπώς αποδέχομαι ότι οι δύο πληθυσμοί έχουν ίσες διακυμάνσεις σε επίπεδο στατιστικής σημαντικότητας 5%.

Με το t=-1,323 και df=28 το **p-value=0,197 > 0,05** και συνεπώς αποδέχομαι τη μηδενική υπόθεση ότι **η μέση πρόσθετη ετήσια αμοιβή των αποφοίτων Λυκείου της επιχείρησης δεν διαφέρει από τη μέση πρόσθετη ετήσια αμοιβή των αποφοίτων ΑΕΙ της επιχείρησης (μ bonus Λυκείου = μ bonus ΑΕΙ)** σε επίπεδο στατιστικής σημαντικότητας 5%

Όπως φαίνεται στον **Πίνακα 49** η μέση διαφορά (Mean Difference) της μέσης πρόσθετης ετήσιας αμοιβής των αποφοίτων Λυκείου σε σχέση με τη μέση πρόσθετη ετήσια αμοιβή των αποφοίτων ΑΕΙ, **με βεβαιότητα 95%, βρίσκεται στο διάστημα (-105,57 € , 22,71 €)**.

Ένας πρόσθετος έλεγχος της αποδοχής ή της απόρριψης της μηδενικής υπόθεσης είναι να εξετάσουμε αν στο ανωτέρω διάστημα, που βρίσκεται (με 95% βεβαιότητα) η διαφορά της μέσης ετήσιας πρόσθετης αμοιβής των γυναικών σε σχέση με την αντίστοιχη των ανδρών, περιλαμβάνεται το μηδέν. Επειδή, λοιπόν, στο διάστημα **(-105,57 € , 22,71 €)** περιλαμβάνεται το μηδέν μπορούμε να αποδεχθούμε την H_o σε επίπεδο στατιστικής σημαντικότητας 5%.

Επειδή τα δύο ανεξάρτητα δείγματα έχουν μικρό μέγεθος και επειδή -προκειμένου να ικανοποιήσουμε την υπόθεση κανονικότητας του δείγματος- κάναμε χρήση του Κεντρικού Οριακού Θεωρήματος, κρίνεται σκόπιμο να εκτελέσουμε τον μη παραμετρικό έλεγχο των **Mann-Whitney**. Σημειώνεται ότι οι μη παραμετρικοί έλεγχοι δεν προϋποθέτουν έλεγχο κανονικότητας.

Εκτελώ τη διαδρομή: **Analyze/Nonparametric Tests/Legacy Dialogs/2 Independent Samples**, από την οποία αντλώ τον Πίνακα 50.

Πίνακας 50

Test Statistics[a]

	BONUS
Mann-Whitney U	81,000
Wilcoxon W	217,000
Z	-1,298
Asymp. Sig. (2-tailed)	,194
Exact Sig. [2*(1-tailed Sig.)]	,208[b]

a. Grouping Variable: ΣΠΟΥΔΕΣ

b. Not corrected for ties.

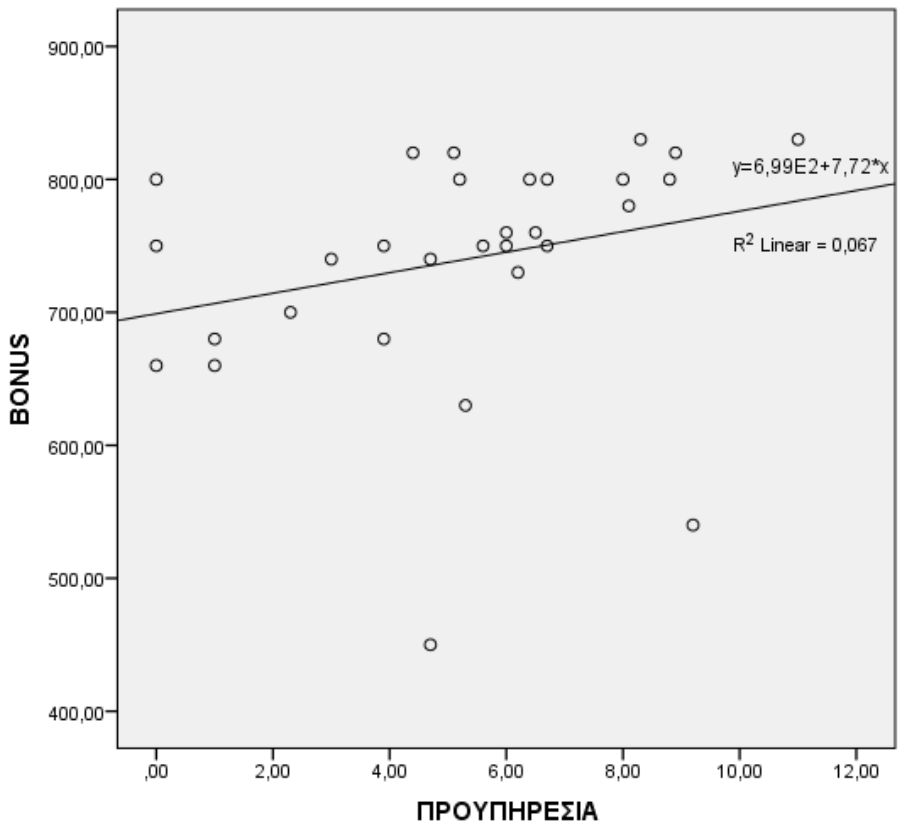
Από τον **Πίνακα 50** το **p-value = 0,194 > 0,05**, και συνεπώς σε επίπεδο στατιστικής σημαντικότητας 5% αποδεχόμαστε τη μηδενική υπόθεση (H_o) ότι η **μέση πρόσθετη ετήσια αμοιβή των αποφοίτων Λυκείου δεν διαφέρει από τη μέση πρόσθετη αμοιβή των αποφοίτων ΑΕΙ ή ομοίως, δεν υπάρχει στατιστικώς σημαντική διαφορά στο μέσο ετήσιο Bonus των αποφοίτων Λυκείου σε σχέση με το αντίστοιχο των αποφοίτων ΑΕΙ.**

13. Εξέταση για το αν το ετήσιο Bonus εξαρτάται γραμμικά από την προϋπηρεσία των υπαλλήλων

Για να μελετήσουμε τη σχέση μεταξύ των δύο μεταβλητών εξετάζουμε το διάγραμμα διασποράς της μιας μτβ (Bonus) συναρτήσει της άλλης (Προϋπηρεσία).

Με τη βοήθεια του SPSS, ακολουθώντας τη διαδρομή, Graphs/Legacy Dialogs/Scatter dot/Simple Scatter και θέτοντας Y axis→ Bonus και X axis→Προϋπηρεσία, λαμβάνουμε το διάγραμμα διασποράς (**Γράφημα 11**) που ακολουθεί.

Γράφημα 11



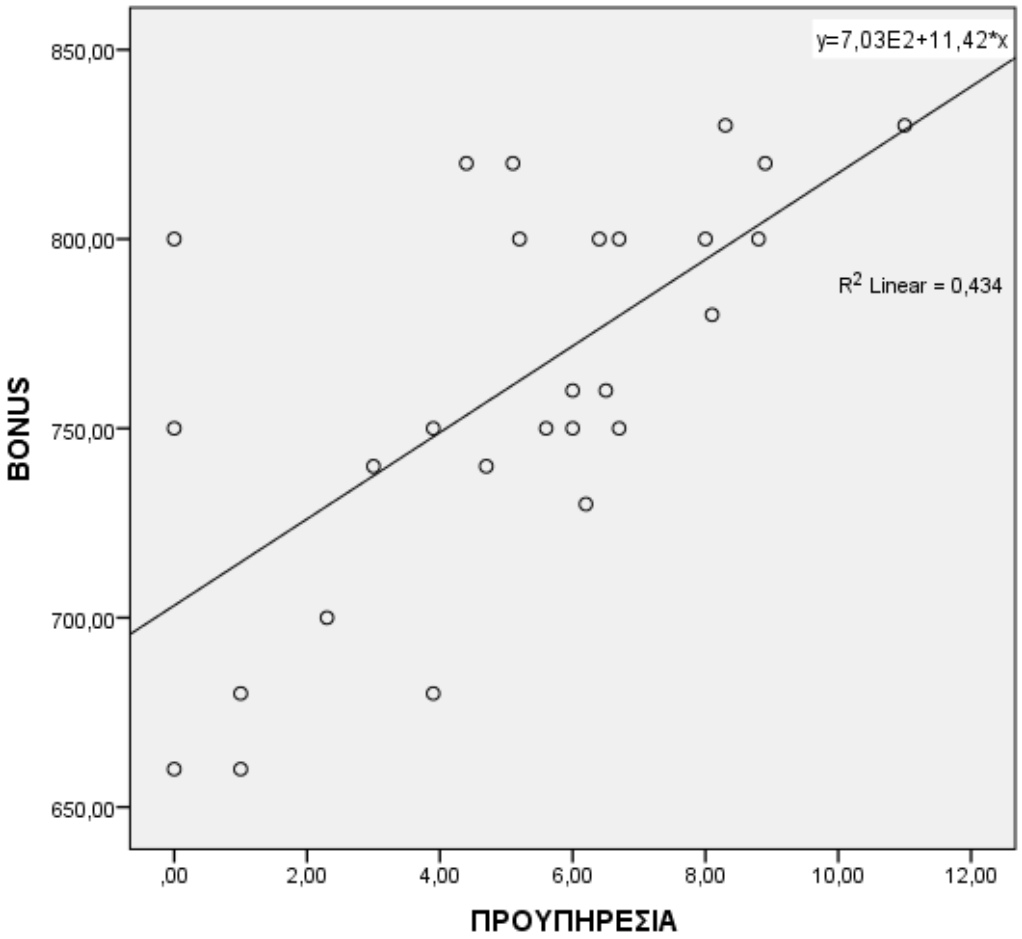
Από το διάγραμμα διασποράς διαπιστώνουμε ότι δύο-τρεις έκτροπες τιμές ενδεχομένως να μεταβάλλουν τη θέση και την κλίση της ευθείας παλινδρόμησης. Επίσης, από τον συντελεστή προσδιορισμού R^2=0,067 συμπεραίνουμε ότι η ευθεία παλινδρόμησης ερμηνεύει μόλις το 6,7% της μεταβλητότητας των τιμών της μτβ Bonus από την προϋπηρεσία.

Οι παρατηρήσεις αυτές, σε συνδυασμό με το γεγονός ότι η μτβ Bonus, για τις παρατηρήσεις του δείγματος, όπως είδαμε στην $8^η$ ενότητα, δεν ακολουθεί την κανονική κατανομή, αλλά και λαμβάνοντας υπόψη ότι για την εύρεση του συντελεστή γραμμικής συσχέτισης του Pearson μια από τις προϋποθέσεις είναι οι μεταβλητές να ακολουθούν την κανονική κατανομή, αποφασίζουμε να «απομακρύνουμε» τις έκτροπες χαμηλές τιμές.

Ειδικότερα, επιλέγουμε τις παρατηρήσεις οι οποίες έχουν τιμή για τη μτβ Bonus πάνω από 640€ (Data/Select Cases/if condition is satisfied/Bonus>640), και έτσι «απομακρύνουμε» τρεις τιμές που είναι μικρότερες ή ίσες των 640€

Το νέο διάγραμμα διασποράς παρουσιάζεται στο **Γράφημα 12**.

Γράφημα 12



Για να προσδιορίσουμε την **ένταση** της γραμμικής συσχέτισης μεταξύ των δύο μτβ θα χρησιμοποιήσουμε το **συντελεστή γραμμικής συσχέτισης του Pearson**. Προϋποθέσεις της χρήσης του εν λόγω συντελεστή είναι:
α) Από το διάγραμμα διασποράς να είναι ορατή μια γραμμική συσχέτιση,
β) Οι μεταβλητές να είναι ποσοτικές και συνεχείς[17], και
γ) Οι μεταβλητές να ακολουθούν την κανονική κατανομή.

Επειδή οι δύο πρώτες προϋποθέσεις ισχύουν, θα εξετάσουμε αν οι μτβ Bonus και Προϋπηρεσία ακολουθούν την κανονική κατανομή.

Για τη μεταβλητή Bonus, διατυπώνω τη μηδενική και την εναλλακτική υπόθεση:
Ho: Η μτβ Bonus του δείγματός μας ακολουθεί την κανονική κατανομή
H_1: Η μτβ Bonus του δείγματός μας ΔΕΝ ακολουθεί την κανονική κατανομή.

Πίνακας 51

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
BONUS	,180	27	,025	,916	27	,032

a. Lilliefors Significance Correction

Από το κριτήριο Shapiro-Wilk[18] και σε επίπεδο στατιστικής σημαντικότητας α=1% διαπιστώνω ότι **p-value=0,032 > 0,01** και συνεπώς αποδέχομαι την Ho ότι η μτβ Bonus ακολουθεί την κανονική κατανομή.
Για τη μεταβλητή Προϋπηρεσία, διατυπώνω τη μηδενική και την εναλλακτική υπόθεση:
Ho: Η μτβ Προϋπηρεσία του δείγματός μας ακολουθεί την κανονική κατανομή
H_1: Η μτβ Προϋπηρεσία του δείγματός μας ΔΕΝ ακολουθεί την κανονική κατανομή.

