ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΣΤΟ SPSS

ΚΕΦΑΛΑΙΟ 7

Αντώνης Κ. Τραυλός (B.A., M.A., Ph.D.)
Καθηγητής
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ
Σχολή Επιστημών Ανθρώπινης Κίνησης και Ποιότητας Ζωής
Τμήμα Οργάνωσης και Διαχείρισης Αθλητισμού

- Μια μεταβλητή για να αναλυθεί παραμετρικά (δηλαδή μέσω εκτίμησης πληθυσμιακών παραμέτρων από κατανομή γνωστής μορφής, μέσου και τυπικής απόκλισης), θα πρέπει να ισχύουν οι εξής **βασικές προϋποθέσεις**, που. τίθενται συνήθως ως στατιστικές παραδοχές (statistical assumptions):

- (i) η μεταβλητή να έχει μετρηθεί στη **διαστημική κλίμακα (interval scale)**, δηλαδή να αποδίδει ποσοτικά δεδομένα (quantitative data),

- (ii) να έχει **ομοιογένεια διασποράς (homogeneity of variance)** σε εξαρτημένα ή ανεξάρτητα δείγματά της (samples) ή ομάδες της (groups),

- (iii) να μην περιέχει **ακραίες τιμές (outliers)**, δηλαδή πολύ μικρές ή πολύ μεγάλες σε σχέση με τις υπόλοιπες τιμές της κατανομής και
- (iv) να **είναι κανονική (normal)**, δηλαδή η κατανομή της να είναι συμμετρική (symmetrical) και μεσόκυρτη (mesokurtic).

Ωστόσο,

- Αν η μεταβλητή έχει χοντρικά, έστω, τα χαρακτηριστικά αυτά, τότε μπορεί να εκπροσωπηθεί από **τον μέσο (M) και την τυπική απόκλιση (s)** και να χρησιμοποιηθεί σε:
  1. σε συσχετίσεις (correlations) της με άλλες μεταβλητές ή
  2. σε συγκρίσεις (comparisons) δειγμάτων της (samples).
- Αν, όμως, η μεταβλητή ανήκει **στη διατακτική (ordinal) ή στην ονομαστική (nominal) κλίμακα** τα δεδομένα της είναι **ποιοτικά (qualitative)** και μπορεί να εκπροσωπηθεί από τον **διάμεσο (median) ή την κορυφή (mode)**, αντίστοιχα.
- Τότε, κατάλληλη στατιστική ανάλυση είναι η **μη παραμετρική (non parametric)**, δηλαδή ελεύθερη κατανομής (distribution free).
- Το ίδιο ισχύει αν η μεταβλητή είναι μεν ποσοτική, αλλά αποκλίνει σαφώς από τις παραδοχές αυτές, εκτός αν μετασχηματισθεί (transformed), για να απαλλαγεί από ακραίες τιμές και τυχόν ασυμμετρία κατανομής.

Μέχρι τώρα …

- Κάναμε χρήση της Διαδικασίας "Frequencies"

```
SPSS: ANALYZE >> DESCRIPTIVE STATISTICS >> FREQUENCIES
FREQUENCIES VARIABLES=ΚΟΙΛΙΑΚΟΙ /NTILES=4
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM
MAXIMUM SEMEAN MEAN MEDIAN MODE SKEWNESS
SESKEW KURTOSIS SEKURT
  /HISTOGRAM NORMAL /ORDER=VARIABLE.
```

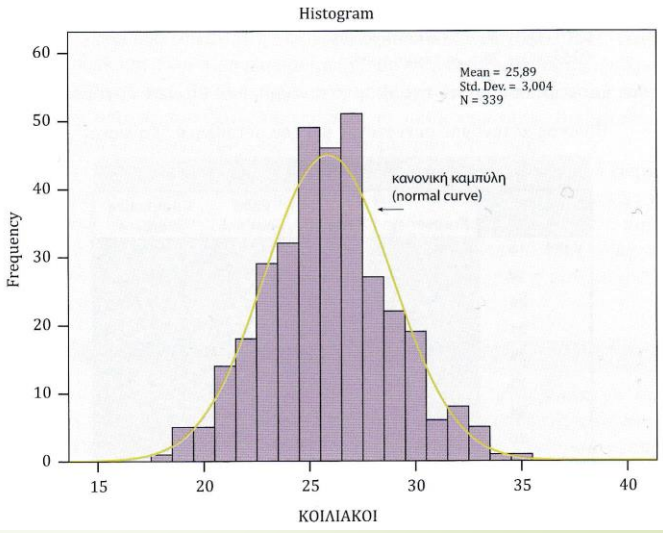
Statistics

ΚΟΙΛΙΑΚΟΙ

N	Valid	339
	Missing	0
Mean		25,89
Std. Error of Mean		,163
Median		26,00
Mode		27
Std. Deviation		3,004
Variance		9,023
Skewness		,136
Std. Error of Skewness		,132
Kurtosis		,047
Std. Error of Kurtosis		,264
Range		17
Minimum		18
Maximum		35
Percentiles	25	24,00
	50	26,00
	75	28,00

