

## Μεθοδολογία των επιστημών του Ανθρώπου : Στατιστική

### Εργαστήριο 3

1. Οι βαθμοί ενός τμήματος 10 φοιτητών σε δύο διαφορετικά μαθήματα, με κλίμακα βαθμολογίας 0 έως 10, δίνονται παρακάτω:

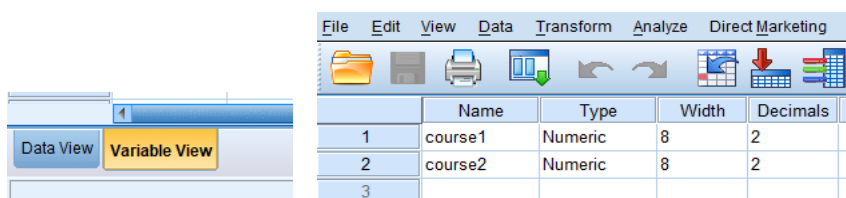
Αυξ. Αριθμός φοιτητή	1	2	3	4	5	6	7	8	9	10
1 <sup>ο</sup> Μάθημα	5	6	5	5	5	10	5	6	9	6
2 <sup>ο</sup> Μάθημα	1	6	5	6	7	8	7	7	8	7

Με τη βοήθεια του SPSS να γίνουν τα παρακάτω:

- i. Σχεδιάστε ένα ιστόγραμμα που να εμφανίζει την κατανομή των βαθμών για το πρώτο μάθημα. ( Να χρησιμοποιήσετε διαστήματα εύρους 1. π.χ. 9-10, 7-8 κ.λ.π.). Το ίδιο και για το δεύτερο μάθημα. Εξετάζοντας τα δύο ιστογράμματα, σε ποιο μάθημα είχε καλύτερη επίδοση το τμήμα;

Αρχικά θα πρέπει να γίνει η εισαγωγή δεδομένων σε ένα νέο αρχείο δεδομένων του SPSS. Τα δεδομένα των βαθμών των 10 φοιτητών θα πρέπει να περάσουν στο αρχείο που έχουμε μπροστά μας μόλις ανοίξουμε το SPSS ακολουθώντας την παρακάτω διαδικασία:

- α. Στον πίνακα **variable view** που επιλέγουμε από το κάτω αριστερό μέρος της οθόνης, πληκτρολογούμε τα ονόματα των δύο μεταβλητών “**course1**” και “**course2**” στο πεδίο **Name** και αφήνουμε τα υπόλοιπα πεδία να πάρουν τις προκαθορισμένες τιμές.

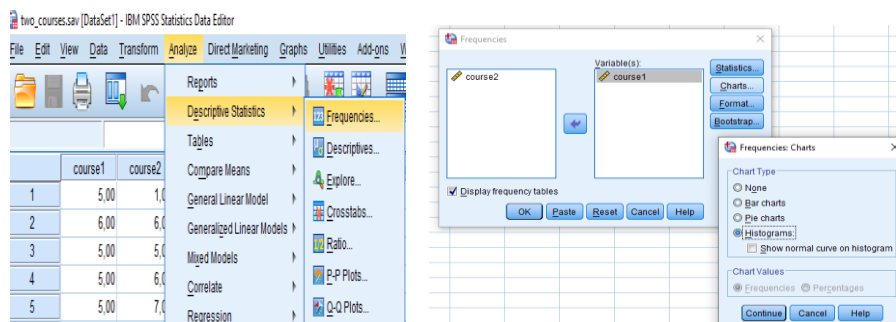


- β. Στη συνέχεια επιλέγουμε το παράθυρο data view που αποτελεί ουσιαστικά τον πίνακα δεδομένων και εισάγουμε τις τιμές για τις δύο μεταβλητές course1 και course2.