Πίνακας 52

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ΠΡΟΥΠΗΡΕΣΙΑ	,101	27	,200*	,957	27	,317

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Από το κριτήριο Shapiro-Wilk και σε επίπεδο στατιστικής σημαντικότητας α=1% διαπιστώνω ότι **p-value= 0,317 > 0,01** και συνεπώς αποδέχομαι την Ho ότι η μτβ Προϋπηρεσία ακολουθεί την κανονική κατανομή.

Εφόσον, λοιπόν, ικανοποιούνται οι προϋποθέσεις χρήσης του συντελεστή συσχέτισης του Pearson, ακολουθώ τη διαδρομή: Analyze/Correlate/Bivariate και Variable→ Bonus και Work.

Πίνακας 53

Correlations

		BONUS	ΠΡΟΥΠΗΡΕΣΙΑ
BONUS	Pearson Correlation	1	,658**
	Sig. (2-tailed)		,000
	N	27	27
ΠΡΟΥΠΗΡΕΣΙΑ	Pearson Correlation	,658**	1
	Sig. (2-tailed)	,000	
	N	27	27

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

Από τον **Πίνακα 53** βλέπουμε ότι ο συντελεστής γραμμικής συσχέτισης του Pearson είναι **0,658** δηλαδή υπάρχει μια **ισχυρή**[19] θετική γραμμική συσχέτιση[20] μεταξύ της προϋπηρεσίας και της πρόσθετης ετήσιας αμοιβής των εργαζομένων της επιχείρηση στο **δείγμα μας**.
Για να ελεγχθεί το επίπεδο στατιστικής σημαντικότητας στον **πληθυσμό** (δηλ. στο σύνολο των εργαζομένων της επιχείρησης), διατυπώνουμε τη μηδενική και την εναλλακτική υπόθεση:
Ho: ρ=0, **δεν** υπάρχει γραμμική συσχέτιση μεταξύ των μεταβλητών στον πληθυσμό (στο σύνολο των εργαζομένων),
H_1: ρ≠ 0, υπάρχει γραμμική συσχέτιση μεταξύ των μεταβλητών στον πληθυσμό.

Επειδή p-value=0,0001<0,01 απορρίπτουμε την Ho και συνεπώς δεχόμαστε ότι υπάρχει **στατιστικώς σημαντική γραμμική συσχέτιση μεταξύ της εργασιακής προϋπηρεσίας και του επιπέδου των ετήσιων πρόσθετων αμοιβών των εργαζομένων της επιχείρησης**.

Αφού διαπιστώσαμε ότι υπάρχει στατιστικώς σημαντική γραμμική συσχέτιση μεταξύ των δύο μεταβλητών στον πληθυσμό, θα προσπαθήσουμε να προσδιορίσουμε τη μαθηματική σχέση που τις συνδέει, δηλ. μια ευθεία της μορφής $\widehat{Y}=\beta_o + \beta_1 X$ (όπου β_o ο σταθερός όρος και β_1 η κλίση της ευθείας), βάσει της οποίας δίνοντας τιμές στην ανεξάρτητη μτβ (Χ) θα μπορούμε να εκτιμήσουμε την τιμή της ανεξάρτητης μτβ (Υ).

Για την εύρεση της ευθείας παλινδρόμησης στο SPSS ακολουθώ τη διαδρομή: Analyze / Regression / Linear, στο πεδίο Dependent→ Bonus, στο πεδίο Independent → Προϋπηρεσίας και στο Regression Coefficients → estimates , confidence intervals (95%), model fit→ descriptives, και λαμβάνουμε τους κατωτέρω Πίνακες.

Πίνακας 54

Descriptive Statistics

	Mean	Std. Deviation	N
BONUS	761,4815	51,56801	27
ΠΡΟΥΠΗΡΕΣΙΑ	5,1000	2,97347	27

Correlations

		BONUS	ΠΡΟΥΠΗΡΕΣΙΑ
Pearson Correlation	BONUS	1,000	**,658**
	ΠΡΟΥΠΗΡΕΣΙΑ	**,658**	1,000
Sig. (1-tailed)	BONUS		,000
	ΠΡΟΥΠΗΡΕΣΙΑ	,000	.
N	BONUS	27	27
	ΠΡΟΥΠΗΡΕΣΙΑ	27	27

Πίνακας 55

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)	703,245	15,339		45,847	,000	671,654	734,836
	ΠΡΟΥΠΗΡΕΣΙΑ	11,419	2,611	,658	4,374	,000	6,042	16,796

a. Dependent Variable: BONUS

Η διαπίστωση της ύπαρξης γραμμικής σχέσης ανάμεσα στην εξαρτημένη και την ανεξάρτητη μτβ, σε επίπεδο στατιστικής σημαντικότητας 5%, γίνεται από τον έλεγχο της υπόθεσης:
Ho: $\beta_1=0$ (δηλ. η μτβ Bonus δεν εξαρτάται γραμμικά από την μτβ Προϋπηρεσία)
H_1 : $\beta_1 \neq 0$ (δηλ. η μτβ Bonus εξαρτάται γραμμικά από τη μτβ Προϋπηρεσία)
Από τον **Πίνακα 55** βλέπουμε ότι t=4,374 και p-value=0,0001 < 0,05, και συνεπώς απορρίπτουμε την Ho και δεχόμαστε ότι η **$\beta_1 \neq 0$,** δηλ. **σε επίπεδο στατιστικής σημαντικότητας 5% η μτβ Bonus εξαρτάται γραμμικά από την μτβ Προϋπηρεσία**.

Από τον **Πίνακα 55** λαμβάνουμε την τιμή του σταθερού όρου της ευθείας (β_o = 703,245), την τιμή της κλίσης (β_1=11,419) και τα σχετικά διαστήματα εμπιστοσύνης [με βεβαιότητα 95% ο σταθερός όρος της ευθείας παίρνει τιμές στο διάστημα (671,654 , 734,836) και **η κλίση (6,042 , 16,796)**].
Συνεπώς, η ευθεία παλινδρόμησης είναι:

$$\widehat{Y}= 703,245 + 11,419\ X$$

Η κλίση της ευθείας μάς δείχνει ότι για κάθε μοναδιαία μεταβολή των ετών προϋπηρεσίας μεταβάλλεται το ετήσιο Bonus κατά 11,419 €. Επιπλέον, από το μαθηματικό τύπο της ευθείας μπορούμε να προβλέψουμε ότι ένας εργαζόμενος με μηδενική προϋπηρεσία μπορεί να λάβει ετήσιο bonus της τάξης των 703 € περίπου.

Από τον κατωτέρω **Πίνακα 56**, λαμβάνουμε τον **συντελεστή προσδιορισμού R^2** ο οποίος μετρά το ποσοστό της ερμηνευόμενης μεταβλητότητας της εξαρτημένης μεταβλητής (Bonus) που οφείλεται στην ευθεία παλινδρόμησης.
Έτσι, **R^2 = 0,434** σημαίνει ότι η ευθεία παλινδρόμησης, που προσδιορίσαμε ανωτέρω, μπορεί να ερμηνεύσει κατά 43,4% τη μεταβλητότητα του ύψους του ετήσιου bonus των εργαζομένων της επιχείρησης.

Πίνακας 56

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,658[a]	,434	,411	39,58073

a. Predictors: (Constant), ΠΡΟΥΠΗΡΕΣΙΑ

Μέρος Β.

1. Εξέταση των γραμμικών μοντέλων που συνδέουν την εξαρτημένη με κάθε μία από τις ανεξάρτητες μτβ ξεχωριστά.

(α). Ακολουθώντας τη διαδρομή Analyze / Regression /Linear και θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → **Ηλικία**, λαμβάνουμε τους εξής Πίνακες:

Πίνακας 57

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,193[a]	,037	,028	780,481895

a. Predictors: (Constant), ΗΛΙΚΙΑ

Πίνακας 58

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	612,590	224,801		2,725	,008
	ΗΛΙΚΙΑ	10,127	5,195	,193	1,949	,054

a. Dependent Variable: ΠΟΣΟΤΡΕΧΟΝΤΟΣΕΤΟΥΣ

Από τον **Πίνακα 57** λαμβάνουμε τον συντελεστή **R=0,193** (ο συντελεστής R για απλή γραμμική παλινδρόμηση είναι ίσος με τον **συντελεστή γραμμικής συσχέτισης του Pearson**), τον συντελεστή προσδιορισμού **R^2=0,037** (ο συντελεστής R^2 μας πληροφορεί για το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής το οποίο εξηγείται από τη γραμμική σχέση που την συνδέει με την ανεξάρτητη μτβ) και τον προσαρμοσμένο συντελεστή προσδιορισμού **Adjusted R^2 = 0,028** (που μας

πληροφορεί για την προσαρμογή του γραμμικού μοντέλου, που προσδιορίσαμε από το δείγμα, στον πληθυσμό -συντελεστής διόρθωσης **R^2**).
Από τον **Πίνακα 58** λαμβάνουμε τον σταθερό όρο και την κλίση της ευθείας παλινδρόμησης, η οποία είναι $\widehat{Y}$**= 612,590 + 10,127 X_1** (όπου X_1, η ηλικία)

(β). Ακολουθώντας τη διαδρομή Analyze / Regression /Linear και θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → **Μισθός**, λαμβάνουμε τους εξής Πίνακες:

Πίνακας 59

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,751[a]	,565	,560	524,835092

a. Predictors: (Constant), ΜΙΣΘΟΙ

Πίνακας 60

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	50,197	101,030		,497	,620
	ΜΙΣΘΟΙ	,018	,002	,751	11,275	,000

a. Dependent Variable: ΠΟΣΟΤΡΕΧΟΝΤΟΣΕΤΟΥΣ

Από τον **Πίνακα 59** λαμβάνουμε τον συντελεστή **R=0,751 ,** τον συντελεστή προσδιορισμού **R^2=0,565** και τον προσαρμοσμένο συντελεστή προσδιορισμού **Adj R^2 = 0,560.**
Από τον **Πίνακα 60** λαμβάνουμε τον σταθερό όρο και την κλίση της ευθείας παλινδρόμησης, η οποία είναι $\widehat{Y}$**= 50,197 + 0,018 X_2** (όπου X_2, ο ετήσιος μισθός)

(γ). Με την ίδια διαδικασία (Analyze / Regression /Linear και θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → **αρ. παιδιών**), λαμβάνουμε τους εξής Πίνακες:

Πίνακας 61

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,157[a]	,025	,015	785,659304

a. Predictors: (Constant), ΑΡΠΑΙΔΙΩΝ

Πίνακας 62

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1138,596	107,452		10,596	,000
	ΑΡΠΑΙΔΙΩΝ	-104,580	66,639	-,157	-1,569	,120

a. Dependent Variable: ΠΟΣΟΤΡΕΧΟΝΤΟΣΕΤΟΥΣ

Από τον **Πίνακα 61** λαμβάνουμε τον συντελεστή **R=0,157 ,** τον συντελεστή προσδιορισμού **R^2=0,025** και τον προσαρμοσμένο συντελεστή προσδιορισμού **Adj R^2 = 0,015.**
Από τον **Πίνακα 62** λαμβάνουμε τον σταθερό όρο και την κλίση της ευθείας παλινδρόμησης, η οποία είναι $\widehat{Y}$**= 1138,6 – 104,6 X_3** (όπου X_3, ο αριθμός των παιδιών.)

(δ). Με την ίδια διαδικασία (Analyze / Regression /Linear και θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → **ποσό που ξόδεψε το προηγούμενο έτος**), λαμβάνουμε τους Πίνακες:

Πίνακας 63

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,100[a]	,010	,000	791,490695

a. Predictors: (Constant), ΠΟΣΟΠΡΟΗΓΟΥΜΕΝΟΥΕΤΟΥΣ

Πίνακας 64

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1086,860	101,589		10,699	,000
	ΠΟΣΟΠΡΟΗΓΟΥΜΕΝΟΥΕΤΟΥΣ	-,100	,100	-,100	-,994	,323

a. Dependent Variable: ΠΟΣΟΤΡΕΧΟΝΤΟΣΕΤΟΥΣ

Από τον **Πίνακα 63** λαμβάνουμε τον συντελεστή **R=0,100 ,** τον συντελεστή προσδιορισμού **R^2=0,010** και τον προσαρμοσμένο συντελεστή προσδιορισμού **Adj R^2 = 0,000.**
Από τον **Πίνακα 64** λαμβάνουμε τον σταθερό όρο και την κλίση της ευθείας παλινδρόμησης, η οποία είναι $\widehat{Y}$**= 1086,9 – 0,1 X_4** (όπου X_4, ποσό που ξοδεύτηκε το προηγούμενο έτος).

