Έλεγχος κανονικότητας

- Έλεγχος κανονικότητας : Επειδή πολλές μέθοδοι της Στατιστικής εφαρμόζονται σε δεδομένα που ακολουθούν την κανονική κατανομή, είναι πολύ χρήσιμο να είμαστε σε θέση να γνωρίζουμε εάν τα δεδομένα μας ακολουθούν την κανονική κατανομή ή ανήκουν σε πληθυσμό που ακολουθεί την κανονική κατανομή.

- Μέθοδοι Ελέγχου κανονικότητας :
- Θεωρητικές: Kolmogorov – Smirnov, Shapiro – Wilk
- Γραφικές: Γραφήματα Normal probability (PP plot), Quantile – Quantile plot (QQ plot)

Έλεγχος Kolmogorov – Smirnov

Έλεγχος Kolmogorov – Smirnov

- Τύπος εξέτασης για ένα δείγμα κυρίως όταν είναι μεγάλο.

- Αρχική Θεώρηση, Έλεγχος συγκεκριμένης υπόθεσης: Το δείγμα μας προέρχεται από πληθυσμό με γνωστή συνάρτηση κατανομής $F_1(X)$, δηλαδή τη συνάρτηση της κανονικής κατανομής.
- $\mathrm{H}_0: F(X) = F_1(X)$
- Εναλλακτικά (εναλλακτική υπόθεση) πρέπει να υποθέσουμε ότι το δείγμα μας δεν προέρχεται από την παραπάνω συνάρτηση αλλά από κάποια άλλη συνάρτηση.
- $\mathrm{H}_1: F(X) \neq F_1(X)$

Έλεγχος Kolmogorov – Smirnov

- Κατασκευάζεται συνάρτηση η οποία βασίζεται στην κατανομή του δείγματος (εμπειρική συνάρτηση)

- Για να ισχύει η υπόθεση ότι το δείγμα μας ακολουθεί την κανονική κατανομή πρέπει η εμπειρική συνάρτηση σχεδόν να ταυτίζεται με τη συνάρτηση κανονικής κατανομής, αλλιώς τα δεδομένα μας δεν προέρχονται από τον συγκεκριμένο πληθυσμό, αφού πήραμε ως δεδομένο ότι ο πληθυσμός ακολουθεί την κανονική κατανομή.

Έλεγχος Kolmogorov – Smirnov

- Από τους Kolmogorov – Smirnov:

1. Ορίστηκε η απόσταση μεταξύ των δύο υποθέσεων ως το μέγιστο της απόλυτης διαφοράς της συνάρτησης κατανομής από την εμπειρική συνάρτηση.

2. Ορίστηκε ένα επίπεδο σημαντικότητας α σύμφωνα με το οποίο η παραπάνω διαφορά είναι αρκετά σημαντική ώστε να γίνει αποδεκτή ή όχι η αρχική υπόθεση (το δείγμα ακολουθεί την κανονική κατανομή)

Έλεγχος Kolmogorov – Smirnov

Σύμφωνα με το επίπεδο σημαντικότητας α γίνεται έλεγχος της τιμής ενός συντελεστή p (p-value ή Sig). Έχει αποδειχθεί και ισχύουν τα εξής:

1. Εάν η τιμή p-value είναι μεγαλύτερη του 0.05, τότε η αρχική υπόθεση γίνεται δεκτή, δηλαδή η τυχαία μεταβλητή από την οποία προήλθε το υπό μελέτη δείγμα ακολουθεί την κανονική κατανομή.
2. Εάν η τιμή p-value είναι μικρότερη του 0.05, τότε η αρχική υπόθεση απορρίπτεται, δηλαδή η τυχαία μεταβλητή από την οποία προήλθε το υπό μελέτη δείγμα δεν ακολουθεί την κανονική κατανομή.

Έλεγχος Shapiro – Wilk

Έλεγχος Shapiro – Wilk

- Τύπος εξέτασης για ένα δείγμα όταν είναι μικρό.

- Αρχική Θεώρηση, Έλεγχος συγκεκριμένης υπόθεσης: Το δείγμα μας προέρχεται από πληθυσμό με γνωστή συνάρτηση κατανομής $F_1(X)$, δηλαδή τη συνάρτηση της κανονικής κατανομής.
- $\mathrm{H}_0: F(X) = F_1(X)$
- Εναλλακτικά (εναλλακτική υπόθεση) πρέπει να υποθέσουμε ότι το δείγμα μας δεν προέρχεται από την παραπάνω συνάρτηση αλλά από κάποια άλλη συνάρτηση.
- $\mathrm{H}_1: F(X) \neq F_1(X)$

Έλεγχος Shapiro – Wilk

- Τύπος εξέτασης για ένα δείγμα όταν είναι μικρό.

- Ο έλεγχος κάνει χρήση της συνάρτησης W όπου

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

- Διατάσσουμε τις τιμές του δείγματος.
- Επίδραση στη συνάρτηση των διασπορών και μέσων τιμών.

Έλεγχος Kolmogorov – Smirnov SPSS

Έλεγχος Kolmogorov – Smirnov SPSS

- Ένας ιατρός θέλει να ελέγξει εάν τα δεδομένα του που αφορούν το χρόνο σε μήνες που ο ασθενής χρειάζεται για να επανέλθει σε φυσιολογική κατάσταση μετά τη διάγνωση της ασθένειας ακολουθούν την κανονική κατανομή.
- Για το λόγο αυτό ο ιατρός καταγράφει το χρόνο που απαιτείται από τη στιγμή της διάγνωσης της ασθένειας μέχρι ο ασθενής να αναρρώσει. Οι χρόνοι σε μήνες για 100 ασθενείς είναι στο επισυναπτόμενο αρχείο.

