



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΠΕΡΙΓΡΑΦΙΚΗ και ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Εισήγηση 4B: Έλεγχοι Κανονικότητας

Διδάσκων: Δαφέρμος Βασίλειος
ΤΜΗΜΑ ΠΟΛΙΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
ΣΧΟΛΗΣ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Άδειες Χρήσης

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



[ή επιλογή ενός άλλου από τους έξι συνδυασμούς]

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

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

Χρηματοδότηση

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



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Δευτέρα 28-4-14

ΠΕΡΙΓΡΑΦΙΚΗ ΚΑΙ ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Εισηγητής: Βασίλης Δαφέρμος, Αναπληρωτής Καθηγητής

ΕΛΕΓΧΟΙ ΚΑΝΟΝΙΚΟΤΗΤΑΣ



Προλεγόμενα και διαπιστώσεις για να εισαχθούμε στους ελέγχους κανονικότητας.

- ένα τυχαίο δείγμα, ποτέ δεν είναι μια τέλεια εικόνα του πληθυσμού.
- έτσι, τα δείγματα που μπορούμε να πάρουμε από ένα κανονικό πληθυσμό, μάλλον περιμένουμε να είναι **κατά προσέγγιση** κανονικώς κατανεμημένα.
- κάθε δείγμα που λαμβάνεται από ένα κανονικό πληθυσμό, δεν ακολουθεί οπωσδήποτε την κανονική κατανομή.
- μια κανονική κατανομή μπορεί να έχει οποιαδήποτε μέση τιμή και οποιαδήποτε τυπική απόκλιση.

- Η κανονικότητα είναι μια βασική παραδοχή για την ΠΑΡΑΜΕΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ. Αυτό σημαίνει ότι τα κλασικά στατιστικά κριτήρια όπως το t-test, η ανάλυση διασποράς, η παλίνδρομη ανάλυση κτλ. για να εφαρμοστούν, θα πρέπει να ικανοποιείται η παραδοχή της κανονικότητας.

- Όταν δεν ικανοποιείται η παραδοχή της κανονικότητας, τότε αναγκαστικά θα περάσουμε στην ΜΗ ΠΑΡΑΜΕΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ, δηλ. σε μια στατιστική που δεν προϋποθέτει την κανονικότητα. Δεν την έχει ανάγκη. Ουσιαστικά στην τελευταία, στην ΜΗ ΠΑΡΑΜΕΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ δεν υπάρχει πρότυπο, σε αντίθεση με την ΠΑΡΑΜΕΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ στην οποία πάντα υπάρχει πρότυπο. Τα δεδομένα στην τελευταία ακολουθούν ένα συγκεκριμένο πρότυπο, ένα συγκεκριμένο δρόμο, μια συγκεκριμένη κατανομή.

- Το πέρασμα όμως από την ΠΑΡΑΜΕΤΡΙΚΗ στην ΜΗ ΠΑΡΑΜΕΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ δεν είναι χωρίς συνέπειες. Οποσδήποτε στη δεύτερη αυτή Στατιστική, στην ΜΗ ΠΑΡΑΜΕΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ, έχουμε λιγότερη ακρίβεια στο επίπεδο των αποτελεσμάτων.

- Και πως αντιμετωπίζουμε αυτό το ζήτημα; Πως μειώνουμε το σφάλμα στις περιπτώσεις που εργαζόμαστε με μη παραμετρικές διαδικασίες;

- Ένας τρόπος είναι οι προσομοιωτικές μέθοδοι (Monte Carlo, Exact, Bootstrapping κτλ. στις οποίες θα αναφερθούμε παρακάτω, αναλυτικά...

Υπάρχουν τουλάχιστον 3 γενικοί τρόποι ελέγχου της κανονικότητας:

- Με τη βοήθεια στατιστικών κριτηρίων. Για παράδειγμα, τα κριτήρια των Kolmogorov-Smirnov, Lilliefors, και Shapiro-Wilk.
- Με τη βοήθεια γραφικών αναπαραστάσεων. Για παράδειγμα, τα ονομαζόμενα από το SPSS, Normal Q-Q Plot, Detrended Q-Q Plot, και Box Plot.