(ε). Με την ίδια διαδικασία (Analyze / Regression /Linear και θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → **αριθμός καταλόγων που έλαβε**), λαμβάνουμε τους Πίνακες:

Πίνακας 65

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,209[a]	,044	,034	777,975690

a. Predictors: (Constant), ΑΡΚΑΤΑΛΟΓΩΝ

Πίνακας 66

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	690,013	176,105		3,918	,000
	ΑΡΚΑΤΑΛΟΓΩΝ	24,065	11,399	,209	2,111	,037

a. Dependent Variable: ΠΟΣΟΤΡΕΧΟΝΤΟΣΕΤΟΥΣ

Από τον **Πίνακα 65** λαμβάνουμε τον συντελεστή **R=0,209 ,** τον συντελεστή προσδιορισμού **R^2=0,044** και τον προσαρμοσμένο συντελεστή προσδιορισμού **Adj R^2 = 0,034.**
Από τον **Πίνακα 66** λαμβάνουμε τον σταθερό όρο και την κλίση της ευθείας παλινδρόμησης, η οποία είναι $\widehat{Y}$**= 690 + 24,06 X_5** (όπου X_5, αριθμός καταλόγων που λαμβάνει ο καταναλωτής).

Από τη σύγκριση των συντελεστών προσδιορισμού **R^2** και ειδικότερα των προσαρμοσμένων συντελεστών προσδιορισμού **adj R^2** διαπιστώνεται ότι το (απλό) παλινδρομικό μοντέλο που προσαρμόζεται καλύτερα στο πληθυσμό (δηλ. που προσδιορίζει συγκριτικά καλύτερα τον τρόπο απόκρισης της εξαρτημένης μτβ στις μεταβολές της ανεξάρτητης) είναι εκείνο που περιγράφει τη γραμμική σχέση μεταξύ του ποσού που ξόδεψαν το τρέχον έτος οι καταναλωτές σε σχέση με τον ετήσιο μισθό τους, δηλ. η εξίσωση: $\widehat{Y}$**= 50,197 + 0,018 X_2** (όπου X_2, ο ετήσιος μισθός)**.** Από το εν λόγω μοντέλο ερμηνεύεται το **56%** (adj R^2 = 0,560) της συνολικής μεταβλητότητας του ποσού που δαπανήθηκε στο συγκεκριμένο κατάστημα βάσει του ετήσιου μισθού του καταναλωτή.

2. Εξέταση του ενδεχομένου παραβίασης των βασικών παραδοχών εγκυρότητας του γραμμικού μοντέλου.

Προϋποθέσεις εφαρμογής της απλής παλινδρομικής ανάλυσης[21], είναι:
α. Τα δεδομένα τόσο της εξαρτημένης όσο και της ανεξάρτητης μτβ θα πρέπει να είναι τύπο **scale** ή **ordinal**.
β. Η ανεξαρτησία των παρατηρήσεων.
γ. Για κάθε τιμή της ανεξάρτητης μτβ η κατανομή των τιμών της εξαρτημένης μτβ πρέπει να είναι κανονική.
δ. Η διασπορά της κατανομής της εξαρτημένης μτβ πρέπει να είναι σταθερή για όλες τις τιμές της ανεξάρτητης μτβ.
ε. Η σχέση μεταξύ της εξαρτημένης και της ανεξάρτητης μτβ στον πληθυσμό πρέπει να είναι γραμμική.
στ. Εξέταση της επίδρασης των έκτροπων παρατηρήσεων (σημείων επιρροής) .

Το παλινδρομικό μοντέλο που, σύμφωνα με το προηγούμενο ερώτημα, εμφανίζεται να προσαρμόζεται καλύτερα στα δεδομένα του δείγματος είναι εκείνο που συνδέει γραμμικά το ποσό που δαπανήθηκε κατά το τρέχον έτος σε σχέση με τον ετήσιο μισθό του καταναλωτή :
$\widehat{Y}$**= 50,197 + 0,018 X** (όπου X, ο ετήσιος μισθός)**.**
Για το εν λόγω μοντέλο θα εξεταστεί σε επίπεδο στατιστικής σημαντικότητας 5% το ενδεχόμενο παραβίασης κάποιας από τις βασικές παραδοχές της παλινδρομικής ανάλυσης.
Ειδικότερα,
(α). Τόσο η εξαρτημένη μτβ («Ποσό που δαπανήθηκε το τρέχον έτος») όσο και η ανεξάρτητη μτβ («ετήσιος μισθός») είναι ποσοτικές και συνεχείς.
(β). Ο έλεγχος της ανεξαρτησίας των παρατηρήσεων θα γίνει με τον δείκτη **Durbin-Watson**.

Πίνακας 67

Model Summary[b]

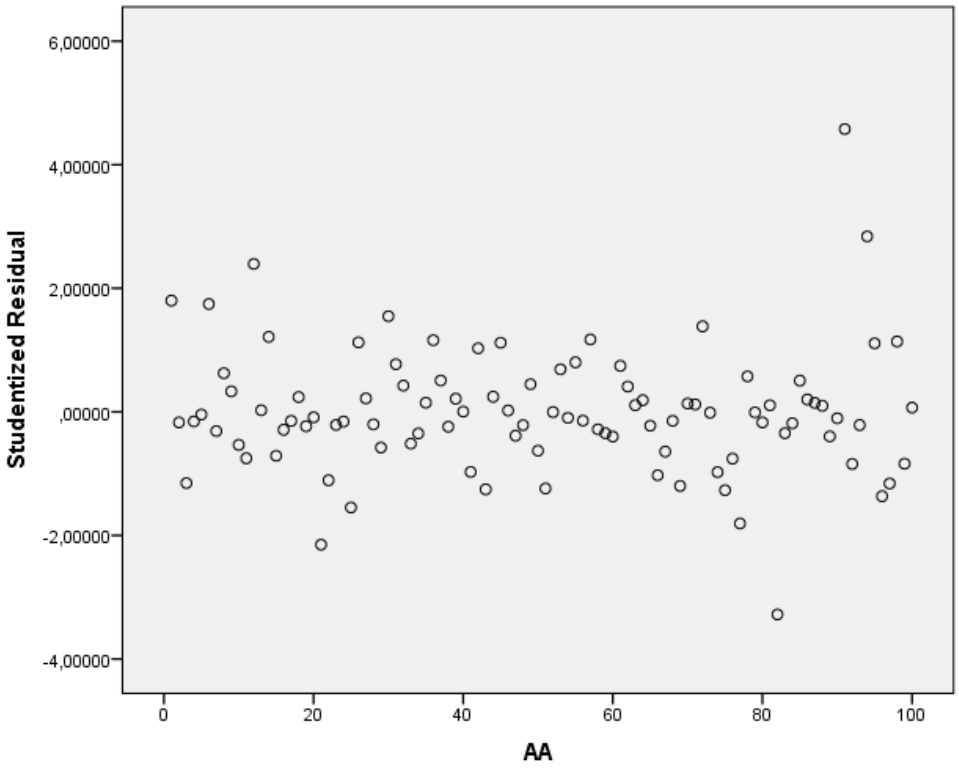
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	**Durbin-Watson**
1	,751[a]	,565	,560	524,835092	**1,996**

a. Predictors: (Constant), ΜΙΣΘΟΙ

b. Dependent Variable: ΠΟΣΟΤΡΕΧΟΝΤΟΣΕΤΟΥΣ

Από τον ανωτέρω **Πίνακα 67** (ο οποίος προκύπτει από τη διαδρομή Analyze / Regression /Linear, θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → μισθοί και στα statistics / residuals επιλέγουμε Durbin-Watson) διαπιστώνουμε ότι ο δείκτης **Durbin-Watson** λαμβάνει την τιμή **1,996**. Από τη θεωρία είναι γνωστό[22] ότι ο δείκτης αυτός λαμβάνει τιμές στο διάστημα (0 , 4) και επιπλέον όταν η τιμή του δείκτη αυτού κυμαίνεται στο διάστημα (**1,5 , 2,5**) μπορούμε να συμπεράνουμε με ασφάλεια ότι **ικανοποιείται η παραδοχή της ανεξαρτησίας** για το εξεταζόμενο μοντέλο. Ομοίως, δεν υπάρχει συσχέτιση μεταξύ των διαδοχικών υπολοίπων.
Η τελευταία παρατήρηση (και συνεπώς η αποδοχή της ανεξαρτησίας), διαπιστώνεται και από το διάγραμμα διασποράς μεταξύ των Studentized Residuals και της αύξουσας σειράς των παρατηρήσεων (α/α).

Γράφημα 13



Από το **Γράφημα 13 δεν** εμφανίζεται κάποιας μορφής πρότυπο ή κάποια ευδιάκριτη συσσώρευση και ως εκ τούτου μπορούμε να συμπεράνουμε ότι **δεν υπάρχει συσχέτιση μεταξύ των διαδοχικών καταλοίπων**[23].
(γ). Ο έλεγχος της κανονικότητας του μοντέλου θα γίνει με τον έλεγχο κανονικότητας των καταλοίπων κατά Student, με τη βοήθεια του ελέγχου Kolmogorov-Smirnov.

Πίνακας 68

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Studentized Residual	,135	100	,000	,926	100	,000

a. Lilliefors Significance Correction

Από τον **Πίνακα 68** (Analyze/ Descriptive Statistics / Explore), με δείγμα n=100 και σε επίπεδο στατιστικής σημαντικότητας α=0,05, το **p-value=0,0001** < 0,05 και ως εκ τούτου απορρίπτεται η υπόθεση της κανονικής κατανομής των Studentized Residuals.
Σημειώνεται ότι από τον μη παραμετρικό έλεγχο των Kolmogorov-Smirnov (nonparametric tests / legacy dialogs/ 1-sample K-S test /test variable →studentized residuals) λαμβάνουμε τον **Πίνακα 69** από τον οποίο συνάγεται ότι σε επίπεδο στατιστικής σημαντικότητας 5% μπορεί να γίνει δεκτή η μηδενική υπόθεση περί κανονικής κατανομής (p-value=0,053 > 0,05), σε αντίθεση με τον προαναφερόμενο παραμετρικό έλεγχο κανονικότητας.

Πίνακας 69

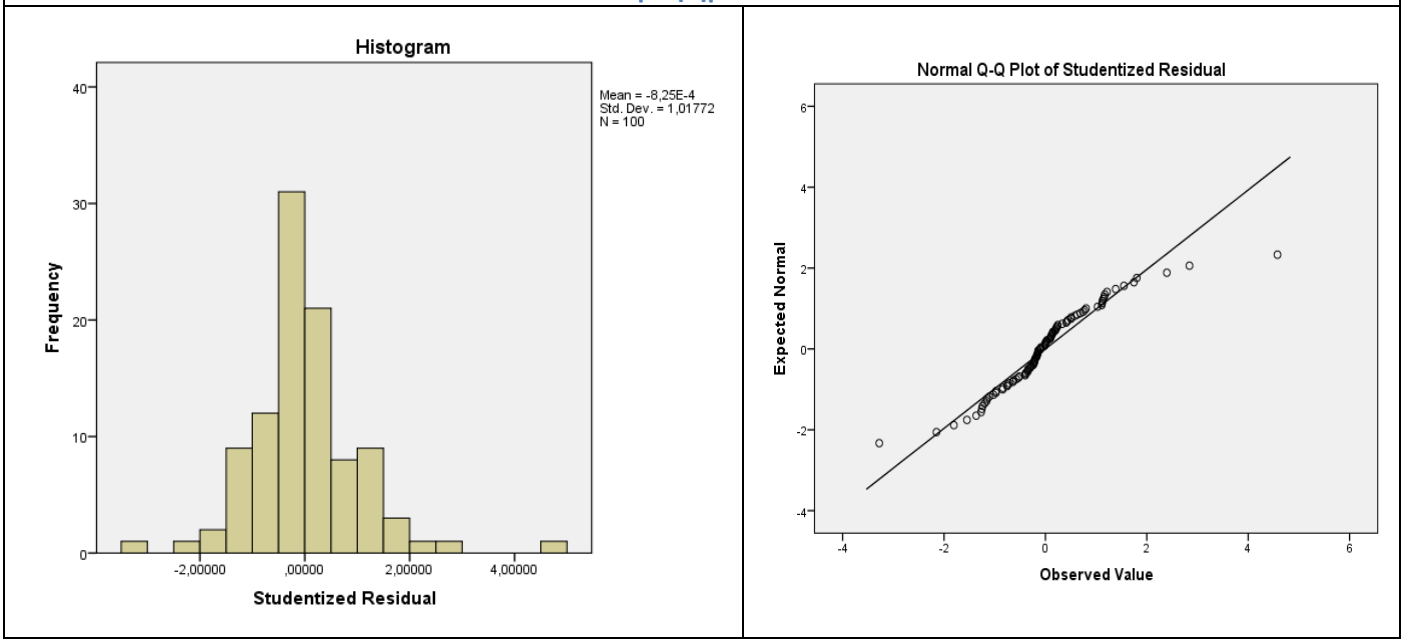
One-Sample Kolmogorov-Smirnov Test

		Studentized Residual
N		100
Normal Parameters[a,b]	Mean	-,0008252
	Std. Deviation	1,01771914
Most Extreme Differences	Absolute	,135
	Positive	,135
	Negative	-,098
Kolmogorov-Smirnov Z		1,347
Asymp. Sig. (2-tailed)		,053

a. Test distribution is Normal.

b. Calculated from data.