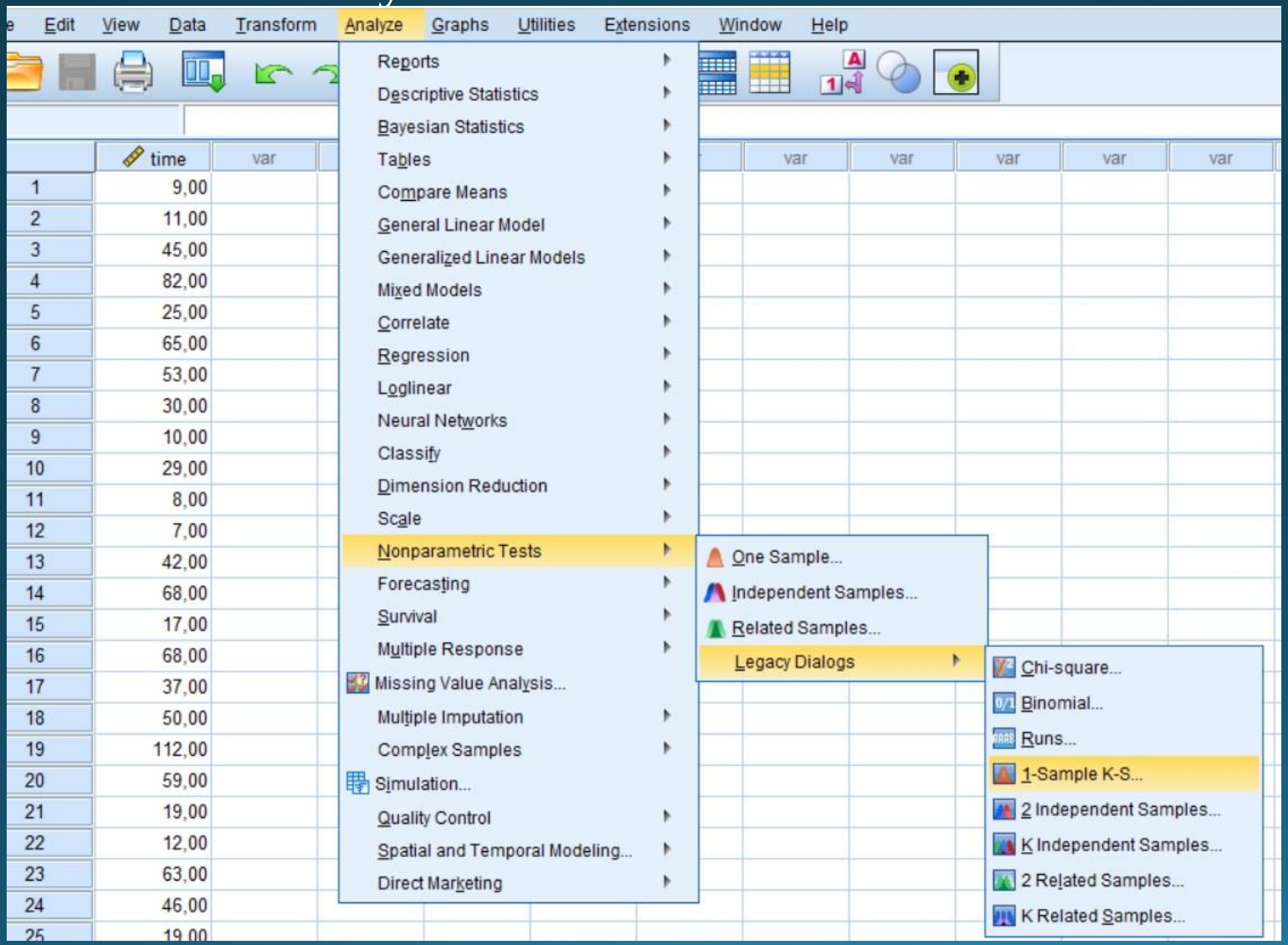
Έλεγχος Kolmogorov – Smirnov SPSS

- Κάνουμε χρήση ελέγχου υπόθεσης ότι τα δεδομένα ακολουθούν την κανονική κατανομή.

- Από το menu Analyze

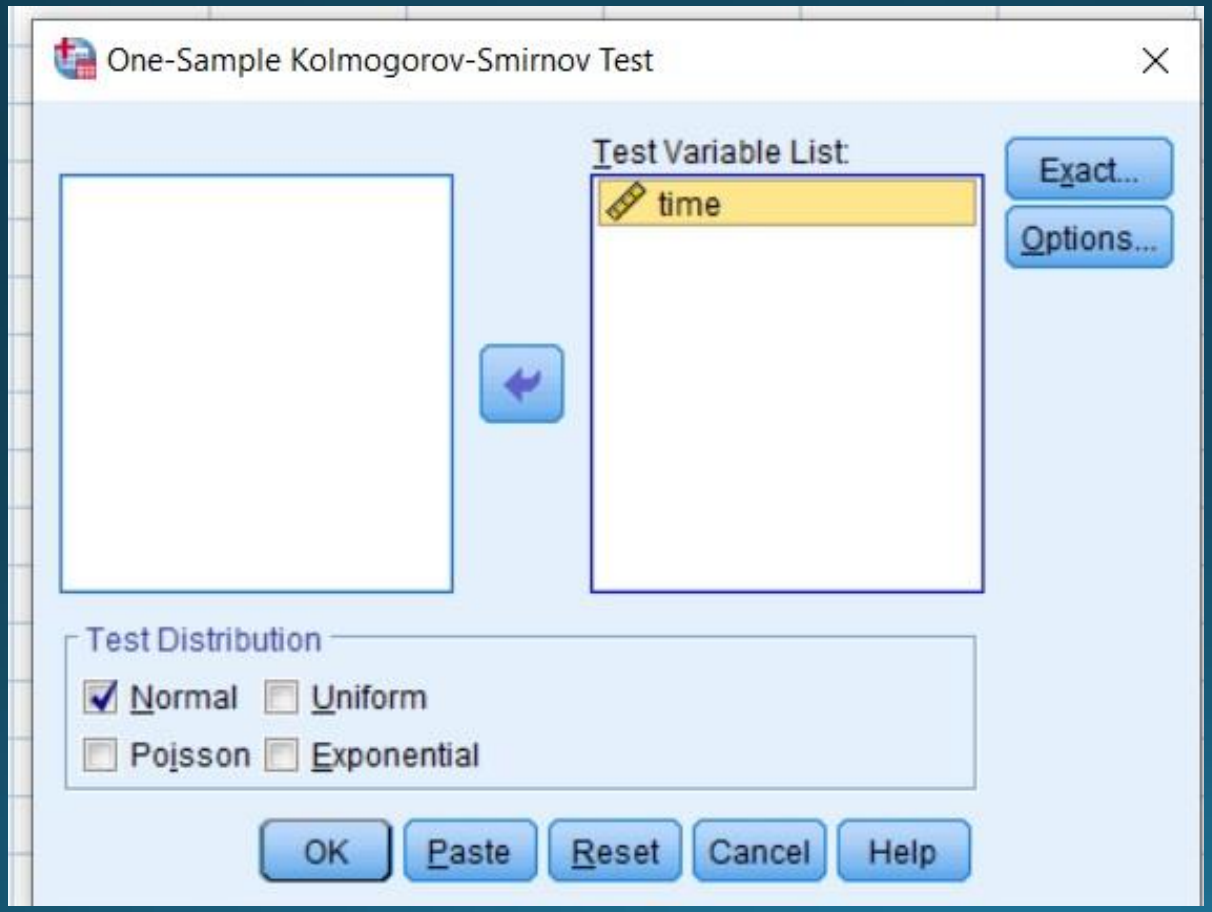
Έλεγχος Kolmogorov – Smirnov SPSS

- Από το menu Analyze



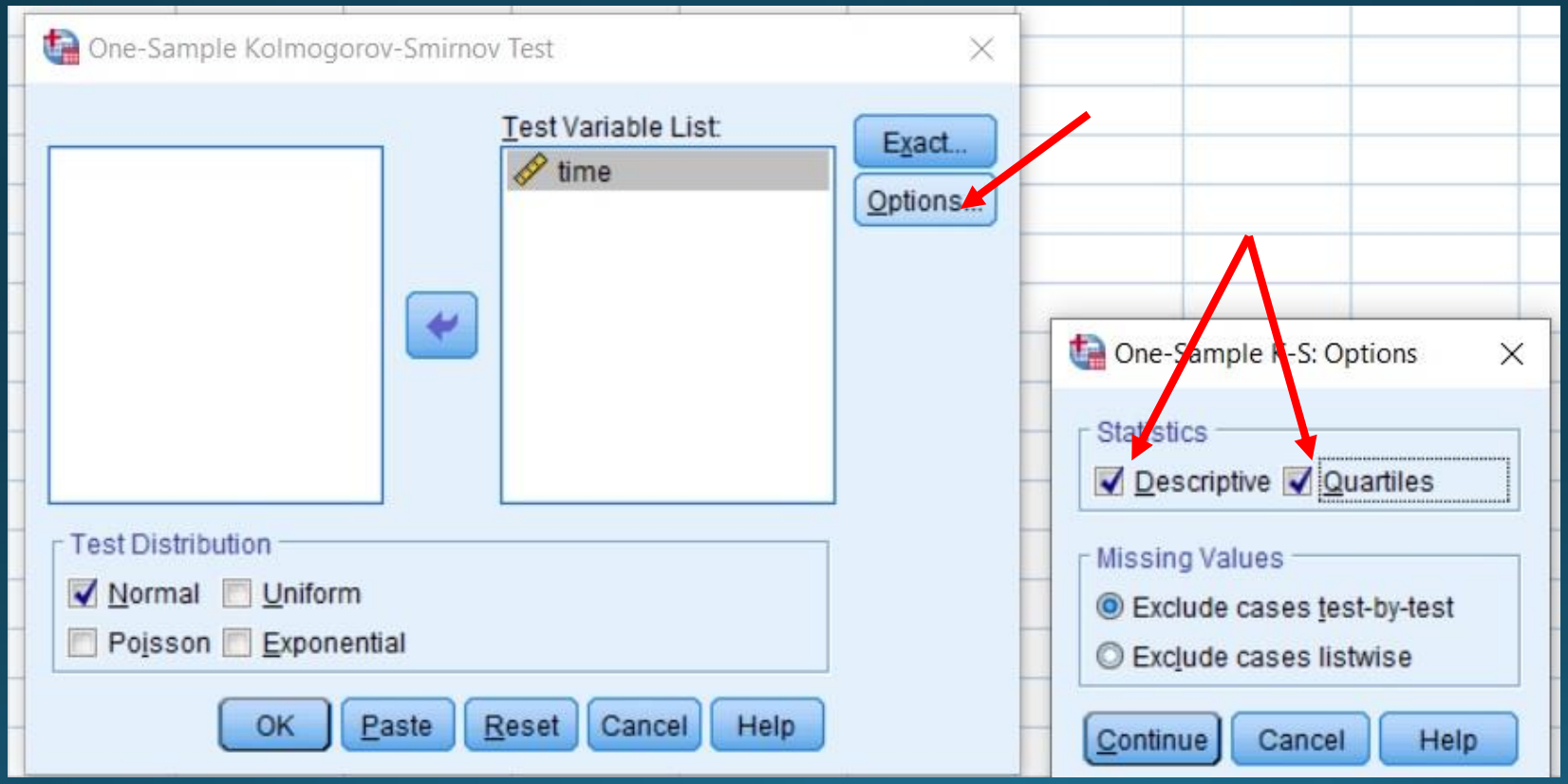
Έλεγχος Kolmogorov – Smirnov SPSS

- Μεταφορά μεταβλητής και Normal



Έλεγχος Kolmogorov – Smirnov SPSS

- Options



Έλεγχος Kolmogorov – Smirnov SPSS

- Περιγραφικά μέτρα

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles 25th	50th (Median)	75th
time	100	50,2900	28,02787	7,00	129,00	27,5000	50,0000	68,0000