ΚΟΙΛΙΑΚΟΙ

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18	1	,3	,3	,3
	19	5	1,5	1,5	1,8
	20	5	1,5	1,5	3,2
	21	14	4,1	4,1	7,4
	22	18	5,3	5,3	12,7
	23	29	8,6	8,6	21,2
	24	32	9,4	9,4	30,7
	25	49	14,5	14,5	45,1
	26	46	13,6	13,6	58,7
	27	51	15,0	15,0	73,7
	28	27	8,0	8,0	81,7
	29	22	6,5	6,5	88,2
	30	19	5,6	5,6	93,8
	31	6	1,8	1,8	95,6
	32	8	2,4	2,4	97,9
	33	5	1,5	1,5	99,4
	34	1	,3	,3	99,7
	35	1	,3	,3	100,0
	Total	339	100,0	100,0	



Διαδικασία **Explore**

Επόμενο εργαστήριο

SPSS: ANALYZE >> DESCRIPTIVE STATISTICS >> EXPLORE

```
EXAMINE VARIABLES=ΚΟΙΛΙΑΚΟΙ
 /PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
 /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
 /STATISTICS DESCRIPTIVES EXTREME /CINTERVAL 95.
```

Περιγραφικά στατιστικά

Descriptives

			Statistic	Std. Error
ΚΟΙΛΙΑΚΟΙ	Mean		25,89	,163
	95% Confidence Interval for Mean	Lower Bound	25,57	
		Upper Bound	26,21	
	5% Trimmed Mean		25,85	
	Median		26,00	
	Variance		9,023	
	Std. Deviation		3,004	
	Minimum		18	
	Maximum		35	
	Range		17	
	Interquartile Range		4	
	Skewness		,136	,132
	Kurtosis		,047	,264

- Ο πίνακας "Descriptives" παραθέτει τα περιγραφικά στατιστικά της μεταβλητής.
- **Mean** (μέσος, M) = ΣX / N = 25.89.
- **Standard error** (of mean) (τυπικό σφάλμα του μέσου) SEM =0.163.
- **95% Confidence Interval for Mean** (διάστημα εμπιστοσύνης 95% για τον μέσο):
  - Mean ± SE_{mean} * $t_{(0, 95)}$ = 25.89 ± 0.163 * 1.968 = 25.89 ± 0.321
- **lower bound** (κατώτερο όριο) = 25.57, **upper bound** (ανώτερο όριο) = 26.21.
- Η τιμή $t_{(0, 95)}$ = 1.968 βρίσκεται από το παράρτημα Δ με δίπλευρο έλεγχο στο α=0.05 και βαθμούς ελευθερίας df = N - 1 = 339 - 1 = 338 ~ 300.
- Η ερμηνεία του "95% CI" είναι η εξής: με πιθανότητα 95% ο πληθυσμιακός μέσος (μ) "πέφτει" στα όρια 25.57 - 26.21, με την πιθανότητα αυτή να είναι ελάχιστη στα δύο άκρα (όρια) και μέγιστη στο κέντρο της κατανομής, που αντιστοιχεί ο μέσος 25.89.

Περιγραφικά στατιστικά

Descriptives

			Statistic	Std. Error
ΚΟΙΛΙΑΚΟΙ	Mean		25,89	,163
	95% Confidence Interval for Mean	Lower Bound	25,57	
		Upper Bound	26,21	
	5% Trimmed Mean		25,85	
	Median		26,00	
	Variance		9,023	
	Std. Deviation		3,004	
	Minimum		18	
	Maximum		35	
	Range		17	
	Interquartile Range		4	
	Skewness		,136	,132
	Kurtosis		,047	,264

- **5% trimmed mean** = (βελτιωμένος) μέσος της κατανομής = 25.85, που υπολογίστηκε αφού αφαιρέθηκαν 5% των τιμών από το κατώτερο και 5% των τιμών από το ανώτερο άκρο της (εύρωστο στατιστικό).
  - Έτσι, **ο μέσος αυτός είναι ανεπηρέαστος από τις ακραίες (πολύ μικρές και πολύ μεγάλες) τιμές της κατανομής**.
- Παρατηρούμε ότι ο μέσος "5% trimmed mean" = 25.85 είναι σχεδόν ίσος με τον μέσο M = 25.89 και αυτό οφείλεται στο γεγονός ότι η κατανομή είναι συμμετρική και ο υπολογισμός των 2 μέσων δεν επηρεάστηκε από τις λίγες και συμμετρικά κατανεμημένες ακραίες τιμές που παρατίθενται στον πίνακα "Extreme Values".

