

# SPSS STATISTICS

ΔΕΣΠΟΙΝΑ ΚΟΠΑΝΑΚΗ



# ΤΙ ΕΙΝΑΙ ΤΟ SPSS?

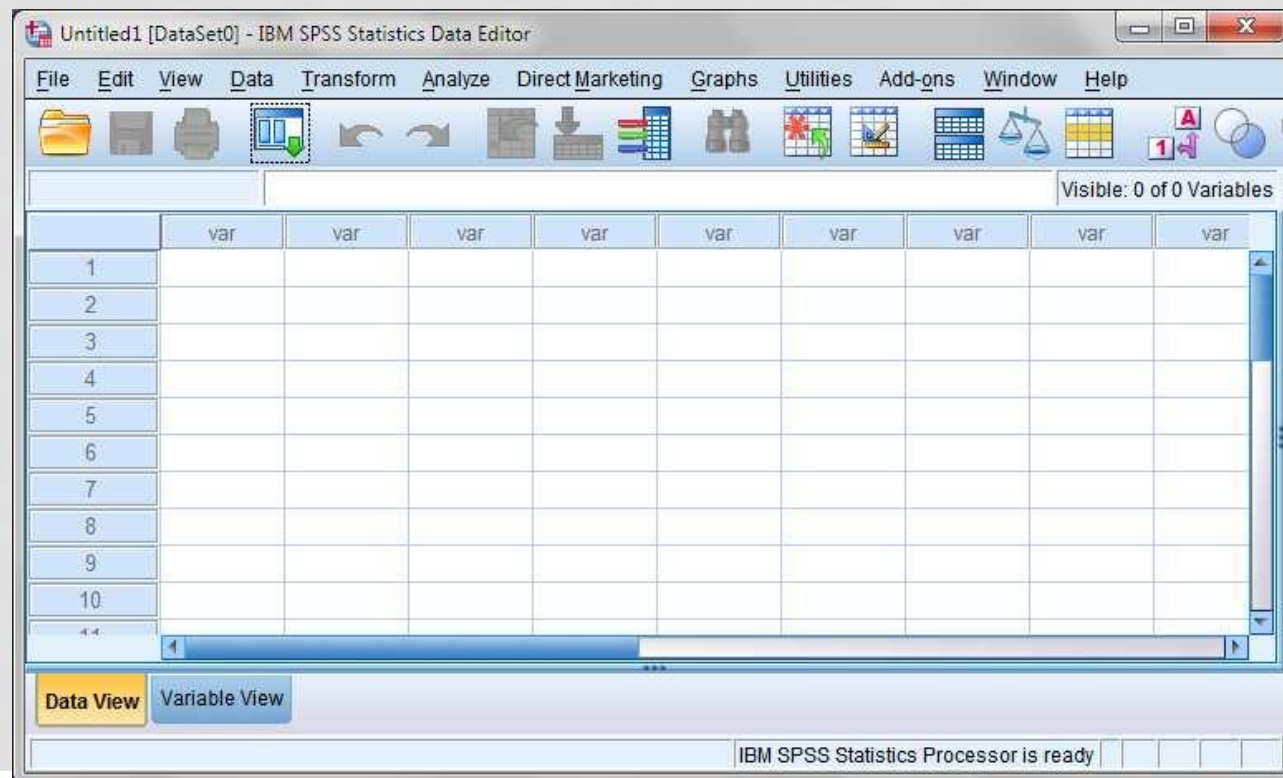
- Ακρωνύμιο του Statistical Package for the Social Sciences.
- Είναι ένα από τα καλύτερα στατιστικά πακέτα και μπορεί να χρησιμοποιηθεί για τη στατιστική ανάλυση τόσο κοινωνικοοικονομικών δεδομένων όσο και δεδομένων των θετικών επιστημών.
- Μπορεί κανείς να εκτελέσει περίπλοκους χειρισμούς δεδομένων και αναλύσεων με πολύ απλούς τρόπους

# ΦΥΛΛΑ ΕΡΓΑΣΙΑΣ ΤΟΥ SPSS

- Υπάρχουν δυο βασικά αρχεία:
- Αρχείο δεδομένων – SPSS Data editor
- Αρχείο αποτελεσμάτων – SPSS Viewer

# SPSS DATA EDITOR

- Φύλλο εργασίας που καταχωρούμε τα δεδομένα που θέλουμε να αναλύσουμε.
- Αποτελείται από δυο παράθυρα
  - **Data view**
  - Variable view

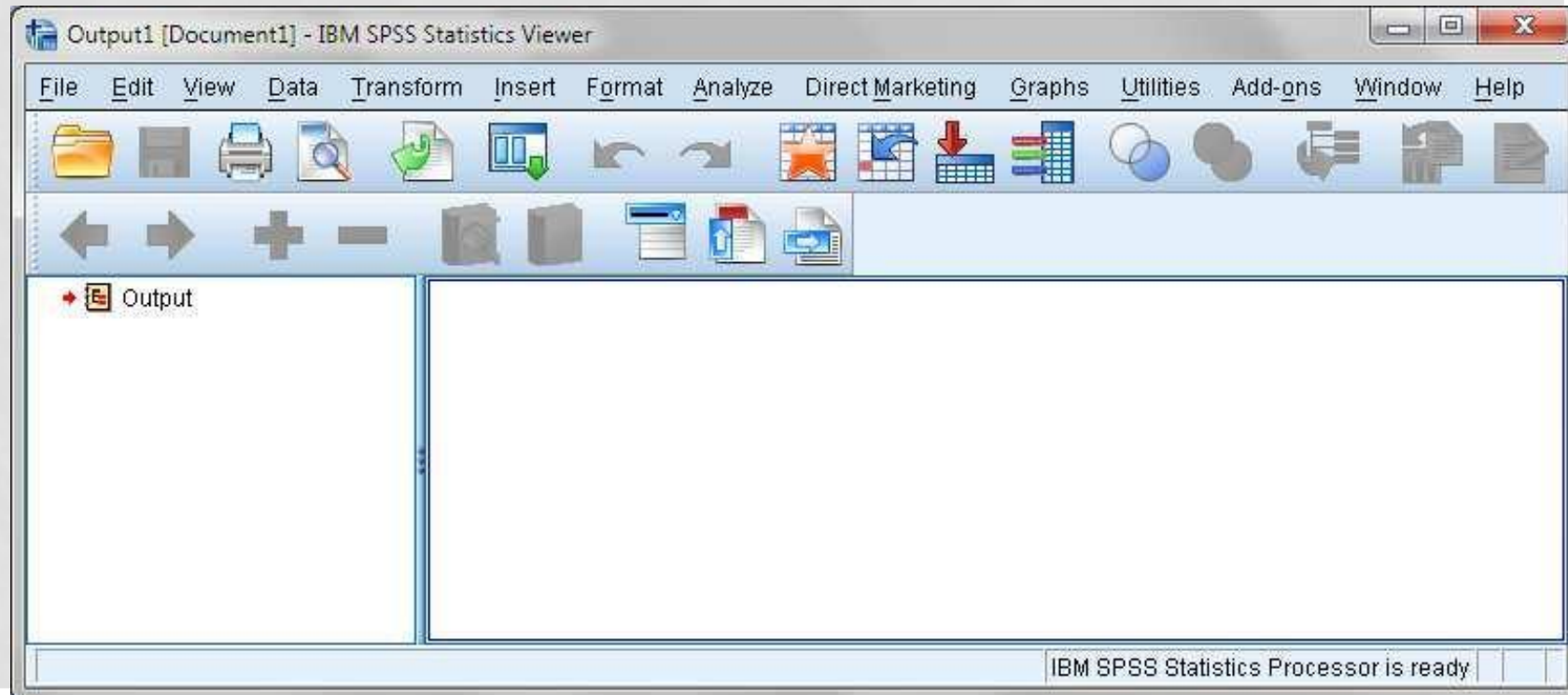






# SPSS VIEWER

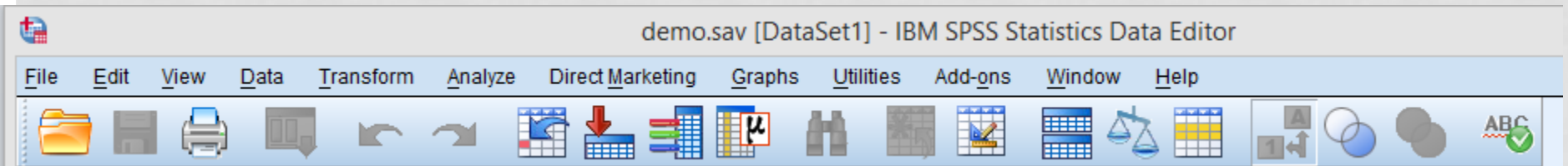
- Είναι το αρχείο αποτελεσμάτων
- Αριστερό μέρος: στατιστικές πράξεις που έχουν γίνει
- Δεξιό μέρος: στατιστικά αποτελέσματα





# MENU BAR

- **File:** άνοιγμα νέου αρχείου (new) ή παλιού (open), αποθήκευση (save), εκτύπωση (print),...
- **Edit:** τροποποίηση ή αντιγραφή τμήματος του αρχείου
- **View:** προσαρμογή στοιχείων ανάλογα με τις επιλογές μας
- **Data:** πραγματοποίηση αλλαγών στα δεδομένα
- **Transform:** πραγματοποίηση αλλαγών στις μεταβλητές
- **Analyze:** στατιστική ανάλυση δεδομένων
- **Graph:** δημιουργία γραφικών παραστάσεων
- .....