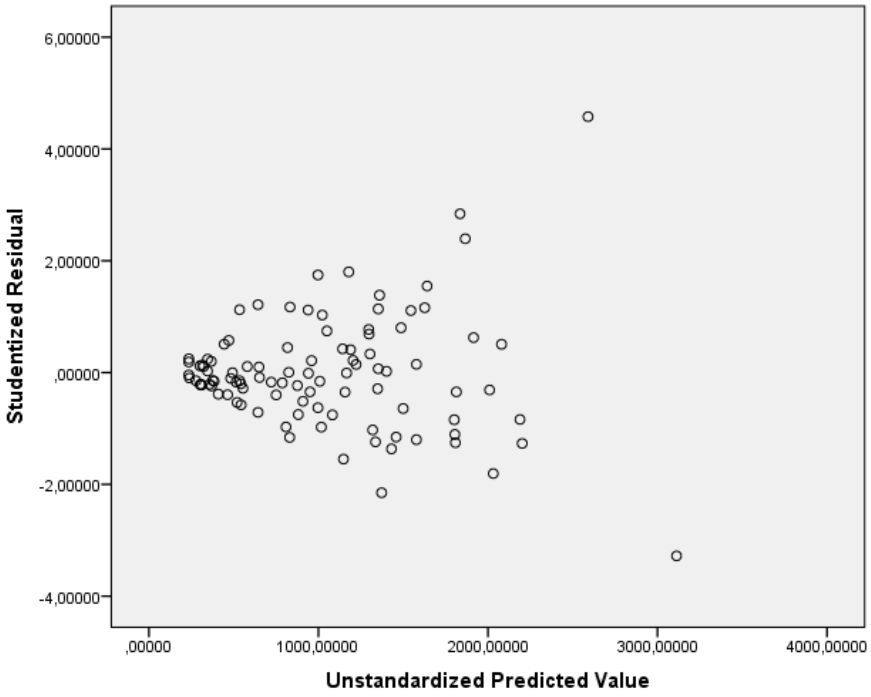
Γράφημα 14



Από το **Γράφημα 14** και ειδικότερα από το ιστόγραμμα φαίνεται ότι η μεγάλη πλειοψηφία των καταλοίπων κατά stundent λαμβάνουν τιμές εντός του διαστήματος μεταξύ δύο τυπικών αποκλίσεων (-2 , 2), το οποίο μας προϊδεάζει για μια κατά προσέγγιση κανονική κατανομή. Αντίστοιχα, και στο Normal Q-Q Plot of Studentized Residual οι τιμές εμφανίζονται πολύ κοντά στην ευθεία κανονικής κατανομής[24].

(δ). Ο έλεγχος της σταθερής διασποράς θα γίνει με τη βοήθεια του διαγράμματος διασποράς μεταξύ των Studentized Residual και Unstandardized Predicted Value (**Γράφημα 15**) :

Γράφημα 15



Από το ανωτέρω γράφημα διαπιστώνεται ότι με την αύξηση των τιμών στον άξονα της ανεξάρτητης μεταβλητής (μη τυποποιημένες προβλεπόμενες τιμές) αυξάνει και η διασπορά των καταλοίπων κατά Student. Αυτό σημαίνει ότι **παραβιάζεται η προϋπόθεση της σταθερής διασποράς** (ομοσκεδαστικότητας)[25].

(ε). Ο έλεγχος της παραδοχής της γραμμικότητας θα γίνει από τον **Πίνακα 70** με τη βοήθεια του στατιστικού δείκτη F.

Πίνακας 70

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	35017498,719	1	35017498,719	127,127	,000[b]
	Residual	26994283,623	98	275451,874		
	Total	62011782,342	99			

a. Dependent Variable: ΠΟΣΟΤΡΕΧΟΝΤΟΣΕΤΟΥΣ

b. Predictors: (Constant), ΜΙΣΘΟΙ

Σε επίπεδο στατιστικής σημαντικότητας 5% διατυπώνω τη μηδενική και την εναλλακτική υπόθεση:

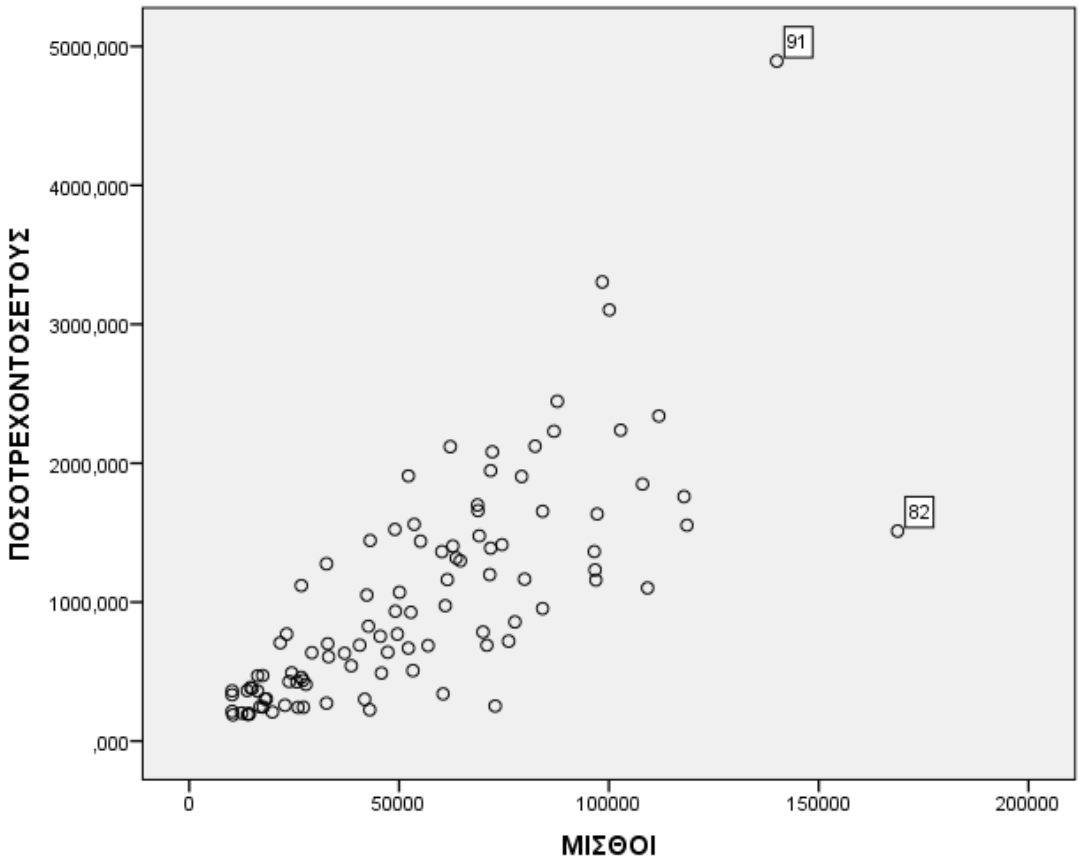
Ho: Η κλίση της ευθείας παλινδρόμησης είναι μηδέν (ομοίως, δεν υπάρχει γραμμική σχέση ανάμεσα στην εξαρτημένη και την ανεξάρτητη μτβ).

H_1: Η κλίση της ευθείας παλινδρόμησης είναι διάφορη του μηδέν (ομοίως, υπάρχει γραμμική σχέση ανάμεσα στην εξαρτημένη και την ανεξάρτητη μτβ)

Επειδή **p-value=0,0001 < 0,05**, απορρίπτουμε την Ho και **συνεπώς δεχόμαστε ότι υπάρχει γραμμική σχέση μεταξύ του ποσού που δαπανήθηκε κατά το τρέχον έτος και του ετήσιου μισθού**.

Ο έλεγχος γραμμικότητας μπορεί να γίνει και από το διάγραμμα διασποράς (scatter dot) των δύο μεταβλητών («Ποσό που δαπανήθηκε το τρέχον έτος» και «Ετήσιος μισθός»).

Γράφημα 16



(στ). Από το **Γράφημα 16** διαπιστώνεται η ύπαρξη κάποιων ακραίων τιμών. Επίσης από τον κατωτέρω **Πίνακα 71** διαπιστώνεται ότι δύο παρατηρήσεις (οι με α/α 82 και 91) έχουν τιμές τυπικών καταλοίπων που υπερβαίνουν τις τρεις τυπικές αποκλίσεις.

Πίνακας 71

Casewise Diagnostics[a]

Case Number	Std. Residual	ΠΟΣΟΤΡΕΧΟΝΤΟΣΕΤΟΥΣ	Predicted Value	Residual
82	-3,050	1511,751	3112,52927	-1600,778268
91	4,390	4894,284	2590,04611	2304,237892

a. Dependent Variable: ΠΟΣΟΤΡΕΧΟΝΤΟΣΕΤΟΥΣ

Ομοίως, από τον **Πίνακα 72** επιβεβαιώνεται η ανωτέρω παρατήρηση μιας και οι ανωτέρω αναφερόμενες παρατηρήσεις (cases 82 & 91) έχουν τιμές καταλοίπων κατά **student** που υπερβαίνουν τις τρεις τυπικές αποκλίσεις.

Πίνακας 72

Casewise Diagnostics[a]

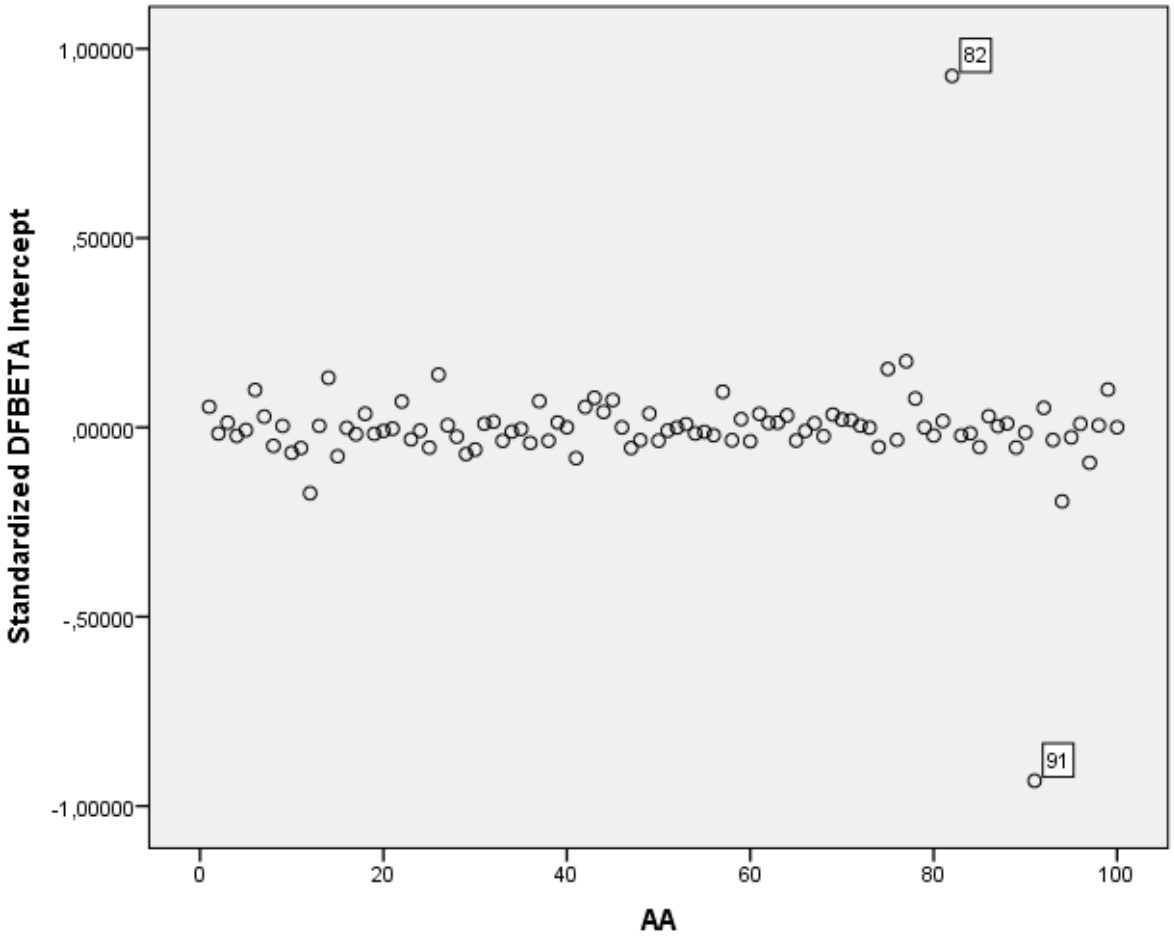
Case Number	Std. Residual	Studentized Residual	Predicted Value	Residual
82	-3,170	-3,27872	-,0380353	-3,24068455
91	4,525	4,57750	-,0486668	4,62616793

a. Dependent Variable: Studentized Residual

Για να ελέγξουμε αν οι ακραίες τιμές επηρεάζουν τις τιμές των παλινδρομικών συντελεστών (σταθερός όρος και/ή συντελεστής κλίσης) και συνεπώς το παλινδρομικό μας μοντέλο, εξετάζουμε τις τιμές της απόστασης **Leverage** για κάθε case. Οι τιμές Leverage εμφανίζονται στο Data View του SPSS κατά την εκτέλεση της παλινδρομικής ανάλυσης (Analyze/Regression/Linear) επιλέγοντας στο SAVE, στα Distances, το πεδίο Leverage Values. Από τον έλεγχο των εν λόγω αποστάσεων διαπιστώνεται ότι **όλες οι τιμές Leverage είναι μικρότερες από 0,2 και ως εκ τούτου μπορούμε να συμπεράνουμε ότι δεν υπάρχουν επιδραστικές τιμές** (Influential Points) στις παρατηρήσεις μας.