- Ενδιάμεσος χρόνος (Median) =50 οι μισοί ασθενείς αναρρώνουν πιο νωρίς από τους 50 μήνες

- Το 75% των ασθενών αναρρώνει μετά από 27,5 μήνες (πρώτο τεταρτημόριο) ενώ το 25% αναρρώνει μετά από 68 μήνες (τρίτο τεταρτημόριο).

Έλεγχος Kolmogorov – Smirnov SPSS

- Τεστ κανονικότητας

One-Sample Kolmogorov-Smirnov Test

		time
N		100
Normal Parameters[a,b]	Mean	50,2900
	Std. Deviation	28,02787
Most Extreme Differences	Absolute	,065
	Positive	,065
	Negative	-,061
Test Statistic		,065
Asymp. Sig. (2-tailed)		,200[c,d]

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

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

- Η μεγαλύτερη διαφορά της εμπειρικής από την αναμενόμενη συνάρτηση κατανομής είναι 0,065

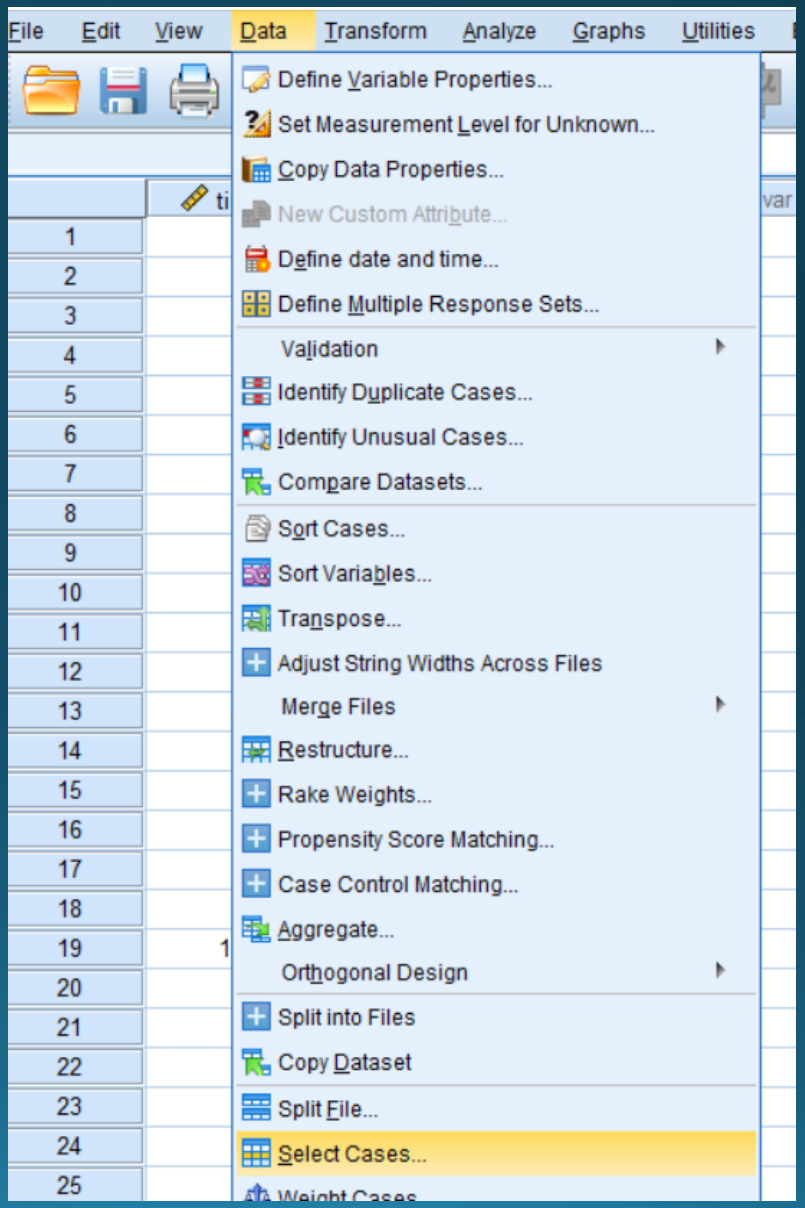
- Με επίπεδο σημαντικότητας 5% Το p-value > 0,2: κρατούμε τη μηδενική υπόθεση - > χρόνος ανάρρωσης ασθενών ακολουθεί κανονική κατανομή

Έλεγχος Shapiro – Wilk SPSS

Έλεγχος Shapiro – Wilk SPSS

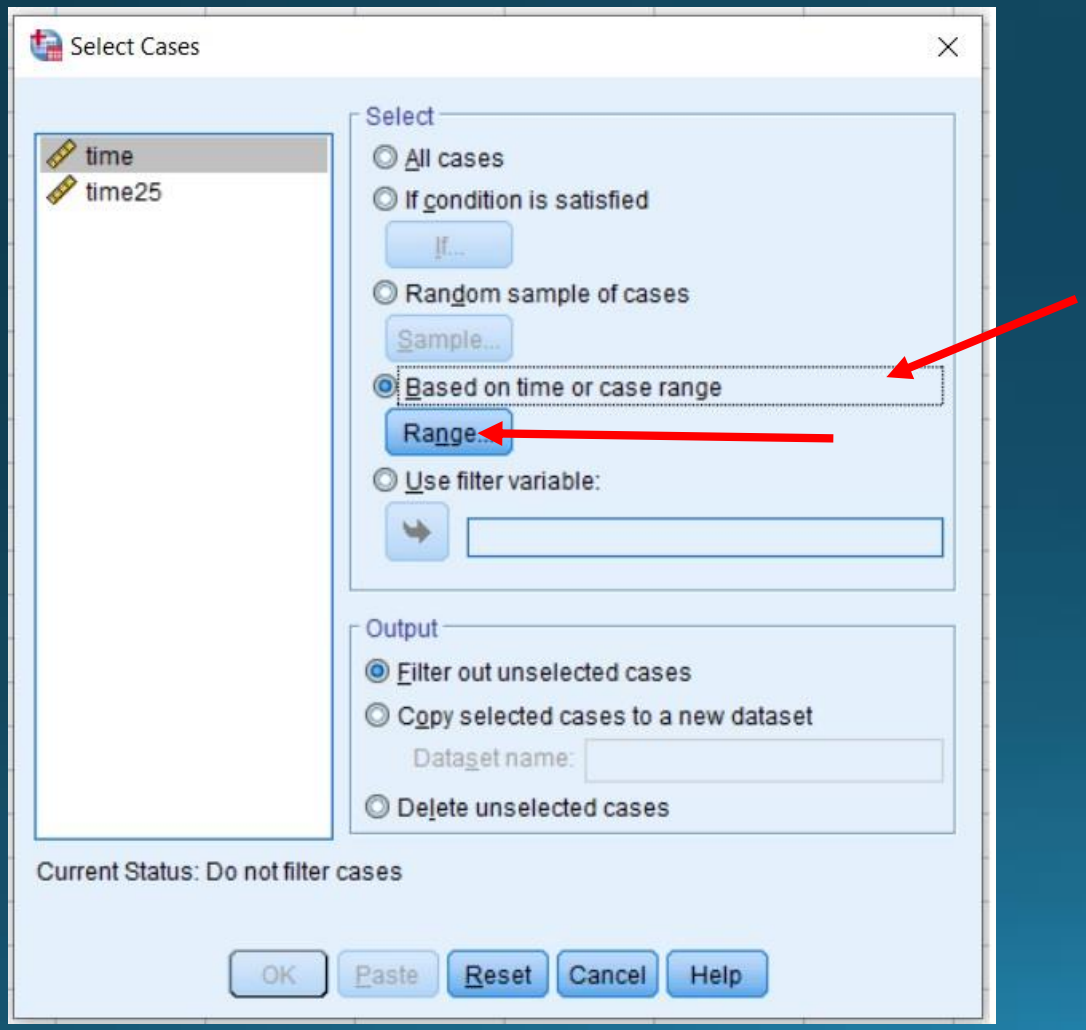
- Έστω ότι το προηγούμενο δείγμα είχε μόνο 25 παρατηρήσεις.

