

# ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

---

- 5. Στατιστική συμπερασματολογία για ποσοτικές μεταβλητές:  
Έλεγχοι υποθέσεων και διαστήματα εμπιστοσύνης**

# Διαστήματα εμπιστοσύνης και έλεγχοι υποθέσεων για τη μέση τιμή

---

- Για μια ποσοτική μεταβλητή μας ενδιαφέρει για την (άγνωστη) μέση τιμή της στον πληθυσμό:
  - Έλεγχος υπόθεσης
    - $H_0: \mu = \mu_0$
    - $H_1: \mu \neq \mu_0$
  - Εύρεση δ.ε. (π.χ. 90%, 95%, 99%)
- Χρησιμοποιούμε:

$$\bar{x}, s^2, n$$

# Παράδειγμα (hobbit restaurant)

---

- Από το δείγμα μας μπορούμε να ισχυριστούμε ότι οι άνθρωποι της περιοχής που μας ενδιαφέρει ξοδεύουν μηνιαία κατά μέσο όρο \$200 για εστιατόρια;
  - ή ακριβέστερα: ο πραγματικός μέσος όρος ( $\mu$ ) διαφέρει σημαντικά από τα \$200;
- Να βρεθούν τα 90%, 95%, 99% δ.ε. Για την πραγματική μέση τιμή

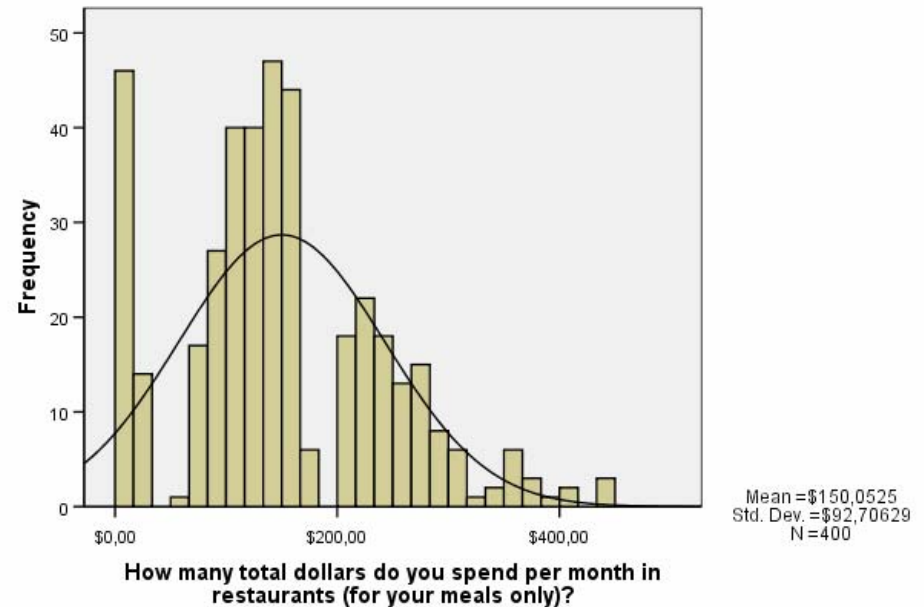
# Στατιστικά μέτρα δείγματος (Frequencies)

