

(2) ΓΡΑΦΙΚΕΣ ΠΑΡΑΣΤΑΣΕΙΣ

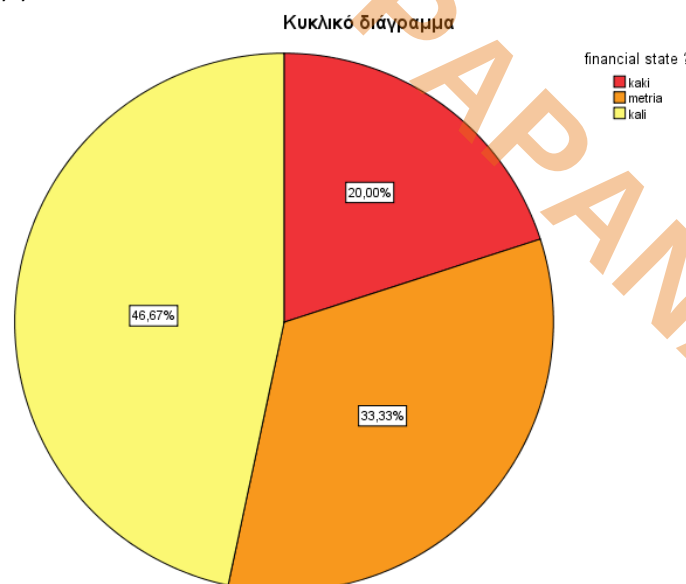
α) Από το μενού → **Graphs**

Επιλέγουμε αναλόγως με την γραφική παράσταση θέλουμε να κατασκευάσουμε:

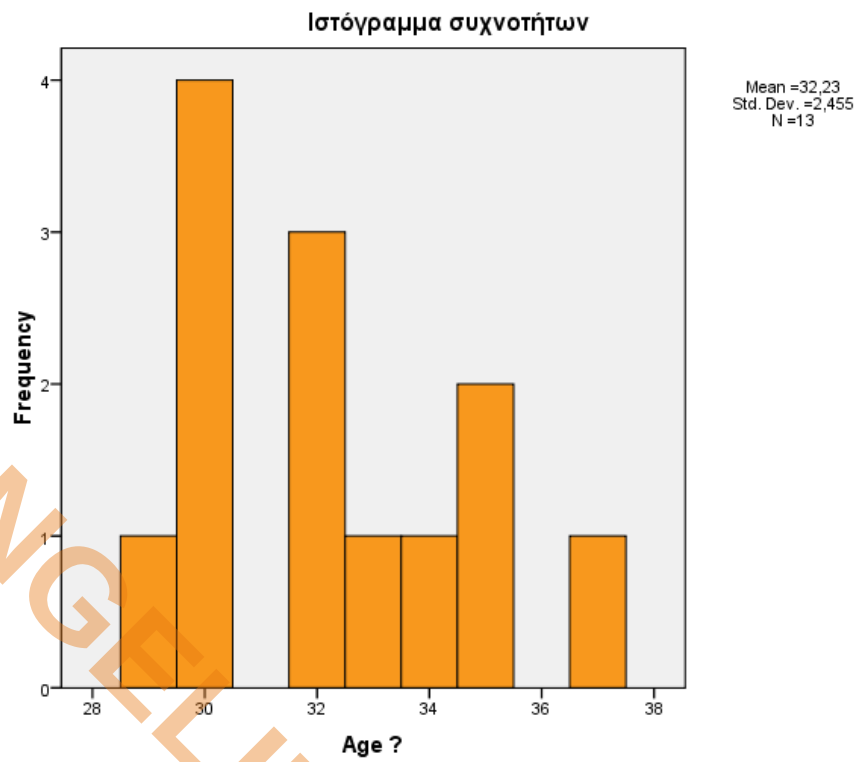
Ραβδόγραμμα : **Bar**



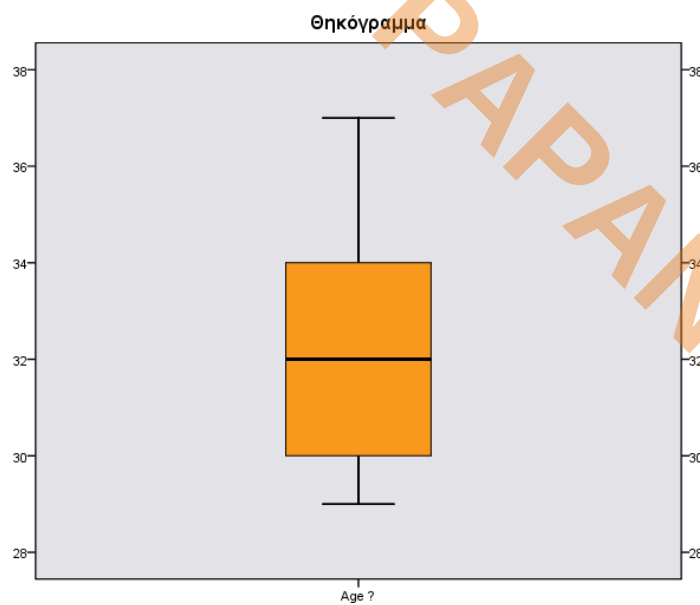
Κυκλικό διάγραμμα : **Pie**



Ιστόγραμμα : Histogram



Θηκόγραμμα : Boxplot



Διάγραμμα διασποράς : Scatter/ Dot



β) Από το μενού → **Analyze**
Frequencies

• **Charts** (Ραβδόγραμμα, Κυκλικό διάγραμμα, Ιστόγραμμα)

γ) Από το μενού → **Analyze**
Explore

• **Charts** (Φυλλογράφημα, Ιστόγραμμα)

Κεφάλαιο 5

ΥΠΟΛΟΓΙΣΜΟΣ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΤΡΩΝ

Τα στατιστικά μέτρα είναι αριθμοί που υπολογίζονται από τα δεδομένα και η τιμή τους αντιπροσωπεύει κάποια τάση ή συμπεριφορά του δείγματος. Τα στατιστικά μέτρα διακρίνονται στα μέτρα κεντρικής τάσης (*measures of central tendency*), στα μέτρα μεταβλητότητας (*measures of dispersion*) και στα μέτρα σχήματος (*measures of shape*). Τις τρεις αυτές κατηγορίες περιγράφουμε αναλυτικά στη συνέχεια.

Μέτρα κεντρικής τάσης

Χρησιμοποιούνται για να περιγράψουν τη θέση του συνόλου των δεδομένων. Δηλαδή, η τιμή τους είναι η πιο αντιπροσωπευτική για να σχηματίσει κάποιος μία πρώτη εικόνα για το ύψος των τιμών των παρατηρήσεων μίας μεταβλητής.

Mean μέση τιμή

Median διάμεσος

Mode επικρατούσα τιμή

Μέτρα μεταβλητότητας / διασποράς

Χρησιμοποιούνται για να περιγράψουν τις διαφορές που υπάρχουν ανάμεσα στις τιμές μίας μεταβλητής. Δηλαδή, προσδιορίζουν αν οι παρατηρήσεις είναι συγκεντρωμένες γύρω από μία αντιπροσωπευτική τιμή ή παρουσιάζουν μεγάλη διασπορά.

range Εύρος

variance Διασπορά/ Διακύμανση

st. deviation Τυπική απόκλιση

quartiles Τεταρτημόρια

interquartile range Ενδοτεταρτημοριακό εύρος

Μέτρα σχήματος

Είναι μέτρα που περιγράφουν το σχήμα της μεταβλητής όταν αυτή παρασταθεί με ιστόγραμμα ή ραβδόγραμμα.