- Data -> Select Cases



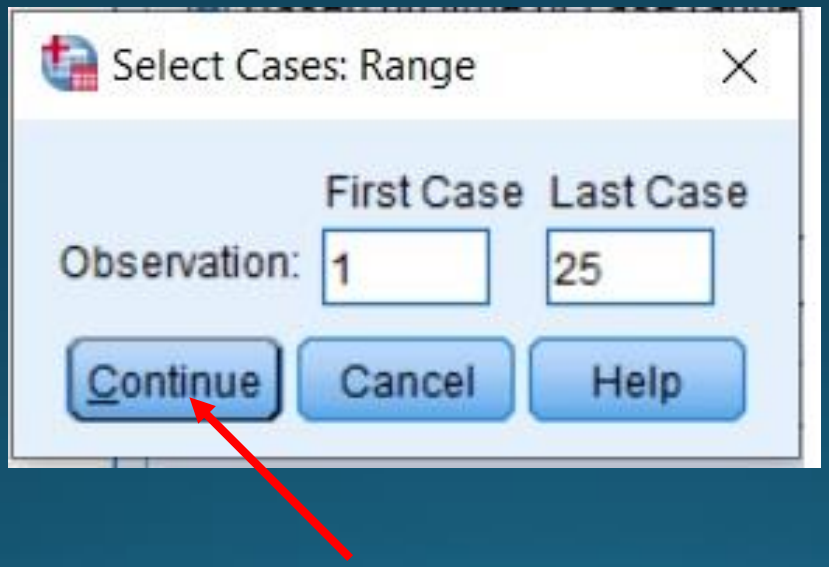
Έλεγχος Shapiro – Wilk SPSS

- Data -> Select Cases



Έλεγχος Shapiro – Wilk SPSS

- Επιλογή 1 έως 25 και continue



Έλεγχος Shapiro – Wilk SPSS

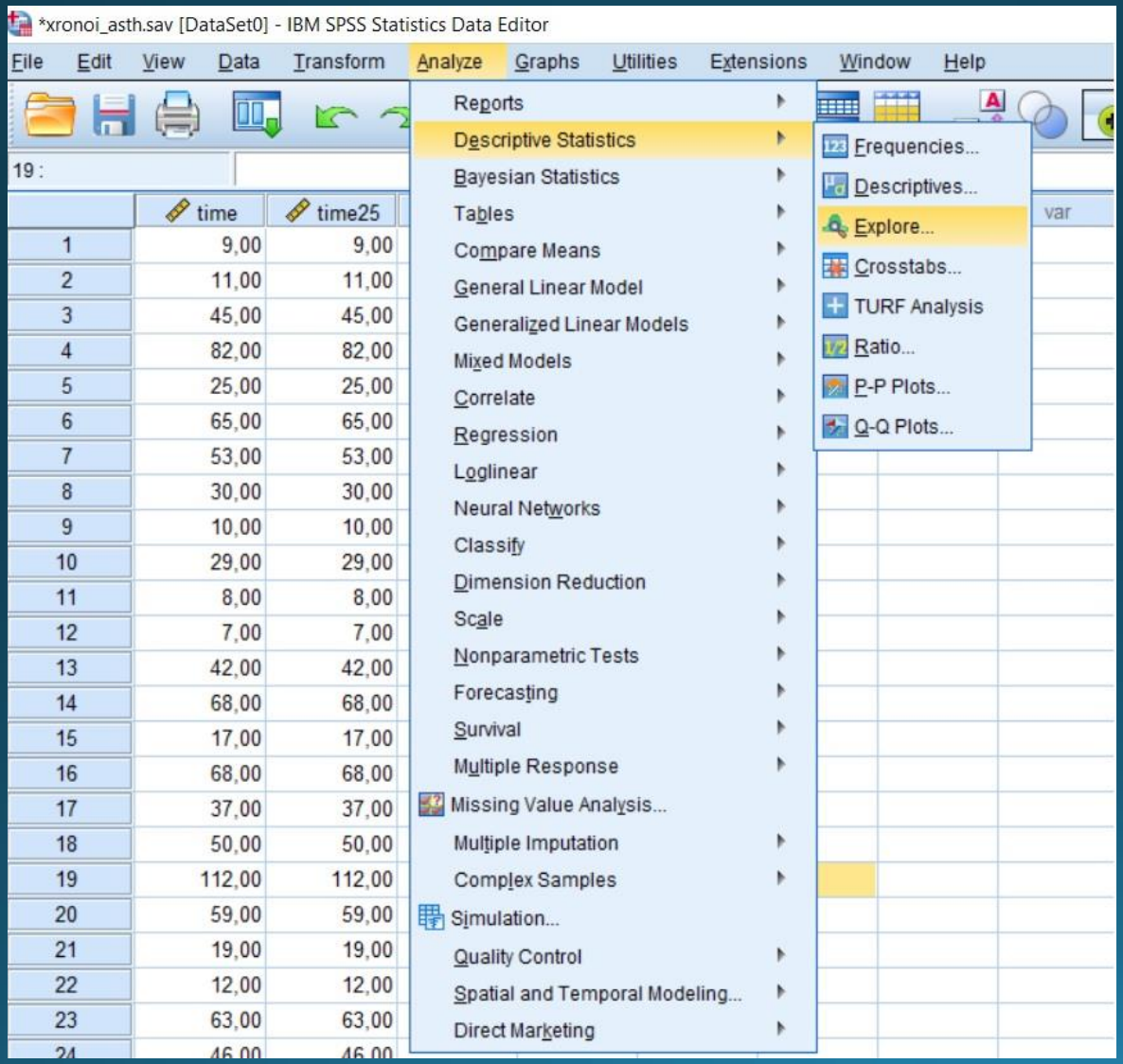
Κάτω του 26 διαγραμμένες

19 :

	time	time25	var
1	9,00	9,00	
2	11,00	11,00	
3	45,00	45,00	
4	82,00	82,00	
5	25,00	25,00	
6	65,00	65,00	
7	53,00	53,00	
8	30,00	30,00	
9	10,00	10,00	
10	29,00	29,00	
11	8,00	8,00	
12	7,00	7,00	
13	42,00	42,00	
14	68,00	68,00	
15	17,00	17,00	
16	68,00	68,00	
17	37,00	37,00	
18	50,00	50,00	
19	112,00	112,00	
20	59,00	59,00	
21	19,00	19,00	
22	12,00	12,00	
23	63,00	63,00	
24	46,00	46,00	
25	19,00	19,00	
~~26~~	68,00	.	
~~27~~	41,00	.	