## Statistics

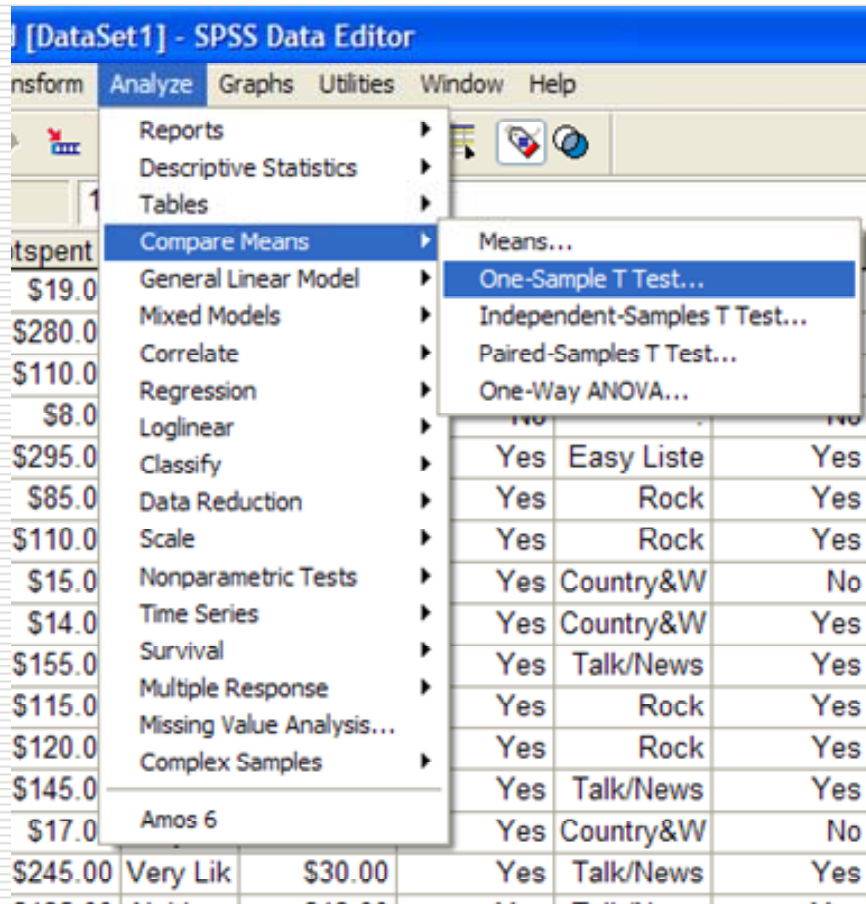
How many total dollars do you spend per month in restaurants (for your meals only)?

N	Valid	400
	Missing	0
Mean		\$150.0525
Std. Error of Mean		\$4.63531
Median		\$135.0000
Std. Deviation		\$92.70629
Variance		8594,456
Minimum		\$5.00
Maximum		\$450.00

## Histogram



# Διαδικασία One-sample T test



**Εκτέλεση του  
στατιστικού  
ελέγχου  
Student's t-test  
για ένα δείγμα**

# Ορισμός τιμής ελέγχου και επιπέδου εμπιστοσύνης

The image shows two dialog boxes from SPSS. The first, 'One-Sample T Test', has 'Test Variable(s): How many total dollars' and 'Test Value: 200'. The second, 'One-Sample T Test: Options', has 'Confidence Interval: 90 %' and 'Exclude cases analysis by analysis' selected. Red arrows point from the '200' and '90 %' fields to yellow callout boxes below.

$H_0: \mu = 200$   
 $H_1: \mu \neq 200$

90% δ.ε. για το  $\mu = 200$

# Αποτελέσματα

## One-Sample Test

	Test Value = 200					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
How many total dollars do you spend per month in restaurants (for your meals only)?	-10,775	399	,000	-\$49.94750	-\$57.5897	-\$42.3053

**Επειδή sig = 0.000... < 0.05  
Απορρίπτεται η  $H_0$   
Δεχόμαστε την  $H_1: \mu \neq 200$**

$$-57.5897 \leq \mu - 200 \leq -42.3053 \Rightarrow 142.4103 \leq \mu \leq 157.6947$$

# Ερμηνεία αποτελεσμάτων

---

- Η (πραγματική) μέση τιμή των χρημάτων που διαθέτουν οι κάτοικοι της περιοχής για εστιατόρια κάθε μήνα διαφέρει σημαντικά από την τιμή ελέγχου (\$200)
- Το 90% δ.ε. για τη μέση τιμή είναι (\$142.41, \$157.69)



# Συνήθως θέτουμε $\mu_0=0$

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
How many total dollars do you spend per month in restaurants (for your meals only)?	32,372	399	,000	\$150.0525	\$142.4103	\$157.6947

**90% δ.ε. για το  $\mu$ :  
(\$142.41,\$157.69)**

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
How many total dollars do you spend per month in restaurants (for your meals only)?	32,372	399	,000	\$150.0525	\$140.9398	\$159.1652

**95% δ.ε. για το  $\mu$ :  
(\$140.94,\$159.17)**

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	99% Confidence Interval of the Difference	
					Lower	Upper
How many total dollars do you spend per month in restaurants (for your meals only)?	32,372	399	,000	\$150.0525	\$138.0553	\$162.0497