Εκτιμητών-Μ (M-estimators)

M-Estimators

	Huber's M-Estimator[a]	Tukey's Biweight[b]	Hampel's M-Estimator[c]	Andrews' Wave[d]
ΚΟΙΛΙΑΚΟΙ	25,88	25,87	25,87	25,87

a. The weighting constant is 1.339.

b. The weighting constant is 4.685.

c. The weighting constants are 1.700, 3.400, and 8.500

d. The weighting constant is 1.340*pi.

- Όταν η κατανομή είναι συμμετρική, οι 4 Μ-εκτιμητές δίνουν σχεδόν ίδια τιμή Μ.
- Από τον πίνακα "Descriptives" βλέπουμε ότι ο (αριθμητικός) μέσος είναι Μ = 25.89. Οι εκτιμήσεις των 4 M-estimators είναι σχεδόν ίδιες (~ 25.9) με τον μέσο, επειδή η κατανομή είναι συμμετρική.
- Σε **περιπτώσεις ασύμμετρων κατανομών (skewed distributions)** κατάλληλος εκτιμητης-Μ είναι, εκτός του 5% trimmed mean, **αυτός του Huber**, που πετυχαίνει καλύτερη εκτίμηση του πληθυσμιακού μέσου (μ). Οι άλλοι 3 Μ-εκτιμητές σε ασύμμετρες κατανομές δίνουν καλύτερη εκτίμηση του πληθυσμιακού διάμεσου (δμ).

Βήματα στην αξιολόγηση της κανονικότητας της κατανομής:

Extreme Values

			Case Number	Value
ΚΟΙΛΙΑΚΟΙ	Highest	1	279	35
		2	228	34
		3	128	33
		4	202	33
		5	227	33[a]
	Lowest	1	66	18
		2	77	19
		3	74	19
		4	63	19
		5	30	19[b]

a. Only a partial list of cases with the value 33 are shown in the table of upper extremes.

b. Only a partial list of cases with the value 19 are shown in the table of lower extremes.

Στον πίνακα "Extreme Values" οι 5 μεγαλύτερες (highest) και οι 5 μικρότερες (lowest) τιμές τής κατανομής.

Οι 5 μεγαλύτερες τιμές είναι 33, 33, 33, 34, 35 και αντιστοιχούν στα άτομα (case) 227, 202, 128, 228, 279.

Οι 5 μικρότερες τιμές είναι 18, 19, 19, 19, 19 και ανήκουν στα άτομα (case) 66, 77, 74, 63, 30.

Case number = περίπτωση (άτομο), Value = τιμή.

1. Ιστόγραμμα & Φυλλόγραμμα,
2. Θηκόγραμμα,
3. **Λοξότητα & Κύρτωση**,
4. Έλεγχος Shapiro-Wilk με σημαντικότητα στο 0.01 ή στο 0.001 ανάλογα με το μέγεθος δείγματος (N).

- Ο πίνακας αυτός περιλαμβάνει πάντα τις 5 ανώτερες και τις 5 κατώτερες τιμές, χωρίς να αξιολογεί αν είναι απόμακρες (outliers) ή ακραίες (extreme).
- Το SPSS αξιολογεί το πόσο ακραία είναι κάποια τιμή X με βάση τα εξής κριτήρια:

  α) αν X > Q3 + 1.5 * (Q3 - Q1) ή X < Q1 - 1.5 * (Q3 - Q1), τότε η τιμή **X** ορίζεται ως **απόμακρη (outlier)**

  β) αν X > Q3 + 3 * (Q3 - Q1) ή X < Q1 - 3 * (Q3 - Q1), τότε η τιμή **X** ορίζεται ως **απόμακρη (outlier)**.

- Στον πίνακα "**Percentiles**" βλέπουμε ότι Q3 = 28 & Q1 = 24, οπότε:

  Q3 + 1.5 * (Q3 - Q1) = 28 + 1.5 * (28 - 24) = **34** και Q1 - 1.5 * (Q3 - Q1) = 24 - 1.5 * (28-24)= **18**,

  δηλαδή, κάθε τιμή **X > 34** ή **X < 18** αποτελεί **μια απόμακρη τιμή (outlier)**,

  Q3 + 3 * (Q3 - Q1) = 28 + 3 * (28 - 24) = **40** και Q1 - 3 * (Q3 - Q1) = 24 - 3 * (28 - 24) = **12**,

  δηλαδή, κάθε τιμή **X > 40** ή **X < 12** αποτελεί **μια ακραία τιμή (extreme)**.
- **Υπάρχει μία μόνο απόμακρη (>34) τιμή, η X=35 (άτομο, case = 279)**.
- Δεν υπάρχουν ακραίες τιμές (> 40 ή < 12) και αυτό φαίνεται και στο θηκόγραμμα (boxplot), όπου αποτυπώνεται μόνο η απόμακρη τιμή X = 35 (case = 279).