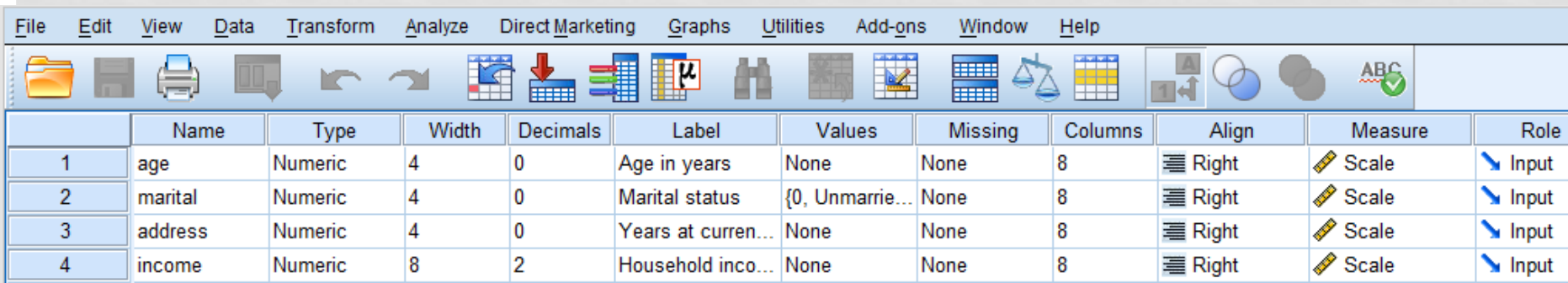


# ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ

- Μεταβλητή (variable): κάθε χαρακτηριστικό του πληθυσμού που μελετάμε
- Εισάγεται σε ξεχωριστή στήλη
- Χωρίζονται σε δυο τύπους
  - Αριθμητικές (numeric): οι τιμές εκφράζονται με αριθμούς
  - Αλφαριθμητικές (string): οι τιμές εκφράζονται με χαρακτήρες
- Χωρίζονται σε δυο κατηγορίες
  - Ποσοτικές μεταβλητές ή μεταβλητές κλίμακας: μεγέθη που μπορούν να μετρηθούν, παίρνουν αριθμητικές τιμές [συνεχείς (continuous)], [διακριτές (discrete)], [ασυνεχείς (discontinuous)]
  - Ποιοτικές μεταβλητές: εκφράζουν ποιοτικά χαρακτηριστικά του πληθυσμού [ονομαστικές (nominal)], [διατεταγμένες (ordinal)]

# ΜΟΡΦΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ

- Μέσω του παράθυρου Data View > Variable view
- Name: όνομα μεταβλητής
- Type: τύπος μεταβλητής (numeric, string,...)
- Width: πόσα γράμματα μπορεί να έχει το όνομα της μεταβλητής
- Decimals: αριθμός δεκαδικών ψηφίων των αριθμητικών μεταβλητών
- Label: σύντομη περιγραφή της κάθε μεταβλητής
- Values: πληροφορίες για τις τιμές όταν είναι κατηγορικές

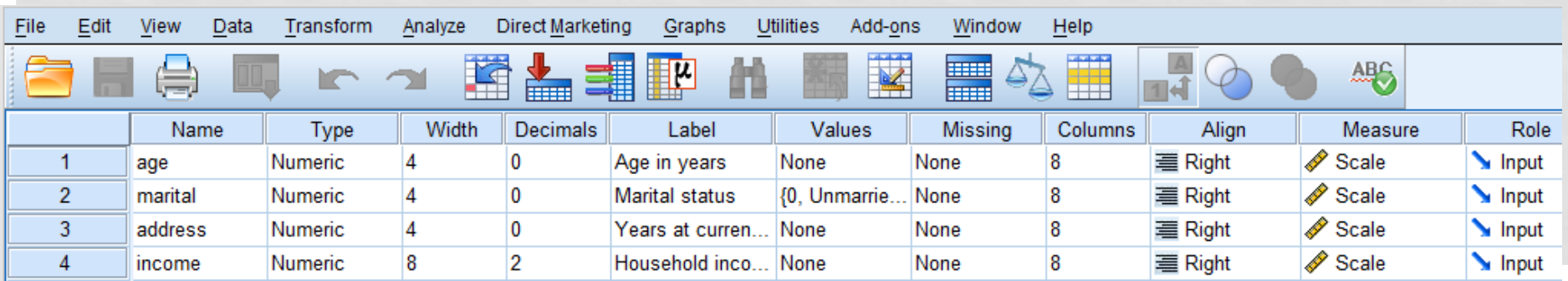


The screenshot shows the SPSS Variable View dialog box. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains icons for file operations, data manipulation, and analysis. The main area displays a table with the following columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	age	Numeric	4	0	Age in years	None	None	8	Right	Scale	Input
2	marital	Numeric	4	0	Marital status	{0, Unmarrie...	None	8	Right	Scale	Input
3	address	Numeric	4	0	Years at curren...	None	None	8	Right	Scale	Input
4	income	Numeric	8	2	Household inco...	None	None	8	Right	Scale	Input

# ΜΟΡΦΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ

- Missing: ελλιπείς τιμές
- Columns: πλάτος – πόσα ψηφία μπορεί να πάρει η μεταβλητή
- Align: στοίχιση των τιμών μιας μεταβλητής
- Measure: ποσοτική (scale), ονομαστική (nominal), σειριακή (ordinal)



The screenshot shows the SPSS Variable View table with the following columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The data is as follows:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	age	Numeric	4	0	Age in years	None	None	8	Right	Scale	Input
2	marital	Numeric	4	0	Marital status	{0, Unmarrie...	None	8	Right	Scale	Input
3	address	Numeric	4	0	Years at curren...	None	None	8	Right	Scale	Input
4	income	Numeric	8	2	Household inco...	None	None	8	Right	Scale	Input

# OPEN DATA

- File > Open > Data...

demo.sav - Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

20 : age 40

	age	marital	address	income	inccat	car
1	55	Marital status	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

Data View Variable View

# TYPE

Numeric	4	0	Marital status	{0, Unmarrie...	None	8	Right	Scale
Numeric	4	0	Years at curren	None	None	8	Right	Scale
Numeric	8							Scale
Numeric	8							Ordinal
Numeric	8							Scale
Numeric	8							Ordinal
Numeric	4							Scale
Numeric	4							Scale
Numeric	4							Scale
Numeric	4							Ordinal
Numeric	4							Scale
String	1							Nominal
Numeric	4							Scale
Numeric	4							Scale
Numeric	4							Scale
Numeric	4							Scale
Numeric	4							Scale
Numeric	4	0	Internet	{0, No}...	8, 9	8	Right	Scale
Numeric	4	0	Caller ID	{0, No}...	None	8	Right	Scale

**Variable Type**

**Numeric**

Comma

Dot

Scientific notation

Date

Dollar


Custom currency

String

Restricted Numeric (integer with leading zeros)

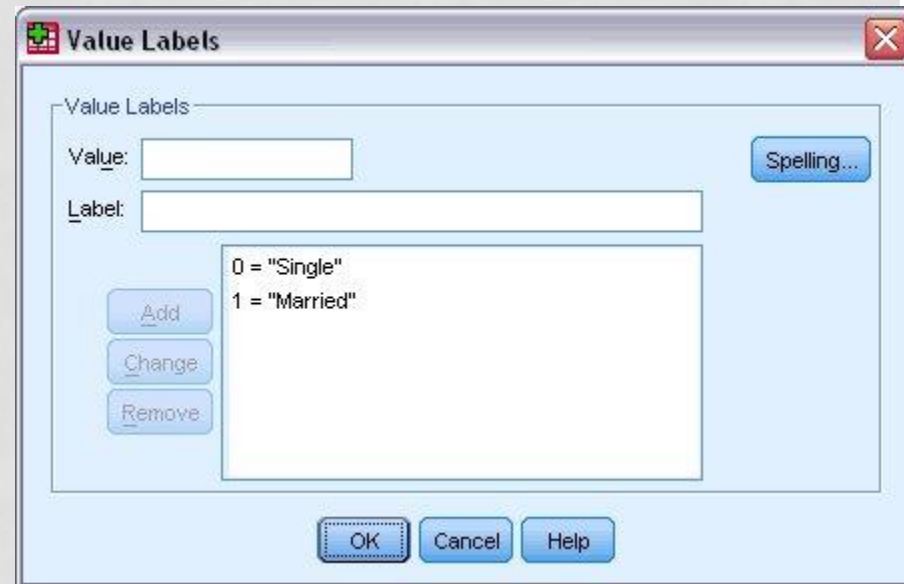
Width:

Decimal Places:

 The Numeric type honors the digit grouping setting, while the Restricted Numeric never uses digit grouping.

# VALUE LABELS

- Επιλέγουμε για την μεταβλητή *marital* την στήλη *Values* και κάνουμε κλικ στο κουμπί που βρίσκεται στην δεξιά πλευρά ώστε να ανοίξει το Value Labels dialog box
- Πληκτρολογούμε την τιμή 0 στο πεδίο Value και Single στο πεδίο Label.
- Επιλέγουμε Add
- Ομοίως και για τις υπόλοιπες ετικέτες



# VIEW LABELS

- **View > Value Labels**



SPSS demo.sav - Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

20 : age 40

	age	marital	address	income	inccat	car
1	55	Married	12	72.00	\$50 - \$74	36.
2	56	Unmarried	29	153.00	\$75+	76.
3	28	Married	9	28.00	\$25 - \$49	13.
4	24	Married	4	26.00	\$25 - \$49	12.
5	25	Unmarried	2	23.00	Under \$25	11.
6	45	Married	9	76.00	\$75+	37.
7	42	Unmarried	19	40.00	\$25 - \$49	19.
8	35	Unmarried	15	57.00	\$50 - \$74	28.
9	46	Unmarried	26	24.00	Under \$25	12.
10	34	Married	0	89.00	\$75+	46.
11	55	Married	17	72.00	\$50 - \$74	35.

Data View Variable View



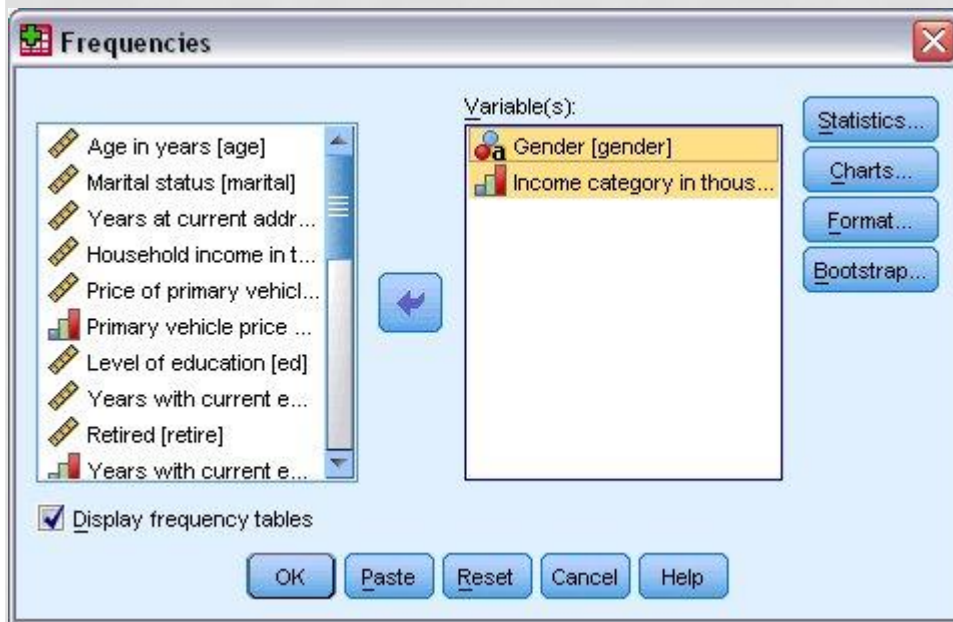
# MISSING DATA

- Δεν απαντούν σε όλες τις ερωτήσεις των ερηματολογίων
- Μπορεί να μην γνωρίζουν την απάντηση
- Μπορεί να απαντήσουν με μη αναμενόμενη μορφή
- Αποτέλεσμα: μη ακριβή αποτελέσματα κατά την ανάλυση των δεδομένων
- Ο λόγος που μια μεταβλητή μπορεί να λείπει μπορεί να είναι σημαντικός για την ανάλυσή σας, π.χ. διαχωρισμός μεταξύ κάποιου που δεν ήθελε να απαντήσει και κάποιου που δεν γνώριζε την απάντηση



# ΔΗΜΙΟΥΡΓΙΑ ΠΙΝΑΚΑ ΣΥΧΝΟΤΗΤΩΝ

- **Analyze > Descriptive Statistics > Frequencies...**
- Επιλέγουμε τις μεταβλητές που θέλουμε να αναλύσουμε



	Numeric	String	Date	Time
Scale (Continuous)		n/a		
Ordinal				
Nominal				

# ΔΗΜΙΟΥΡΓΙΑ ΠΙΝΑΚΑ ΣΥΧΝΟΤΗΤΩΝ

The image shows the SPSS Output2 - Viewer window. The left pane displays a tree view of the output, with 'Income category in thousands' selected. The main pane shows two frequency tables. The first table is titled 'Gender' and the second is titled 'Income category in thousands'. Both tables have columns for Frequency, Percent, Valid Percent, and Cumulative Percent. A red arrow points to the 'Income category in thousands' table.

