

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Τι είναι η στατιστική

Κύριο αντικείμενο έρευνας και μελέτης της Στατιστικής είναι η συλλογή, ταξινόμηση, επεξεργασία, παρουσίαση, ανάλυση και ερμηνεία διαφόρων δεδομένων με απώτερο στόχο την εξαγωγή ασφαλών συμπερασμάτων για λήψη ορθών αποφάσεων.

- Η Στατιστική χωρίζεται σε δύο μεγάλους τομείς:
 - την Περιγραφική Στατιστική
 - την Επαγωγική Στατιστική

Τι πρέπει να περιγράψουμε

1. Απαραίτητα το μέγεθος του δείγματος και αν το δείγμα κατανέμεται σε δύο ή περισσότερες ομάδες τότε αναφέρουμε το μέγεθος της κάθε ομάδας και με βάση ποιο ή ποια κριτήρια έγινε ο διαχωρισμός των ομάδων.
2. Απαραίτητα όλες τις σημαντικές μεταβλητές οι οποίες είναι:
 - I. Τουλάχιστον (και προτιμητέο) μια εξαρτημένη μεταβλητή που είναι η κύρια έκβαση.
 - II. Μια ή περισσότερες ανεξάρτητες μεταβλητές που ενδεχόμενα να συσχετίζονται ή να επηρεάζουν την εξαρτημένη μεταβλητή.

Πως πρέπει να περιγράψουμε τις μεταβλητές

Αυτό εξαρτάται από το είδος της μεταβλητής.

Είδη μεταβλητών

1. Συνεχείς
2. Ταξινομικές
3. Κατηγορικές

1. Συνεχείς μεταβλητές

Παραδείγματα

1. Ηλικία, ύψος, βάρος, θερμοκρασία, συγκέντρωση ασβεστίου.
2. Πάντα πρέπει να αναφέρουμε τις μονάδες μέτρησης.
3. Μια ιδανική συνεχής μεταβλητή χαρακτηρίζεται ότι έχει μηδενική τιμή αναφοράς και κλίμακα σύγκρισης.

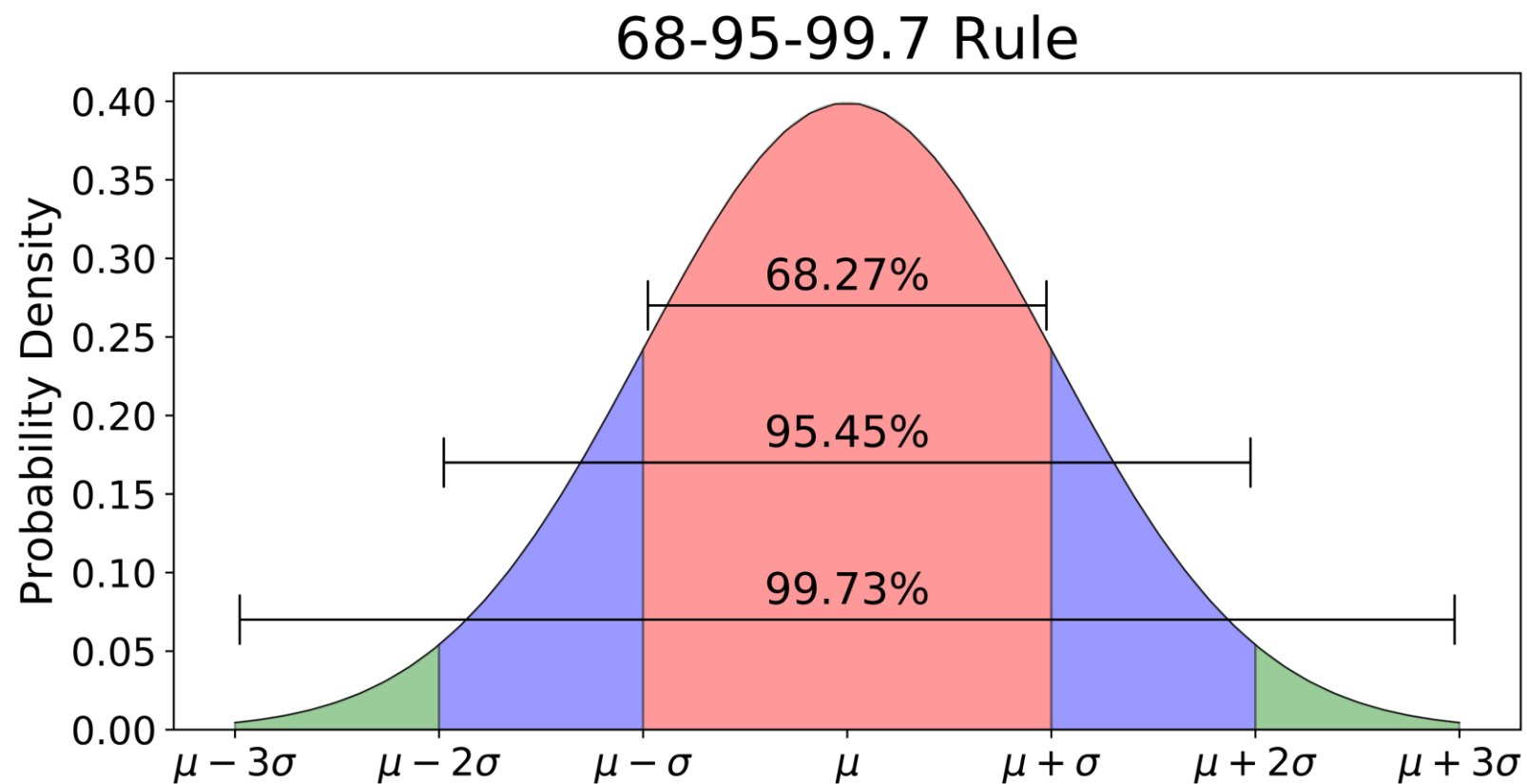
Πως περιγράφονται οι συνεχείς μεταβλητές