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average (Definition 1)	ΚΟΙΛΙΑΚΟΙ	21,00	22,00	24,00	26,00	28,00	30,00	31,00
Tukey's Hinges	ΚΟΙΛΙΑΚΟΙ			24,00	26,00	28,00		

Έλεγχος κανονικότητας κατανομής ... 1

- Ο έλεγχος κανονικότητας της κατανομής μπορεί να γίνει και πιθανολογικά μέσω των ελέγχων Kolmogorov-Smirnov και Shapiro-Wilk

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ΚΟΙΛΙΑΚΟΙ	,093	339	,000	,987	339	,004

a. Lilliefors Significance Correction

- Οι έλεγχοι αυτοί συγκρίνουν την κατανομή του δείγματος με την κανονική αλλά τείνουν να **δίνουν κανονικότητα σε μικρά δείγματα (π.χ. N < 30)** και **μη κανονικότητα σε μεγάλα δείγματα (π.χ. N > 100)**. Για τον λόγο αυτό η σημαντικότητά τους μπορεί να ελεγχθεί στο α = 0.01 ή ακόμα και στο α = 0.001 ή αν το δείγμα είναι πολύ μεγάλο, μπορεί η αξιολόγηση της κατανομής να γίνει με το ιστογράμμα και τα μέτρα λοξότητας και κύρτωσης.
- Ο **έλεγχος κανονικότητας Shapiro-Wilk** είναι ισχυρότερος του **Kolmogorov-Smimov** για **δείγματα N < 50** και είναι γενικά προτιμητέος (Razali & Wah, 2011).

Έλεγχος κανονικότητας κατανομής ... 2

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ΚΟΙΛΙΑΚΟΙ	,093	339	,000	,987	339	,004

a. Lilliefors Significance Correction

- Με βάση τα αποτελέσματα των 2 αυτών ελέγχων θα έπρεπε να συμπεράνουμε ότι η κατανομή δεν είναι κανονική (συμμετρική και μεσόκυρτη), επειδή και οι δύο έλεγχοι έδειξαν ότι υπάρχει στατιστικώς σημαντική διαφορά μεταξύ της δειγματικής κατανομής και του υποδείγματος της κανονικής κατανομής (normal curve).
- Όμως, **στο παράδειγμά μας έχουμε ένα αρκετά μεγάλο δείγμα (N = 339)** και, έτσι, οι 2 αυτοί έλεγχοι μπορούν να **αγνοηθούν** και να βασίσουμε την αξιολόγηση της κανονικότητας της κατανομής σε άλλα σχετικά στατιστικά, όπως στο **ιστόγραμμα (το οποίο είναι συμμετρικό)** και στις τιμές της **λοξότητας και της κύρτωσης σε σχέση με τα τυπικά τους σφάλματα**, που όπως είδαμε ήδη συνιστούν κανονικότητα.