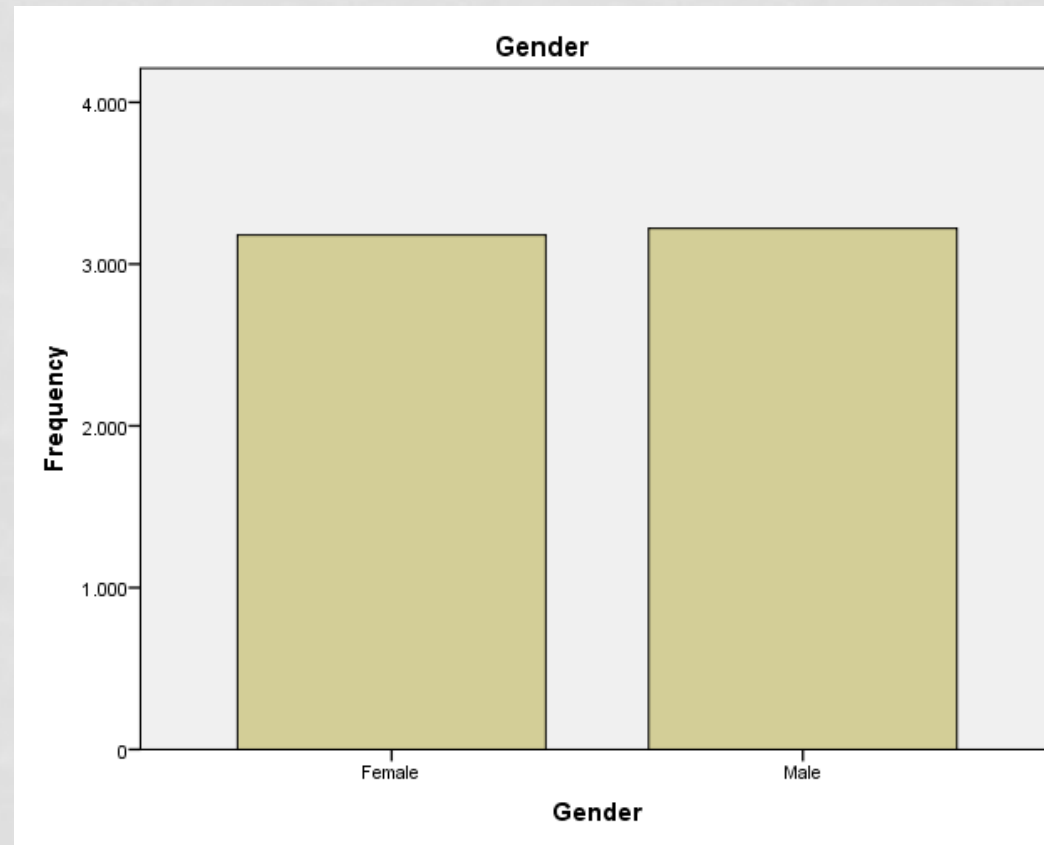
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	3179	49.7	49.7	
	Male	3221	50.3	50.3	
	Total	6400	100.0	100.0	

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under \$25	1174	18.3	18.3	
	\$25 - \$49	2388	37.3	37.3	
	\$50 - \$74	1120	17.5	17.5	
	\$75+	1718	26.8	26.8	
	Total	6400	100.0	100.0	

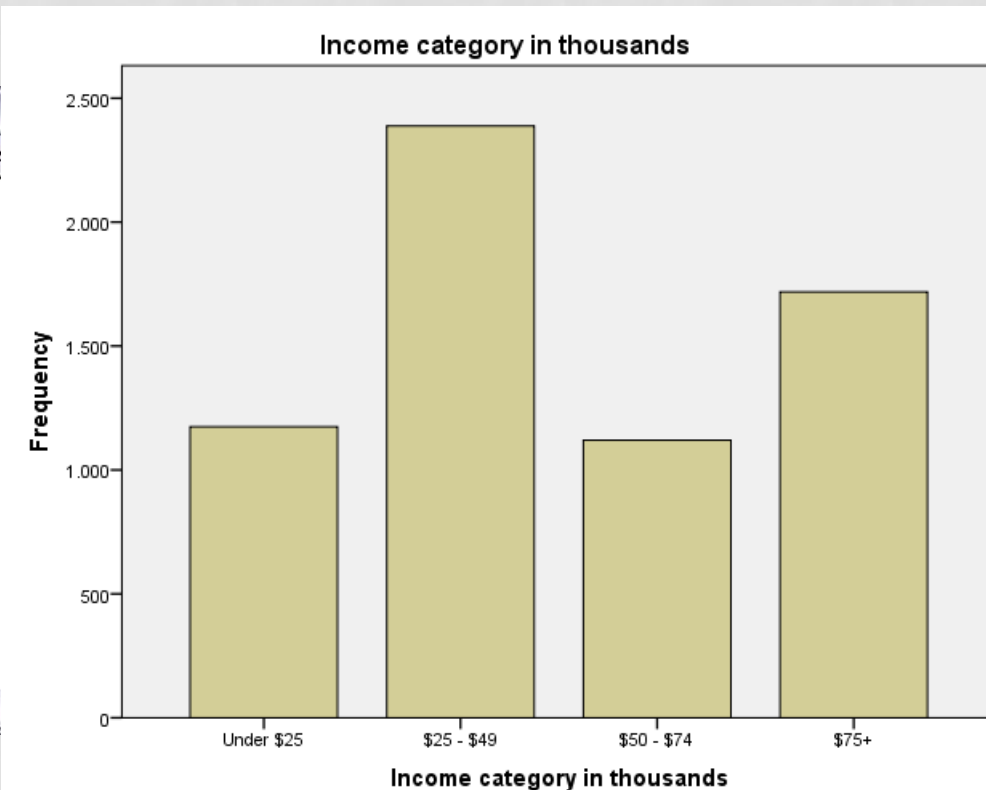
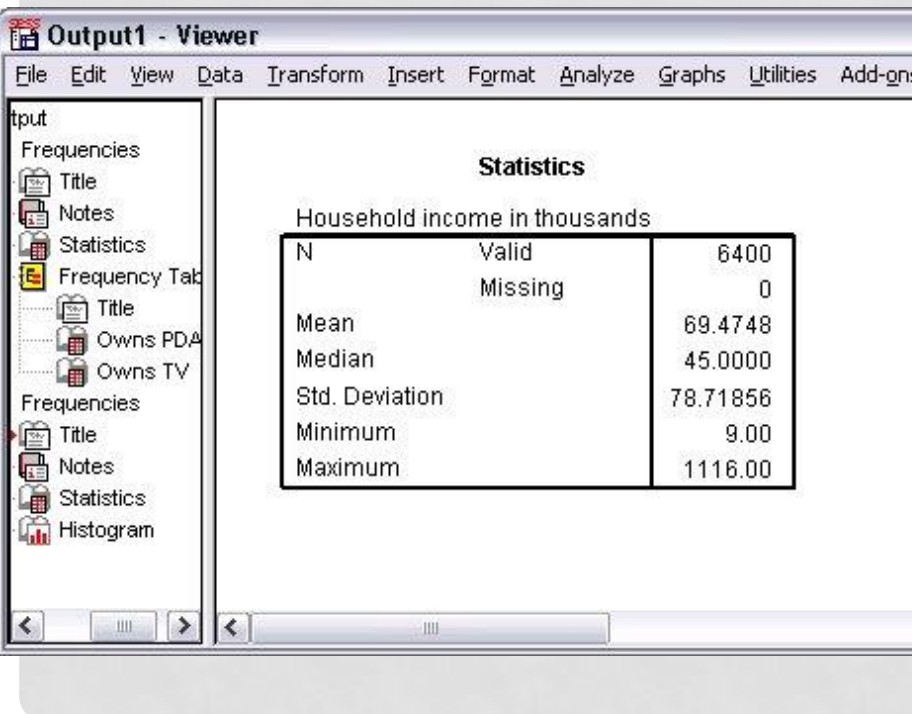
# ΔΗΜΙΟΥΡΓΙΑ ΓΡΑΦΗΜΑΤΟΣ ΓΙΑ ΚΑΤΗΓΟΡΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ

- **Analyze > Descriptive Statistics > Frequencies...**
- Επιλέγουμε **Charts > Bar charts**
- Εκτός από τους πίνακες συχνοτήτων εμφανίζονται και τα bar charts των επιλεγμένων μεταβλητών