Αντίστοιχος έλεγχος για την ύπαρξη επιδραστικών τιμών (που μεταβάλουν σημαντικά το παλινδρομικό μας μοντέλο) μπορεί να γίνει και με το διάγραμμα διασποράς μεταξύ των Standardized DfBetas και της αύξουσας σειράς των παρατηρήσεων (**Γράφημα 17**). Τις τιμές Standardized DfBetas για κάθε case τις συγκρίνουμε με την ποσότητα $\frac{2}{\sqrt{n}}$, όπου n=100 (n, ο αριθμός των cases). Εάν κάποιες ακραίες παρατηρήσεις εμφανίζουν τιμές Std DfBetas, κατ' απόλυτη τιμή, μεγαλύτερες από την συγκρινόμενη ποσότητα ($\frac{2}{\sqrt{100}} = 0{,}2$), ενδεχομένως να μεταβάλουν τον συντελεστή διεύθυνσης (κλίση) της ευθείας παλινδρόμησης. Επειδή, η παρατήρηση με α/α 82 έχει τιμή Std DfBetas=-0,02 = |0,02| < 0,2 και η παρατήρηση με α/α 91 έχει τιμή Std DfBetas= 0,02 < 0,2, δεν είναι Influential Points.

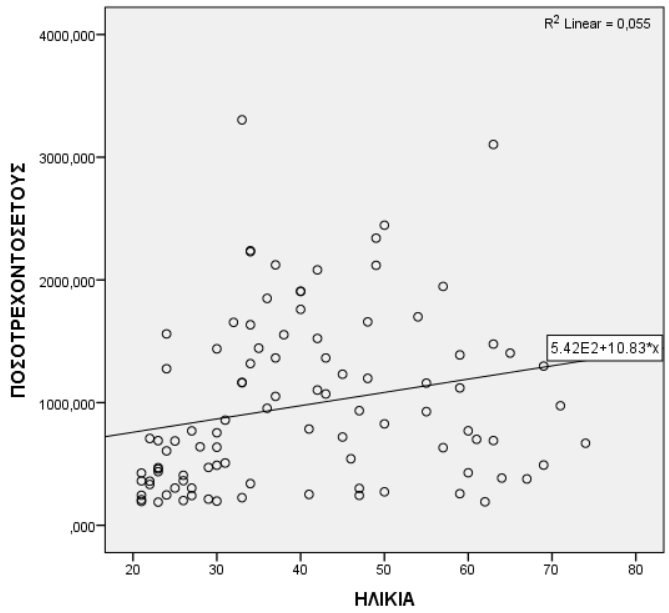
Γράφημα 17



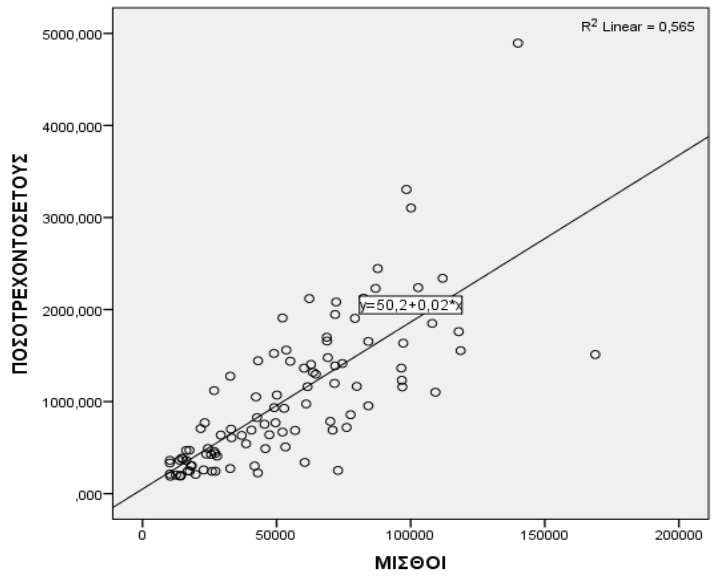
Συνοπτικά από τους ανωτέρω ελέγχους διαπιστώνεται ότι παραβιάζεται η προϋπόθεση της σταθερής διασποράς ενώ η υπόθεση κανονικότητας γίνεται οριακά αποδεκτή.

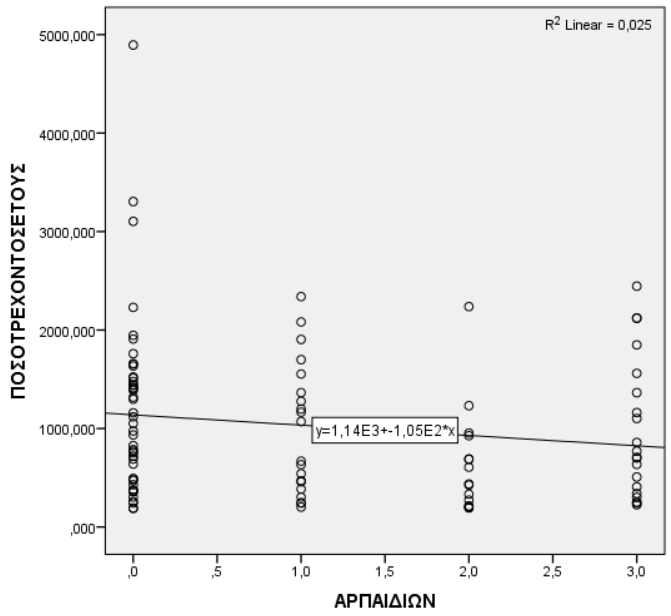
3. Διαγράμματα διασποράς και ευθείες παλινδρόμηση μεταξύ της εξαρτημένης και των ανεξάρτητων μτβ.

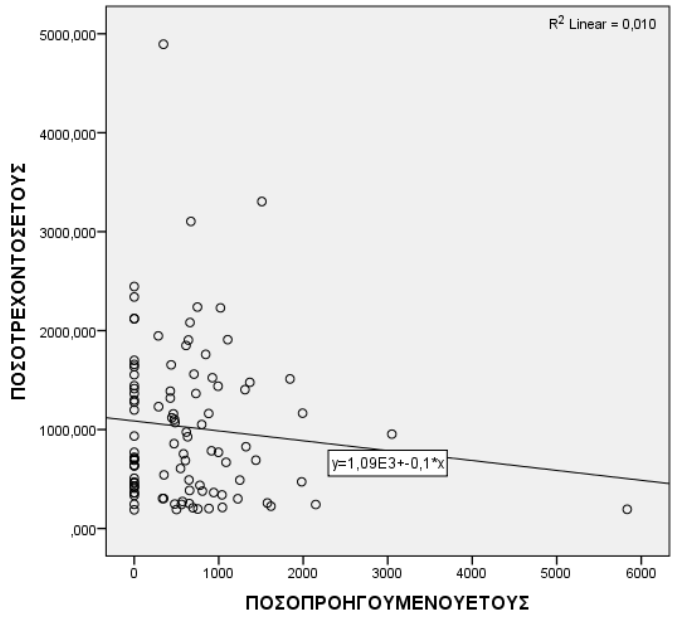
Με τη διαδικασία Graphs/Legacy Dialogs/Scatter dot λαμβάνουμε το διάγραμμα διασποράς μεταξύ της εξαρτημένης μτβ («Ποσό που ξόδεψε το τρέχον έτος») και της ανεξάρτητης μτβ **Ηλικία**.

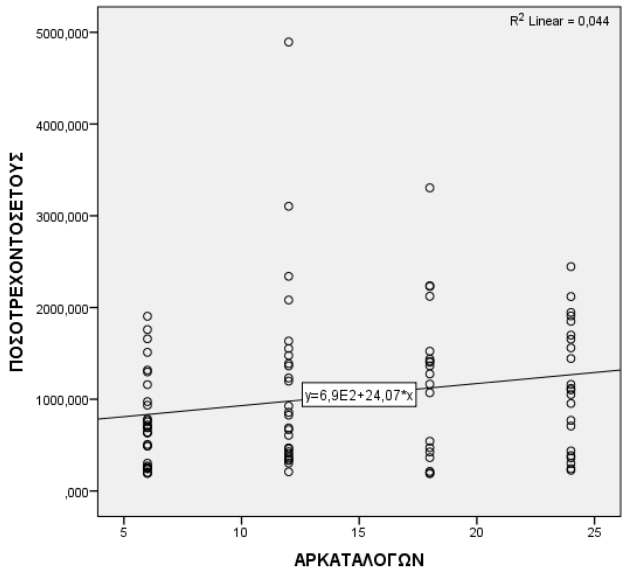
Γράφημα 18	Ανεξάρτητη μτβ «**ηλικία**»
	$\hat{Y}$**= 542,17 + 10,826 X_1** **R^2=0,055** **adj R^2 = 0,045 και** **R= 0,234** Επειδή 0,10 < R < 0,30 η γραμμική συσχέτιση θεωρείται **μέτρια** Σημειώνεται ότι η τιμή του adj R^2 και η ακρίβεια στα δεκαδικά ψηφία των συντελεστών στην εξίσωση της ευθείας, λαμβάνονται από τη διαδρομή Analyze / Regression /Linear, θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → μισθοί

Με την ίδια διαδικασία (Graphs/ Scatter Dot) λαμβάνουμε και τα γραφήματα που ακολουθούν.

Γράφημα 19	Ανεξάρτητη μτβ «**ΜΙΣΘΟΙ**»
	$\hat{Y}$**= 50,197 + 0,018 X_2** **R^2=0,565** **adj R^2 = 0,560 και** **R= 0,751** Επειδή R > 0,50 η γραμμική συσχέτιση θεωρείται **ισχυρή** Σημειώνεται ότι η τιμή του adj R^2 και η ακρίβεια στο τρίτο δεκαδικό των συντελεστών στην εξίσωση της ευθείας λαμβάνονται από τη διαδρομή Analyze / Regression /Linear, θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → μισθοί

Γράφημα 20	Ανεξάρτητη μτβ «**αρ. παιδιών**»
	$\hat{Y}$**= 1138,60 - 104,58 X_3** **R^2=0,025** **adj R^2 = 0,015 και** **R= 0,157** Επειδή 0,10 < R < 0,30 η γραμμική συσχέτιση θεωρείται **μέτρια**. Σημειώνεται ότι η τιμή του adj R^2 και η ακρίβεια των συντελεστών στην εξίσωση της ευθείας λαμβάνονται από τη διαδρομή Analyze / Regression /Linear, θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → αρ. παιδιών Όπως φαίνεται από το γράφημα αλλά και από το πρόσημο του συντελεστή κλίσης, η συσχέτιση είναι αρνητική

Γράφημα 21	Ανεξάρτητη μτβ «**ποσό προηγούμενο έτος**»
	$\hat{Y}$**= 1086,86 – 0,1 X_4** **R^2=0,01** **adj R^2 = 0,0 και** **R= 0,10** Επειδή R = 0,10 η γραμμική συσχέτιση θεωρείται **ασθενής**. Σημειώνεται ότι η τιμή του adj R^2 και η ακρίβεια των συντελεστών στην εξίσωση της ευθείας λαμβάνονται από τη διαδρομή Analyze / Regression /Linear, θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → ποσό προηγούμενου έτους. Όπως φαίνεται από το γράφημα αλλά και από το πρόσημο του συντελεστή κλίσης, η συσχέτιση είναι αρνητική

Γράφημα 22	Ανεξάρτητη μτβ «**αριθμός καταλόγων**»
	$\widehat{Y}$**= 690,013 + 24,065 X_5** **R^2=0,044** **adj R^2 = 0,034 και** **R= 0,209** Επειδή 0,10 < R < 0,30 η γραμμική συσχέτιση θεωρείται **μέτρια**. Σημειώνεται ότι η τιμή του adj R^2 και η ακρίβεια των συντελεστών στην εξίσωση της ευθείας λαμβάνονται από τη διαδρομή Analyze / Regression /Linear, θέτοντας ως Dependent → Ποσό τρέχοντος έτους και Independent → αριθμός καταλόγων.