Και τέλος,

- Με τη βοήθεια του λόγου t , όπου,

$$t = \frac{\text{statistic}}{\text{standard error of statistic}}$$

Ας πάρουμε και πάλι το παράδειγμα των 15 φοιτητών του Δεληβοριά και ας κάνουμε έλεγχο κανονικότητας με ένα-έναν από τους παραπάνω τρόπους ξεχωριστά.

Πρώτος τρόπος ελέγχου Κανονικότητας

Με τη βοήθεια των στατιστικών κριτηρίων Kolmogorov-Smirnov, Lilliefors, και Shapiro-Wilk.

Πίνακας 1. Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
v1	,183	15	,191	,964	15	,768

a. Lilliefors Significance Correction

Στη θέα του Πίνακα1 διατυπώνουμε τη μηδενική και την εναλλακτική μας υπόθεση...

H_0 : Η κατανομή του δείγματός μας ΔΕΝ απέχει και πολύ από την κανονική.

H_1 : Η κατανομή του δείγματός μας, απέχει πολύ από την κανονική.

ή ισοδύναμα:

H_0 : Το δείγμα μας προέρχεται από κανονικά κατανεμημένο πληθυσμό.

H_1 : Το δείγμα μας ΔΕΝ προέρχεται από κανονικά κατανεμημένο πληθυσμό.

- Ο Πίνακας 1., ουσιαστικά είναι ένας διπλός πίνακας. Το δεξί του μέρος αφορά το στατιστικό κριτήριο των Shapiro-Wilk, ενώ το αριστερό του μέρος αφορά το στατιστικό κριτήριο των Kolmogorov- Smirnov. Εμείς ωστόσο, **ποιο από τα δύο** θα χρησιμοποιήσουμε;
- Η απάντηση είναι ότι αυτό θα εξαρτηθεί από τη φύση των δεδομένων μας και από το μέγεθος του δείγματος.
- να προτείνουμε μια πολιτική για να αποφανθούμε για την ύπαρξη ή μη κανονικότητας, χωρίς να απομακρυνθούμε από την επιστημονική αλήθεια :
 - Αν το δείγμα μας έχει μέγεθος μικρότερο ή ίσο του 50 (**$n \leq 50$**), τότε το στατιστικό κριτήριο που είναι κατάλληλο να ελέγξει την ύπαρξη ή μη κανονικότητας είναι εκείνο των Shapiro-Wilk. (Carver & Nash, 2000, p.138, SPSS 6.1, Guide to Data Analysis, 1993, p. 283, Coakes, Steed, 1999, p. 30).
 - Αν το δείγμα μας έχει μέγεθος μικρότερο ή ίσο του 50 (**$n \leq 50$**), τότε το στατιστικό κριτήριο που είναι κατάλληλο να ελέγξει την ύπαρξη ή μη κανονικότητας είναι εκείνο των Shapiro-Wilk. (Carver & Nash, 2000, p.138, SPSS 6.1, Guide to Data Analysis, 1993, p. 283, Coakes, Steed, 1999, p. 30).
 - Επειδή το δείγμα μας έχει μέγεθος $n=15 < 50$, είναι σαφές ότι το κατάλληλο στατιστικό κριτήριο για τον έλεγχο κανονικότητας, είναι το κριτήριο των Shapiro-Wilk. Επομένως, το δεξί και μόνο μέρος του Πίνακα 1, μας ενδιαφέρει να σχολιάσουμε.

Εκτελούμε έλεγχο κανονικότητας σε επίπεδο στατιστικής σημαντικότητας 5%.

Καταγράφουμε και τα στοιχεία για το στατιστικό κριτήριο των Shapiro –Wilk:

S-W=0,964

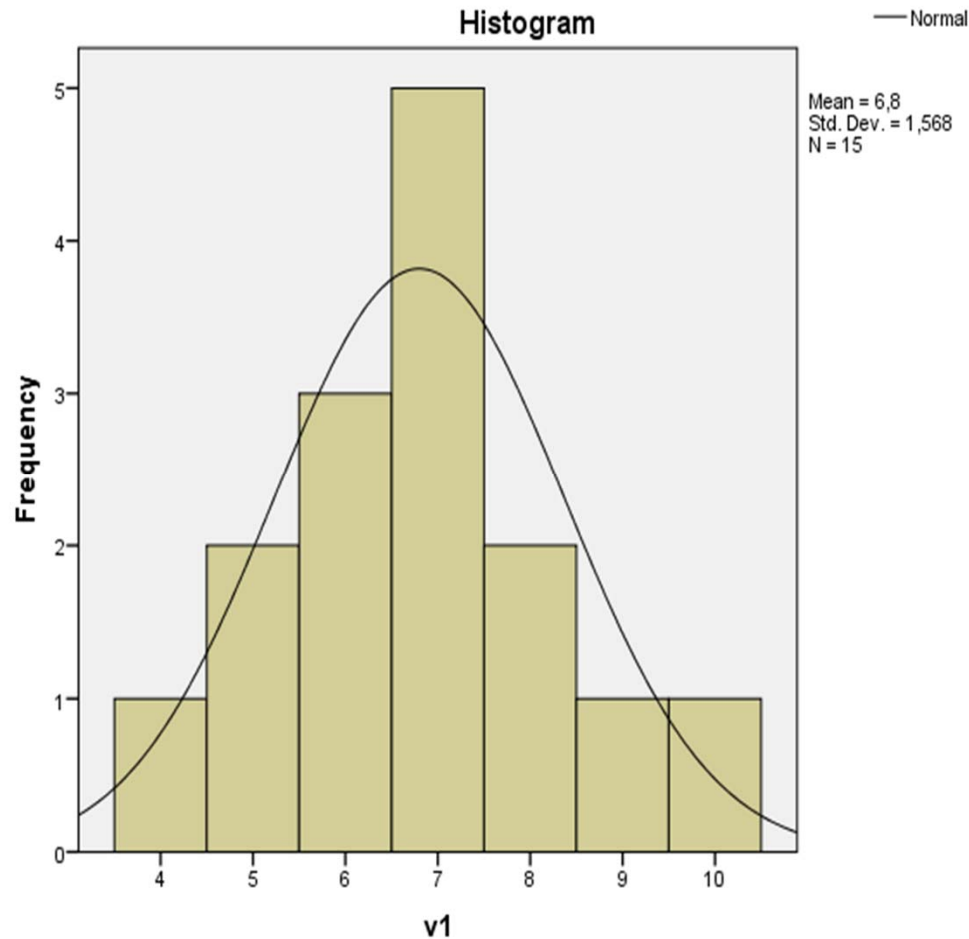
Df=15

Sign.=0,768=76,8 % > 5 %

Συλλογισμός-Συμπέρασμα:

Επειδή το παρατηρούμενο επίπεδο στατιστικής σημαντικότητας, αυτό δηλ. που μας προσφέρει στο πιάτο το SPSS, είναι sign.= 76,8 % > 5 % (=πιθανοθεωρητικό όριο που εμείς θέτουμε για να κρίνουμε τη μηδενική μας υπόθεση), οδηγούμαστε στο συμπέρασμα, ότι δεν διαθέτουμε αρκετές πληροφορίες, ώστε να απορρίψουμε τη μηδενική μας υπόθεση. Επομένως αυτή ισχύει. Με άλλα λόγια, η κατανομή του πληθυσμού από τον οποίο προέρχεται το δείγμα μας είναι, προσεγγιστικά, κανονική.