Πάντα πρέπει να δώσουμε δύο μέτρα

1. Μέτρο κεντρικής θέσης
2. Μέτρο διασποράς

Τα μέτρα που θα δώσουμε εξαρτώνται από το αν η μεταβλητή ακολουθεί την κανονική κατανομή (και μερικές φορές αν το δείγμα είναι αρκετά μεγάλο).

Κανονική κατανομή (The bell curve)



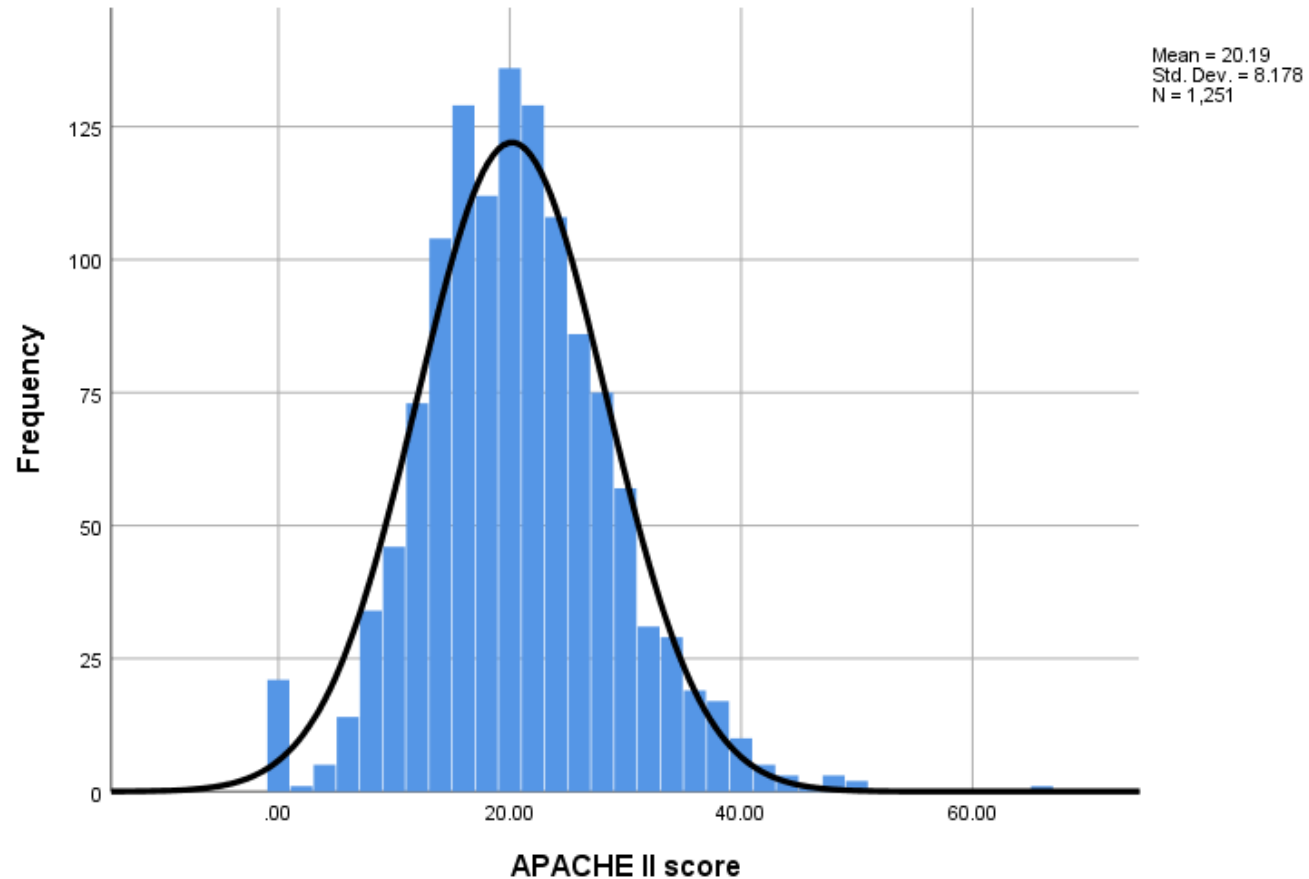
Περιγραφή συνεχούς μεταβλητής που ακολουθεί την κανονική κατανομή