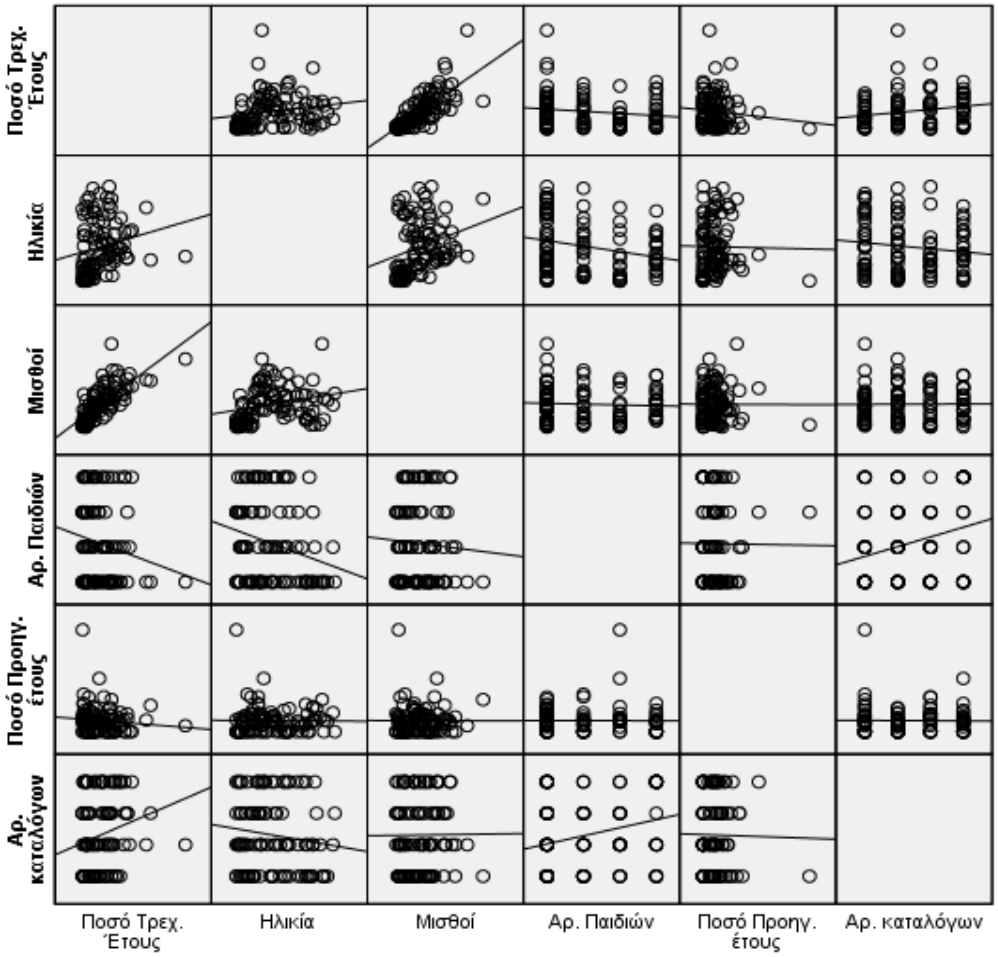
Από τη σύγκριση των συντελεστών προσδιορισμού **R^2** και ειδικότερα των προσαρμοσμένων συντελεστών προσδιορισμού **adj R^2** διαπιστώνεται ότι το (απλό) παλινδρομικό μοντέλο που προσαρμόζεται καλύτερα στο πληθυσμό (δηλ. που προσδιορίζει συγκριτικά καλύτερα τον τρόπο απόκρισης της εξαρτημένης μτβ στις μεταβολές της ανεξάρτητης) είναι εκείνο που περιγράφει τη γραμμική σχέση μεταξύ του ποσού που ξόδεψαν το τρέχον έτος οι καταναλωτές σε σχέση με τον ετήσιο μισθό τους, δηλ. η εξίσωση: $\widehat{Y}$**= 50,197 + 0,018 X_2** (όπου X_2, ο ετήσιος μισθός)**.** Από το εν λόγω μοντέλο ερμηνεύεται το **56%** (adj R^2 = 0,560) της συνολικής μεταβλητότητας του ποσού που δαπανήθηκε στο συγκεκριμένο κατάστημα βάσει του ετήσιου μισθού του καταναλωτή.

Το εν λόγω γραμμικό μοντέλο είναι και το μόνο που χαρακτηρίζεται από ισχυρή γραμμική συσχέτιση μεταξύ της εξαρτημένης και της ανεξάρτητης μεταβλητής. Αντίθετα τα υπόλοιπα μοντέλα παρουσιάζουν μέτρια ή ασθενή γραμμική συσχέτιση, όπως άλλωστε διαπιστώνεται και από τους αντίστοιχους συντελεστές συσχέτισης R.

4. Χρησιμοποιώντας όλες τις ανεξάρτητες μτβ να εκτιμηθεί και να ερμηνευθεί το μοντέλο πολλαπλής παλινδρόμησης και ο συντελεστής προσδιορισμού της.

Πριν προχωρήσουμε στην ανάλυση παλινδρόμησης κρίνεται σκόπιμο να δημιουργήσουμε όλα τα ανά δύο διαγράμματα διασποράς (Graph/Scatter/Matrix, **Γράφημα 23**) και τον Πίνακα συσχετίσεων (**Πίνακας 73**) προκειμένου να μορφώσουμε εικόνα για το ποια είναι η γραμμική συσχέτιση τόσο μεταξύ της ανεξάρτητης και των εξαρτημένων μεταβλητών, όσο και μεταξύ των εξαρτημένων μεταβλητών μεταξύ τους.

Γράφημα 23



Πίνακας 73

Correlations[c]

		Ποσό Τρεχ. Έτους	Ηλικία	Μισθοί	Αρ. Παιδιών	Ποσό Προηγ. έτους	Αρ. καταλόγων
Ποσό Τρεχ. Έτους	Pearson Correlation	1	,193	,751**	-,157	-,100	,209*
	Sig. (2-tailed)		,054	,000	,120	,323	,037
Ηλικία	Pearson Correlation	,193	1	,264**	-,247*	-,013	-,133
	Sig. (2-tailed)	,054		,008	,013	,898	,188
Μισθοί	Pearson Correlation	,751**	,264**	1	-,055	-,004	,007
	Sig. (2-tailed)	,000	,008		,588	,970	,945
Αρ. Παιδιών	Pearson Correlation	-,157	-,247*	-,055	1	-,006	,268**
	Sig. (2-tailed)	,120	,013	,588		,951	,007
Ποσό Προηγ. έτους	Pearson Correlation	-,100	-,013	-,004	-,006	1	-,014
	Sig. (2-tailed)	,323	,898	,970	,951		,891
Αρ. καταλόγων	Pearson Correlation	,209*	-,133	,007	,268**	-,014	1
	Sig. (2-tailed)	,037	,188	,945	,007	,891	

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

*. Correlation is significant at the 0.05 level (2-tailed).

c. Listwise N=100

Το ανωτέρω Γράφημα και ο σχετικός Πίνακας μας επιτρέπουν έναν αδρομερή έλεγχο από τον οποίο συνάγονται οι εξής αρχικές ενδείξεις:
(α). Υπάρχει στατιστικώς σημαντική **θετική και ισχυρή** γραμμική συσχέτιση μεταξύ της εξαρτημένης και της ανεξάρτητης μτβ «Μισθοί» (Pearson Correlation=0,751 και p-value=0,0001 <0,01).
(β). Σε επίπεδο στατιστικής σημαντικότητας 5% υπάρχει **μέτρια** γραμμική συσχέτιση μεταξύ της εξαρτημένης και της ανεξάρτητης μτβ «Αριθμός καταλόγων» (Pearson Correlation=0,209 και p-value=0,037 <0,05).
(γ). Μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών «Αριθμός Παιδιών» και «Ποσό Προηγούμενου Έτους» υπάρχει μια στατιστικώς μη σημαντική αρνητική γραμμική συσχέτιση.
(δ). Σε επίπεδο στατιστικής σημαντικότητας 5% υπάρχει μια στατιστικώς μη σημαντική θετική γραμμική συσχέτιση (Pearson Correlation=0,193 και p-value=0,054 > 0,050) μεταξύ της εξαρτημένης και της ανεξάρτητης μτβ «Ηλικία».
(ε). Σε επίπεδο στατιστικής σημαντικότητας 5% υπάρχει μια στατιστικώς σημαντική αλλά μέτρια γραμμική συσχέτιση μεταξύ της ανεξάρτητης μτβ «Ηλικία» και των ανεξάρτητων μτβ «Μισθοί» (Pearson Correlation=0,264 και p-value=0,008 < 0,05) και «αριθμός Παιδιών» (Pearson Correlation=-0,247 και p-value=0,013 < 0,05).
(στ). Στατιστικώς σημαντική γραμμική συσχέτιση εμφανίζεται να υπάρχει και μεταξύ των ανεξάρτητων μεταβλητών «Αριθμός καταλόγων» και «αριθμός παιδιών» (Pearson Correlation=-0,268 και p-value=0,007 < 0,05).

Μετά τις ανωτέρω ενδείξεις προχωρούμε στην παλινδρομική ανάλυση με τη μέθοδο Enter (Analyze/Regression/Linear) με εξαρτημένη μεταβλητή το «Ποσό του Τρέχοντος Έτους» και ανεξάρτητες μτβ όλες τις υπόλοιπες, από την οποία λαμβάνουμε τους Πίνακες που ακολουθούν.

Πίνακας 74

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	**-104,633**	198,112		-,528	,599	-497,988	288,722		
	Ηλικία	**-,905**	3,445	-,017	-,263	,793	-7,746	5,936	,871	1,149
	Μισθοί	**,018**	,002	,744	**11,684**	**,000**	,015	,021	,929	1,077
	Αρ. Παιδιών	**-125,292**	43,593	-,188	**-2,874**	**,005**	-211,847	-38,737	,883	1,133
	Ποσό Προηγ. έτους	**-,095**	,061	-,095	-1,549	,125	-,217	,027	1,000	1,000
	Αρ. καταλόγων	**28,840**	7,370	,250	**3,913**	**,000**	14,206	43,474	,922	1,085

a. Dependent Variable: Ποσό Τρεχ. Έτους

Από τον **Πίνακα 74** δημιουργούμε την παλινδρομική μας εξίσωση:

$$\widehat{Y} = -104{,}633 - 0{,}905\, X_1 + 0{,}018\, X_2 - 125{,}292\, X_3 - 0{,}095\, X_4 + 28{,}840\, X_5$$

όπου X_1: Ηλικία, X_2:Μισθοί, X_3:Αρ. παιδιών, X_4:Ποσό προηγούμενου έτους και X_5:Αρ. καταλόγων

Από την ανωτέρω εξίσωση διαπιστώνουμε ότι οι μεταβλητές «Ηλικία», «αρ. παιδιών» και «Ποσό προηγούμενου έτους» ασκούν αρνητική επίδραση στην τιμή της εξαρτημένης μεταβλητής (δηλ. η

αύξηση της τιμής τους μειώνει την τιμή της εξαρτημένης μεταβλητής), ενώ οι μεταβλητές «Μισθοί» και «αρ. καταλόγων» ασκούν θετική επίδραση (δηλ. η αύξηση της τιμής του προκαλεί αύξηση της τιμής της εξαρτημένης μεταβλητής). Επιπλέον, η ανεξάρτητη μεταβλητή που ασκεί την μεγαλύτερη επίδραση (κατ' απόλυτη τιμή) στην εξαρτημένη μεταβλητή, είναι η X_3 (αρ. παιδιών) καθώς έχει τον μεγαλύτερο (κατ' απόλυτη τιμή) συντελεστή παλινδρόμησης (125,292). Αυτό σημαίνει ότι η μοναδιαία μεταβολή της τιμής της X_3 θα προκαλέσει τη μεγαλύτερη μεταβολή στην τιμής της εξαρτημένης μτβ (αν όλες οι άλλες παραμείνουν αμετάβλητες), σε σχέση με την μεταβολή που θα της προκαλέσουν οι μοναδιαίες μεταβολές οποιασδήποτε άλλης ανεξάρτητης μεταβλητής.

Ο **συντελεστής πολλαπλού προσδιορισμού R^2** που δείχνει πόσο καλά προσαρμόζεται το μοντέλο στα δεδομένα μας, φαίνεται στον **Πίνακα 75** και έχει τιμή: **R^2=0,646**

Πίνακας 75

Model Summary[b]

		Model
		1
R		,804[a]
R Square		,646
Adjusted R Square		,628
Std. Error of the Estimate		482,941264
Change Statistics	R Square Change	,646
	F Change	34,376
	df1	5
	df2	94
	Sig. F Change	,000
Durbin-Watson		2,181

a. Predictors: (Constant), Αρ. καταλόγων, Μισθοί, Ποσό Προηγ. έτους, Αρ. Παιδιών, Ηλικία

b. Dependent Variable: Ποσό Τρεχ. Έτους

Ο συντελεστής πολλαπλής συσχέτισης R=0,804, δείχνει τη συσχέτιση ανάμεσα στις παρατηρούμενες και στις προβλεπόμενες τιμές της εξαρτημένης μεταβλητής.