skewness Λοξότητα

kurtosis κυρτότητα

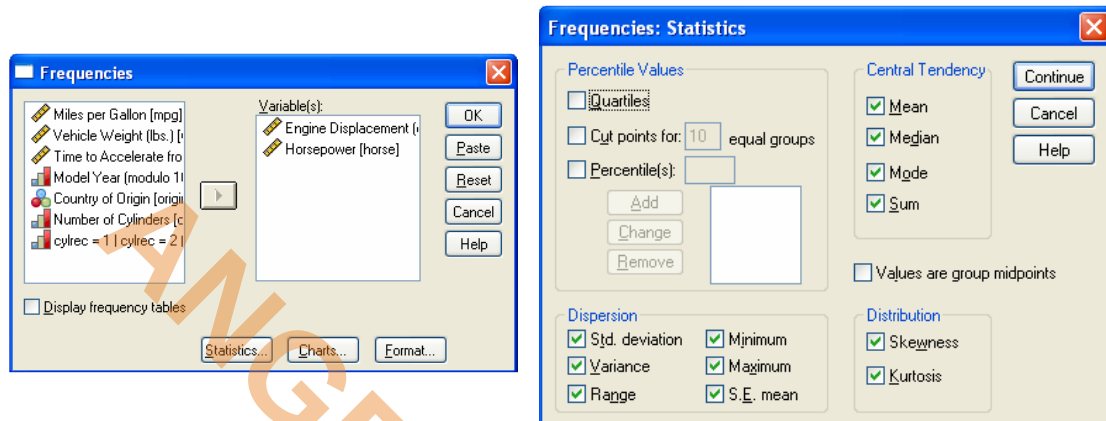
(1) Για μια απλή ποσοτική μεταβλητή

Από το μενού επιλέγουμε

Analyze

Descriptive statistics

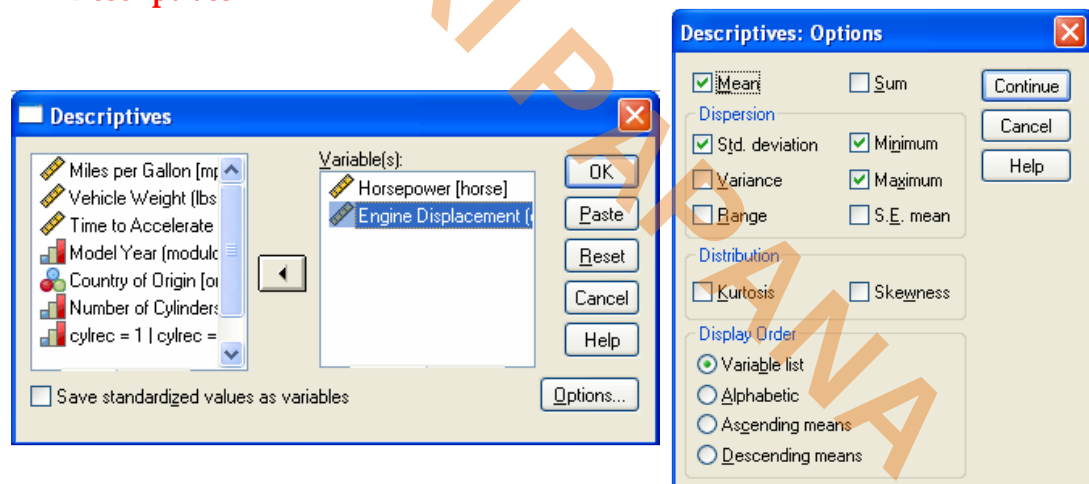
Frequencies



Analyze

Descriptive statistics

Descriptives



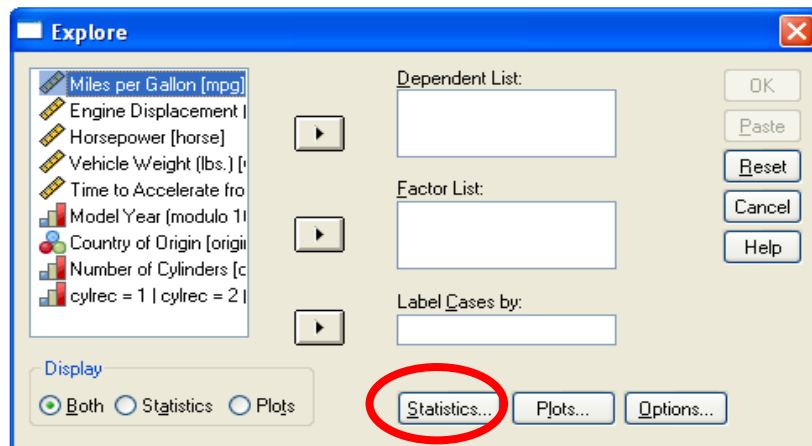
(2) Για μια απλή ποσοτική μεταβλητή κατηγοριοποιημένη ως προς μια ποιοτική μεταβλητή