**99% δ.ε. για το  $\mu$ :  
(\$138.06,\$162.05)**

# Συμπέρασμα

---

- ❑ Θέτοντας τιμή ελέγχου  $\theta$  και ορίζοντας το επίπεδο εμπιστοσύνης παίρνουμε το αντίστοιχο δ.ε. για το  $\mu$
- ❑ Από το δ.ε. μπορούμε να ελέγξουμε και την υπόθεση που μας ενδιαφέρει
- ❑ Π.χ. Αφού το 99% δ.ε. δεν περιλαμβάνει το 200, μπορούμε να απορρίψουμε την  $H_0$  με «σιγουριά» 99%
- ❑ Το εύρος των δ.ε. αυξάνει όσο αυξάνει και το επίπεδο εμπιστοσύνης

# Εφαρμογές

---

- Να βρεθούν 95% και 99% δ.ε. για όλες τις ποσοτικές συνεχείς μεταβλητές των δεδομένων
  - anprice
  - ηλικία (προκύπτει από τη yrborn)
  - reincome

# Σύγκριση μέσων τιμών 2 ανεξάρτητων δειγμάτων

---

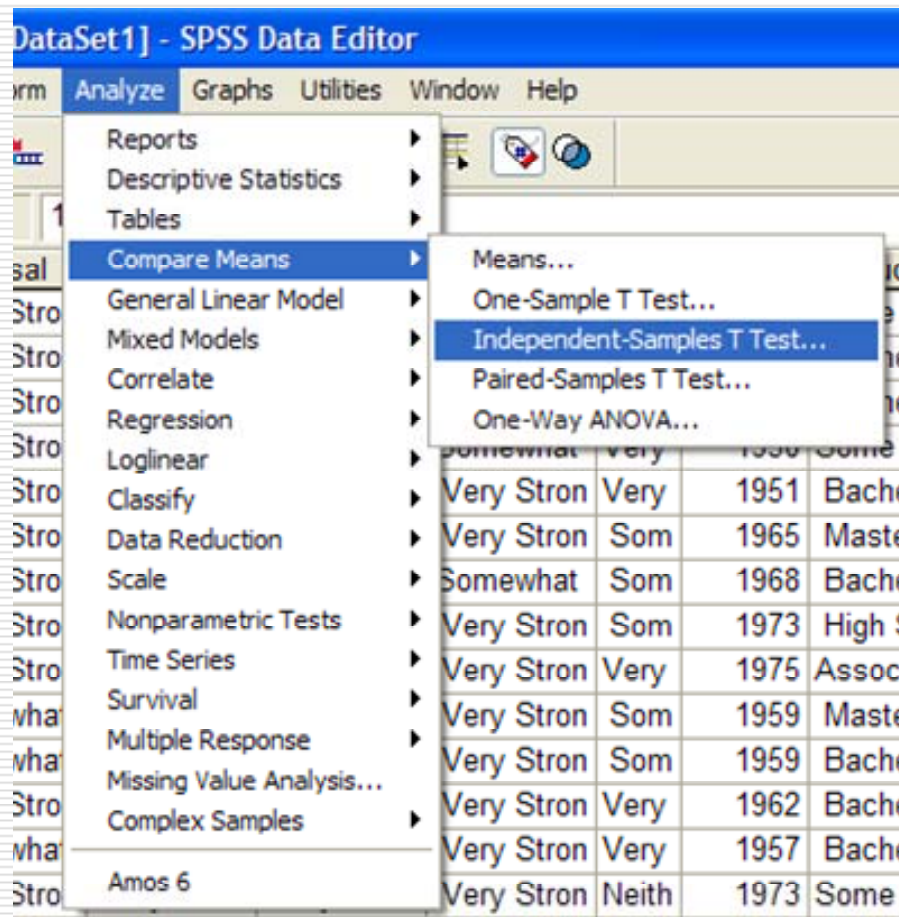
- Για μια ποσοτική μεταβλητή σε δύο ανεξάρτητους πληθυσμούς μας ενδιαφέρει για τις (άγνωστες) μέσες τιμές τους:
  - Έλεγχος υπόθεσης
$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$
  - Εύρεση δ.ε. για τη διαφορά  $\mu_1 - \mu_2$  (90%, 95%, 99%)