1. Εφόσον έχει πραγματοποιηθεί ο στατιστικός έλεγχος και έχουμε αποδείξει ότι η μεταβλητή ακολουθεί την κανονική κατανομή.
2. Η μεταβλητή περιγράφεται με τη μέση τιμή και την τυπική απόκλιση.
3. Για παράδειγμα η ηλικία του δείγματος είναι 68.7 ± 13.1 έτη.

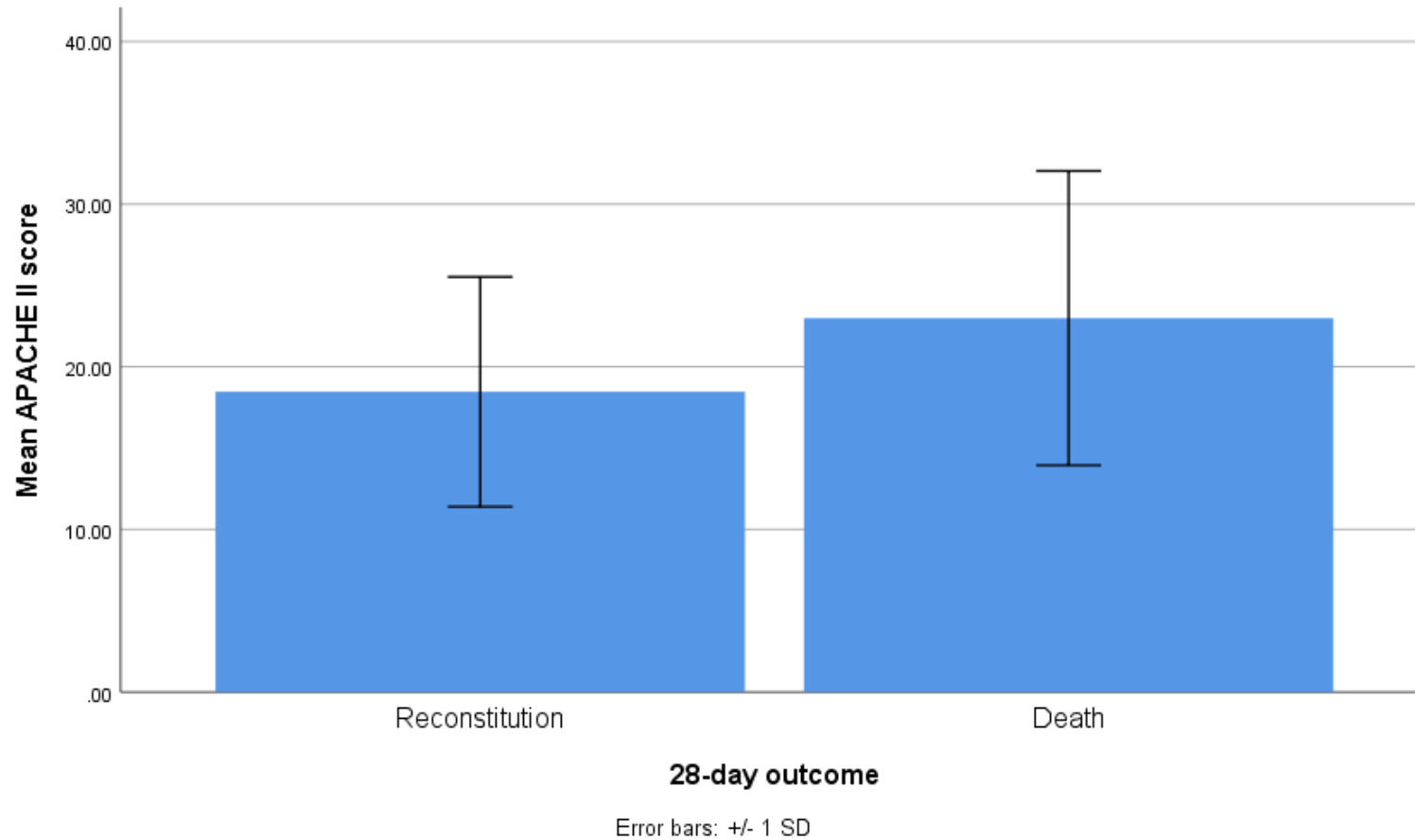
Παρατηρήσεις:

1. Η ακρίβεια με την οποία θα δώσουμε τα στατιστικά μεγέθη δεν πρέπει να υπερβαίνει την ακρίβεια με την οποία έχουμε καταγράψει τη μεταβλητή. Π.χ. αν η ηλικία έχει καταγραφεί σε ακέραιο αριθμό ετών (34, 67 κλπ.) τότε πιο σωστά η πάνω αναφορά πρέπει να είναι 69 ± 13 έτη.
2. Μερικές φορές αντί της τυπικής απόκλισης δίνονται τα 95% διαστήματα εμπιστοσύνης (ΔE – confidence intervals, CI), που προκύπτουν από τα τυπικά σφάλματα (standard errors – SE)
3. $SE=SD/\sqrt{n}$, όπου SD-τυπική απόκλιση και n-μέγεθος του δείγματος.
4. Διαστήματα εμπιστοσύνης Μέση τιμή $\pm 1.96 \times SE$
5. Για παράδειγμα αν το μέγεθος του παραπάνω δείγματος είναι 100 άτομα τότε $SE=13.1/\sqrt{100} \approx 1.3$ και $1.96 \times 1.31 =$ τότε γράφουμε ότι η μέση τιμή της ηλικίας σε έτη είναι 68.7 (95% $\Delta E = 66.1-71.3$)