Από το μενού επιλέγουμε

Analyze

Descriptive statistics

Explore



ή

Από το μενού επιλέγουμε

Data
Split File

και έπειτα

Analyze
Descriptive statistics
Frequencies ή **Descriptives**

Κεφάλαιο 6

ΤΕΣΤ ΑΝΕΞΑΡΤΗΣΙΑΣ – ΣΥΣΧΕΤΙΣΗ ΜΕΤΑΒΛΗΤΩΝ

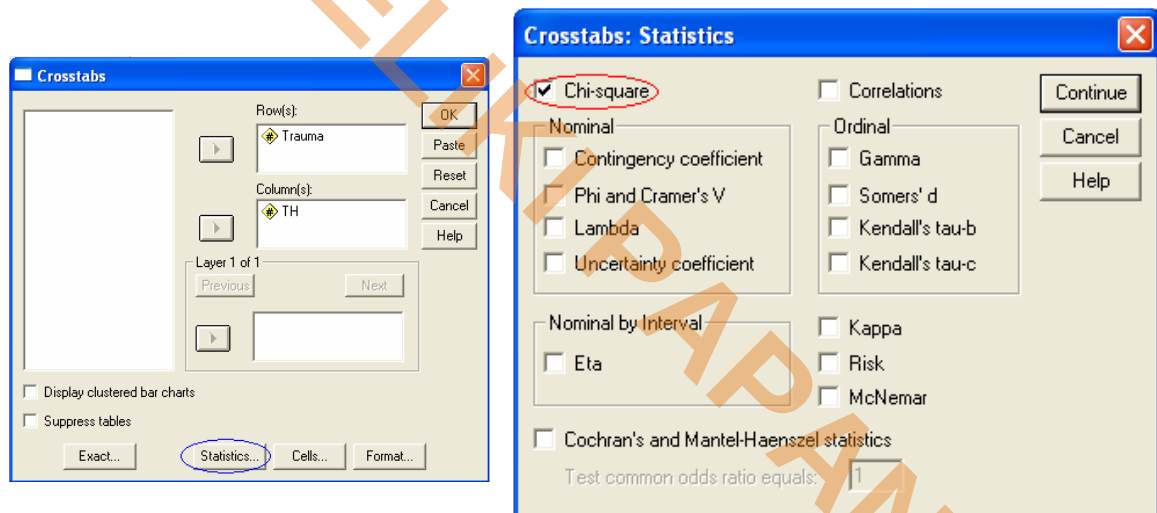
(1) Δυο ποιοτικές μεταβλητές ή μια ποιοτική και μια ποσοτική μεταβλητή

Ελέγχουμε την υπόθεση: οι δύο μεταβλητές δεν συσχετίζονται.

Analyze

Descriptive Statistics

Crosstabs



Output

- Πίνακας: Chi-Square Tests

Pearson Chi –Square (γραμμικός συντελεστής συσχέτισης Pearson)

Αν η σημαντικότητα (significance) του συντελεστή συσχέτισης Pearson < 0.05 , τότε η υπόθεση H_0 απορρίπτεται, δηλαδή οι δύο μεταβλητές είναι συσχετισμένες.

(2) Δυο ποσοτικές μεταβλητές: ΠΑΛΙΝΔΡΟΜΗΣΗ

Ο γενικός όρος 'Ανάλυση Παλινδρόμησης' περιγράφει μία σειρά αναλύσεων που είναι πλέον πολύ δημοφιλείς στην επιστημονική κοινότητα. Γενικά, η ανάλυση παλινδρόμησης μας βοηθά να εξετάσουμε σχέσεις αιτιότητας (causal relationships), να ερμηνεύσουμε δηλαδή τη διακύμανση των τιμών της εξαρτημένης μεταβλητής με βάση μία ή περισσότερες ανεξάρτητες μεταβλητές.