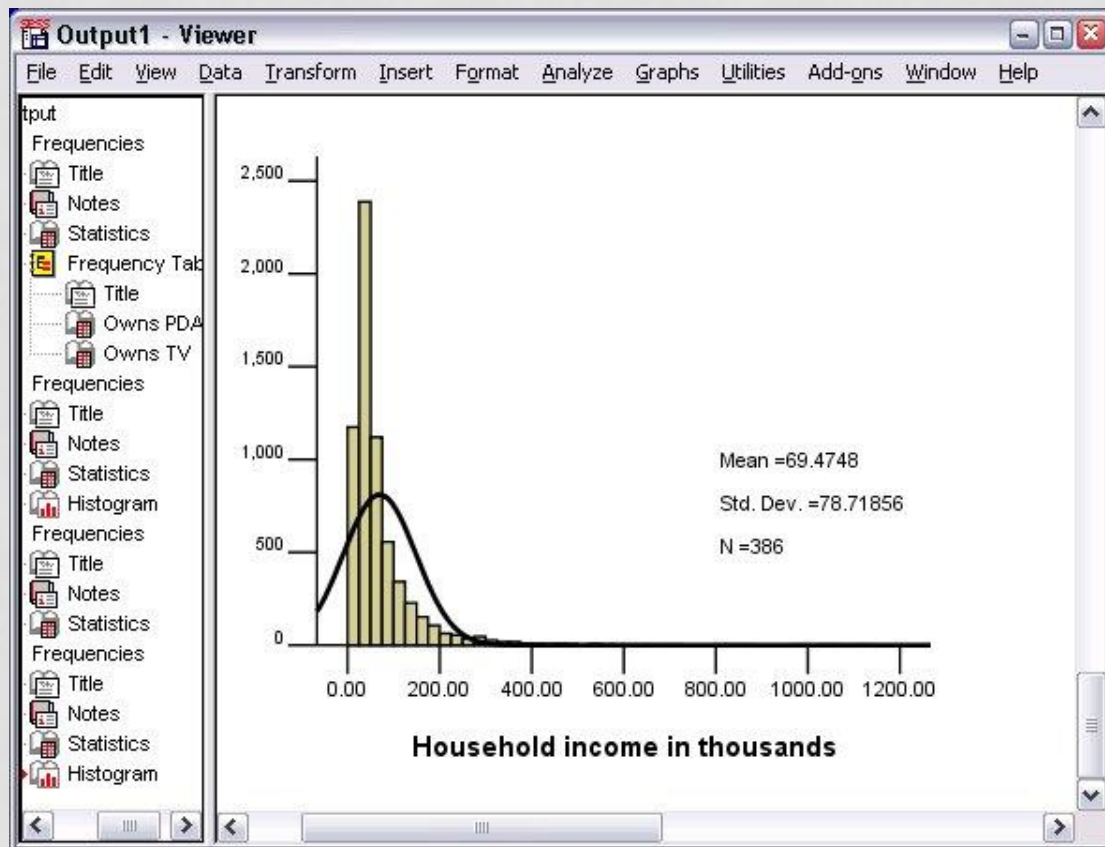
# ΔΗΜΙΟΥΡΓΙΑ ΓΡΑΦΗΜΑΤΟΣ ΓΙΑ ΣΥΝΕΧΕΙΣ ΜΕΤΑΒΛΗΤΕΣ

- **Analyze > Descriptive Statistics > Frequencies...**
- Επιλέγουμε **Statistics > Mean, Median, Std. deviation, Minimum, and Maximum**



# ΔΗΜΙΟΥΡΓΙΑ ΓΡΑΦΗΜΑΤΟΣ ΓΙΑ ΣΥΝΕΧΕΙΣ ΜΕΤΑΒΛΗΤΕΣ

- **Analyze > Descriptive Statistics > Frequencies...**
- Επιλέγουμε **Charts > Histograms, With normal curve**





# ΠΙΝΑΚΕΣ ΔΙΑΣΤΑΥΡΩΣΗΣ

- Crosstabulation tables: δείχνουν τη συσχέτιση μεταξύ δυο ή περισσότερων κατηγορικών μεταβλητών
- Πληθώρα στατιστικών τεστ είναι διαθέσιμα για να προσδιορίσουν την ύπαρξη ή μη σχέσης μεταξύ των μεταβλητών
- Ποιοι είναι οι παράγοντες που επηρεάζουν τα προϊόντα που αγοράζουν οι καταναλωτές? Π.χ. εισόδημα

# ΠΙΝΑΚΕΣ ΔΙΑΣΤΑΥΡΩΣΗΣ

- Analyze > Descriptive Statistics > Crosstabs

Income category in thousands \* Owns PDA Crosstabulation

Count		Owns PDA		Total
		No	Yes	
Income category in thousands	Under \$25	983	191	1174
	\$25 - \$49	1933	455	2388
	\$50 - \$74	889	231	1120
	\$75+	1288	430	1718
Total		5093	1307	6400

The screenshot shows the SPSS Crosstabs dialog box. The 'Row(s)' field contains 'Income category in thou...' and the 'Column(s)' field contains 'Owns PDA [ownpda]'. The 'Display clustered bar charts' checkbox is checked. The 'Display layer variables in table layers' checkbox is also checked. The 'Suppress tables' checkbox is unchecked. The 'OK' button is highlighted.

# ΠΙΝΑΚΕΣ ΔΙΑΣΤΑΥΡΩΣΗΣ

- Analyze > Descriptive Statistics > Crosstabs > Cells

Income category in thousands \* Owns PDA Crosstabulation

		Owns PDA		Total	
		No	Yes		
Income category in thousands	Under \$25	Count	983	191	1174
		% within Income category in thousands	83.7%	16.3%	100.0%
	\$25 - \$49	Count	1933	455	2388
		% within Income category in thousands	80.9%	19.1%	100.0%
\$50 - \$74	Count	889	231	1120	
	% within Income category in thousands	79.4%	20.6%	100.0%	
\$75+	Count	1288	430	1718	
	% within Income category in thousands	75.0%	25.0%	100.0%	
Total	Count	5093	1307	6400	
	% within Income category in thousands	79.6%	20.4%	100.0%	

**Crosstabs: Cell Display**

Counts

Observed

Expected

Hide small counts

Less than

z-test

Compare column proportions

Adjust p-values (Bonferroni method)

Percentages

Row

Column

Total

Residuals

Unstandardized

Standardized

Adjusted standardized

Noninteger Weights

Round cell counts

Round case weights

Truncate cell counts

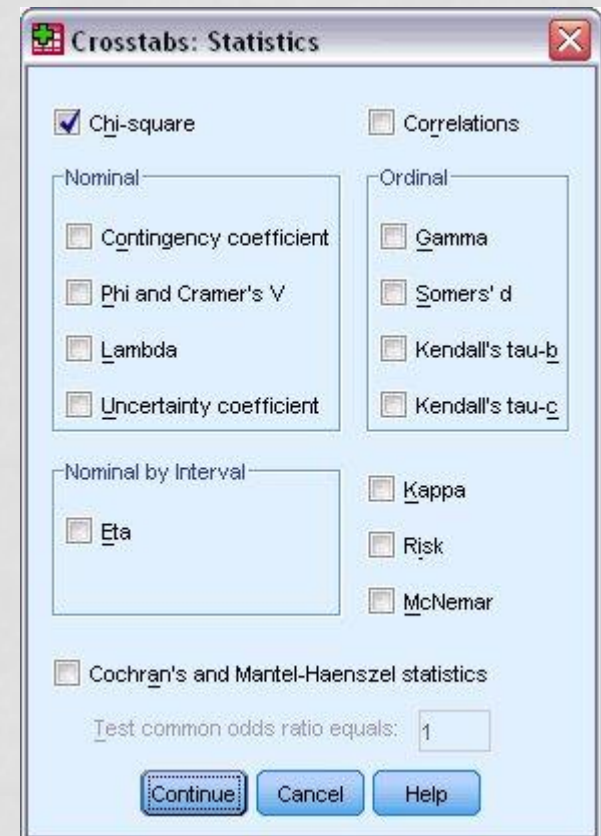
Truncate case weights

No adjustments

Continue Cancel Help

# ΠΙΝΑΚΕΣ ΔΙΑΣΤΑΥΡΩΣΗΣ

- Είναι σημαντική η συσχέτιση μεταξύ των μεταβλητών?
- Έλεγχος με *Pearson chi-square* test
- **Analyze > Descriptive Statistics > Crosstabs > Statistics**



# ΠΙΝΑΚΕΣ ΔΙΑΣΤΑΥΡΩΣΗΣ

- $H_0$ : rows and columns are independent
  - $H_1$ : otherwise
- Απορρίπτουμε την μηδενική υπόθεση (significance value = 0)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	37.677 <sup>a</sup>	3	.000
Likelihood Ratio	37.313	3	.000
Linear-by-Linear Association	36.537	1	.000
N of Valid Cases	6400		

# ΔΗΜΙΟΥΡΓΙΑ ΓΡΑΦΗΜΑΤΩΝ

- **Graphs > Chart Builder...**
- Δημιουργία γραφήματος που δείχνει την σχέση μεταξύ των μεταβλητών wireless telephone service και Owns PDA

**Chart Builder**

Variables:

- Faxing service [pay...]
- Internet [internet]
- Caller ID [callid]
- Call waiting [callwait]
- Owns TV [ownstv]
- Owns VCR [ownvcr]
- Owns stereo/CD pl...
- Owns PDA [ownpda]**
- Owns computer [o...]
- Owns fax machine ...
- Newspaper subscri...
- Response [response]

Cluster on X: set color

Owns PDA

Count

No Yes

Wireless service

**Gallery** Basic Elements Groups/Point ID Titles/Footnotes

Choose from:

- Favorites
- Bar
- Line
- Area
- Pie/Polar
- Scatter/Dot
- Histogram
- High-Low
- Boxplot
- Dual Axes

Element Properties...

Options...

OK Paste Reset Cancel Help



# ΔΗΜΙΟΥΡΓΙΑ ΓΡΑΦΗΜΑΤΩΝ

- Επιλέγουμε το **Gallery** tab και στη συνέχεια **Bar**
- Τραβάμε μέσα στον λευκό καμβά το γράφημα που θέλουμε να δημιουργήσουμε
- Τραβάμε την μεταβλητή *Wireless service* στον άξονα x
- Τραβάμε την μεταβλητή *Owns PDA* στην επάνω δεξιά μεριά του καμβά