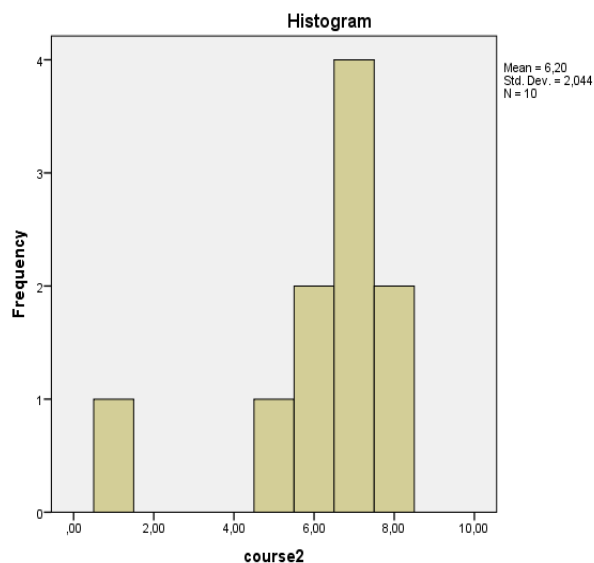
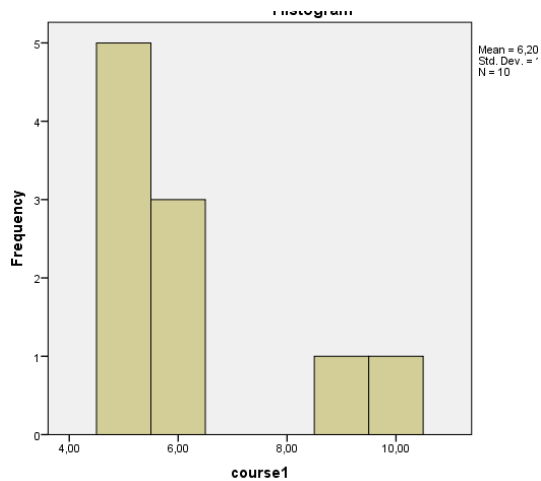
	course1	course2	var	var	var
1	5,00	1,00			
2	6,00	6,00			
3	5,00	5,00			
4	5,00	6,00			
5	5,00	7,00			
6	10,00	8,00			
7	5,00	7,00			
8	6,00	7,00			
9	9,00	8,00			
10	6,00	7,00			
11					
12					

γ. Τέλος για να ολοκληρωθεί η δημιουργία του αρχείου το αποθηκεύουμε με το όνομα **'two\_courses'** στον επιθυμητό φάκελο από την επιλογή **file=> save**.

Το ιστόγραμμα της μεταβλητής **'course1'** μπορεί να γίνει από την επιλογή **Analyze=>descriptive statistics=>frequencies** ως εξής: στο δεξί παράθυρο επιλέγουμε το όνομα της μεταβλητής που θέλουμε να επεξεργαστούμε. Στη συνέχεια από την υποεπιλογή **Charts** επιλέγουμε **histograms**.



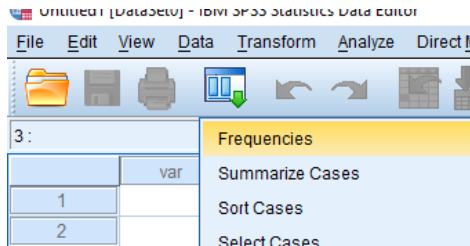
Το αποτέλεσμα της διαδικασίας εγγράφεται στο αρχείο αποτελεσμάτων του SPSS. Εκτελώντας ξανά τη δημιουργία γραφήματος αλλά τώρα για την **"course2"** έχω τα παρακάτω αποτελέσματα:



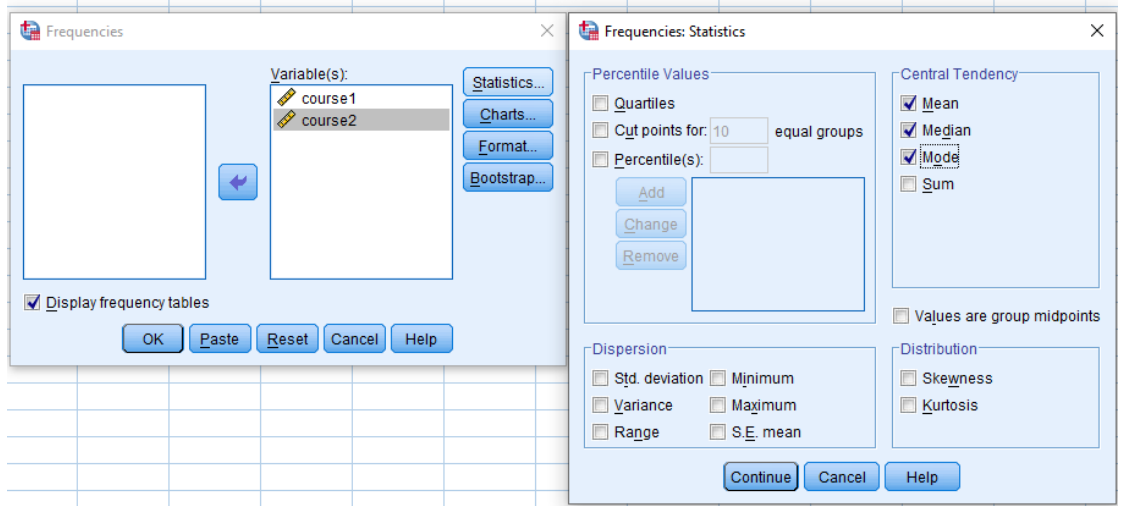
Παρατηρούμε ότι η επικρατέστερη τιμή είναι 5 (την προσεγγίζουμε με το κέντρο του διαστήματος με την μεγαλύτερη συχνότητα) για το πρώτο μάθημα, και 7 για το δεύτερο. Επίσης είναι φανερό ότι στο πρώτο μάθημα ο κύριος όγκος των δεδομένων βρίσκεται σε χαμηλές τιμές σε αντίθεση με το δεύτερο μάθημα στο οποίο ο κύριος όγκος βαθμών βρίσκεται σε μια περιοχή υψηλών τιμών.

- ii. Υπολογίστε το μέσο όρο και τη διάμεσο για κάθε μάθημα. Ποιος δείκτης περιγράφει καλύτερα την διαφορά των δύο κατανομών; Γιατί συμβαίνει αυτό κατά την γνώμη σας;

Από την ίδια επιλογή του SPSS **Analyze=>descriptive statistics=>frequencies** ή ισοδύναμα από το εικονίδιο **recall recently used dialog** επιλέγεται απλά η επιλογή **frequencies**



Στην περίπτωση αυτή μπορούμε από τη υποεπιλογή **Statistics** να επιλέξουμε όλα τα μέτρα κεντρικής θέσης, Μέση τιμή (mean), Διάμεσος (median) και επικρατέστερη τιμή (mode) όπως φαίνεται παρακάτω:



Το αποτέλεσμα που έχει ενδιαφέρον μετά το **Continue** και το **Ok** στο βασικό πλαίσιο **frequencies** είναι:

**Statistics**

		course1	course2
N	Valid	10	10
	Missing	0	0
Mean		6,2000	6,2000
Median		5,5000	7,0000
Mode		5,00	7,00

Προφανώς η σύγκριση των διαμέσων (5,5 και 7) της επίδοσης στα δύο μαθήματα οδηγεί στο ίδιο συμπέρασμα με αυτό που είδαμε μελετώντας τα ιστογράμματα και συγκρίνοντας επικρατέστερες τιμές. Αντίθετα στηριζόμενοι στη μέση τιμή θα συμπεραίναμε ότι η επίδοση στα δύο μαθήματα είναι ίδια (6,2). Το τελευταίο συμπέρασμα δεν ευσταθεί και οφείλεται στην παρουσία ακραίων τιμών ή/και ασυμμετρίας αντίθετης κατεύθυνσης στις δύο κατανομές. Η μέση τιμή επηρεάζεται από την ασυμμετρία η/και τις ακραίες τιμές αφού όλες οι τιμές της μεταβλητής