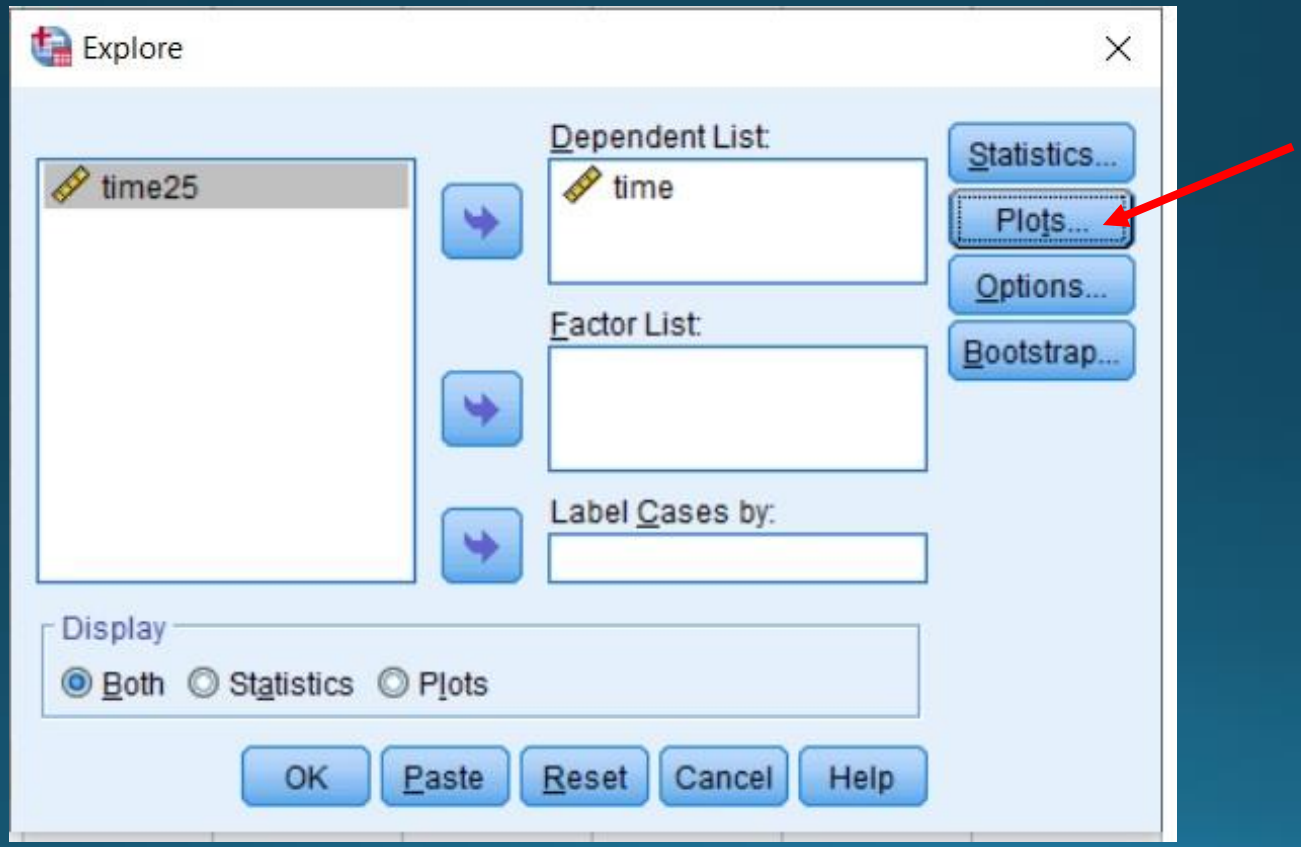
Έλεγχος Shapiro – Wilk SPSS

Menu Analyze

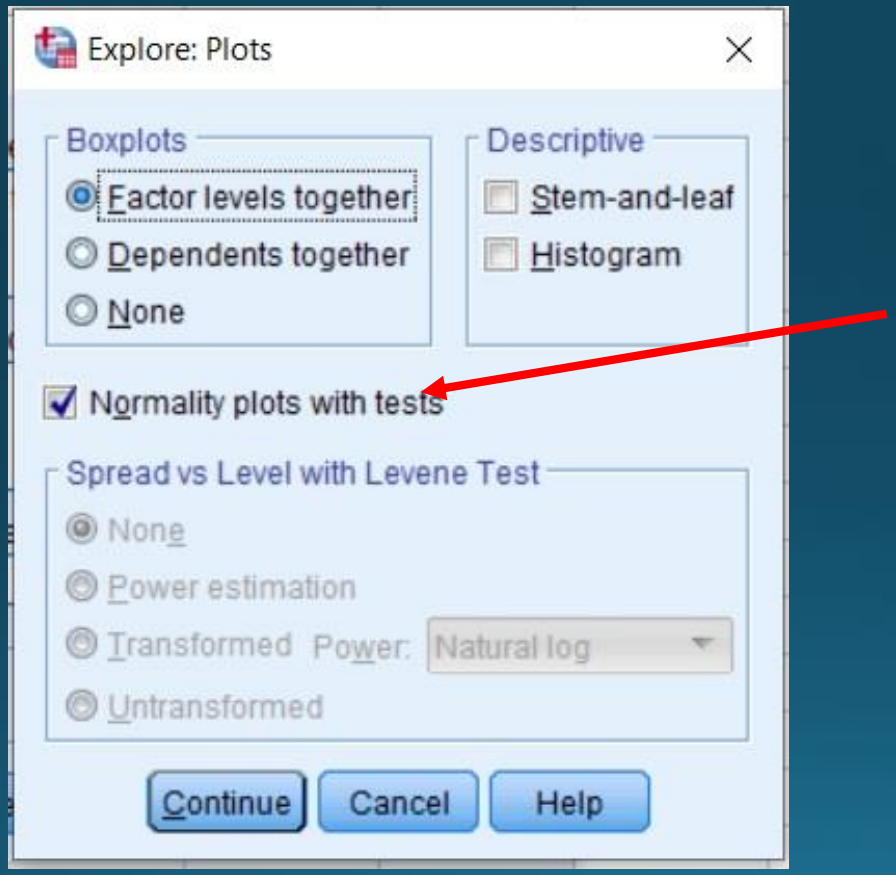


Έλεγχος Shapiro – Wilk SPSS

Θέτουμε την μεταβλητή και πιέζουμε το plots



Έλεγχος Shapiro – Wilk SPSS



Έλεγχος Shapiro – Wilk SPSS

Descriptives

			Statistic	Std. Error
time	Mean		39,4400	5,45010
	95% Confidence Interval for Mean	Lower Bound	28,1915	
		Upper Bound	50,6885	
	5% Trimmed Mean		37,5333	
	Median		37,0000	
	Variance		742,590	
	Std. Deviation		27,25050	
	Minimum		7,00	
	Maximum		112,00	
	Range		105,00	
	Interquartile Range		46,50	
	Skewness		,786	,464
	Kurtosis		,355	,902

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
time	,133	25	,200*	,923	25	,062

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

a. Lilliefors Significance Correction

P-value οριακά άνω του 0.05. Ίσως ακολουθεί κανονική κατανομή. Είναι μικρό το δείγμα.

Διαγράμματα πιθανότητας

Διαγράμματα πιθανότητας

Θεωρούμε τυχαίο δείγμα $Y_1, Y_2, \dots, Y_n$ με συνεχή αθροιστική συνάρτηση κατανομής $F(y)$

Θεωρούμε διατεταγμένο δείγμα της τυχαίας μεταβλητής $X_i = F(Y_i)$ με αναμενόμενη μέση τιμή $E[F(Y_i)] = \frac{i}{n+1}$ για κάθε i $= 1,2, \dots, n$

Παίρνουμε ένα μεγάλο πλήθος παρατηρήσεων $x_1, x_2, \dots, x_n$ και τοποθετουμε σε γράφημα τα σημεία

$$F(y_i), \frac{i}{n+1}$$

Πρέπει να βρίσκονται κοντά σε ευθεία $y = x$

Σχηματίζεται Γράφημα με άξονα x την παρατηρούμενη αθροιστική συνάρτηση πιθανότητας και άξονα y την αναμενόμενη αθροιστική συνάρτηση πιθανότητας: Διάγραμμα P-P

Παρόμοιο γράφημα είναι αυτό που στον άξονα x έχουμε την παρατηρούμενη τιμή και στον άξονα y την αναμενόμενη κανονικοποιημένη τιμή: Διάγραμμα Q-Q

Έλεγχος κανονικότητας
Διαγράμματα πιθανότητας περιληπτικά

- Αρχικά βρίσκουμε τη συνάρτηση πιθανότητας. Αποτελεί τη συνάρτηση που περιγράφει μαθηματικά τον πληθυσμό από τον οποίο προέρχεται το προς εξέταση δείγμα.