Χρησιμοποιείται λοιπόν προκειμένου να:

- ☐ Προβλέψουμε τις τιμές μιας μεταβλητής με βάση τις τιμές μίας ή περισσότερων άλλων μεταβλητών.
- ☐ Αποφασίσουμε αν κάποια μεταβλητή είναι 'καλή' για την πρόβλεψη κάποιας άλλης μεταβλητής.
- ☐ Να βρούμε ποιο είναι το ποσοστό της διακύμανσης των τιμών μιας μεταβλητής που μπορεί να εξηγηθεί από τις τιμές μίας μεμονωμένης ή ενός συνόλου μεταβλητών.
- ☐ Να φτιάξουμε καινούρια μοντέλα και να ελέγξουμε υπάρχουσες θεωρίες

Απλή γραμμική παλινδρόμηση:

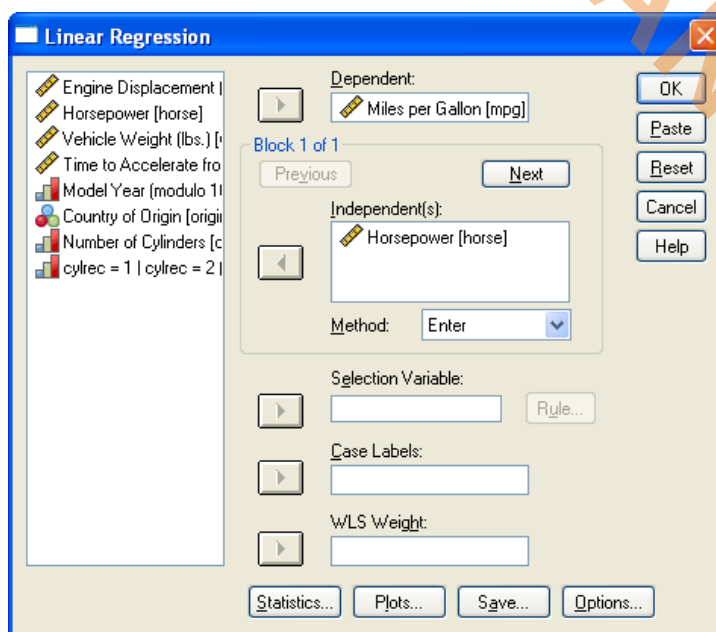
Τα ζεύγη των τιμών των δυο μεταβλητών (x, y) προσαρμόζονται σε μια ευθεία. Ψάχνουμε τους συντελεστές της ευθείας και αν γίνεται καλή προσαρμογή.

Εξετάζουμε την υπόθεση H_0 : δεν υπάρχει γραμμική σχέση μεταξύ των μεταβλητών X, Y.

Analyze

Regression

Curve Estimation ή Linear Regression

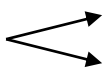


Output

Πίνακας: Model Summary

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,001 ^a	,000	-,002	\$1,945.610

a. Predictors: (Constant), Job time (months)

R : συντελεστής προσδιορισμού 

0 : κακή προσαρμογή
1 : καλή προσαρμογή

R Square : συντελεστής πολλαπλής συσχέτισης

Adjusted R Square (για μικρό δείγμα)

Πίνακας: ANOVA

Αν Sig. F < 0.05, τότε η H_0 απορρίπτεται, δηλ. οι μεταβλητές X και Y είναι γραμμικά συσχετισμένες.

ANOVA ^b						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2908,070	1	2908,070	,001	,978 ^a
	Residual	1,790E9	473	3785397,579		
	Total	1,790E9	474			

a. Predictors: (Constant), Job time (months)

b. Dependent Variable: Current Salary

Πίνακας: Coefficients

Δίνονται οι συντελεστές της ευθείας.

Αν Sig. < 0.05 τότε απορρίπτεται η υπόθεση ότι οι συντελεστές της ευθείας είναι μηδέν.

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	45109,696	277,435		,000
	Job time (months)	-,245	8,855	-,001	,978

a. Dependent Variable: Current Salary

Πολλαπλή γραμμική παλινδρόμηση

Επέκταση της απλής γραμμικής παλινδρόμησης για δύο ή περισσότερες ανεξάρτητες μεταβλητές.