Ο συντελεστής που φανερώνει την προβλεπτική δύναμη του παλινδρομικού μας μοντέλου είναι ο προσαρμοσμένος συντελεστής προσδιορισμού **Adjusted R^2** = 0,628. Από τον εν λόγω συντελεστή μπορούμε να συμπεράνουμε ότι το παλινδρομικό μας μοντέλο είναι σε θέση να ερμηνεύσει/προβλέψει την τιμή που λαμβάνει η εξαρτημένη μεταβλητή στον πληθυσμό μας (πελάτες της επιχείρησης), βάσει των τιμών που κάθε φορά λαμβάνουν οι ανεξάρτητες μτβ. **Από την τιμή του προσαρμοσμένου συντελεστή (62,8%) φαίνεται ότι το μοντέλο μας έχει αρκετά καλή προβλεπτική ικανότητα[26]**.

Από τον στατιστικό δείκτη F change=34,376 με p-value=0,0001 μπορούμε να συμπεράνουμε ότι, σε επίπεδο στατιστικής σημαντικότητας 5%, τουλάχιστον ένας από τους συντελεστές των ανεξάρτητων μεταβλητών είναι διάφορος του μηδέν στον πληθυσμό (πελάτες επιχείρησης).

5. Ποιες από τις ανεξάρτητες μτβ του παλινδρομικού μας μοντέλου μπορούμε να εξαιρέσουμε από την εξίσωση παλινδρόμησης.

Από τον έλεγχο του στατιστικού δείκτη t μπορούμε να βγάλουμε συμπέρασμα για το ποιες από τις ανεξάρτητες μεταβλητές είναι σημαντικές για την προβλεπτική ικανότητα του μοντέλου μας. Ειδικότερα, όσο οι τιμές t των ανεξάρτητων μτβ απομακρύνονται από το διάστημα **(-2, 2)** τόσο

σημαντικότερη είναι η συνεισφορά της εκάστοτε μτβ στο παλινδρομικό μας μοντέλο, ενώ το αν είναι στατιστικώς σημαντική ελέγχεται από το αντίστοιχο p-value.
Από τον **Πίνακα 74** βλέπουμε ότι οι μτβ X_2 , X_3 και X_5 έχουν απόλυτες τιμές t μεγαλύτερες του 2 και ταυτόχρονα είναι στατιστικώς σημαντικές σε επίπεδο στατιστικής σημαντικότητας 5% (καθώς έχουν p-value < 0,05). Ταξινομώντας με βάση τη σημαντικότητά της την κάθε μια από τις μτβ αυτές, διαπιστώνεται ότι η πλέον σημαντική είναι η **X_2 (Μισθοί)**, έπεται η **X_5 (αρ. καταλόγων)** και ακολουθεί η **X_3 (αρ. παιδιών)**.

Συνεπώς, από την εξίσωση παλινδρόμησης ($\widehat{Y}$**= -104,633 – 0,905 X_1+ 0,018 X_2 – 125,292 X_3 – 0,095 X_4 + 28,840 X_5**) μπορούμε να απαλείψουμε τις μεταβλητές X_1 «Ηλικία» και X_4 «ποσό προηγούμενου έτους», οι οποίες έχουν την μικρότερη συνεισφορά (t_1=-0,263 και t_4=-1,549) και ταυτόχρονα δεν είναι στατιστικώς σημαντικές σε επίπεδο στατιστικής σημαντικότητας 5% (p_1-value=0,793 >0,05 και p_4-value=0,125 >0,05).

Εκτελώντας εκ νέου παλινδρομική ανάλυση χωρίς τις μεταβλητές «Ηλικία» και «Ποσό προηγούμενου έτους», λαμβάνουμε τους **Πίνακες 76 και 77**.

Πίνακας 76

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)	-203,397	137,210		-1,482	,142	-475,757	68,963
	Μισθοί	,018	,001	,740	12,011	,000	,015	,021
	Αρ. Παιδιών	-122,664	42,689	-,184	-2,873	,005	-207,400	-37,927
	ΑΡΚΑΤΑΛΟΓΩΝ	29,138	7,363	,253	3,957	,000	14,522	43,754

a. Dependent Variable: Ποσό Τρεχ. Έτους

Πίνακας 77

Model Summary[b]

		Model
		1
R		,798[a]
R Square		,637
Adjusted R Square		,626
Std. Error of the Estimate		484,084499
Change Statistics	R Square Change	,637
	F Change	56,209
	df1	3
	df2	96
	Sig. F Change	,000
Durbin-Watson		2,130

a. Predictors: (Constant), ΑΡΚΑΤΑΛΟΓΩΝ, Μισθοί, Αρ. Παιδιών

b. Dependent Variable: Ποσό Τρεχ. Έτους

Συνοπτικά, η εξίσωση παλινδρόμησης μπορεί να πάρει την μορφή:

$\hat{Y}$**= -203,397 + 0,018 X_2 – 122,664 X_3 + 29,138 X_5 ,** χωρίς ουσιαστική μείωση στην προβλεπτική της ικανότητα (**Adjusted R^2 = 0,626**) σε σχέση με την εξίσωση που περιλαμβάνει το σύνολο των ανεξάρτητων μτβ (**Adjusted R^2 = 0,628**).

6. Παλινδρομική ανάλυση με την μέθοδο Stepwise.

(i) Εφαρμόζοντας το μοντέλο παλινδρομικής ανάλυσης Stepwise, με όλες τις ανεξάρτητες μεταβλητές, λαμβάνουμε τους Πίνακες:

Πίνακας 78

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Μισθοί	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
2	Αρ. καταλόγων	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
3	Αρ. Παιδιών	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).

a. Dependent Variable: Ποσό Τρεχ. Έτους

Στον **Πίνακα 78** φαίνεται η σειρά εισαγωγής των ανεξάρτητων μεταβλητών στο παλινδρομικό μας μοντέλο (ανάλογα με την σπουδαιότητά τους) και τα κριτήρια εισόδου και εξόδου από αυτό.

Από τον εν λόγω πίνακα επιβεβαιώνεται η ανάλυση που έγινε για τη σχετική σπουδαιότητα (σημαντικότητα) των ανεξάρτητων μεταβλητών στην παλινδρομική ανάλυση που έγινε στην προηγούμενη ενότητα με την μέθοδο παλινδρόμησης Enter.

Έτσι, οι σημαντικές μτβ για το παλινδρομικό μας μοντέλο είναι (κατά σειρά σημαντικότητας): η **X_2 (Μισθοί)**, έπεται η **X_5 (αρ. καταλόγων)** και ακολουθεί η **X_3 (αρ. παιδιών).**

Από τον πίνακα που ακολουθεί (**Πίνακας 79**) λαμβάνουμε το παλινδρομικό μας μοντέλο με την μέθοδο Stepwise**:**

Πίνακας 79

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	50,197	101,030		,497	,620	-150,294	250,689		
	Μισθοί	,018	,002	,751	11,275	,000	,015	,021	1,000	1,000
2	(Constant)	-273,075	140,011		-1,950	,054	-550,958	4,808		
	Μισθοί	,018	,002	,750	11,768	,000	,015	,021	1,000	1,000
	Αρ. καταλόγων	23,458	7,354	,203	3,190	,002	8,863	38,052	1,000	1,000

3	**(Constant)**	**-203,397**	**137,210**		**-1,482**	**,142**	**-475,757**	**68,963**		
	Μισθοί	**,018**	**,001**	**,740**	**12,011**	**,000**	**,015**	**,021**	**,996**	**1,004**
	Αρ. καταλόγων	**29,138**	**7,363**	**,253**	**3,957**	**,000**	**14,522**	**43,754**	**,928**	**1,078**
	Αρ. Παιδιών	**-122,664**	**42,689**	**-,184**	**-2,873**	**,005**	**-207,400**	**-37,927**	**,925**	**1,081**

a. Dependent Variable: Ποσό Τρεχ. Έτους

Ειδικότερα, η εξίσωση παλινδρόμησης έχει την μορφή:

$$\widehat{Y} = -203{,}397 + 0{,}018\ X_2 + 29{,}138\ X_5 - 122{,}664\ X_3$$

Όπου, X_2:Μισθοί, X_5:Αρ. καταλόγων και X_3:Αρ. παιδιών.

Ο συντελεστής προσδιορισμού του μοντέλου προκύπτει από τον **Πίνακα 80**.

Πίνακας 80

Model Summary[d]

		Model		
		1	2	**3**
R		,751[a]	,778[b]	**,798[c]**
R Square		,565	,606	**,637**
Adjusted R Square		,560	,598	**,626**
Std. Error of the Estimate		524,835092	501,865457	**484,084499**
Change Statistics	**R Square Change**	**,565**	**,041**	**,031**
	F Change	127,127	10,176	**8,257**
	df1	1	1	**1**
	df2	98	97	**96**
	Sig. F Change	,000	,002	**,005**
Durbin-Watson				**2,130**

a. Predictors: (Constant), Μισθοί

b. Predictors: (Constant), Μισθοί, Αρ. καταλόγων

c. Predictors: (Constant), Μισθοί, Αρ. καταλόγων, Αρ. Παιδιών

d. Dependent Variable: Ποσό Τρεχ. Έτους

Έτσι, ο **συντελεστής πολλαπλού προσδιορισμού R^2** που δείχνει πόσο καλά προσαρμόζεται το μοντέλο μας (model 3) στα δεδομένα μας, έχει τιμή: **R^2=0,637**

Από την τιμή **R^2 Change = 0,041** συμπεραίνουμε ότι η προβλεπτική ικανότητα του 2ου προτεινόμενου μοντέλου (model 2) βελτιώνεται κατά 4,1% εφόσον προσθέσουμε τη μτβ «αρ. καταλόγων» στο αρχικό μοντέλο (model 1), που περιλαμβάνει μόνο τον σταθερό όρο και την μτβ «μισθοί». Επίσης, από την τιμή **R^2 Change=0,031** συμπεραίνουμε ότι η προβλεπτική ικανότητα του τελικού μας μοντέλου (model 3) βελτιώνεται επιπλέον κατά 3,1% εφόσον προσθέσουμε και τη μτβ «αρ. παιδιών».

Η προβλεπτική ικανότητα του μοντέλου μας στο επίπεδο του πληθυσμού είναι 62,6% (Adjusted R^2 =0,626)

(ii) Τα διαστήματα εμπιστοσύνης για τον σταθερό όρο και τους συντελεστές των ανεξάρτητων μεταβλητών δίνονται από τον **Πίνακα 79.**

Ειδικότερα, **με βεβαιότητα 95%**:

ο σταθερός όρος παίρνει τιμές στο διάστημα (**-475,757 , 68,963**),

ο συντελεστής της μτβ «ηλικία» (X_1) παίρνει τιμές στο διάστημα (**0,015 , 0,021**),

ο συντελεστής της μτβ «αρ. καταλόγων» (X_5) παίρνει τιμές στο διάστημα (**14,522 , 43,754**), και

ο συντελεστής της μτβ «αρ. παιδιών» (X_3) παίρνει τιμές στο διάστημα (**-207,400 , -37,927**),

7. Παλινδρομική ανάλυση για τους πελάτες που κατοικούν κοντά στο κατάστημα.

(i) Για να προσδιορίσουμε σε επίπεδο στατιστικής σημαντικότητας 5%, αν το ποσό που ξόδεψαν (εξαρτημένη μτβ) οι πελάτες που διαμένουν κοντά στο κατάστημα επηρεάζεται από τον αριθμό των καταλόγων που έλαβαν (ανεξάρτητη μτβ), προβαίνουμε σε απλή παλινδρομική ανάλυση με τη μέθοδο stepwise.

Αρχικά επιλέγουμε από το δείγμα των 100 πελατών **μόνο** εκείνους που διαμένουν κοντά στο κατάστημα (Data / Select Cases / If condition is satisfied → «Περιοχή» =1) και προβαίνουμε σε παλινδρομικό έλεγχο με εξαρτημένη μτβ το «Ποσό τρέχοντος έτους» και ανεξάρτητη μτβ τον «αριθμό καταλόγων που έλαβαν».