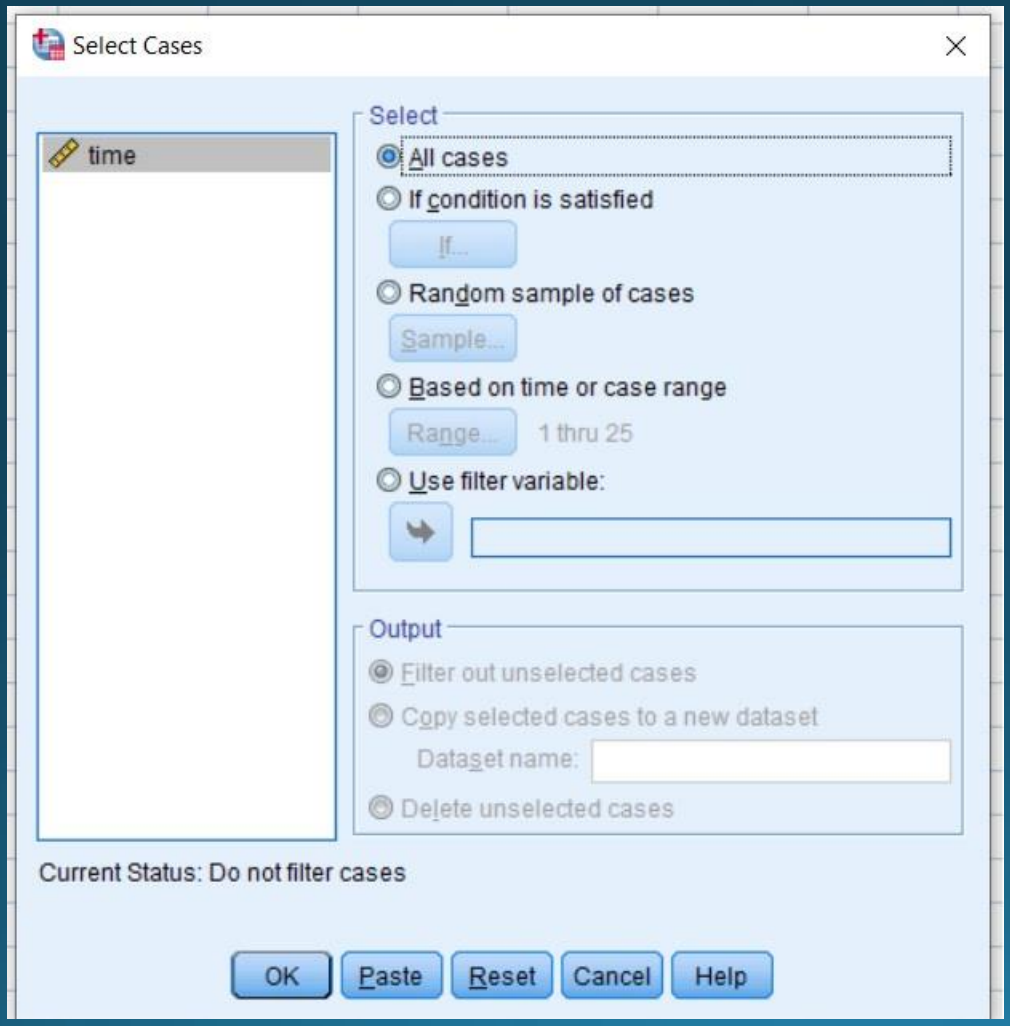
- Για τον σχεδιασμό των γραφημάτων πιθανότητας χρησιμοποιούμε τους άξονες του επιπέδου βαθμονομημένους με τη δεδομένη κατανομή, δηλαδή την κανονική.

- Διατάσσουμε τις παρατηρήσεις από την μικρότερη στη μεγαλύτερη.
- Σχεδιάζουμε τα γραφήματα.