Δεύτερος τρόπος ελέγχου Κανονικότητας: **με τη βοήθεια γραφικών αναπαραστάσεων**

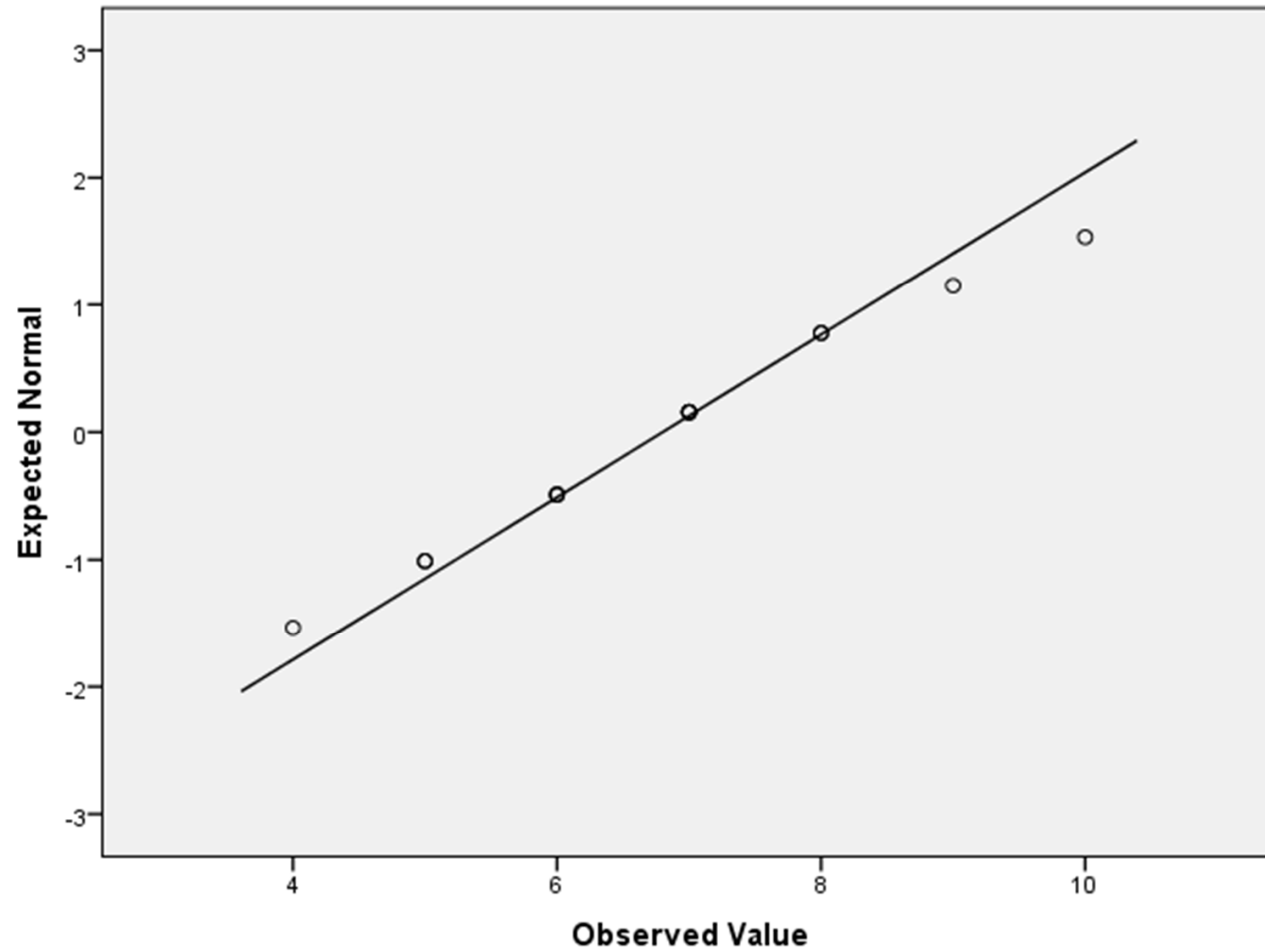


v1 Stem-and-Leaf Plot