# Παράδειγμα

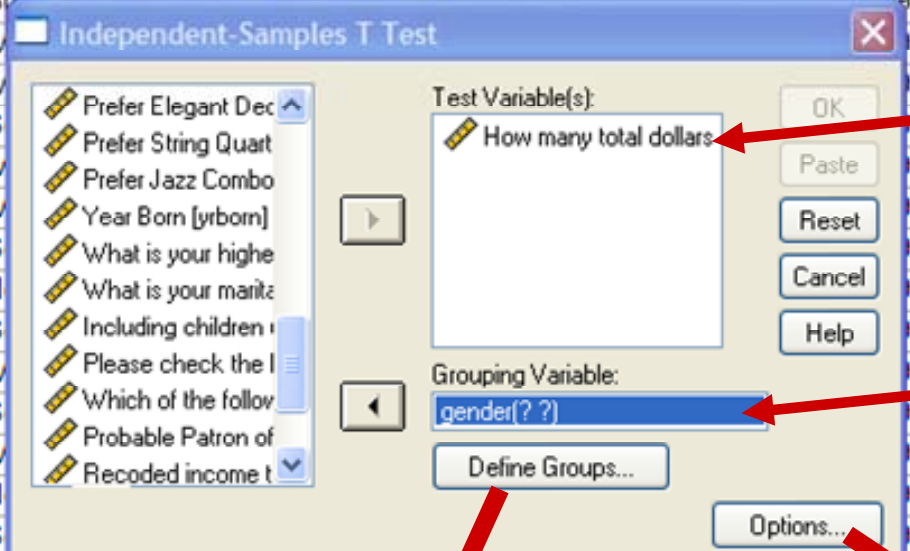
---

- Διαφέρουν οι μέσες τιμές των μηνιαίων χρημάτων που ξοδεύουν οι άντρες από αυτά που ξοδεύουν οι γυναίκες;
- Προσοχή! Απαιτείται ύπαρξη 2 μεταβλητών (στηλών): μια ποσοτική (αυτή που ενδιαφέρει να συγκρίνουμε) και μια κατηγορική (αυτή που ορίζει τους πληθυσμούς)
- Στο παράδειγμα θα χρησιμοποιηθούν οι `totspent` και `gender`