Διαγράμματα πιθανότητας SPSS

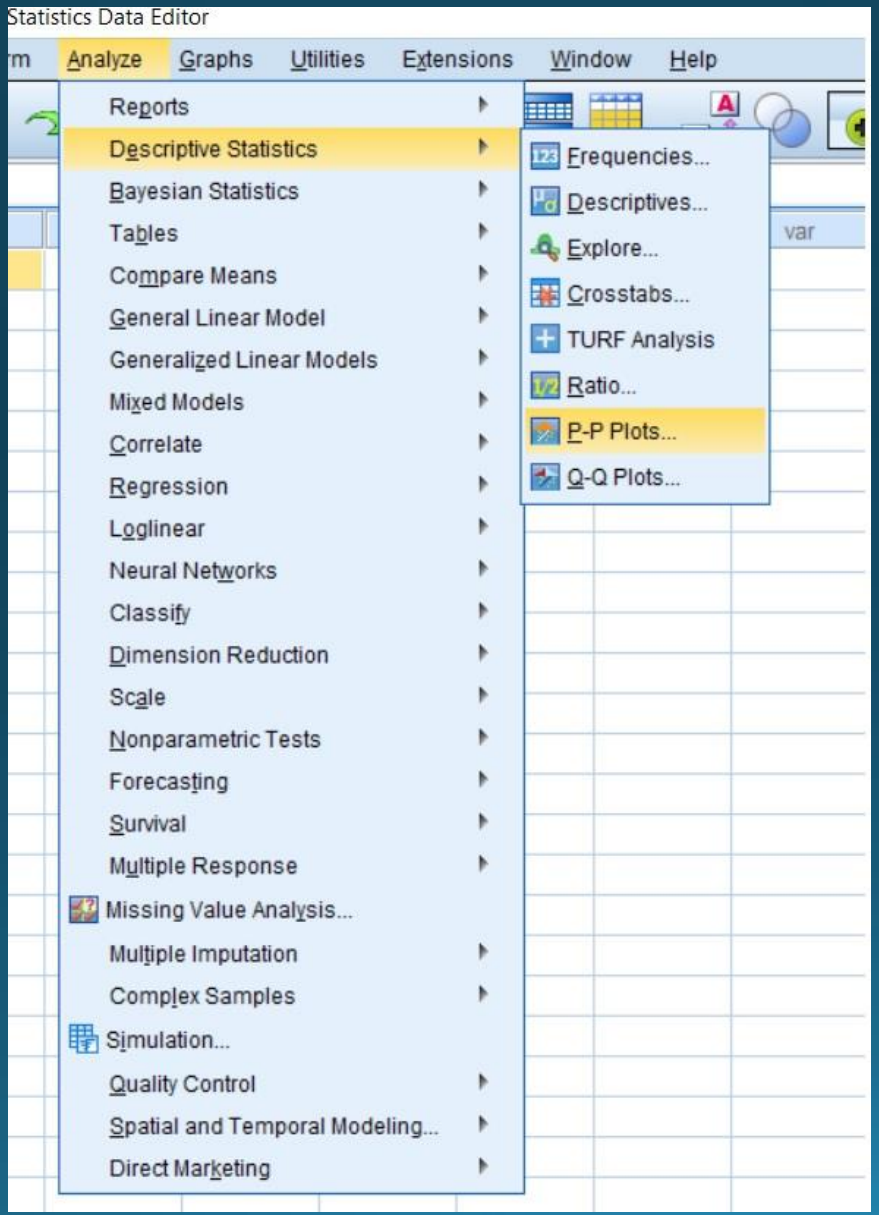
Διαγράμματα πιθανότητας SPSS

- Data -> Select cases -> All cases

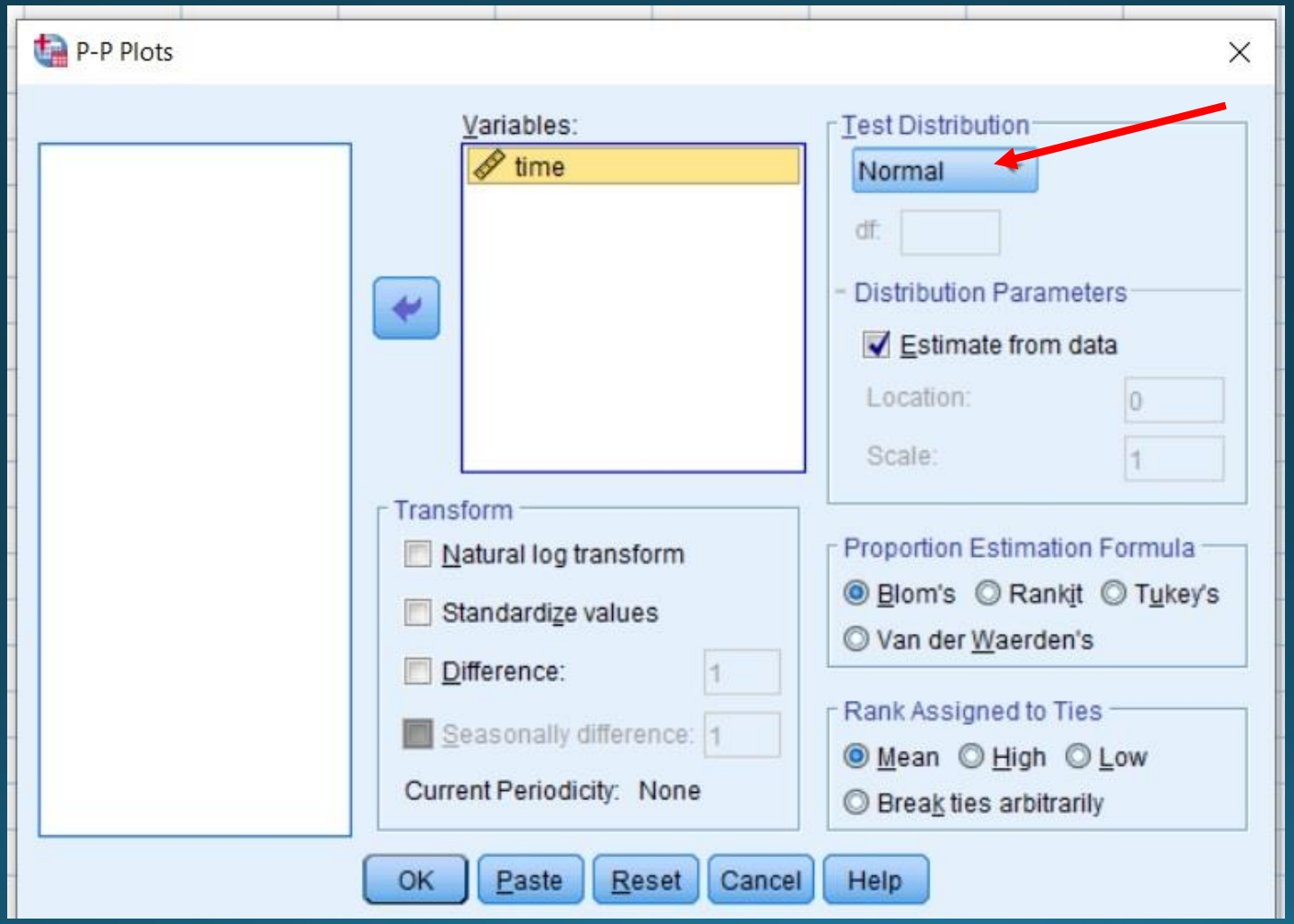


Διαγράμματα πιθανότητας SPSS

- Analyze



Διαγράμματα πιθανότητας SPSS



Διαγράμματα πιθανότητας SPSS

Εκτίμηση μέσης τιμής και διασποράς από την κανονική κατανομή

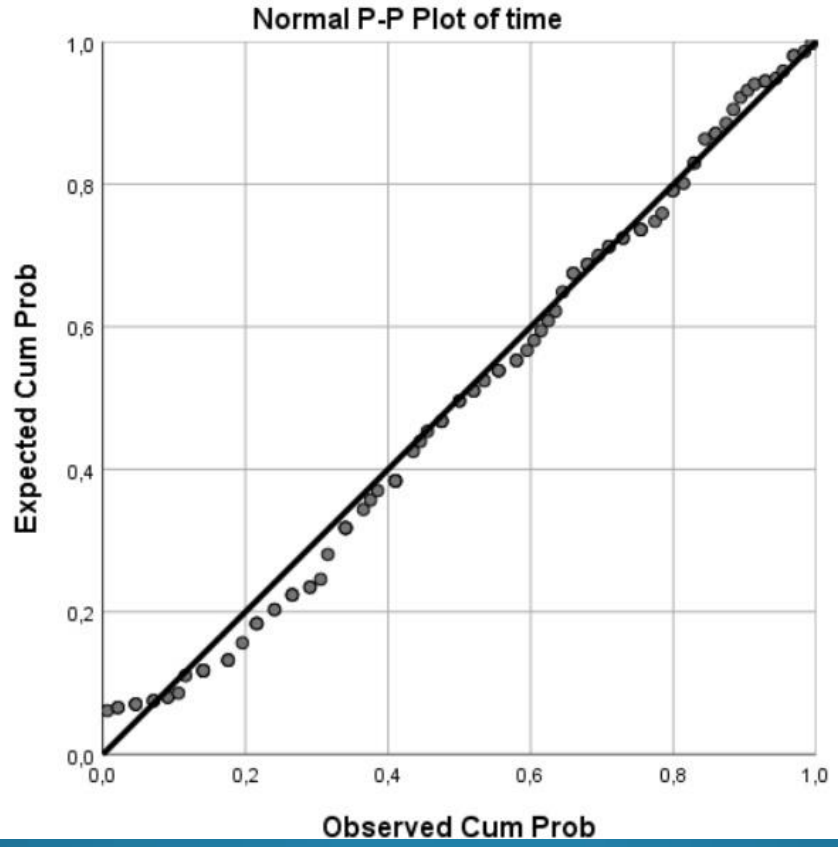
Όλες οι τιμές γύρω από την ευθεία άρα ακολουθεί την κανονική κατανομή

Estimated Distribution Parameters

		time
Normal Distribution	Location	50,2900
	Scale	28,02787

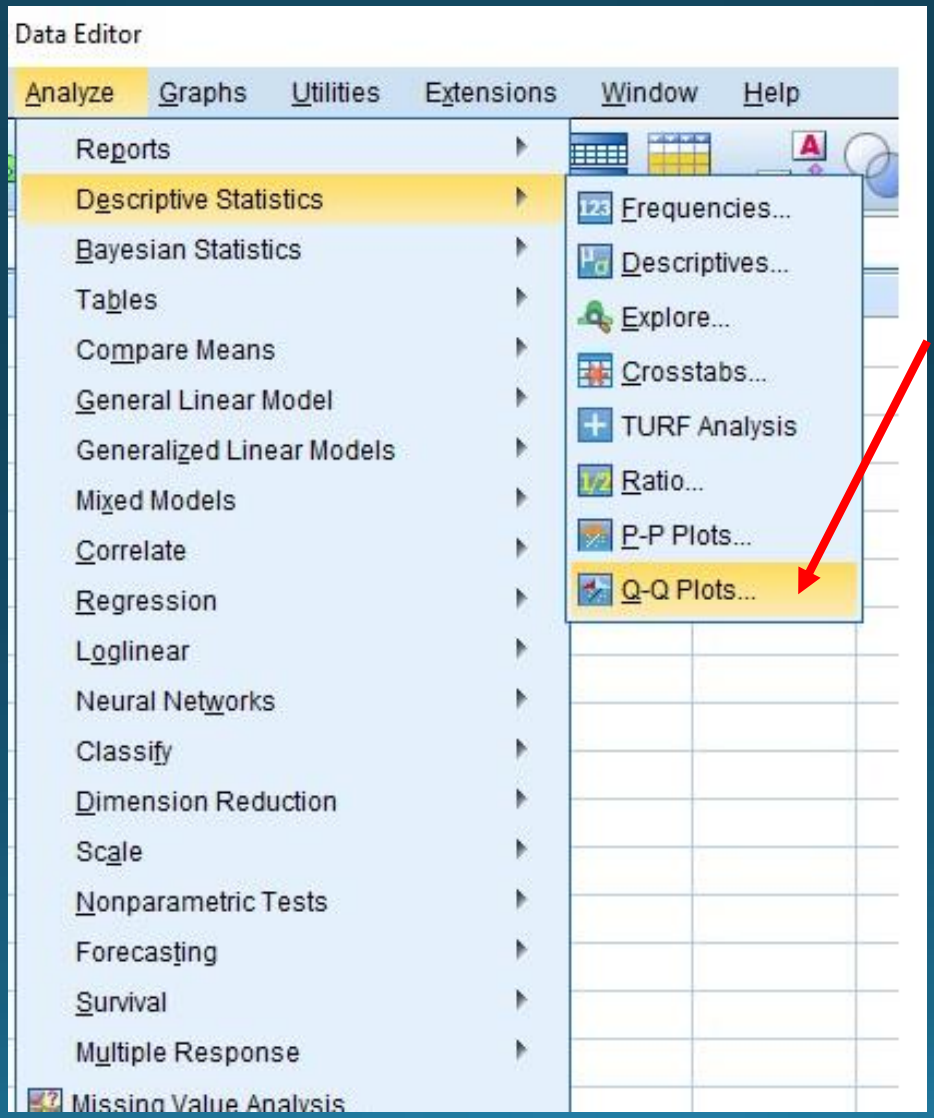
The cases are unweighted.