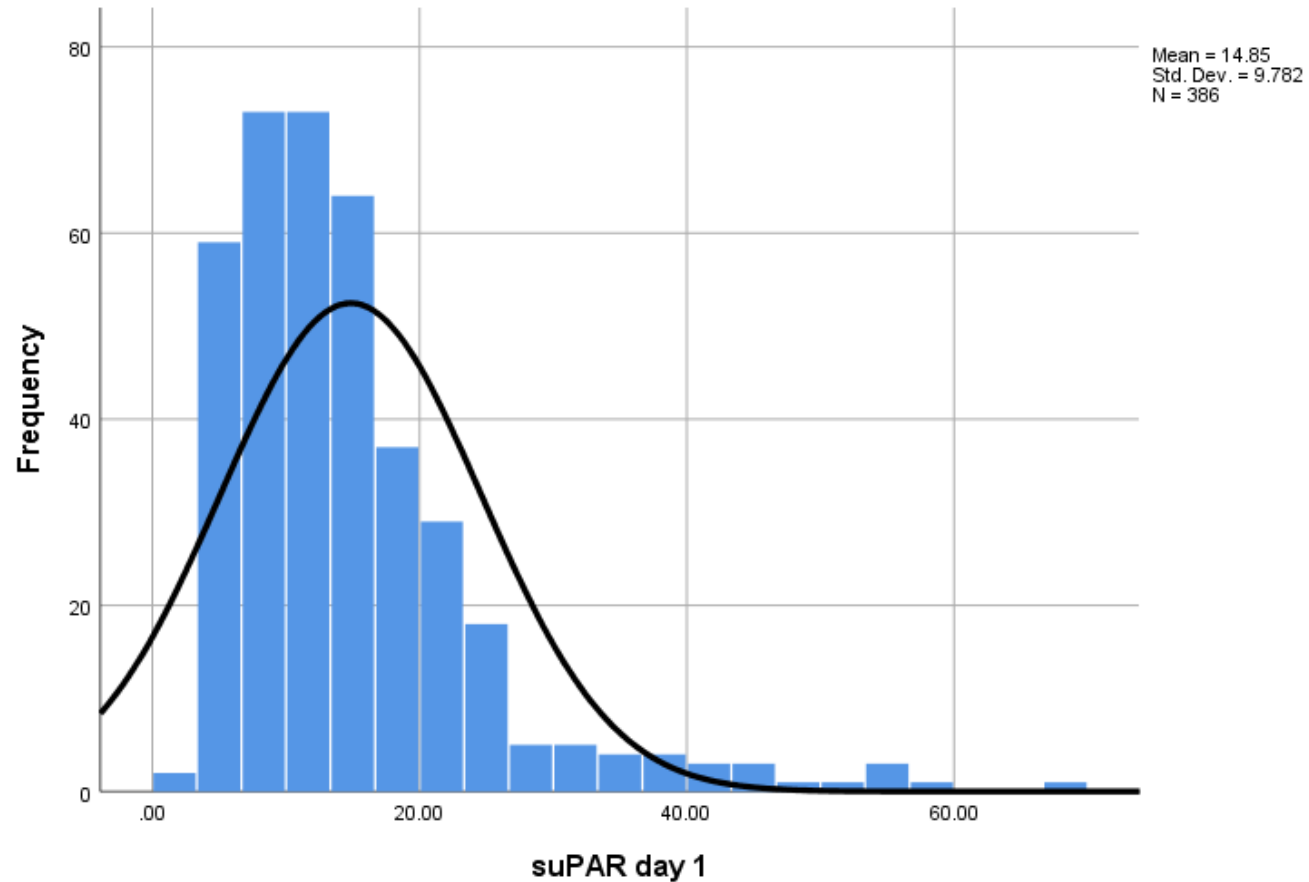
Γραφική αναπαράσταση μια συνεχούς μεταβλητής που “ακολουθεί” την κανονική κατανομή.



Γραφική παρουσίαση περιγραφικών δεδομένων για δύο ομάδες (κανονική κατανομή).



Γραφική αναπαράσταση μια συνεχούς μεταβλητής που δεν ακολουθεί την κανονική κατανομή.