# Διαδικασία: Independent-Samples T Test

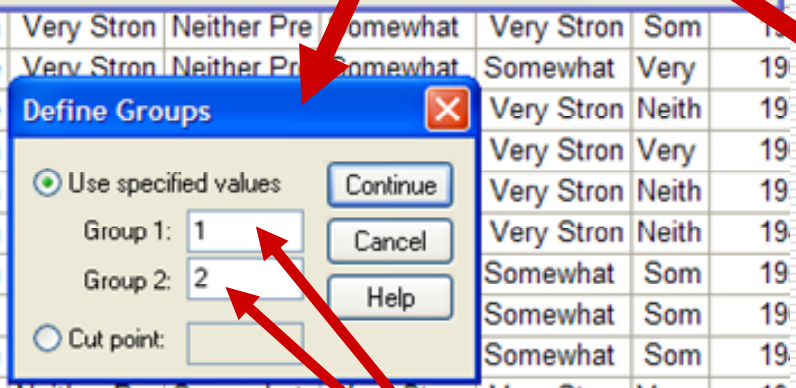


**Εκτέλεση του  
στατιστικού  
ελέγχου  
Student's t-test  
για δύο  
ανεξάρτητα  
δειγμάτα**

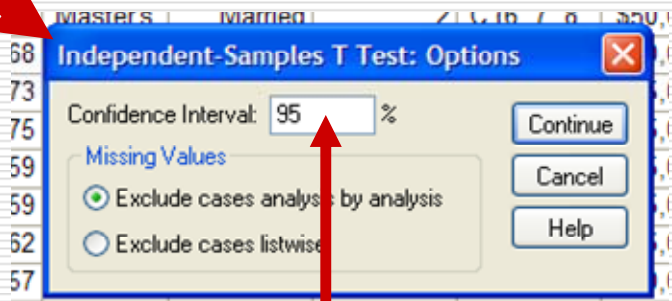


**ποσοτική μεταβλητή**

**κατηγορική μεταβλητή**



**τιμές που ορίζουν τους πληθυσμούς (ομάδες)**



**ορισμός δ.ε.**

# Αποτελέσματα - περιγραφικά

Group Statistics

	What is your gender?	N	Mean	Std. Deviation	Std. Error Mean
How many total dollars do you spend per month in restaurants (for your meals only)?	Male	204	\$154.1275	\$96.82936	\$6.77941
	Female	196	\$145.8112	\$88.25842	\$6.30417

**Περιγραφικά στατιστικά της ποσοτικής μεταβλητής στα δύο δείγματα**

**Είναι στατιστικά σημαντική η διαφορά αυτή;**



# Αποτελέσματα - συμπερασματολογία

Independent Samples Test

		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	95% Confidence Interval of the Difference	
									Lower	Upper
How many total dollars do you spend per month in restaurants (for your meals only)?	Equal variances assumed	,565	,453	,897	398	,370	\$8.31623	\$9.27476	-\$9.91742	\$26.54987
	Equal variances not assumed			,898	396,908	,370	\$8.31623	\$9.25759	-\$9.88382	\$26.51627

**1<sup>ο</sup> βήμα: Είναι οι διασπορές των πληθυσμών ίσες;**

**sig. = 0.453 > 0.05**  
**Οι διασπορές δεν έχουν διαφορά (θεωρούνται ίσες)**

# Αποτελέσματα - συμπερασματολογία

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
How many total dollars do you spend per month in restaurants (for your meals only)?		Equal variances assumed	,897	398	,370	\$8.31623	\$9.27476	-\$9.91742	\$26.54987
		Equal variances not assumed	,898	396,908	,370	\$8.31623	\$9.25759	-\$9.88382	\$26.51627

**2<sup>ο</sup> βήμα: Θεωρούμε το t-test για έλεγχο μέσων τιμών υποθέτοντας ίσες διασπορές**

**sig. = 0.370 > 0.05  
Οι μέσες τιμές δεν έχουν διαφορά (θεωρούνται ίσες)**

# Αποτελέσματα - συμπερασματολογία

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
How many total dollars do you spend per month in restaurants (for your meals only)?	Equal variances assumed	,565	,897	398	,370	\$8.31623	\$9.27476	-\$9.91742	\$26.54987
	Equal variances not assumed		,898	396,908	,370	\$8.31623	\$9.25759	-\$9.88387	\$26.51627