Πίνακας 81

Correlations

		Ποσό Τρεχ. Έτους	Αρ. καταλόγων
Pearson Correlation	Ποσό Τρεχ. Έτους	1,000	,218
	Αρ. καταλόγων	,218	1,000
Sig. (1-tailed)	Ποσό Τρεχ. Έτους	.	,035
	Αρ. καταλόγων	,035	.
N	Ποσό Τρεχ. Έτους	70	70
	Αρ. καταλόγων	70	70

Από τον Πίνακα 81 συμπεραίνουμε ότι υπάρχει μια μέτρια γραμμική συσχέτιση (Pearson Correlation = 0,218) μεταξύ της εξαρτημένης και της ανεξάρτητης μεταβλητής «αρ. καταλόγων». Μάλιστα σε επίπεδο στατιστικής σημαντικότητας 5%, επειδή p-value=0,035<0,05, δεχόμαστε ότι στον πληθυσμό (δηλ. στο σύνολο των πελατών που κατοικούν κοντά στο κατάστημα) αυτή η γραμμική συσχέτιση είναι στατιστικώς σημαντική.

Από τον **Πίνακα 82** που ακολουθεί, λαμβάνουμε την ευθεία παλινδρόμησης του μοντέλου μας η οποία έχει την μορφή: $\widehat{Y}$**= 664,397 + 20,827 X**

Πίνακας 82

Coefficients[a]

		Model 1	
		(Constant)	Αρ. καταλόγων
Unstandardized Coefficients	**B**	**664,397**	**20,827**
	Std. Error	172,145	11,291
Standardized Coefficients	Beta		,218
t		3,860	**1,845**
Sig.		,000	**,069**
95,0% Confidence Interval for B	Lower Bound	320,888	-1,703
	Upper Bound	1007,907	43,358
Collinearity Statistics	Tolerance		1,000
	VIF		1,000

a. Dependent Variable: Ποσό Τρεχ. Έτους

Από τον έλεγχο του στατιστικού δείκτη t μπορούμε να βγάλουμε συμπέρασμα για το εάν η ανεξάρτητη μτβ είναι σημαντική για την προβλεπτική ικανότητα του μοντέλου μας. Ειδικότερα, όταν η τιμή t της ανεξάρτητης μτβ βρίσκεται στο διάστημα **(-2, 2)** τότε **έχει μικρή συνεισφορά** στο παλινδρομικό μας μοντέλο, ενώ το αν είναι στατιστικώς σημαντική ελέγχεται από το αντίστοιχο p-value.

Από τον **Πίνακα 82** βλέπουμε ότι, σε επίπεδο στατιστικής σημαντικότητας 5%, η μτβ Χ («αρ. καταλόγων») έχει τιμή t=1,845 [δηλ. εντός του διαστήματος (-2 , 2)] και επειδή p-value =0,069 > 0,05, συμπεραίνουμε ότι η συνεισφορά της στο γραμμικό μοντέλο είναι μικρή και **ΔΕΝ** κρίνεται στατιστικώς σημαντική.

Συνοπτικά, σε επίπεδο στατιστικής σημαντικότητας 5%, το ποσό που ξόδεψαν στο κατάστημα οι πελάτες που διαμένουν κοντά στο κατάστημα, δεν επηρεάζεται από τον αριθμό των καταλόγων που έλαβαν.

(ii) Από τον **Πίνακα 83** λαμβάνουμε τον συντελεστή προσδιορισμού **R^2 = 0,048** που μας υποδεικνύει ότι το μοντέλο μας προσαρμόζεται μόλις κατά 4,8% στα δεδομένα μας, ενώ ο διορθωμένος συντελεστής προσδιορισμού **Adjusted R^2 = 0,034** καταδεικνύει ότι το μοντέλο μας **δεν** ενδείκνυται για σχετικές προβλέψεις κατανάλωσης στο επίπεδο του πληθυσμού των πελατών που κατοικούν κοντά στο κατάστημα.

Πίνακας 83

Model Summary[b]

		Model
		1
R		,218[a]
R Square		,048
Adjusted R Square		,034
Std. Error of the Estimate		661,580298
Change Statistics	R Square Change	,048
	F Change	3,403
	df1	1
	df2	68
	Sig. F Change	,069
Durbin-Watson		2,016

a. Predictors: (Constant), Αρ. καταλόγων

b. Dependent Variable: Ποσό Τρεχ. Έτους

Υποσημειώσεις

[1] Υφαντόπουλος Γ – Νικολαΐδου Κ, Η Στατιστική στην Κοινωνική Έρευνα, εκδόσεις Gutenberg, Αθήνα 2008, σελ. 265

[2] Το φυλλογράφημα εξάγεται με τη βοήθεια του SPSS (Analyze/Descriptive Statistics/Explore)

[3] Οι δύο ακραίες τιμές βρίσκονται στην 25η και 29η μη διατεταγμένες παρατηρήσεις, όπως φαίνεται στο θηκόγραμμα, και οι τιμές είναι 450€ και 540€, αντίστοιχα

[4] Η κατηγοριοποίηση των τιμών της μεταβλητής Bonus στη νέα μεταβλητή Bonus_Class γίνεται στο SPSS με τη διαδικασία Transform/Recode into Different Variables

[5] Δαφέρμος Βασίλης, Κοινωνική Στατιστική με το SPSS, εκδόσεις ΖΗΤΗ, Θεσσαλονίκη 2005, σελ.107

[6] Μια πρόσθετη χρησιμότητα του τυπικού σφάλματος του μέσου, υπό την προϋπόθεση ότι η κατανομή της μεταβλητής bonus είναι κανονική, αφορά στον προσδιορισμό της κατανομής των δειγματικών μέσων τιμών. Έτσι, γνωρίζοντας τη μέση τιμή ενός δείγματος και το τυπικό σφάλμα του μέσου (15,82€) μπορούμε να πούμε ότι το 95,44% όλων των δειγματικών μέσων (του ίδιου μεγέθους, n=30) θα βρίσκεται στο διάστημα μεταξύ δύο τυπικών σφαλμάτων [739,33 ± 2* 15,82] ή ακριβέστερα το 95% θα βρίσκεται στο διάστημα [**739,33± 2,045*15,82**] , όπου το 2,045 βρίσκεται από τον πίνακα της t κατανομής για n-1=30-1=**29 βαθμούς ελευθερίας** και στατιστική σημαντικότητα 1-0,05/2=**0,975 (Πίνακες t κατανομής, Δαφέρμος Β. ό.π. σελ. 681)**

[7] Δαφέρμος Β, ό.π. σελ. 115

[8] Υφαντόπουλος Γ – Νικολαΐδου, ό.π. σελ.301

[9] Δαφέρμος Β, ό.π. σελ. 112

[10] Τον Πίνακα 10 τον αντλήσαμε για κάθε φύλο ξεχωριστά, από το Data/Select Cases (Άνδρες =1 /Γυναίκες =0) και κατόπιν Analyze/Descriptive Statistics/Frequencies. Έτσι αντλήσαμε και τον Πίνακα 11 προκειμένου να καθορίσουμε την επικρατούσα τιμή της μεταβλητής Bonus για κάθε φύλο ξεχωριστά.

[11] Δαφέρμος Β, ό.π. σελ. 297

[12] Δαφέρμος Β. ό.π. σελ. 226, όπου αναφέρεται ότι εφόσον το μέγεθος τους δείγματος είναι μεγαλύτερο του 50 (n>50) είναι προτιμότερο ο έλεγχος κανονικότητας να γίνεται με το κριτήριο Kolmogorov-Smirnov, ενώ αν το μέγεθος του δείγματος είναι μικρότερο του 50 (n<50) το στατιστικό κριτήριο που είναι κατάλληλο να ελέγξει την ύπαρξη ή μη κανονικότητας είναι εκείνο των Shapiro-Wilk.

[13] Δαφέρμος Β. ό.π. σελ. 230-233

[14] Δαφέρμος Β. ό.π. σελ. 233

[15] Δαφέρμος Β. ό.π. σελ. 359

[16] Ο κανόνας απόφασης είναι: απέρριψε την Ho αν η τιμή του t* είναι μεγαλύτερη του $+ t_{\alpha/2}$ **ή μικρότερη του - $t_{\alpha/2}$**

[17] Κατά τον Δαφέρμο (ό.π. σελ. 382) «*…Μόνο η μία από τις συσχετιζόμενες μεταβλητές είναι υποχρεωτικό να είναι συνεχής. Η άλλη μπορεί να είναι κατηγορική, αλλά με την προϋπόθεση να είναι οπωσδήποτε διχοτομική, να έχει το ίδιο ποσοστό τιμών σε κάθε μία από τις δύο κατηγορίες και οπωσδήποτε να έχει κωδικοποιηθεί το ένα της επίπεδο με τον αριθμό μηδέν και το άλλο με τον αριθμό 1 (Coakes and Steed, 1999). Για παράδειγμα, αν μία μεταβλητή είναι αριθμητική και η άλλη είναι το φύλο θα πρέπει να κωδικοποιήσουμε τη μεταβλητή SEX ως εξής: 0=Αγόρι και 1= Κορίτσι ή το αντίθετο».*

[18] Το πλήθος των παρατηρήσεων μετά την απομάκρυνση των έκτροπων χαμηλών τιμών, είναι 27 και για το λόγο αυτό δεν χρησιμοποιώ το κριτήριο των Kolmogorov – Smirnov.

[19] «Σύμφωνα με τον Cohen (εργασίες 1969 και 1988) **αν ο συντελεστής του Pearson έχει τιμές r=± 0,50, τότε η συσχέτιση θεωρείται ισχυρή, ενώ αν r= ±0,30, η συσχέτιση θεωρείται μέτρια και τέλος αν r=±0,10 η συσχέτιση θεωρείται ασθενής»,** αναφορά Δαφέρμος Β. ό.π. σελ. 379.

[20] Θετική γραμμική συσχέτιση σημαίνει ότι υψηλές τιμές τις μιας μτβ συνδέονται με υψηλές τιμές της άλλης μτβ.

[21] Δαφέρμος Β. ό.π. σελ. 425.

[22] Δαφέρμος Β. ό.π. σελ. 447 και 485-486.

[23] Δαφέρμος Β. ό.π. σελ. 487 «αν θέλουμε η παραδοχή της ανεξαρτησίας να ικανοποιείται, η τιμή των residuals δεν θα πρέπει να σχετίζεται με την σειρά, με την οποία τα δεδομένα ελήφθησαν και καταχωρήθηκαν».

[24] Σύμφωνα με την Norusis (2000), αναφορικά με τον έλεγχο της παραδοχής της κανονικότητας στην παλινδρομική ανάλυση, «εάν το δείγμα μας είναι μεγάλο (n>30) τα τεστ κανονικότητας ίσως μας οδηγήσουν να απορρίψουμε την παραδοχή της κανονικότητας βασιζόμενοι σε μικρές απομακρύνσεις από την κανονικότητα». Αναφορά στο Δαφέρμος Β. ό.π. 448. Ο εν λόγω συγγραφέας αναφέρει ότι μόνο μεγάλες παραβιάσεις της παραδοχής της κανονικότητας στις περιπτώσεις μεγάλων δειγμάτων είναι ανησυχητικές, «οι μικρές παραβιάσεις της κανονικότητας δεν μπορούν να επηρεάσουν την παλινδρομική ανάλυση».

[25] Από το Γράφημα 15 διαπιστώνεται ότι η μεταβλητότητα (διασπορά) των residuals αυξάνει, καθώς αυξάνουν οι προβλεπόμενες από το μοντέλο τιμές (predicted values). Αυτό σημαίνει ότι η διασπορά των residuals είναι μικρότερη για μικρότερες προβλεπόμενες τιμές της ανεξάρτητης μεταβλητής, και μεγαλύτερη για μεγαλύτερες. Αλλά κάτι τέτοιο, σύμφωνα με την Norusis (2000), δεν είναι ασυνήθιστο, καθώς η μεταβλητότητα της εξαρτημένης μεταβλητής συχνά

αυξάνει με την αύξηση των τιμών της ανεξάρτητης μεταβλητής. Πάντως, στις περιπτώσεις που η διασπορά της εξαρτημένης μεταβλητής αυξάνει γραμμικά με την αύξηση των τιμών της ανεξάρτητης μεταβλητής, και όλες οι τιμές της εξαρτημένης μτβ είναι θετικές, τότε ενδείκνυται ο μετασχηματισμός της τετραγωνικής ρίζας. Δηλ. μετασχηματίζουμε την Υ μτβ σε $\sqrt{Y}$ και εκτελούμε ξανά regression. Αναφορά στο Δαφέρμος Β. ό.π. σελ 541

[26] Κατά τον Stevens (2002) «*η προβλεπτική δύναμη ενός παλινδρομικού μοντέλου, πάνω σε άλλα δείγματα, πέρα από το εξεταζόμενο και βέβαια προερχόμενα όλα από τον ίδιο πληθυσμό, εξαρτάται σε σημαντικό βαθμό από την τιμή του λόγου n/k όπου n: το μέγεθος του εξεταζόμενου δείγματος και k: ο αριθμός των ανεξάρτητων μεταβλητών. Για κάθε πολλαπλή παλινδρομική ανάλυση κατώτατο όριο για την τιμή του λόγου αυτού είναι το 5*». Αναφορά στο Δαφέρμος Β. ό.π. σελ 478-479. Σημειώνεται ότι η τιμή του λόγου αυτού στο δικό μας δείγμα είναι 100/5=20 > 5.