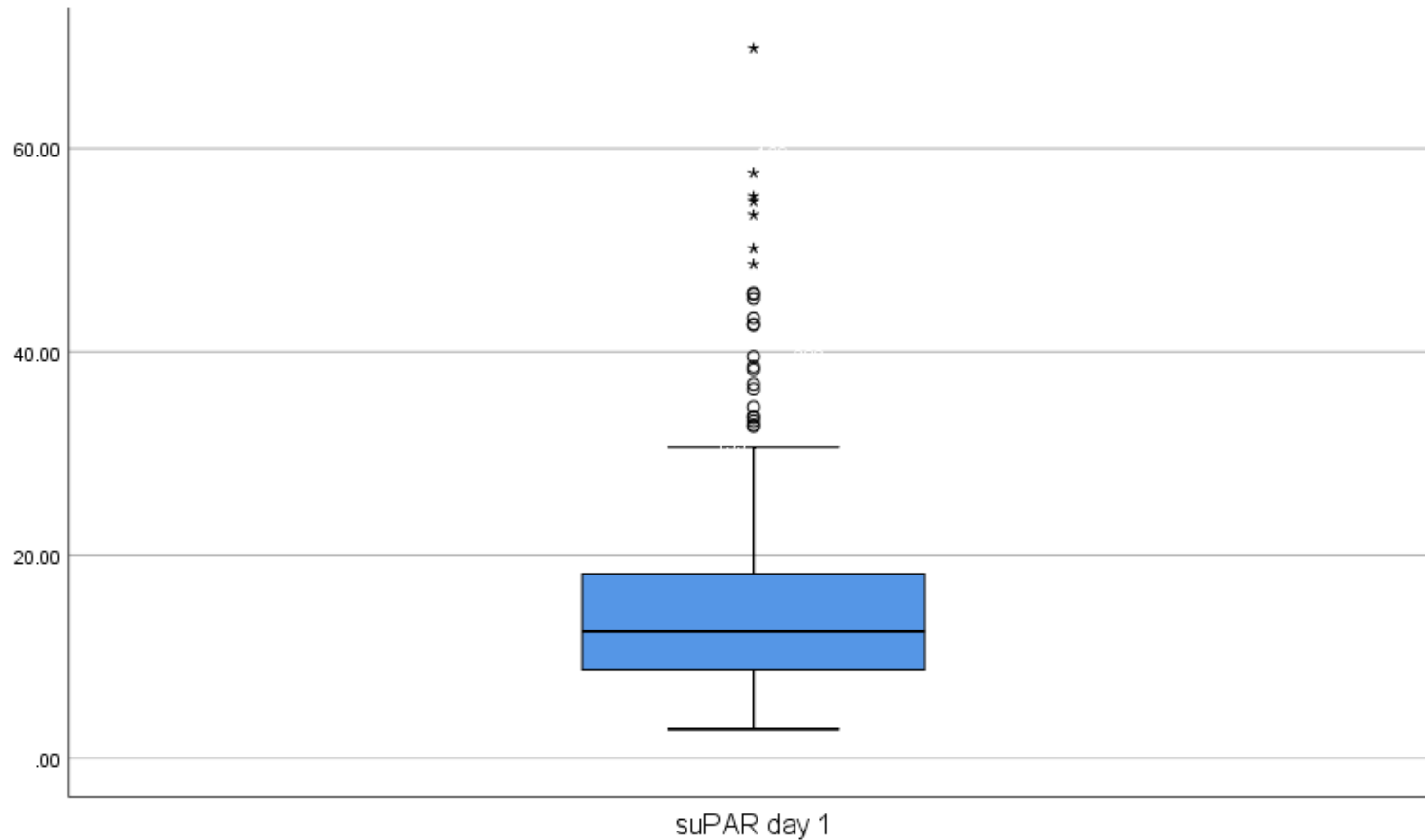
Περιγραφή συνεχούς μεταβλητής που **δεν** ακολουθεί την κανονική κατανομή

1. Εφόσον έχει πραγματοποιηθεί ο στατιστικός έλεγχος και έχουμε αποδείξει ότι η μεταβλητή δεν ακολουθεί την κανονική κατανομή.
2. Η μεταβλητή περιγράφεται με τη διάμεσο και το ενδοτεταρτημοριακό εύρος.
3. Για παράδειγμα για την τιμή suPAR του δείγματος έχουμε 12.5 (8.6–18.1) ng/ml, ή 12.5 (9.5).