**3<sup>ο</sup> βήμα: Θεωρούμε το δ.ε. της διαφοράς των μέσων τιμών υποθέτοντας ίσες διασπορές**

**95% δ.ε. για τη διαφορά  $\mu_1 - \mu_2$ : (-9.92, 26.55)**

# Συμπεράσματα

---

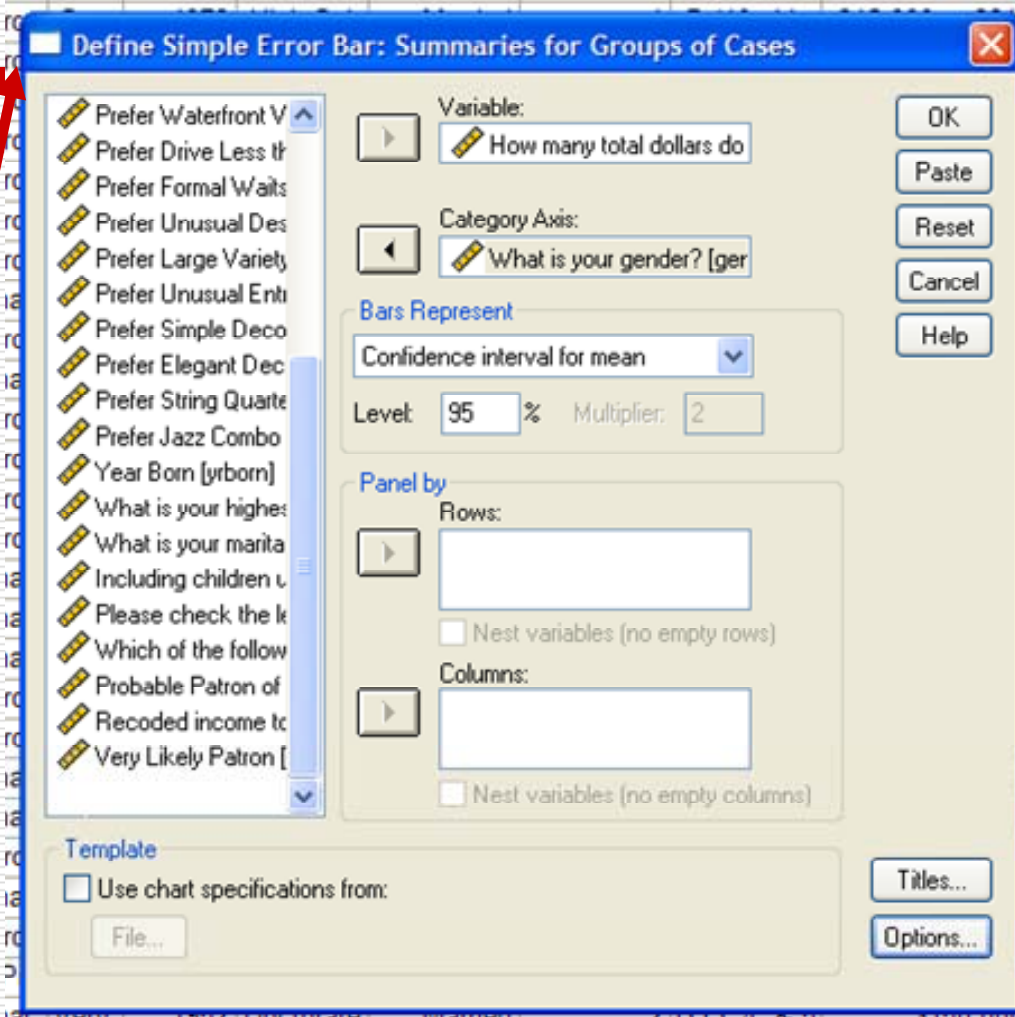
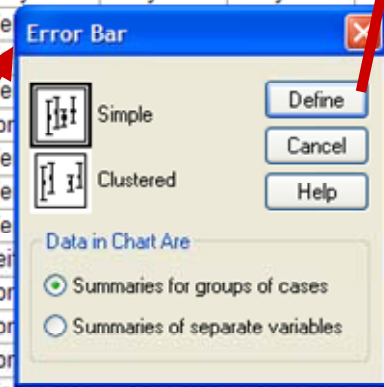
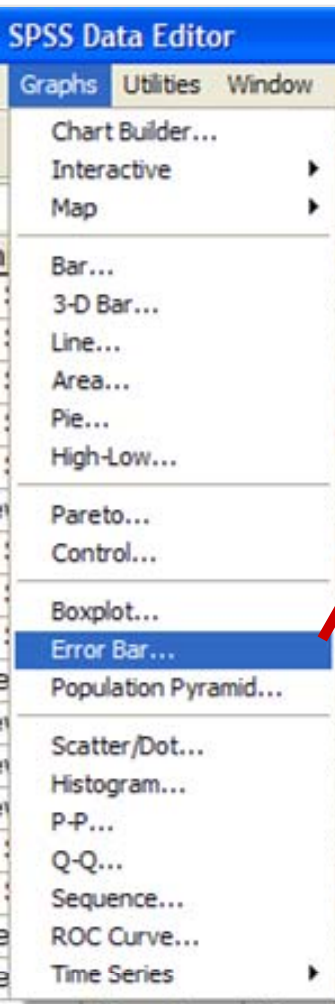
- Πρώτα εξετάζουμε αν είναι οι διασπορές των πληθυσμών ίσες
- Τα αποτελέσματα κάτω από την υπόθεση ίσων διασπορών γενικά διαφέρουν από αυτά κάτω από την υπόθεση άνισων διασπορών
- Εδώ οι διασπορές μπορούν να θεωρηθούν ίσες