Γράφημα Stem-and-Leaf (στέλεχος-&-φύλλο)

```
ΚΟΙΛΙΑΚΟΙ Stem-and-Leaf Plot

 Frequency
          Stem &  Leaf
     1.00   18 .  0
     5.00   19 .  00000
     5.00   20 .  00000
    14.00   21 .  00000000000000
    18.00   22 .  000000000000000000
    29.00   23 .  00000000000000000000000000000
    32.00   24 .  00000000000000000000000000000000
    49.00   25 .  0000000000000000000000000000000000000000000000000000
    46.00   26 .  00000000000000000000000000000000000000000000000000
    51.00   27 .  000000000000000000000000000000000000000000000000000000
    27.00   28 .  000000000000000000000000000
    22.00   29 .  0000000000000000000000
    19.00   30 .  000000000000000000
     6.00   31 .  000000
     8.00   32 .  00000000
     5.00   33 .  00000
     1.00   34 .  0
     1.00 Extremes    (>=35.0)

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

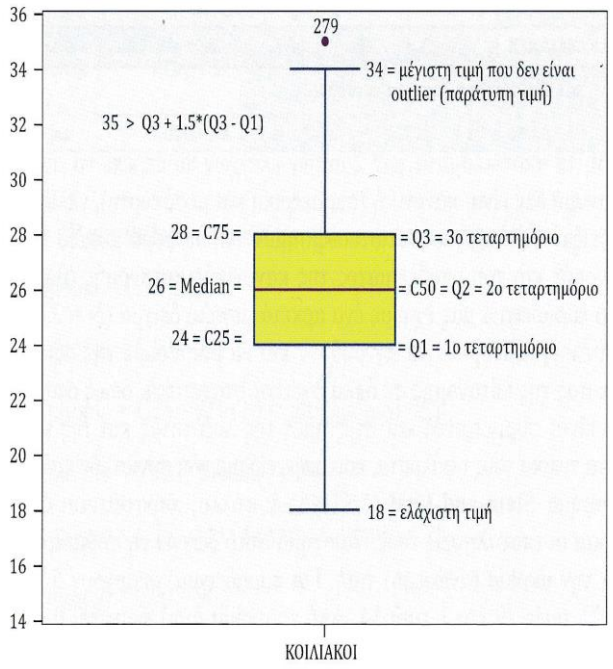
- Στο γράφημα Stem-and-Leaf (στέλεχος-&-φύλλο) παρατίθενται όλες οι αρχικές τιμές και οι επαναλήψεις τους. Το σχήμα αυτό δείχνει τη συμμετρία της κατανομής και την ακραία (ανώτερη) τιμή.
- Για παράδειγμα, υπάρχουν 5 τιμές 20, 32 τιμές 24, 22 τιμές 29 και 1 τιμή 34. Από το σχήμα αυτό φαίνεται η κορυφή (27) καθότι έχει συχνότητα f = 51 (όπως φαίνεται και από το μέγιστο μήκος της αντίστοιχης στήλης).
- Το γράφημα αυτό έχει όλα τα στοιχεία του ιστογράμματος και επιπλέον δίνει όλες τις τιμές σε διαδοχικές ομάδες και τυχόν ακραίες τιμές (extremes) με βάση τα κριτήρια που αναπτύχθηκαν πιο πάνω.

ΚΟΙΛΙΑΚΟΙ Stem-and-Leaf Plot

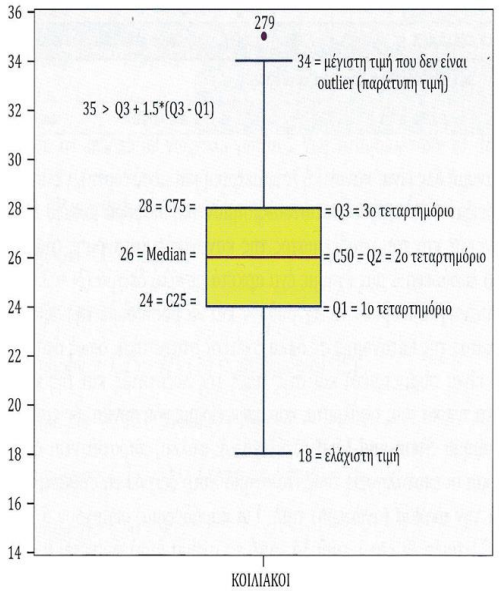
```
Frequency
        Stem &  Leaf
    1.00   18 .  0
    5.00   19 .  00000
    5.00   20 .  00000
   14.00   21 .  00000000000000
   18.00   22 .  000000000000000000
   29.00   23 .  00000000000000000000000000000
   32.00   24 .  00000000000000000000000000000000
   49.00   25 .  0000000000000000000000000000000000000000000000000000000
   46.00   26 .  0000000000000000000000000000000000000000000000000000
   51.00   27 .  000000000000000000000000000000000000000000000000000000(
   27.00   28 .  000000000000000000000000000
   22.00   29 .  0000000000000000000000
   19.00   30 .  0000000000000000000
    6.00   31 .  000000
    8.00   32 .  00000000
    5.00   33 .  00000
    1.00   34 .  0
    1.00 Extremes    (>=35.0)

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

Θηκόγραμμα (boxplot) της κατανομής των 339 τιμών.



Θηκόγραμμα (Boxplot)

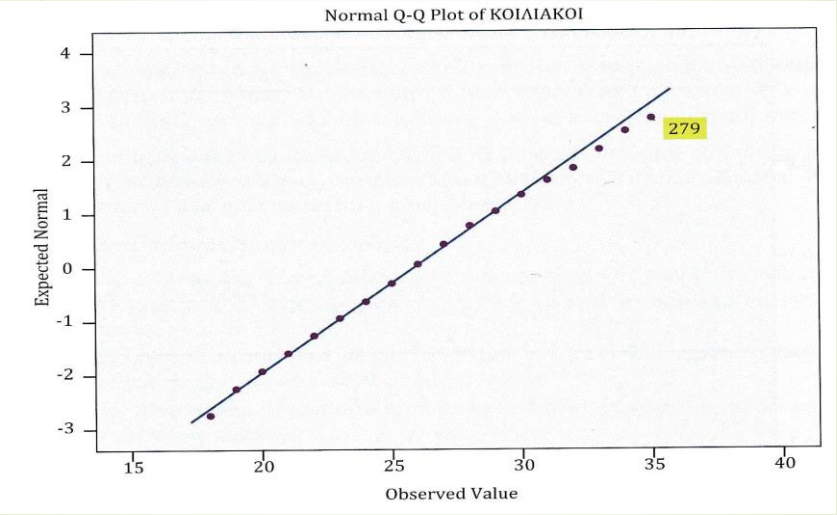


- Υπάρχει μια απόμακρη (outlier) τιμή ίση με 35, που απέχει πάνω από ενάμιση μήκος θήκης από το 3ο τεταρτημόριο Q3 = 34, και ανήκει στο άτομο (case) 279