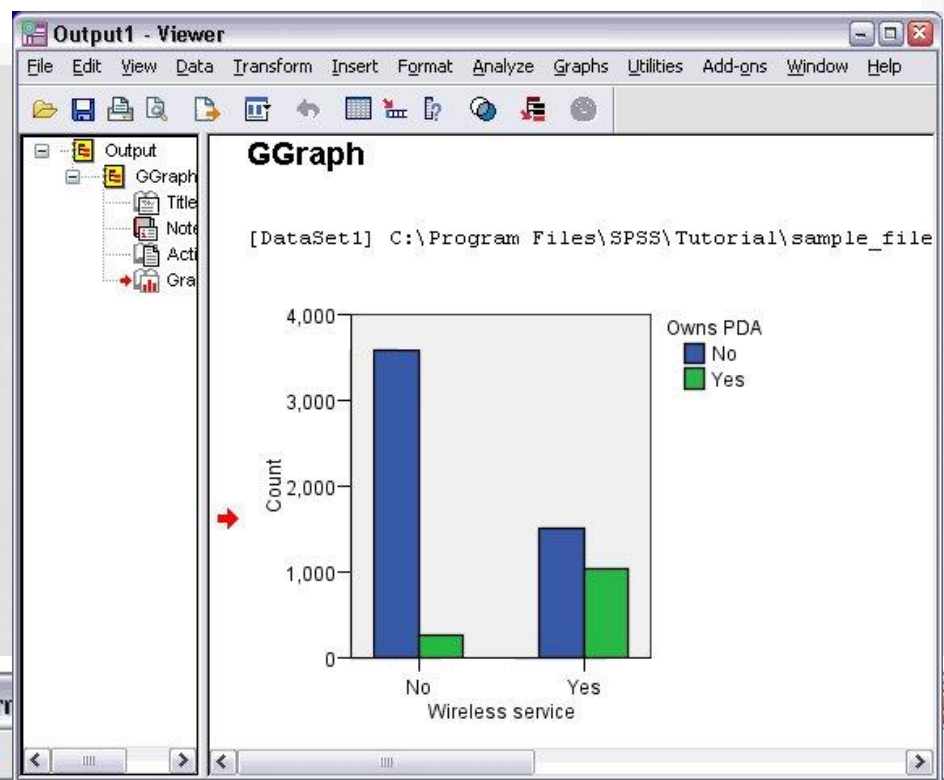
The screenshot shows the 'Chart Builder' window with the following elements:

- Variables:** A list of variables including 'Faging service [pay...', 'Internet [internet]', 'Caller ID [callid]', 'Call waiting [callwait]', 'Owns TV [ownstv]', 'Owns VCR [ownvcr]', 'Owns stereo/CD pl...', 'Owns PDA [ownpda]', 'Owns computer [o...', 'Owns fax machine ...', 'Newspaper subscri...', and 'Response [response]'. The 'Owns PDA [ownpda]' variable is selected.
- Chart Preview:** A bar chart showing the relationship between 'Wireless service' (x-axis) and 'Owns PDA' (y-axis). The x-axis has two categories: 'No' and 'Yes'. The y-axis is labeled 'Count'. The chart shows that for 'No' wireless service, there are approximately 4 'No' PDA owners and 5 'Yes' PDA owners. For 'Yes' wireless service, there are approximately 2 'No' PDA owners and 6 'Yes' PDA owners. A legend indicates 'Cluster on X: set color' and 'Owns PDA'.
- Gallery:** A tabbed interface with 'Gallery' selected. It shows a list of chart types: Favorites, Bar, Line, Area, Pie/Polar, Scatter/Dot, Histogram, High-Low, Boxplot, and Dual Axes. The 'Bar' chart type is selected.
- Buttons:** 'Element Properties...', 'Options...', 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.



# ΔΗΜΙΟΥΡΓΙΑ ΓΡΑΦΗΜΑΤΩΝ

- Το γράφημα εμφανίζεται στον SPSS Viewer
- Κάνοντας διπλό κλικ πάνω στο γράφημα μπορείτε να το επεξεργαστείτε
- Μπορείτε να επεξεργαστείτε να αναδιατάξετε όλα τα γραφήματα και τους πίνακες που έχουν προκύψει στον Output Viewer



The screenshot shows the SPSS Output Viewer window displaying a pivot table titled "Marital status". The table has columns for Frequency, Percent, Valid Percent, and Cumulative Percent. The rows are Valid Unmarried, Married, and Total.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Unmarried	3224	50.4	50.4	50.4
Married	3176	49.6	49.6	100.0
Total	6400	100.0	100.0	

The screenshot shows the SPSS Output Viewer window displaying a pivot table titled "Level of education". The table has columns for Frequency, Percent, Valid Percent, and Cumulative Percent. The rows are Valid Did not complete high school, High school degree, Some college, College degree, Post-undergraduate degree, and Total.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Did not complete high school	1390	21.7	21.7	21.7
High school degree	1936	30.3	30.3	52.0
Some college	1360	21.3	21.3	73.2
College degree	1355	21.2	21.2	94.4
Post-undergraduate degree	359	5.6	5.6	100.0
Total	6400	100.0	100.0	

# ΤΡΟΠΟΠΟΙΗΣΗ ΣΥΝΕΧΩΝ ΜΕΤΑΒΛΗΤΩΝ

- Transform > Visual Binning
- Make Cutpoints > Equal Width Intervals

**Make Cutpoints**

**Equal Width Intervals**

Intervals - fill in at least two fields

First Cutpoint Location:

Number of Cutpoints:

Width:

Last Cutpoint Location:

**Equal Percentiles Based on Scanned Cases**

Intervals - fill in either field

Number of Cutpoints:

Width(%):

**Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases**

+/- 1 Std. Deviation

+/- 2 Std. Deviation

+/- 3 Std. Deviation

**i** Apply will replace the current cutpoint definitions with this specification. A final interval will include all remaining values: N cutpoints produce N+1 intervals.

**Visual Binning**

Scanned Variable List:

- Household income in th...

Current Variable:  Name:

Binned Variable:  Label:

Minimum:  Nonmissing Values:  Maximum:

Grid: **i** Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value of 10, for example, defines an interval starting above the previous interval and ending at 10.

	Value	Label
1		HIGH
2		

Cases Scanned:  Missing Values:

Copy Bins:

Upper Endpoints:

Included (=>)

Excluded (<=)

Reverse scale

# ΤΡΟΠΟΠΟΙΗΣΗ ΣΥΝΕΧΩΝ ΜΕΤΑΒΛΗΤΩΝ

- **Επιλέξτε Make Labels** για αυτόματη δημιουργία ετικετών
- Μπορείτε να σύρετε τις κάθετες μπάρες ώστε να προσδιορίσετε τα cutpoints
- Δημιουργήσατε μια νέα μεταβλητή που εμφανίζεται στην τελευταία γραμμή του Variable View

**Visual Binning**

Scanned Variable List:

- Household income in th...

Name: Household income in thousands

Label: Household income in thousands (Binned)

Current Variable: income

Binned Variable: inccat2

Minimum: 9.00 Nonmissing Values Maximum: 1116.00

Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value of 10, for example, defines an interval starting above the previous interval and ending at 10.

	Value	Label
1	25.000 < 25.00	
2	50.000 25.00 - 49.00	
3	75.000 50.00 - 74.00	
4	HIGH 75.00+	
5		

Upper Endpoints:

- Included (<=)
- Excluded (<)

Buttons: Make Cutpoints..., Make Labels, Reverse scale

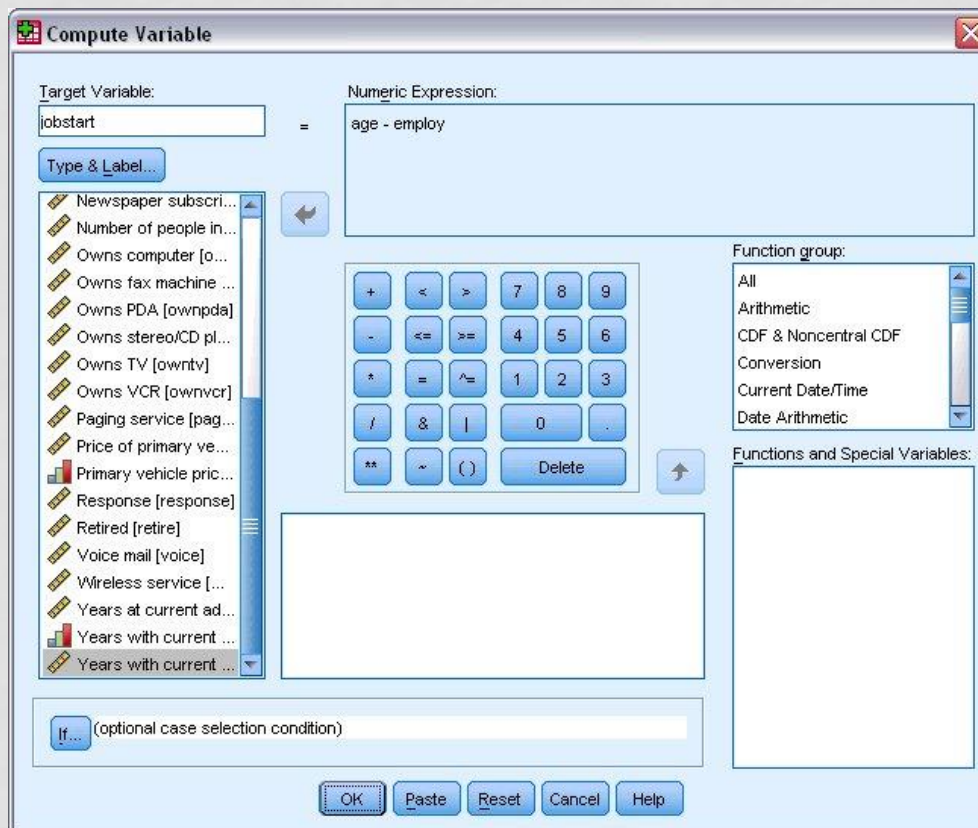
Buttons: OK, Paste, Reset, Cancel, Help

# ΔΗΜΙΟΥΡΓΙΑ ΝΕΑΣ ΜΕΤΑΒΛΗΤΗΣ

- Μεγάλο εύρος μαθηματικών συναρτήσεων για τη δημιουργία νέων μεταβλητών από τις υπάρχουσες
- Π.χ. γνωρίζουμε την ηλικία των συμμετεχόντων και τον αριθμό των ετών που βρίσκονται στην τρέχουσα δουλειά
- Μπορούμε να δημιουργήσουμε μια νέα μεταβλητή που να δείχνει προσεγγιστικά την ηλικία που ξεκίνησαν την τρέχουσα δουλειά