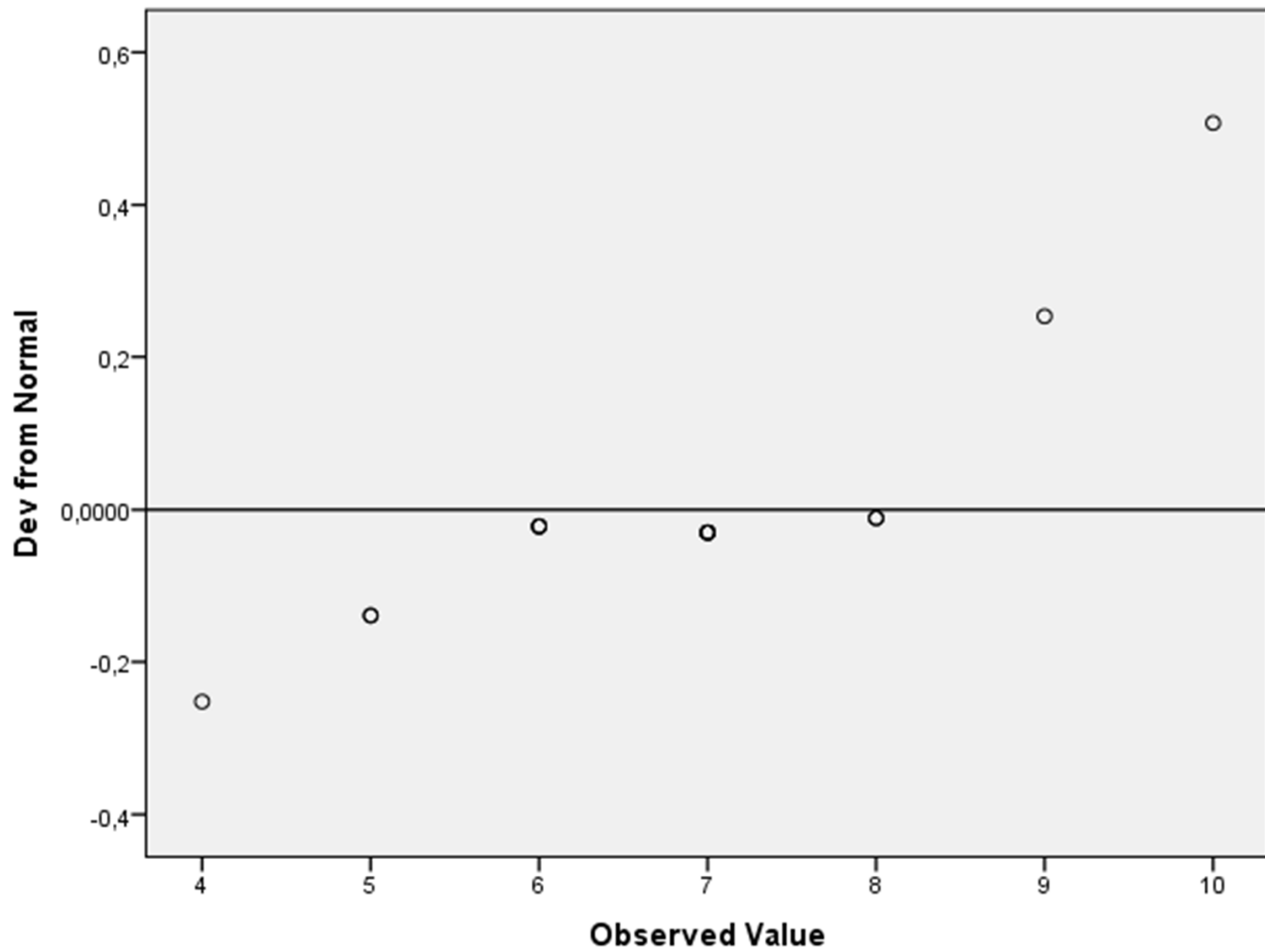
Frequency	Stem & Leaf
1,00	4 . 0
2,00	5 . 00
3,00	6 . 000
5,00	7 . 00000
2,00	8 . 00
1,00	9 . 0
1,00	Extremes (>=10,0)