Εξετάζουμε την υπόθεση H_0 : δεν υπάρχει γραμμική σχέση μεταξύ των μεταβλητών.

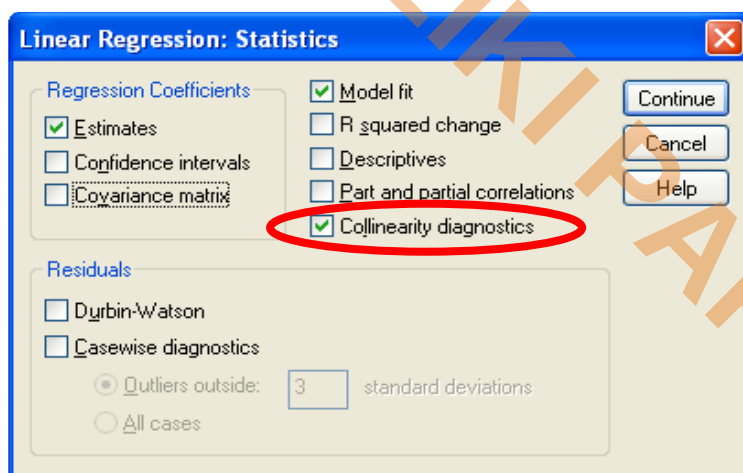
Από το μενού επιλέγουμε πάλι:

Analyze

Regression

Linear

Μία απαραίτητη προϋπόθεση η οποία είναι απαραίτητη γενικά σε όλα τα μοντέλα με περισσότερες εκ της μίας ανεξάρτητων μεταβλητών είναι η έλλειψη συγγραμμικότητας. Η συγγραμμικότητα είναι ένα σοβαρό πρόβλημα για την πολλαπλή γραμμική παλινδρόμηση. Όταν μία ανεξάρτητη μεταβλητή συσχετίζεται με μία άλλη ανεξάρτητη, δηλαδή μέσω της μίας μπορούμε να εκτιμήσουμε τις τιμές της άλλης τότε μιλάμε για πρόβλημα συγγραμμικότητας. Επομένως η ύπαρξη και των δύο μεταβλητών στο μοντέλο δεν είναι δυνατή.



Output

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Job time (months)	Education (years)
1	1	2,839	1,000	,01	,01	,01
	2	,123	4,801	,00	,37	,57
	3	,038	8,685	,99	,61	,42

a. Dependent Variable: Current Salary

Η στήλη του παραπάνω πίνακα **Condition Index** αποτελεί ένα διαγνωστικό του προβλήματος. Τιμές μεγαλύτερες του 15 φανερώνουν πιθανό πρόβλημα συγγραμμικότητας και τιμές άνω του 30 σοβαρό πρόβλημα συγγραμμικότητας.

Coefficients ^a								
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	45241,999	373,487		121,134	,000		
	Job time (months)	-,528	8,878	-,003	-,059	,953	,996	1,004
	Education (years)	-24,887	46,991	-,024	-,530	,597	,996	1,004