# Συμπεράσματα (συν.)

---

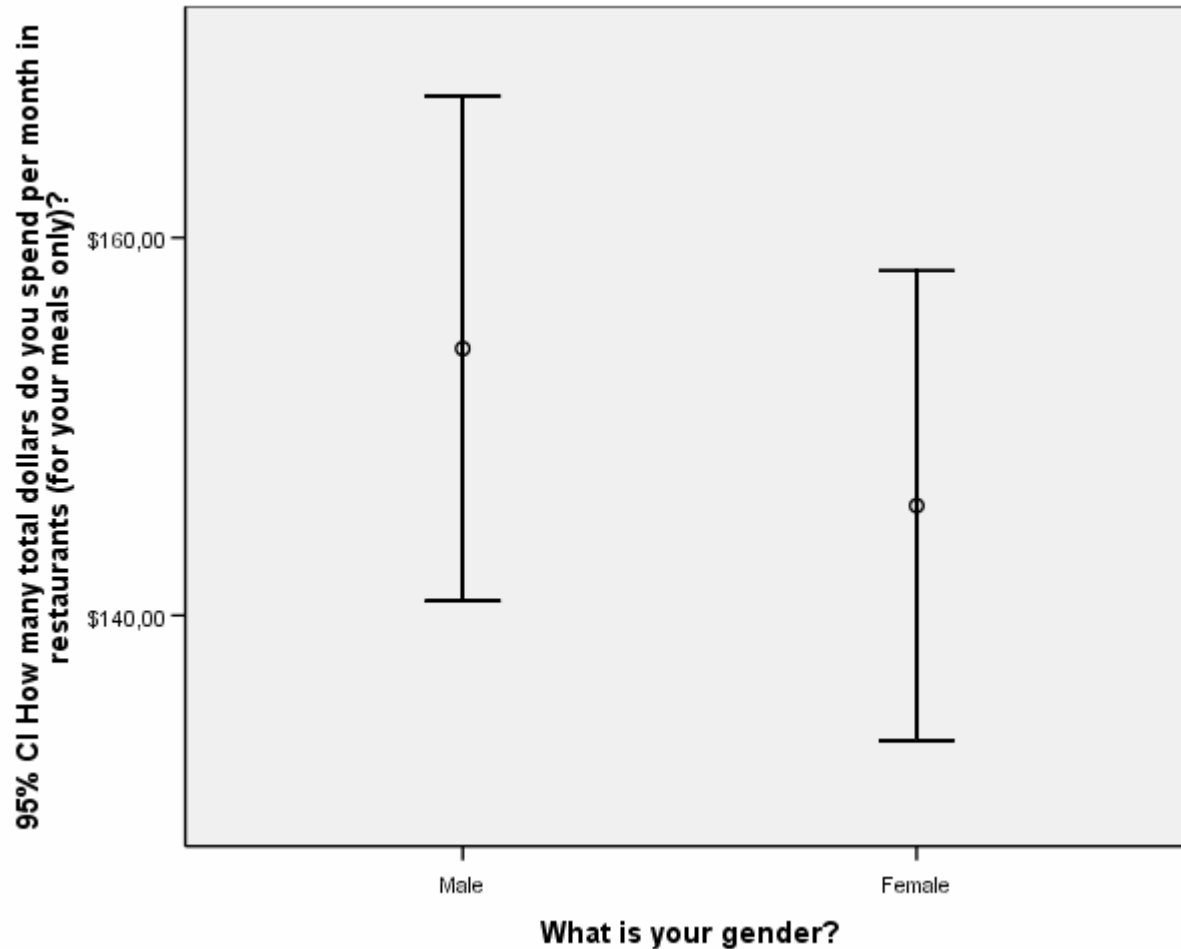
- Οι μέσες τιμές των χρημάτων που ξοδεύουν οι άνδρες και οι γυναίκες δεν παρουσιάζουν στατιστικά σημαντική διαφορά ( $\mu_1 = \mu_2$ )
- Το παραπάνω συμπέρασμα προκύπτει και από το 95% δ.ε. το οποίο περιέχει το 0