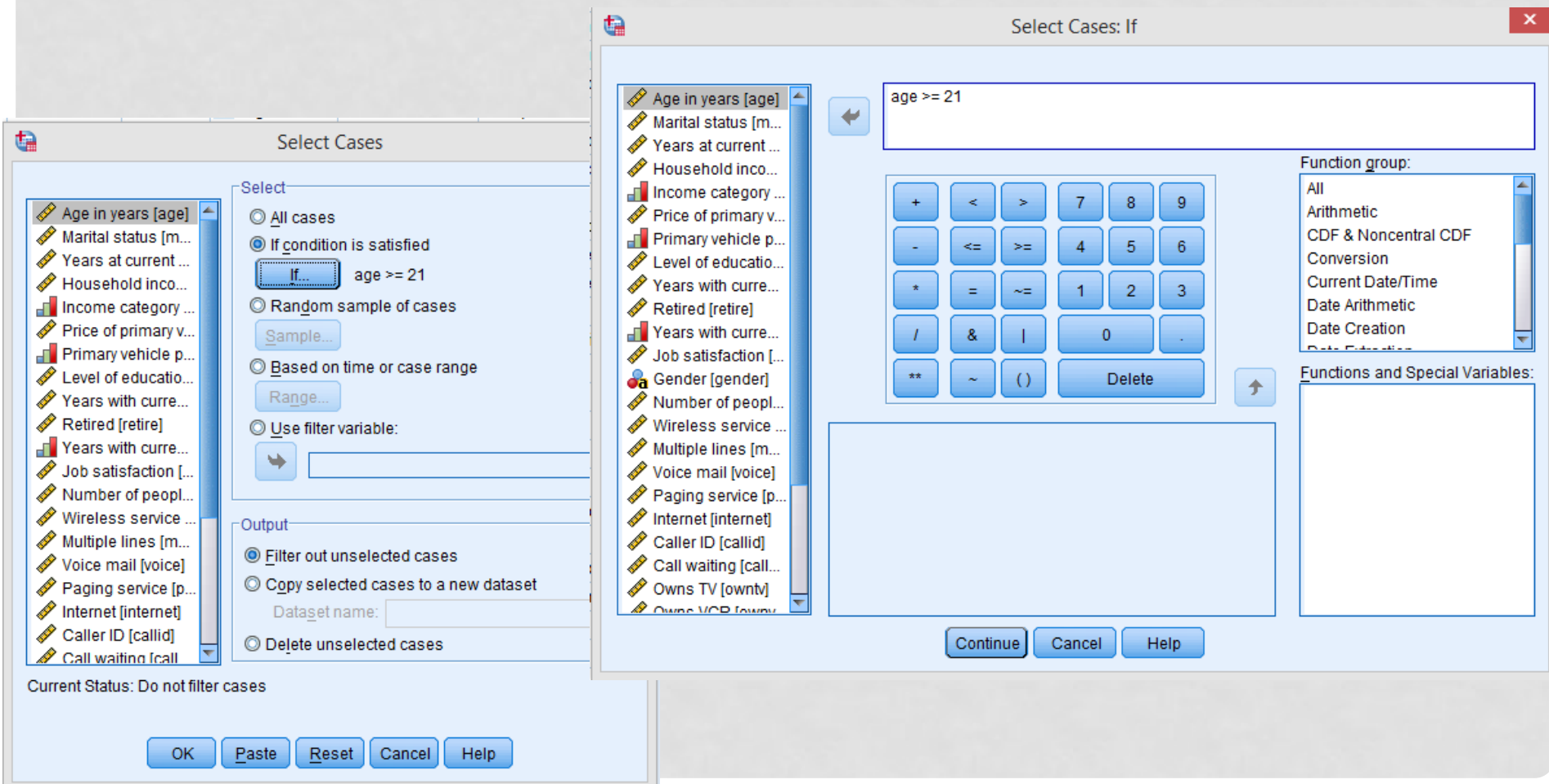
# ΔΗΜΙΟΥΡΓΙΑ ΝΕΑΣ ΜΕΤΑΒΛΗΤΗΣ

- Transform > Compute Variable...



# ΕΠΙΛΟΓΗ ΠΕΡΙΠΤΩΣΕΩΝ

- Data > Select Cases...



The image shows two overlapping dialog boxes from the SPSS software. The background dialog is the 'Select Cases' dialog, and the foreground dialog is the 'Select Cases: If' dialog.

**Select Cases Dialog (Background):**

- Select:**
  - All cases
  - If condition is satisfied
    - if... age >= 21
  - Random sample of cases
    - Sample...
  - Based on time or case range
    - Range...
  - Use filter variable:
- Output:**
  - Filter out unselected cases
  - Copy selected cases to a new dataset
    - Dataset name:
  - Delete unselected cases

Current Status: Do not filter cases

Buttons: OK, Paste, Reset, Cancel, Help

**Select Cases: If Dialog (Foreground):**

- Function group:**
  - All
  - Arithmetic
  - CDF & Noncentral CDF
  - Conversion
  - Current Date/Time
  - Date Arithmetic
  - Date Creation
  - Date Extraction
- Functions and Special Variables:**

Buttons: Continue, Cancel, Help



# ΕΠΙΛΟΓΗ ΠΕΡΙΠΤΩΣΕΩΝ

\*osteological data.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 10 of 10 Variables

	sex	bday	arthritis	osteophytosis	height	bmass	heightm	heightcat	arthritis2	filter_\$	var
1	m	05.05.1958	6	3	182	80	1,82	4	3	0	
2	f	16.02.1951	4	3	168	59	1,68	2	2	1	
3	m	22.03.1971	2	1	178	85	1,78	3	1	0	
4	f	15.04.1975	1	1	163	65	1,63	2	1	1	
5	f	09.08.1965	2	2	160	0	1,60	1	1	1	
6	m	23.02.1952	4	2	168	75	1,68	2	2	0	
7	m	06.08.1956	2	1	172	80	1,72	3	1	0	
8	f	06.05.1976	1	1	154	50	1,54	1	1	1	
9	f	23.11.1970	2	1	156	55	1,56	1	1	1	
10	f	03.12.1970	1	1	163	69	1,63	2	1	1	
11	f	07.02.1959	2	3	170	80	1,70	2	1	1	
12	m	11.11.1966	1	1	181	92	1,81	4	1	0	
13	f	15.01.1969	3	2	158	49	1,58	1	2	1	
14	m	26.12.1949	2	1	176	83	1,76	3	1	0	

Data View Variable View

IBM SPSS Statistics Processor is ready Filter On

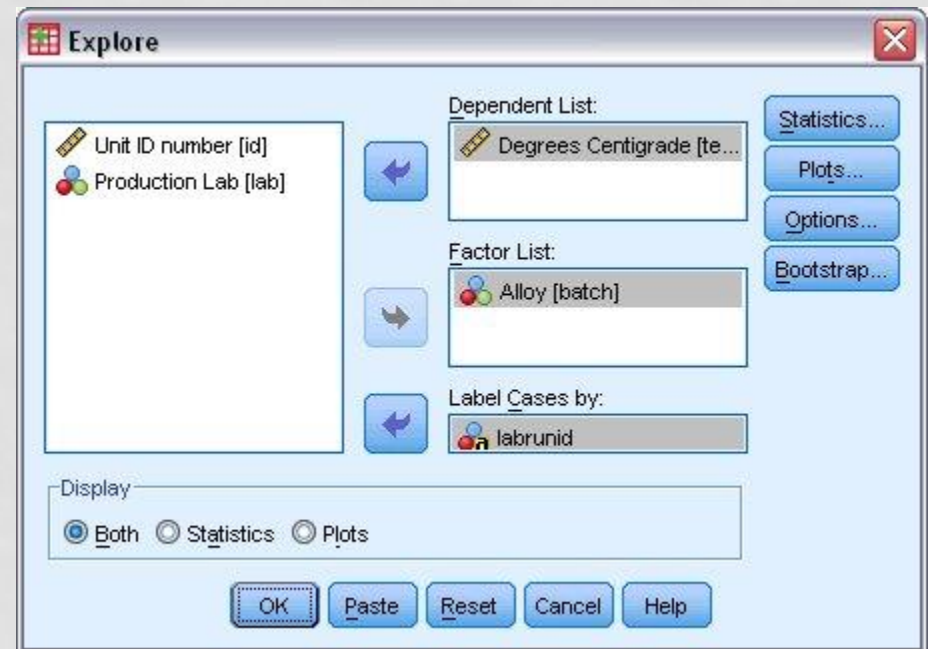
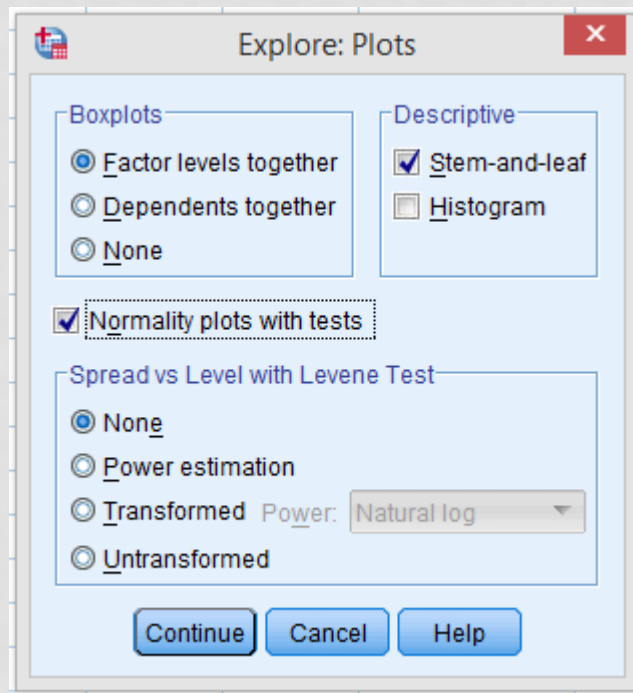


# ΚΑΤΑΝΟΜΕΣ

- Μια κατασκευαστική επιχείρηση χρησιμοποιεί ασημένιο νιτρίδιο για να δημιουργήσει κεραμικά ρουλεμάν που αντέχουν σε θερμοκρασία 1500 βαθμών Κελσίου ή υψηλότερη. Η θερμική αντίσταση ενός προτύπου κράματος είναι γνωστό ότι είναι κανονικά κατανομημένη. Ωστόσο, ένα νέο κράμα πριμοδότησης είναι υπό δοκιμή, και η κατανομή του είναι άγνωστη.

# ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

- **Analyze > Descriptive Statistics > Explore...**



# ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

Alloy: Premium

			Statistic	Std. Error
Degrees	Mean		1542.0787	.61165
Centigrade	95% Confidence Interval for Mean	Lower Bound	1540.8738	
		Upper Bound	1543.2836	
	5% Trimmed Mean		1541.2805	
	Median		1539.7181	
	Variance		89.789	
	Std. Deviation		9.47569	
	Minimum		1530.44	
	Maximum		1591.04	
	Range		60.61	
	Interquartile Range		11.5051	
	Skewness		1.439	.157
	Kurtosis		3.036	.313

Alloy: Standard

			Statistic	Std. Error
Degrees	Mean		1514.6564	.62004
Centigrade	95% Confidence Interval for Mean	Lower Bound	1513.4350	
		Upper Bound	1515.8779	
	5% Trimmed Mean		1514.7302	
	Median		1514.5317	
	Variance		92.269	
	Std. Deviation		9.60566	
	Minimum		1488.30	
	Maximum		1537.99	
	Range		49.69	
	Interquartile Range		13.5098	
	Skewness		-.078	.157
	Kurtosis		-.343	.313

# ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

- $H_0$ : normal distribution
- $H_1$ : otherwise

	Alloy	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Degrees Centigrade	Premium	.123	240	.000	.888	240	.000
	Standard	.027	240	.200*	.995	240	.602

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

a. Lilliefors Significance Correction

## Stem-and-Leaf Plots

Degrees Centigrade Stem-and-Leaf Plot for  
batch= Premium

```
Frequency  Stem & Leaf
24,00      153 . 0000000111111111111111
22,00      153 . 22222222233333333333
26,00      153 . 4444444444555555555555
26,00      153 . 666666666666666777777777
24,00      153 . 8888888888888899999999
19,00      154 . 000000000111111111
25,00      154 . 222222222222222333333333
10,00      154 . 4444455555
12,00      154 . 666666667777
10,00      154 . 8888999999
 8,00      155 . 00111111
 4,00      155 . 2223
 6,00      155 . 445555
 6,00      155 . 666667
 6,00      155 . 888899
 3,00      156 . 011
 3,00      156 . 223
 6,00      Extremes (>=1566)
```

Stem width: 10,00  
Each leaf: 1 case(s)

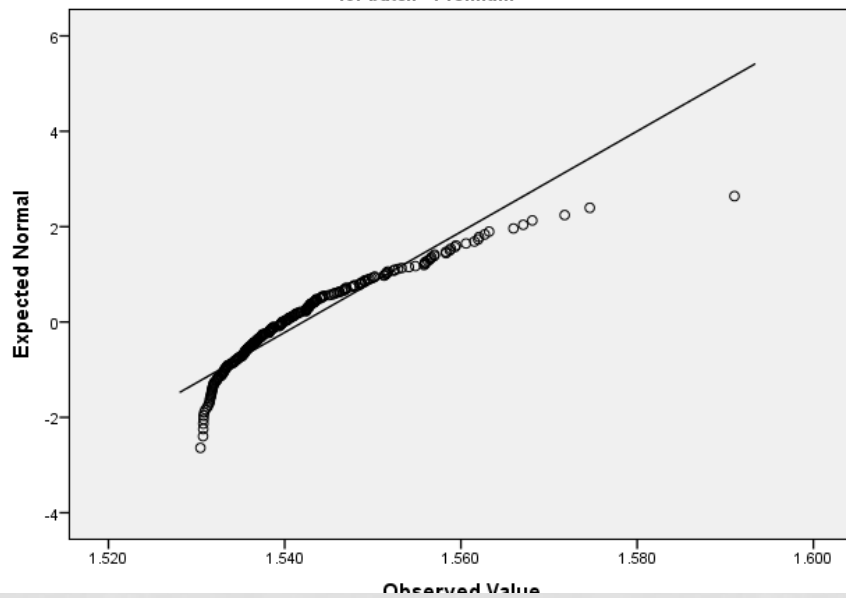
Degrees Centigrade Stem-and-Leaf Plot for  
batch= Standard

```
Frequency  Stem & Leaf
 2,00      148 . 88
 2,00      149 . 44
12,00      149 . 566677788999
22,00      150 . 000111112233333333444
35,00      150 . 555555566666777777777788888999
54,00      151 . 00000001111111111122222233333333333333334444444444
43,00      151 . 55555666666666777777777777888899999999
32,00      152 . 00000000111111122223333333334444
22,00      152 . 5555556666667777888899
15,00      153 . 000000122223344
 1,00      153 . 7
```

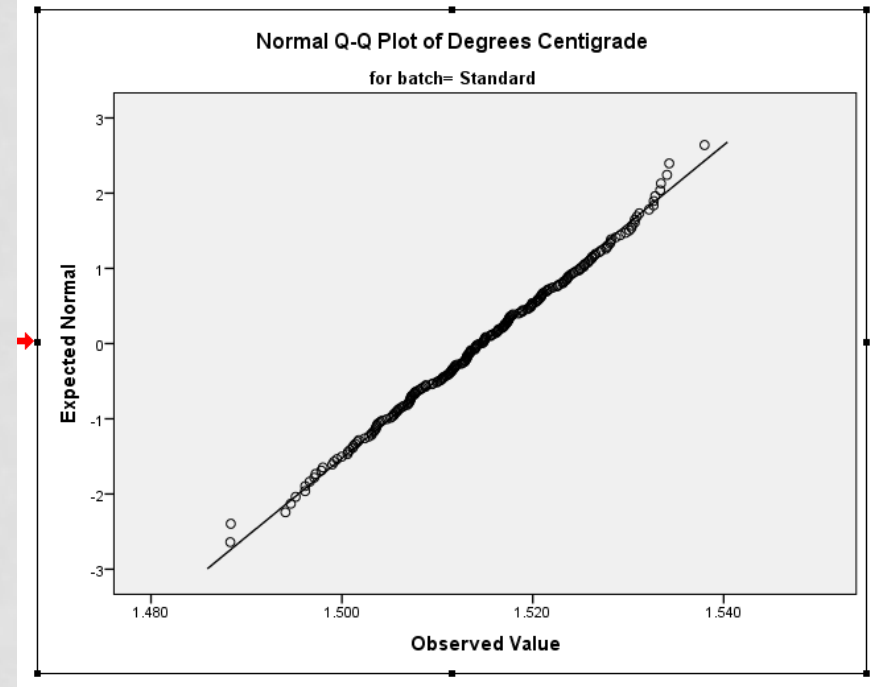
Stem width: 10,00  
Each leaf: 1 case(s)

# ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

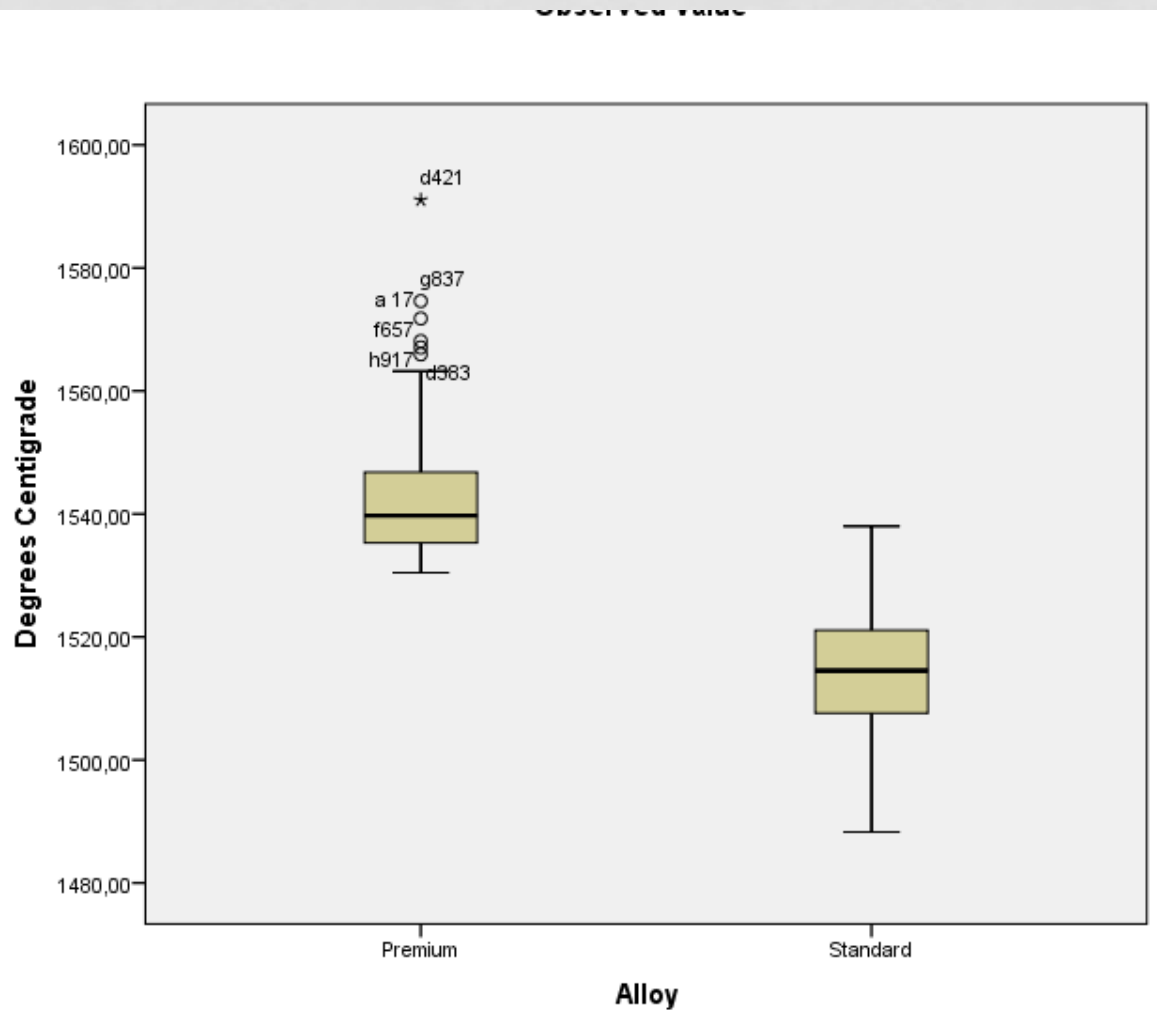
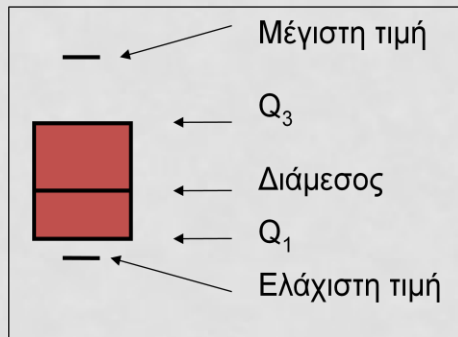
Normal Q-Q Plot of Degrees Centigrade  
for batch= Premium



Normal Q-Q Plot of Degrees Centigrade  
for batch= Standard



# ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ



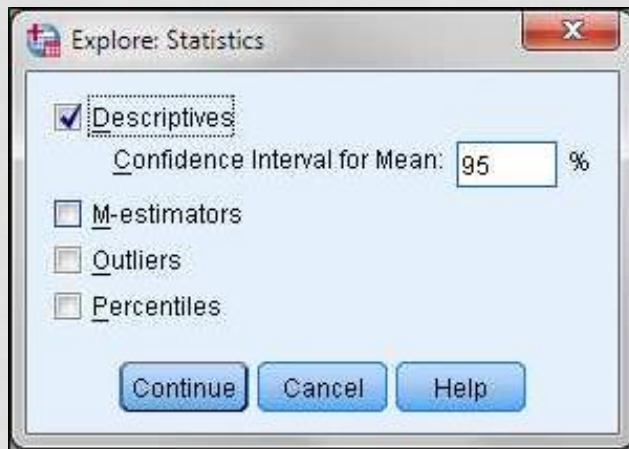
# ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ