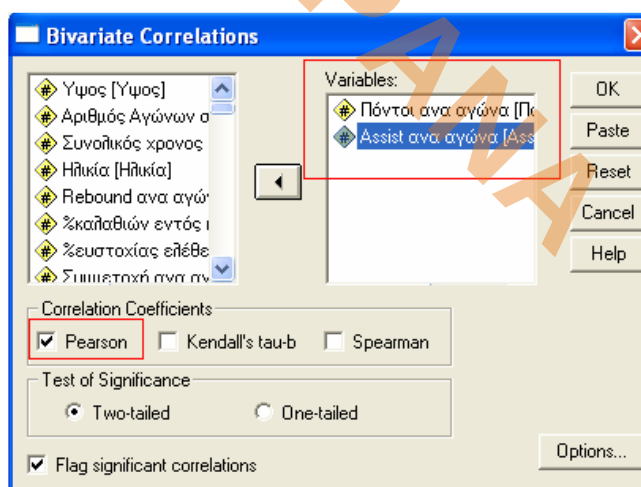
a. Dependent Variable: Current Salary

Οι δύο τελευταίες στήλες του παραπάνω πίνακα αναφέρονται σε διαγνωστικά συγγραμμικότητας, όπως και ο παρακάτω πίνακας. Το VIF (Variation Inflation Factor) είναι μέτρο διάγνωσης συγγραμμικότητας. Τιμές μεγαλύτερες του δύο αποτελούν ένδειξη ότι έχουμε πρόβλημα συγγραμμικότητας. Οι τιμή της Tolerance για μία τιμή φανερώνει το ποσοστό της διακύμανσης της μεταβλητής που εξηγείται από τις υπόλοιπες ανεξάρτητες μεταβλητές του μοντέλου. Πιο συγκεκριμένα ισχύει ότι το ποσοστό αυτό είναι ίσο με $(1 - \text{Tolerance})\%$. Τιμές της Tolerance μικρότερες του 0.5 αποτελούν ένδειξη του προβλήματος.

Η διαδικασία "Correlate"

Με την διαδικασία "Correlate" υπολογίζουμε περιγραφικά στατιστικά μέτρα που μας επιτρέπουν να εξετάσουμε την ένταση και την φύση της σχέσης μεταξύ δύο ποσοτικών μεταβλητών.

Analyze
Correlate
Bivariate



Ο γραμμικός συντελεστής **Pearson** παίρνει τιμές στο $[-1,1]$, όπου το 1 σημαίνει τέλεια θετική γραμμική συσχέτιση, το -1 σημαίνει τέλεια αρνητική γραμμική συσχέτιση, και το 0 σημαίνει ότι δεν υπάρχει γραμμική συσχέτιση μεταξύ των μεταβλητών.

Κεφάλαιο 7

ΣΥΓΚΡΙΣΗ ΜΕΣΩΝ ΟΡΩΝ ΔΕΙΓΜΑΤΩΝ/ ΠΛΗΘΥΣΜΟΥ

Το δείγματα πρέπει να κατανέμονται κανονικά ή να είναι ≥ 30 .

(1) Σύγκριση μέσου όρου ενός δείγματος ή ενός πληθυσμού με μια σταθερή τιμή.

Μηδενική υπόθεση

H_0 : η μέση τιμή ενός δείγματος είναι ίση με μια δοσμένη τιμή.

Analyze

Compare Means

One Sample T-Test

Μεταφέρουμε στο πλαίσιο Test Variable(s) τη μεταβλητή της οποίας τη μέση τιμή θέλουμε να συγκρίνουμε με δεδομένη τιμή.

Την τιμή για την οποία θα γίνει η σύγκριση εισάγουμε στη συνέχεια στη θέση Test Value.

Output

One-Sample Test

	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Education (years)	11,199	474	,000	,9797	,808	1,152

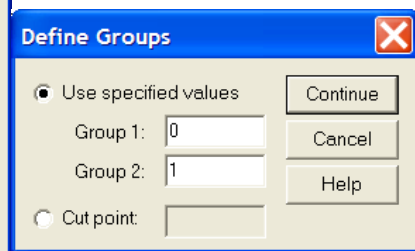
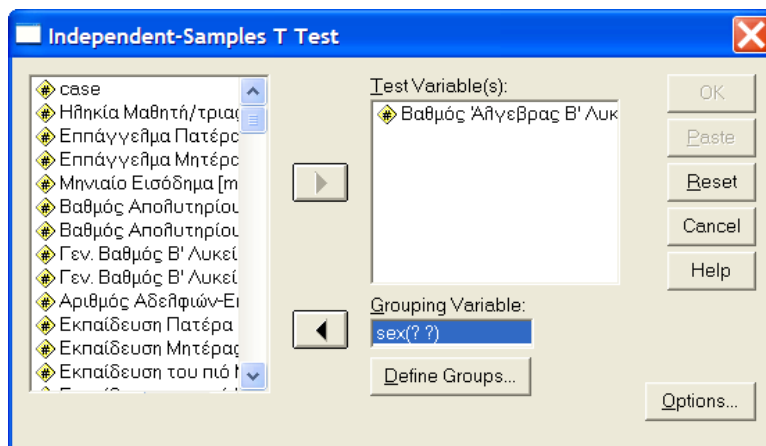
(2) Σύγκριση των μέσων όρων δύο ανεξάρτητων δειγμάτων – Κατηγοριοποίηση των τιμών μιας ποσοτικής μεταβλητής με βάση μια ποιοτική μεταβλητή.