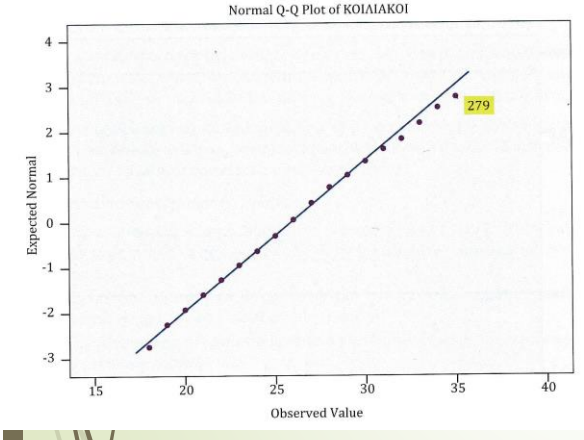
  35 > Q3 + 1.5 * (Q3 - Q1) = 28 + (1.5 * (28 - 24)) = 28 + (1.5 * 4) = 28 + 6 = **34**

- Το 3ο τεταρτημόριο (Q3, 3st quartile) είναι η τιμή 28 και διαχωρίζει το πρώτο 75% των τιμών (C75) από το υπόλοιπο 25% των τιμών της κατανομής.
- Ο διάμεσος (Median) είναι η τιμή 26 χωρίζει την κατανομή σε δύο ίσα τμήματα από 50% των τιμών σε κάθε ένα.
- Το 1ο τεταρτημόριο (Q1, 1st quartile) είναι η τιμή 18 και διαχωρίζει το πρώτο 25% των τιμών (C25) από το υπόλοιπο 75% των τιμών της κατανομής.
- Ο διάμεσος (26) χωρίζει τη θήκη (πλαίσιο) σε δύο σχεδόν ίσα μισά.
- Υπάρχει ίση απόσταση από το διάμεσο (26) των 2 whisker (2 οριζόντιων γραμμών), δηλαδή της μεγαλύτερης (34) και της μικρότερης (18) τιμής που δεν είναι απόμακρη (outlier).
- Με βάση τα στοιχεία αυτά συμπεραίνουμε ότι η κατανομή είναι συμμετρική.

Γράφημα "Normal Q-Q Plot" της κατανομής των 339 τιμών.

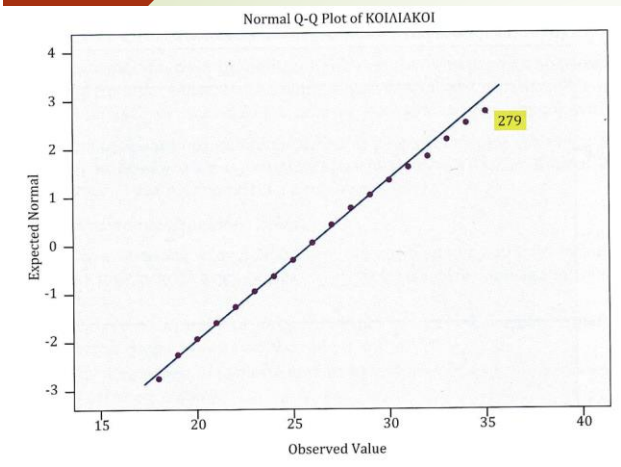


Γράφημα "Normal Q-Q Plot" της κατανομής των 339 τιμών.



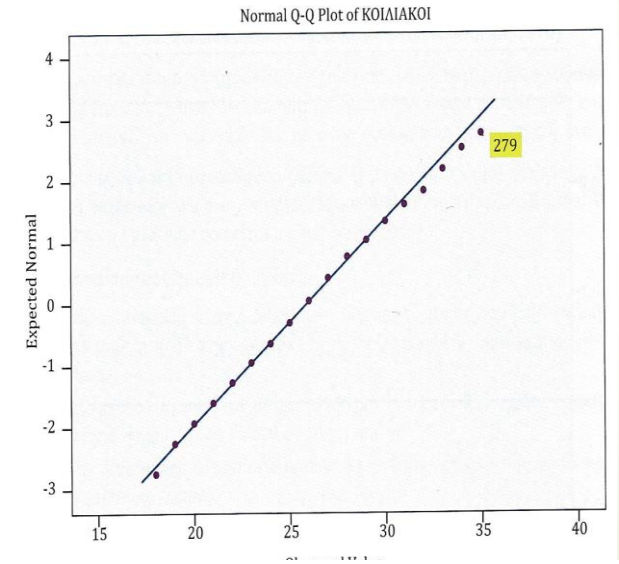
- **Quantile** (**q**): ποσοστιαία σημεία ή ποσοστημόριο
- **Quantile** (**q**) είναι η τιμή X_i που ορίζει το πρώτο q% των τιμών της κατανομής: οι N τιμές X σε σειρά μεγέθους διαιρούν την κλίμακα X σε N + 1 μέρη.
- Έτσι, η αναλογία των τιμών που πέφτουν πριν την X_i είναι η i/(N + 1) και για κάθε q η $i = q\ (N + 1)$.
- Π.χ. σε N=50 τιμές η quantile q = 0.21 είναι η X που ορίζει το πρώτο 21% των τιμών: η i = q (N + 1) = 0.21 (50 + 1) = 10.71η, δηλαδή η X μεταξύ $10^{ης}$ και $11^{ης}$ σε σειρά τιμής.