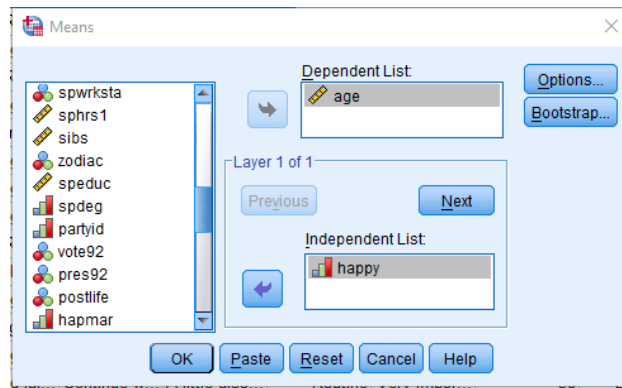
συμμετέχουν στον υπολογισμό της. Στα μικρά δείγματα ιδιαίτερα (πχ.  $N < 30$ ) είμαστε πολύ προσεκτικοί αναφορικά με τη χρήση της μέσης τιμής. Πάντα πρέπει να προηγείται η μελέτη της μορφής της κατανομής μιας μεταβλητής. Αν όπως στο παράδειγμά μας παρουσιάζεται ισχυρή ασυμμετρία η/και ακραίες τιμές, θα προτιμηθεί η διάμεσος.

2. Να χρησιμοποιηθεί το αρχείο `gss.sav` για να απαντηθούν τα ακόλουθα ερωτήματα:

- i. Τι είδους άνθρωποι είναι οι ευτυχισμένοι άνθρωποι (μεταβλητή **happy**); Να γραφεί μια σύντομη απάντηση που συγκρίνει μεταξύ τους ανθρώπους που είναι πολύ ευτυχισμένοι, αρκετά ευτυχισμένοι και όχι τόσο ευτυχισμένοι. Η σύγκριση θα γίνει ως προς τα χαρακτηριστικά ηλικία, φύλο, εκπαίδευση, ώρες τηλεόρασης κ.λ.π. (**Compare Means > Means...**). Να κατασκευαστούν κατάλληλα γραφήματα (**Graph > Bar... > Simple**)

Από τον φάκελο **'guide to data analysis'**, τον οποίο πρέπει να κατεβάσετε από το φάκελο [Δεδομένα](#) της e-class και να αποσυμπιέσετε στον υπολογιστή σας, θα ανοίξουμε το αρχείο **'gss.sav'** από την επιλογή **file=>open=>data**. Θα χρησιμοποιηθεί μόνο το παράδειγμα της ηλικίας (**age**). Θα διερευνηθεί περιγραφικά πως η ηλικία συνδέεται με τη δήλωση ευτυχίας.

α. Χρησιμοποιώντας πίνακα μέσων τιμών: Από την επιλογή **Analyze=> Compare Means=> Means** βάζουμε τη μεταβλητή **'age'** στο πλαίσιο **Dependent list** και την μεταβλητή **happy** στο πλαίσιο **Independent list**. Η σύγκριση των μέσων τιμών (ή κεντρικών θέσεων γενικότερα) διερευνά τη σχέση ποσοτικής με κατηγορική (ή κατηγορική με διάταξη) και οι μεταβλητές πρέπει εισέρχονται στα πλαίσια μ' αυτή την αντιστοίχιση: **Dependent list** η ποσοτική και **Independent list** η κατηγορική.



Το αποτέλεσμα στον αρχείου αποτελεσμάτων είναι ο παρακάτω πίνακας:

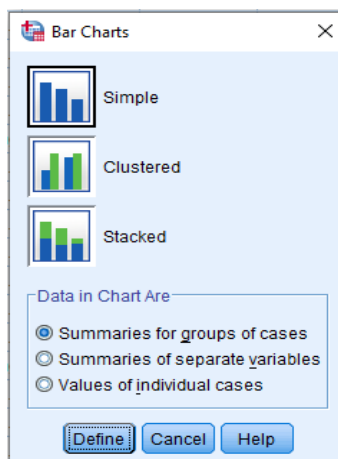
### Report

age Age of Respondent

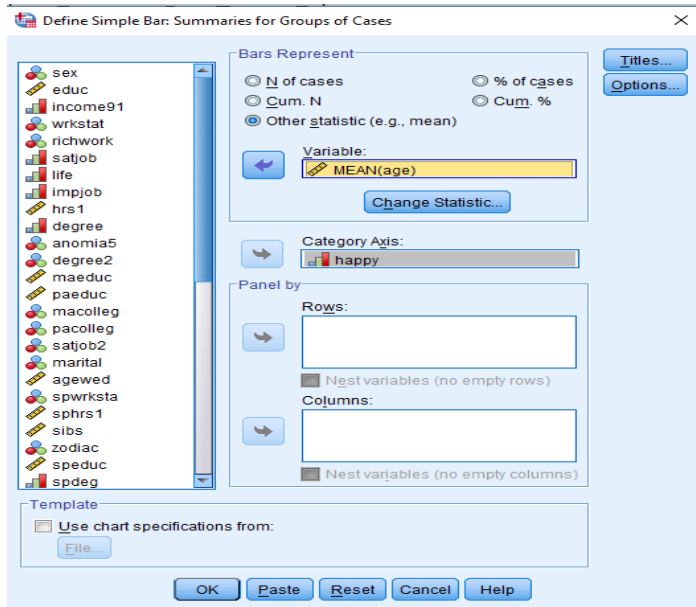
happy GENERAL HAPPINESS	Mean	N	Std. Deviation
1 VERY HAPPY	47,39	474	17,513
2 PRETTY HAPPY	44,92	847	17,336
3 NOT TOO HAPPY	49,20	169	17,102
Total	46,19	1490	17,423

Συγκρίνοντας τις μέσες τιμές οι οποίες δεν παρουσιάζουν αξιόλογες διαφορές μεταξύ τους θα μπορούσαμε να συμπεράνουμε ότι: Οι υψηλότερης ευτυχίας άνθρωποι βρίσκονται σε μια ενδιάμεση ηλικιακή ομάδα (47,39) σε σχέση με τους μέτριας ευτυχίας που είναι οι νεότεροι (44,92) και τους χαμηλότερης ευτυχίας συμμετέχοντες που είναι οι πιο ηλικιωμένοι από τις υπόλοιπες κατηγορίες (49,2)

β. χρήση του ραβδογράμματος μέσω τιμών. Από την επιλογή **Graph=>legacy dialogs=>Bars** στο πλαίσιο που παρουσιάζεται επιλέγω **Simple** και **summaries for groups of cases** πατάμε **define**

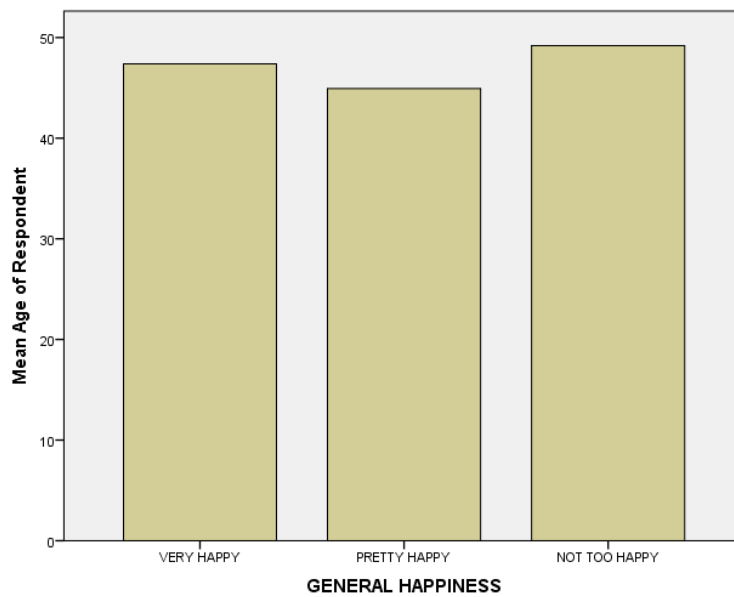


Στη συνέχεια τοποθετώ κατάλληλα τις δύο μεταβλητές όπως φαίνεται παρακάτω:



Στην περιοχή **Bars Represent** του πλαισίου επιλέγουμε **other statistics** που ως προεπιλογή είναι η μέση τιμή (MEAN(age)). Η επιλογή αυτή δηλώνει ότι το ύψος της κάθε ράβδου (μια για κάθε κατηγορία ευτυχίας) είναι ίσο με την μέση τιμή της ηλικίας των ατόμων της κατηγορίας ευτυχίας που αντιστοιχεί στη ράβδο.

Το αποτέλεσμα των παραπάνω επιλογών οδηγούν στο παρακάτω γράφημα:



Η ερμηνεία είναι ίδια με εκείνη που δόθηκε στον προηγούμενο πίνακα μέσω των τιμών.