time



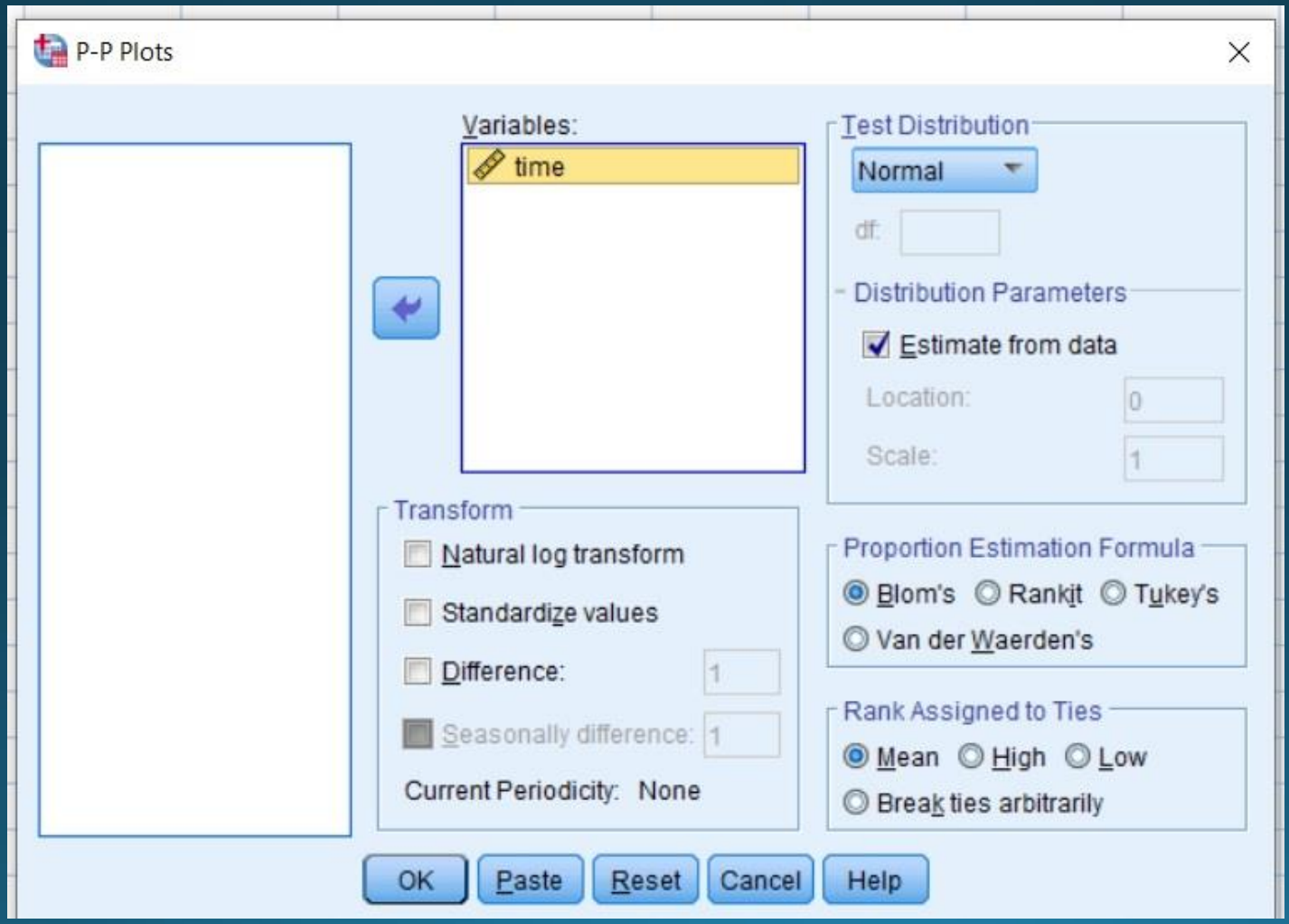
Γραφική απεικόνιση ποσοτικών δεδομένων
Q-Q Plot

- Επιλογή: Analyze -> Descriptive -> QQ plot



Γραφική απεικόνιση ποσοτικών δεδομένων
Q-Q Plot

- Επιλογή της μεταβλητής και της κατανομής από το menu κατανομών



Γραφική απεικόνιση ποσοτικών δεδομένων
Q-Q Plot

- Πίνακας εκτιμημένων παραμέτρων της κανονικής κατανομής

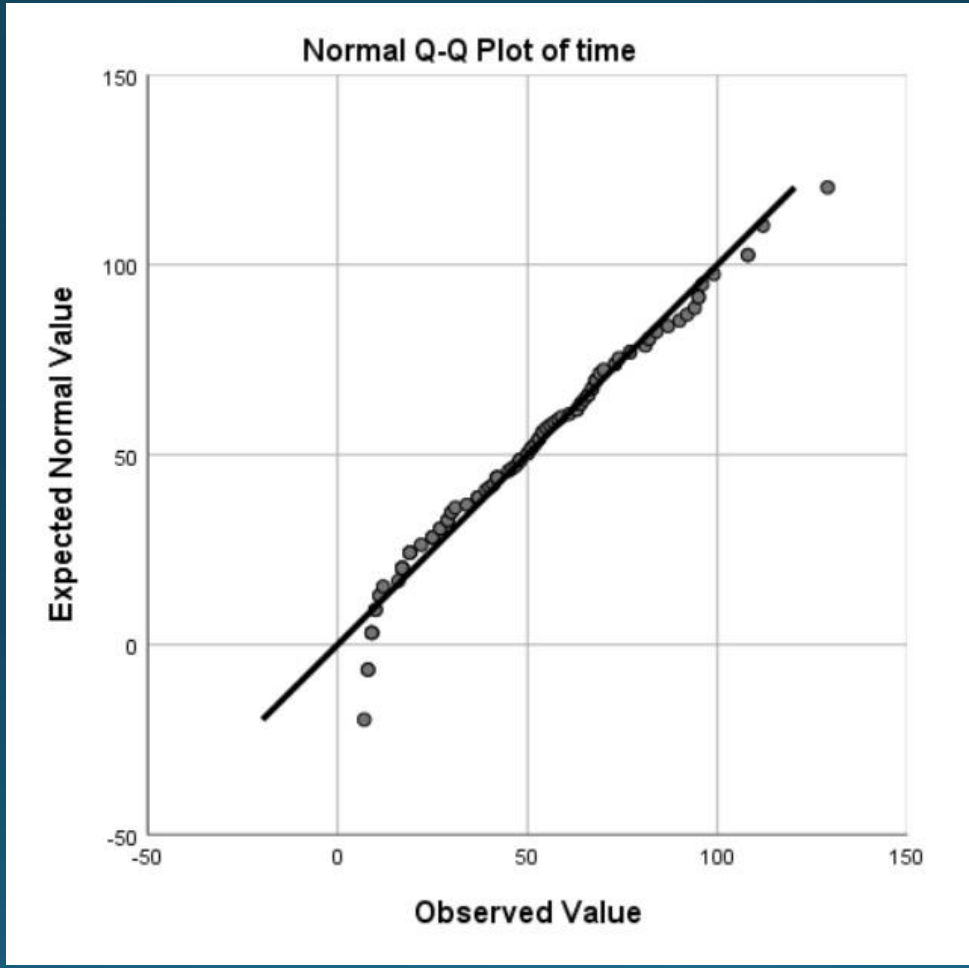
Estimated Distribution Parameters

		time
Normal Distribution	Location	50,2900
	Scale	28,02787

The cases are unweighted.

- Η μέση τιμή (location) εκτιμήθηκε 50,29
- Η τυπική απόκλιση (Scale) εκτιμήθηκε 28,027

Γραφική απεικόνιση ποσοτικών δεδομένων
Q-Q Plot



- Όλες οι τιμές βρίσκονται κοντά στη διαγώνιο -> τα δεδομένα επαληθεύεται ότι ακολουθούν την κανονική κατανομή