Normal Q-Q Plot



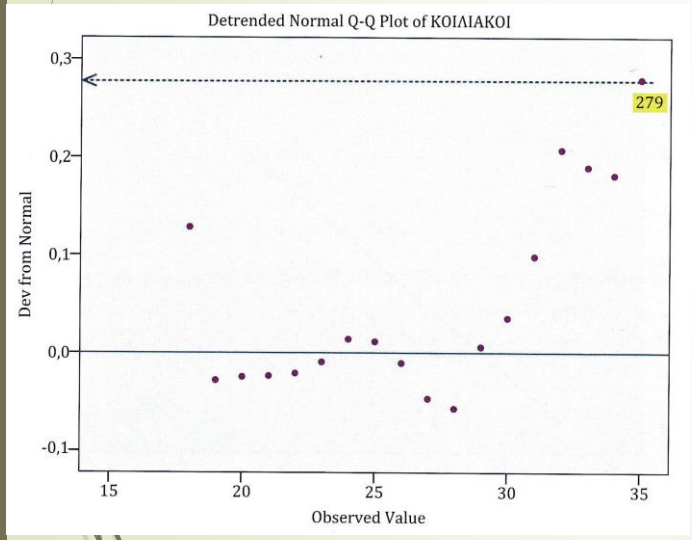
- **Normal Q-Q Plot**: *Quantile-Quantile Plot*. Τα quantiles των παρατηρήσεων υποτυπώνονται έναντι τω αντίστοιχων της κανονικής (normal expected).
- Το υποτύπωμα (plot) δίνει τα σημεία (ο) που αντιπροσωπεύουν την πραγματική κατανομή και μια ευθεία γραμμή που συμβολίζει την τέλεια κανονικότητα.
- Στην κατανομή αυτή υπάρχουν 18 διαδοχικές ακέραιες τιμές X (observed value) και έτσι **παρήχθησαν 18 αντίστοιχες quantiles που ως τυπικές τιμές z αντιπαραβάλλονται με τις τυπικές τιμές z των αντίστοιχων quantiles της κανονικής κατανομής**.

Normal Q-Q Plot



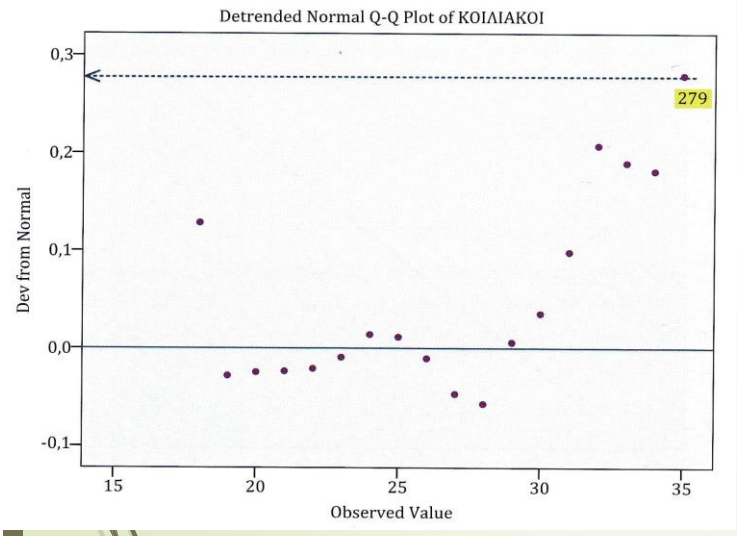
- Η ερμηνεία του γραφήματος αυτού είναι η εξής:
- Όσο τα σημεία που προέκυψαν από τις τιμές (observed values) είναι κοντά ή και πάνω στην γραμμή, τόσο η κατανομή τείνει να είναι κανονική (normal).
- Το συγκεκριμένο υποτύπωμα δείχνει ότι με εξαίρεση την περίπτωση (case) 279 (που όπως είδαμε είναι η απόμακρη τιμή 35) τα σημεία πέφτουν πολύ κοντά στη γραμμή της κανονικότητας και επομένως η κατανομή είναι κανονική.