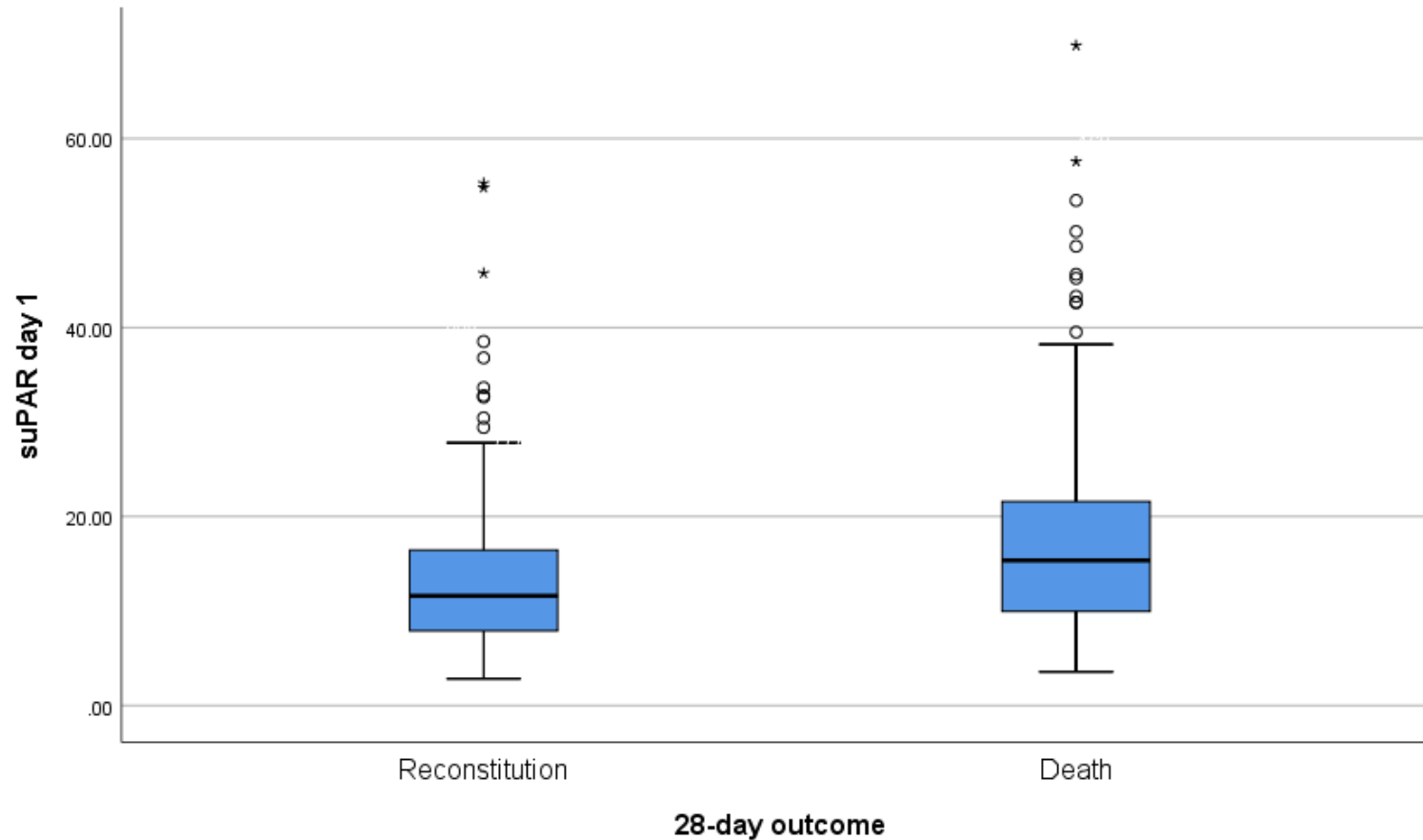
Παρατηρήσεις:

1. Σύγκριση με το 14.9 ± 9.8 ng/ml (Μέση τιμή και τυπική απόκλιση).
2. Μερικές φορές είναι χρήσιμο να δίνουμε εκτός από τη διάμεσο και τα τεταρτημόρια, τα 5% και 95% εκατοστημόρια, ή ακόμα και την ελάχιστη και μέγιστη τιμή. Εδώ, για παράδειγμα έχουμε 2.8-69.8. Έτσι μπορεί να γίνει και έλεγχος αν υπάρχουν λανθασμένες τιμές εκτός του εύρους των μετρήσεων.

Γραφική παρουσίαση μια συνεχούς μεταβλητής που δεν ακολουθεί την κανονική κατανομή.



Γραφική παρουσίαση περιγραφικών δεδομένων για δύο ομάδες (μη κανονική κατανομή).