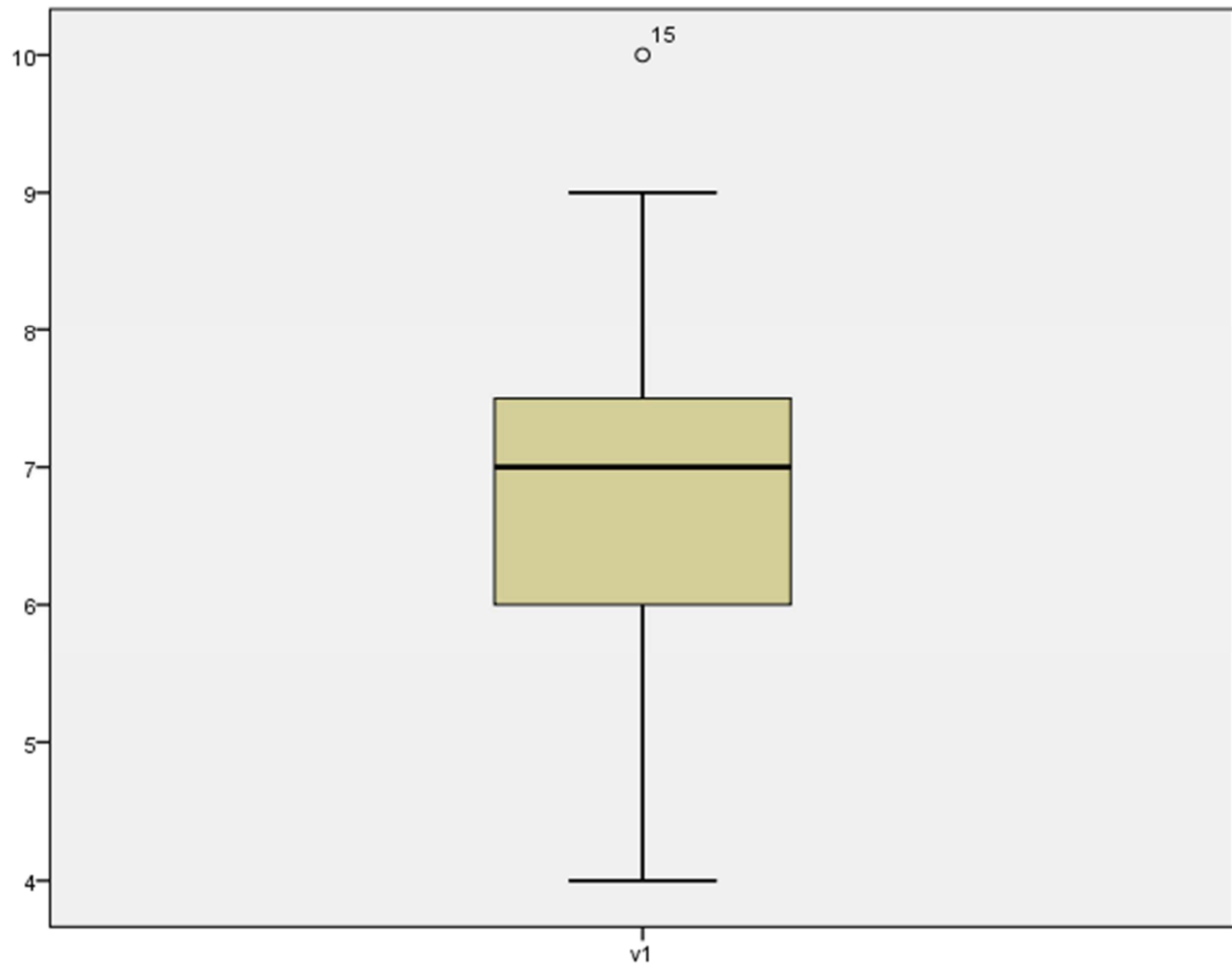
Stem width: 1
Each leaf: 1 case(s)

Normal Q-Q Plot of v1



Detrended Normal Q-Q Plot of v1





Extreme Values				
			Case Number	Value
v1	Highest	1	15	10
		2	14	9
		3	6	8
		4	13	8
		5	4	7 ^a
	Lowest	1	7	4
		2	8	5
		3	1	5
		4	9	6
		5	3	6 ^b
a. Only a partial list of cases with the value 7 are shown in the table of upper extremes.				
b. Only a partial list of cases with the value 6 are shown in the table of lower extremes.				

Τρίτος τρόπος ελέγχου κανονικότητας: Με τη βοήθεια του λόγου t, όπου,

$$t = \frac{\text{statistic}}{\text{standard error of statistic}}$$

Statistics

v1		
N	Valid	15
	Missing	0
Mean		6,80
Skewness		,252
Std. Error of Skewness		,580
Kurtosis		,165
Std. Error of Kurtosis		1,121

Πρακτικός κανόνας από τη Θεωρία:

Εάν ισχύει $-2 \leq t \leq +2 \Rightarrow$ Υπάρχει Κανονικότητα

Αντίθετα αν

$t < -2$ ή αν $t > -2$
τότε \Rightarrow ΔΕΝ υπάρχει Κανονικότητα.

Εδώ έχουμε για τη στατιστική που ονομάζεται στρεβλότητα:

$$t = \frac{\textit{skewness}}{\textit{st.error _ of _ skewness}} = \frac{0.252}{0.580} = 0.435$$

Αυτό το νούμερο 0.435 είναι προφανώς πιο μεγάλο από το -2 και
Ταυτόχρονα πιο μικρό από το +2.

Δηλ. παίζει μπάλα ανάμεσα σε αυτά τα όρια που προσδιορίσαμε για να έχουμε κανονικότητα.

Η αγωνία του τερματοφύλακα πριν από το πέναλτι....

Προφανώς έχουμε κανονικότητα...

Τέλος εισήγησης...

Τέλος Ενότητας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