Γράφημα "Detrended Normal Q-Q Plot" της κατανομής των 339 τιμών



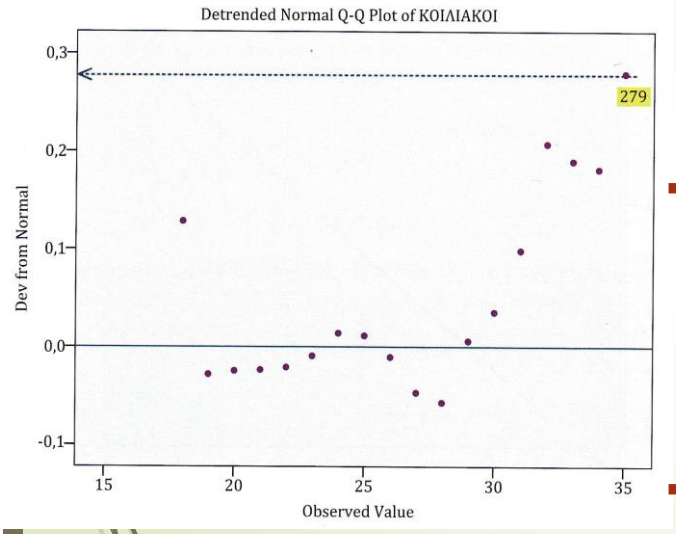
- **Detrended Normal Q-Q Plot**: Detrended *Quantile-Quantile Plot*. Οι τυποποιημένες (z) αποκλίσεις (**Dev from Normal**) των quantiles από την κανονική κατανομή υποτυπώνονται έναντι της μηδενικής απόκλισης (γραμμή).
- Το γράφημα αυτό αποτελεί προέκταση του "Normal Q-Q Plot" και αξιολογεί σε τυπικές τιμές z πόσο αποκλίνουν από την κανονικότητα οι τιμές X της κατανομής.

"Detrended Normal Q-Q Plot"

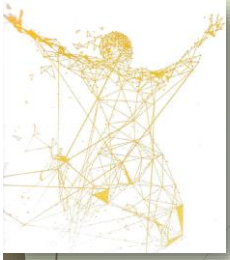


- Τα σημεία (ο) του γραφήματος αυτού προκύπτουν ως εξής: κάθε quantile X_i μετασχηματίζεται σε z-τιμή (απόκλιση / τυπική απόκλιση) και αφαιρείται από την z-τιμή της αντίστοιχης προσδοκώμενης quantile της κανονικής (expected normal).
- Οι αποκλίσεις αυτές υποτυπώνονται έναντι των αντίστοιχων αρχικών (observed).

"Detrended Normal Q-Q Plot"



- Η ερμηνεία του γραφήματος αυτού είναι η εξής: Τα σημεία που απέχουν πολύ από τη γραμμή κανονικότητας συνεισφέρουν περισσότερο στην όποια απόκλιση της συνολικής κατανομής από την κανονικότητα.
- Στο παράδειγμα των 339 τιμών λίγες μόνο τιμές αποκλίνουν από την κανονικότητα με κύρια αυτή της περίπτωσης 279 με τιμή X = 35 που απέχει ~ 0.28z από την κανονικότητα, ενώ οι άλλες απέχουν το πολύ μέχρι 0.2z και αφορούν όσες είναι > 30. Όμως, οι αποκλίσεις αυτές δεν επαρκούν για να αμφισβητηθεί η κανονικότητα της κατανομής.
- Έτσι, συνάγεται ότι η συγκεκριμένη κατανομή είναι αδρά κανονική.



Τι γίνεται σε περίπτωση μεγάλης ασυμμετρίας στην κατανομή;

- Αν σε μια ανάλυση παρατηρηθεί μεγάλη ασυμμετρία, τότε έχουμε 3 επιλογές:
  1. να μετασχηματισθούν οι τιμές X ώστε να επιτευχθεί η συμμετρία της κατανομής και να συνεχίσουμε με κάποια παραμετρική στατιστική ανάλυση, όπως π.χ. t-test, ANOVA, Pearson correlation, regression (Bland & Altman, 1996).
  2. να διερευνηθεί η εκδοχή της αφαίρεσης από την ανάλυση μερικών πολύ ακραίων τιμών (μετά από επαρκή αιτιολόγηση), ώστε να βελτιωθεί περαιτέρω η κατανομική δομή προς τη συμμετρία και μετά να γίνει κάποια παραμετρική ανάλυση.
  3. να εφαρμοσθεί στα αρχικά δεδομένα (χωρίς μετασχηματισμό) κάποια μη παραμετρική στατιστική ανάλυση, όπως π.χ. Mann-Whitney ή Wilcoxon, Kruskal-Wallis ή Friedman, Spearman correlation κ.ά (Siegel, 1956).

Σας ευχαριστώ για την προσοχή σας. … έχουμε και συνέχεια

ΑΚΤ