- Σκοπός είναι να προσδιορίσουμε ή να εκτιμήσουμε την πραγματική τιμή  $\mu$  της μεταβλητής  $x$  που μελετάμε και την τυπική απόκλιση  $\sigma$
- $p\%$  διάστημα εμπιστοσύνης μιας παραμέτρου  $\theta$  του πληθυσμού είναι το διάστημα μέσα στο οποίο αναμένεται να υπάρχει η  $\theta$  με πιθανότητα  $p\%$
- Το 95% διάστημα εμπιστοσύνης είναι αυτό που χρησιμοποιείται ευρέως και υπολογίζεται σε δείγματα που ακολουθούν κανονική κατανομή



# ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ

- Analyze > Descriptive Statistics > Explore
- Επιλέγουμε Statistics



			Statistic	Std. Error
sex				
height	f	Mean	161,52	2,246
		95% Confidence Interval for Mean	Lower Bound 156,84 Upper Bound 166,21	
		5% Trimmed Mean	160,98	
		Median	162,00	
		Variance	105,962	
		Std. Deviation	10,294	
		Minimum	148	
		Maximum	185	
		Range	37	
		Interquartile Range	15	
		Skewness	,707	,501
		Kurtosis	,208	,972
		m	m	Mean
95% Confidence Interval for Mean	Lower Bound 172,42 Upper Bound 179,37			
5% Trimmed Mean	176,03			
Median	178,00			
Variance	83,596			
Std. Deviation	9,143			
Minimum	159			
Maximum	190			
Range	31			
Interquartile Range	15			
Skewness	-,274			,434
Kurtosis	-,989			,845

# ΕΛΕΓΧΟΙ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ

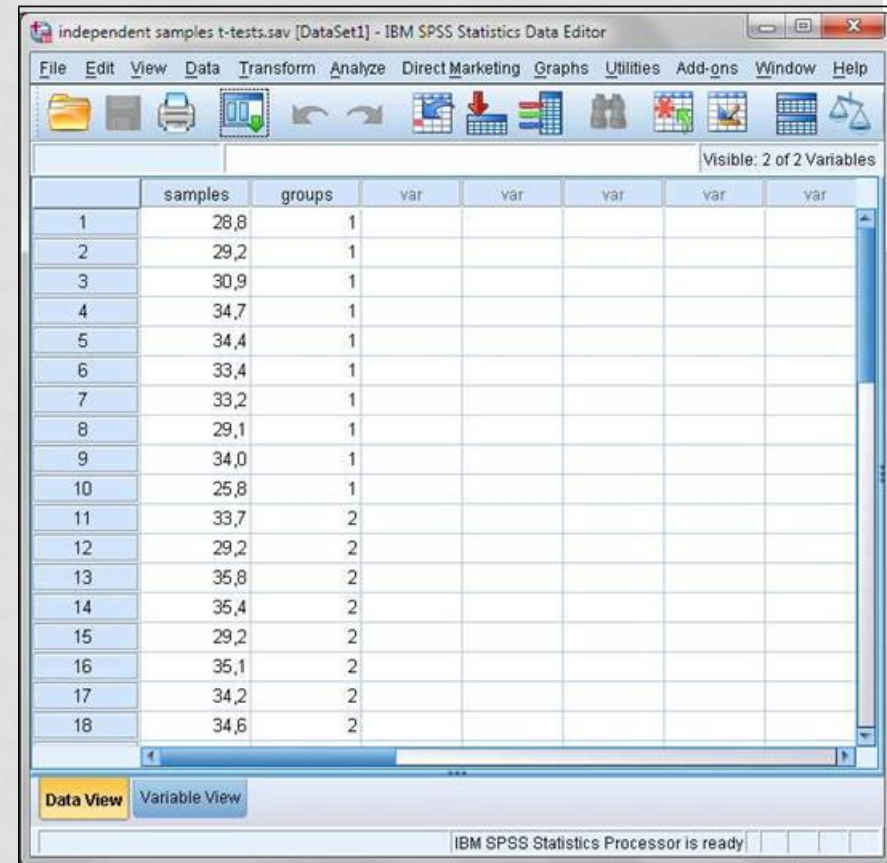
- Η πιθανότητα να συμβεί ή όχι ένα γεγονός
- Για να λάβουμε αποφάσεις πρέπει να κάνουμε υποθέσεις
- Η βασική υπόθεση ονομάζεται μηδενική απόφαση,
  - $H_0$ : δεν υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των δειγμάτων
    - $H_1$ : εναλλακτική
- Σφάλμα τύπου I: απορρίψουμε μια υπόθεση που είναι αληθινή
- Σφάλμα τύπου II: δεχθούμε μια λανθασμένη υπόθεση

# ΕΛΕΓΧΟΙ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ

- Επίπεδο σημαντικότητας: μέγιστη πιθανότητα με την οποία δεχόμαστε να κάνουμε σφάλμα τύπου I
- $p$ -value: υπολογίζεται από το SPSS, δείχνει την πιθανότητα να κάνουμε λάθος απορρίπτοντας την  $H_0$ .
- Αν  $p < \alpha$  τότε η  $H_0$  απορρίπτεται
- Αν  $p > \alpha$  τότε η  $H_0$  δεν απορρίπτεται

# ΔΙΑΦΟΡΕΣ ΜΕΣΩΝ ΤΙΜΩΝ ΔΕΙΓΜΑΤΩΝ

- Συγκρίνουμε δυο ανεξάρτητα δείγματα για να δούμε αν υπάρχει στατιστική διαφορά μεταξύ τους. Π.χ. σύγκριση ύψους μαθητών μιας τάξης με μαθητές άλλης τάξης
- Μεταφέρουμε τα δυο δείγματα που θέλουμε να αναλύσουμε σε μια στήλη
- Στην διπλανή στήλη χρησιμοποιούμε τις τιμές 1 και 2 για να διαχωρίσουμε τα δείγματα



The screenshot shows the IBM SPSS Statistics Data Editor window. The title bar reads "independent samples t-tests.sav [DataSet1] - IBM SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and analysis. The main data grid shows two columns: "samples" and "groups". The "samples" column contains 18 rows of numerical values, and the "groups" column contains 18 rows of values (1 or 2). The status bar at the bottom indicates "Visible: 2 of 2 Variables" and "Data View".

	samples	groups	var	var	var	var	var
1	28,8	1					
2	29,2	1					
3	30,9	1					
4	34,7	1					
5	34,4	1					
6	33,4	1					
7	33,2	1					
8	29,1	1					
9	34,0	1					
10	25,8	1					
11	33,7	2					
12	29,2	2					
13	35,8	2					
14	35,4	2					
15	29,2	2					
16	35,1	2					
17	34,2	2					
18	34,6	2					

# ΔΙΑΦΟΡΕΣ ΜΕΣΩΝ ΤΙΜΩΝ ΔΕΙΓΜΑΤΩΝ

- Analyze>  
Compare  
Means>  
Independent-  
Samples T-Test

groups	N	Mean	Std. Deviation	Std. Error Mean
samples 1	10	31,350	3,0252	,9566
2	8	33,400	2,6753	,9459

- Η  $H_0$  δεν απορρίπτεται άρα δεν υπάρχει στατιστικά σημαντική διαφορά στις διασπορές των δυο δειγμάτων

	Levene's Test for Equality of Variances		t-test for Equality of Means				
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
	Equal variances assumed	,567	,462	-1,502	16	,153	-2,0500
Equal variances not assumed			-1,524	15,793	,147	-2,0500	1,3453

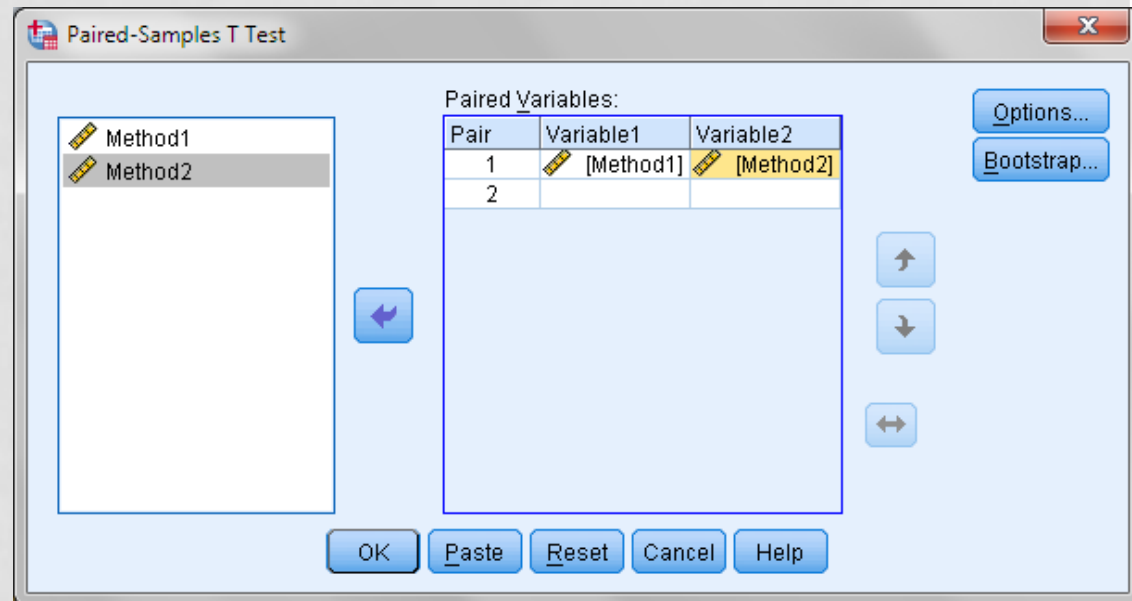
# ΣΥΓΚΡΙΣΗ ΖΕΥΓΩΝ ΔΕΙΓΜΑΤΩΝ

- Δυο δείγματα σχηματίζουν ένα ζεύγος αν υπάρχει ένα προς ένα αντιστοιχία μεταξύ των τιμών των δειγμάτων
- Π.χ. χρονολόγηση 8 ταφικών αντικειμένων με δυο διαφορετικές μεθόδους
- Υπάρχει στατιστικά σημαντική διαφορά μεταξύ των μετρήσεων των δυο μεθόδων?

	Method1	Method2	var	var	var	var	var
1	3,9	4,3					
2	3,4	3,5					
3	3,1	3,3					
4	2,8	3,0					
5	4,1	3,9					
6	3,3	3,1					
7	3,5	3,8					
8	3,4	3,6					
9							

# ΣΥΓΚΡΙΣΗ ΖΕΥΓΩΝ ΔΕΙΓΜΑΤΩΝ

- Analyze >  
Compare Means >  
Paired-Samples T-  
Test



**Paired Samples Test**

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Method1 - Method2	-,1250	,2188	,0773	-,3079	,0579	-1,616	7	,150