Μηδενική υπόθεση H_0 : οι μέσες τιμές δύο δειγμάτων είναι ίσες.

Analyze

Compare Means

Independent Samples t-Test



Output

		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference		Lower	Upper
Βαθμός Άλγεβρας Β' Λυκείου	Equal variances assumed	,094	,760	-,085	343	,932	-,03	,346		-,711	,652
	Equal variances not assumed			-,086	327,28	,932	-,03	,345		-,709	,650

Γίνεται έλεγχος για τη σύγκριση των διασπορών των δειγμάτων: **Levene's Test for Equality of Variances** (H_0 : οι διασπορές δεν διαφέρουν). Από τη σημαντικότητα αυτού του ελέγχου

$$\text{Sig.} = 0.760 > 0.05$$

συμπεραίνουμε ότι δεν υπάρχει σημαντική διαφορά στις διασπορές και επομένως συμπεραίνουμε ότι οι διασπορές δεν διαφέρουν.

Επειδή η σημαντικότητα του ελέγχου είναι

$$\text{Sig} = 0.932 > 0.05,$$

αποτυγχάνουμε να απορρίψουμε τη μηδενική υπόθεση και συμπεραίνουμε ότι δεν υπάρχει στατιστικά σημαντική διαφορά στο βαθμό Άλγεβρας μεταξύ αγοριών και κοριτσιών.

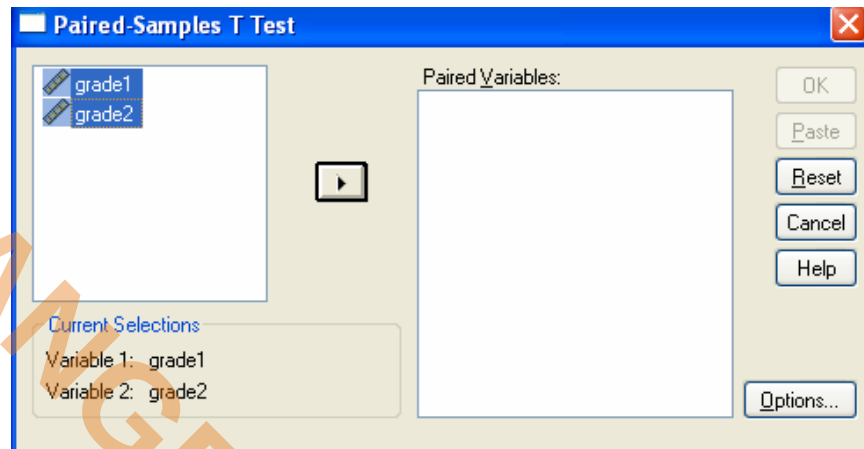
(3) Σύγκριση των μέσων όρων δύο εξαρτημένων δειγμάτων : Πολλές φορές για την πραγματοποίηση μιας μελέτης, είναι απαραίτητο να συλλέξουμε παρατηρήσεις πάνω στα ίδια αντικείμενα σε δυο διαφορετικές χρονικές στιγμές μετρώντας την ίδια παράμετρο/ μεταβλητή.

Μηδενική υπόθεση H_0 : οι μέσες τιμές δύο δειγμάτων είναι ίσες.

Analyze

Compare Means

Paired Samples t-Test



Αν η σημαντικότητα (significance) κάθε ελέγχου < 0.05 τότε η μηδενική υπόθεση της ισότητας των μέσων τιμών απορρίπτεται και άρα συμπεραίνουμε ότι τα δείγματα δεν έχουν ίσες μέσες τιμές.

Output

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Current Salary - Beginning Salary	*****	\$2,281.797	\$104.696	*****	*****	191,358	474	,000