2. Ταξινομικές μεταβλητές

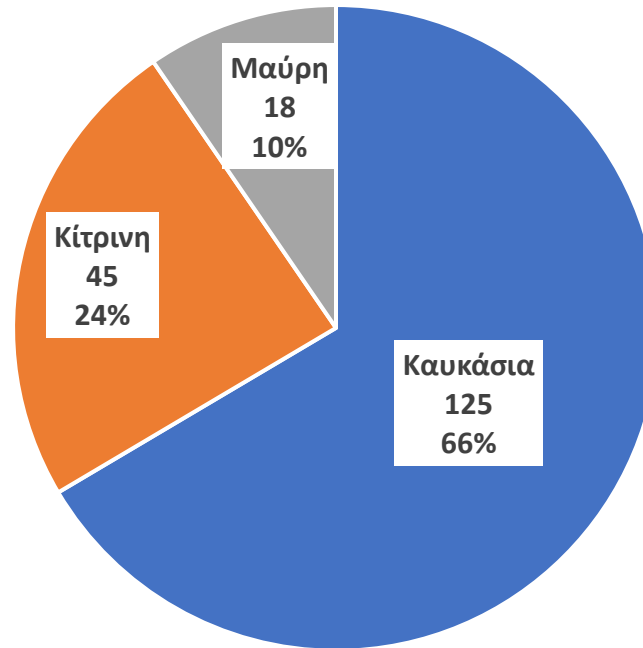
1. Παράδειγμα οι απαντήσεις σε ερωτηματολόγια, η κλίμακα VAS (1-10).
2. Διαφορά ταξινομικών μεταβλητών από τις συνεχείς.
3. Όταν έχουμε ικανοποιητικό αριθμό ταξινομήσεων (≥ 5) τότε χρησιμοποιούμε και πάλι τη διάμεσο και τα τεταρτημόρια (ποτέ τη μέση τιμή και την τυπική απόκλιση).
4. Όταν έχουμε μικρό αριθμό απαντήσεων τότε τις περιγράφουμε με τον ίδιο τρόπο που περιγράφουμε τις κατηγορικές μεταβλητές.

3. Κατηγορικές μεταβλητές

1. Παραδείγματα, χρώμα ή μάρκα αυτοκινήτου, φυλή, εθνικότητα.
2. Η κωδικοποίηση τους μπορεί να γίνεται με αριθμούς, οι οποίοι όμως δεν μεταφέρουν καμιά πληροφορία και δεν είναι συγκρίσιμοι μεταξύ τους.
3. Περιγράφονται μόνο με απόλυτη και ποσοστιαία κατανομή συχνοτήτων.
4. Το άθροισμα των απόλυτων αριθμών ισούται με το μέγεθος του δείγματος και τα ποσοστά αθροίζουν στο 100%

Γραφική αναπαράσταση κατηγορικής μεταβλητής.

Κατανομή της φυλής



Παρατηρήσεις: Και εδώ καλό είναι όταν αναφέρουμε τα ποσοστά να αναφέρουμε και τα 95% ΔΕ
Π.χ. εδώ Καυκάσιοι 125/188 – 66.5% (95% ΔΕ 59.5%-72.9%).

3α. Διωνυμικές μεταβλητές

1. Οι διωνυμικές μεταβλητές είναι μια ειδική περίπτωση των κατηγορικών μεταβλητών, όπου η μεταβλητή μπορεί να πάρει μια από ΔΥΟ δυνατές τιμές.
2. Παραδείγματα:
 1. Στρίψιμο του νομίσματος (κορώνα-γράμματα)
 2. Φύλο (Άρρεν-Θήλυ)
3. Μια ειδική και πολύ σημαντική περίπτωση διωνυμικών μεταβλητών είναι αυτές που οι τιμές που μπορούν να πάρουν είναι Όχι-Ναι (ή Ψευδές-Αληθές).
 1. Θάνατος μέχρι την 28^η ημέρα
 2. Έγκυος
 3. Σακχαρώδης διαβήτης

Καταγραφή και περιγραφή διωνυμικών μεταβλητών.

1. Οι διωνυμικές μεταβλητές καταγράφονται σαν 0-Όχι και 1-Ναι.
2. Περιγράφονται όπως και οι άλλες κατηγορικές μεταβλητές με απόλυτους αριθμούς και ποσοστά.
3. Επειδή, όπως είπαμε τα ποσοστά αθροίζουν στο 100%, καθώς έχουμε μόνο δύο δυνατότητες, πολλές φορές είναι αρκετό να δώσουμε το ένα ποσοστό.
4. Για παράδειγμα, έχουμε καταγράψει ένα αριθμό 1288 ασθενών στις ΜΕΘ και σε 307 (23.8%) έχουμε χορηγήσει κολιστίνη. Αυτό σημαίνει ότι σε 981 (76.2%) δεν έχει χορηγηθεί το αντιβιοτικό, αλλά αυτό δεν χρειάζεται να το αναφέρουμε.