# Σχετικές γραφικές παραστάσεις: Error Bars



# Σχετικές γραφικές παραστάσεις: Error Bars

**δ.ε. για τη μέση τιμή  
κάθε φύλου  
ξεχωριστά: Παρόλο  
που για τους άνδρες  
φαίνονται  
υψηλότερες τιμές, η  
διαφορά δεν είναι  
στατιστικά σημαντική  
(προσέξτε την  
επικάλυψη των 2  
δ.ε.)**



# Εφαρμογές

---

- Να γίνουν οι έλεγχοι για διαφορές των μέσων τιμών όλων των ποσοτικών συνεχών μεταβλητών των δεδομένων (*totspent, avprice, ηλικία, reincome*) ανάμεσα σε 2 ομάδες που ορίζονται από δίτιμες κατηγορικές μεταβλητές (Yes/No) ή ακόμα και από άλλες κατηγορικές (π.χ. στην marital – married/single)



# Σύγκριση μέσων τιμών 2 εξαρτημένων δειγμάτων

---

- Για μια ποσοτική μεταβλητή (ή δύο παρόμοιες) που μετριέται στα ίδια ακριβώς άτομα μας ενδιαφέρει για τις (άγνωστες) μέσες τιμές τους:
  - Έλεγχος υπόθεσης
$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$
  - Εύρεση δ.ε. για τη διαφορά  $\mu_1 - \mu_2$  (90%, 95%, 99%)

# Παράδειγμα

---

- Διαφέρουν οι μέσες τιμές των βαθμολογιών στις μεταβλητές *variety* και *unusual*;
- Οι μετρήσεις έχουν γίνει στα ίδια άτομα (επομένως αναφερόμαστε σε έναν πληθυσμό)
- Προσοχή! Η θεώρηση της βαθμολογίας 1-5 ως ποσοτική συνεχής μεταβλητή γίνεται για τις ανάγκες του παραδείγματος. Για αυτές τις μεταβλητές καλύτεροι οι μη-παραμετρικοί έλεγχοι
- Απαραίτητη η ύπαρξη των δύο μεταβλητών σε δύο διαφορετικές στήλες

# Η διαδικασία Paired-Samples T Test

The screenshot illustrates the steps to perform a Paired-Samples T Test in SPSS. The 'Analyze' menu is open, showing the path: Analyze > Compare Means > Paired-Samples T Test... The 'Paired-Samples T Test' dialog box is open, showing the following variables selected:

- Variable 1: Do you eat at this time
- Variable 2: How many total dollars

The 'Paired-Samples T Test: Options' dialog box is also open, showing the following settings:

- Confidence Interval: 95 %
- Missing Values:  Exclude cases analysis by analysis

The background shows a data table with columns labeled 'Very Stron', 'Very Stron', and 'Somewhat'.

# Αποτελέσματα – αρχικά στατιστικά

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Prefer Large Variety of Entrees	2,48	400	1,466	,073
	Prefer Unusual Entrees	2,41	400	1,550	,077

**η διαφορά δε φαίνεται μεγάλη (πρέπει να επιβεβαιωθεί με t-test)**

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Prefer Large Variety of Entrees & Prefer Unusual Entrees	400	,831	,000

**η συσχέτιση είναι σημαντική (sig=0.000...<0.05)  
Δικαιολογείται η χρήση Paired t-test.**

# Αποτελέσματα – paired 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			
Pair 1	Prefer Large Variety of Entrees - Prefer Unusual Entrees	,070	,882	,044	-,017	,157	1,588	399	,113

**95% δ.ε. για τη διαφορά  $\mu_1 - \mu_2$ :  
(-0.017, 0.157)**

**sig. = 0.113 > 0.05  
Οι μέσες τιμές δεν έχουν διαφορά  
(θεωρούνται ίσες)**

# Συμπεράσματα

---

- ❑ Οι μέσες τιμές των βαθμολογιών που δίνουν τα ίδια άτομα σε 2 διαφορετικές ερωτήσεις δεν παρουσιάζουν στατιστικά σημαντική διαφορά ( $\mu_1 = \mu_2$ )
- ❑ Το παραπάνω συμπέρασμα προκύπτει και από το 95% δ.ε. το οποίο περιέχει το 0

# Εφαρμογές

---

- Να γίνουν παρόμοιοι έλεγχοι και για άλλα ζευγάρια βαθμολογιών:
- water-simple
- water-jazz
- ....