

5° ΕΡΓΑΣΤΗΡΙΟ

Σε αυτό το εργαστήριο θα παρουσιάσουμε κάποιες απλές τεχνικές επαγωγικής στατιστικής. Η επαγωγική στατιστική μας επιτρέπει να ελέγξουμε εάν τα αποτελέσματα από την περιγραφική ανάλυση των δεδομένων ενός δείγματος μπορούν να γενικευτούν στον ευρύτερο πληθυσμό από τον οποίο προέρχεται το δείγμα. Συγκεκριμένα, θα παρουσιάσουμε το στατιστικό κριτήριο (τεστ) χ^2 , μια δοκιμασία που χρησιμοποιείται για να ελέγξουμε εάν δύο μεταβλητές τύπου nominal ή ordinal είναι ανεξάρτητες ή εξαρτημένες (δηλ. οι τιμές της μίας μεταβλητής εξαρτώνται από τις τιμές της άλλης). Επίσης, θα παρουσιάσουμε τους δείκτες συσχέτισης Pearson r και Spearman Rho που χρησιμοποιούνται όταν θέλουμε να εξετάσουμε το βαθμό αλληλεξάρτησης (τη συσχέτιση) μεταξύ δύο μεταβλητών τύπου scale ή δύο μεταβλητών τύπου ordinal, αντίστοιχα.

1. Το στατιστικό κριτήριο χ^2 (chi-square test): Σύγκριση συχνοτήτων κατηγοριών

Ο χ^2 έλεγχος ανεξαρτησίας εφαρμόζεται για τη διερεύνηση της εξάρτησης μεταξύ δυο μεταβλητών τύπου NOMINAL ή ORDINAL. Τα βήματα για τη διεξαγωγή οποιουδήποτε στατιστικού τεστ είναι τα εξής:

1. Διατύπωση του ερευνητικού στόχου.
2. Διατύπωση των ερευνητικών υποθέσεων.
 - H_0 : Η μηδενική υπόθεση υποστηρίζει ότι δεν υπάρχει στατιστικά σημαντική σχέση (εξάρτηση) μεταξύ των μεταβλητών.
 - H_1 : Η εναλλακτική υπόθεση αναφέρεται στην πρόβλεψη που κάνει ο ερευνητής αναφορικά με τη στατιστική σχέση (εξάρτηση) που υπάρχει μεταξύ των μεταβλητών που μελετά. Για να αποδεχτούμε την εναλλακτική υπόθεση πρέπει οπωσδήποτε να απορριφθεί η μηδενική υπόθεση.
3. Ορισμός του επιπέδου στατιστικής σημαντικότητας. Με άλλα λόγια, ορίζουμε την μέγιστη επιτρεπτή τιμή της πιθανότητας το αποτέλεσμα της ανάλυσης των δεδομένων μας να έχει προκύψει τυχαία, πάνω από την οποία η μηδενική υπόθεση θεωρείται αληθινή και επομένως, η εναλλακτική υπόθεση δεν μπορεί να υποστηριχθεί. Στις κοινωνικές επιστήμες χρησιμοποιούμε, συνήθως, **το επίπεδο στατιστικής σημαντικότητας $\alpha=0,05$ ως ανώτατο επιτρεπτό όριο** που σημαίνει ότι, για να απορρίψουμε την μηδενική υπόθεση, θα πρέπει η πιθανότητα να οφείλεται το αποτέλεσμα της περιγραφικής ανάλυσης σε τυχαίους παράγοντες να είναι ίση με 5% ή χαμηλότερη.
4. Επιλογή του στατιστικού κριτηρίου (τεστ) ανάλογα με τον τύπο των μεταβλητών.
5. Εφαρμογή του κριτηρίου στο δείγμα μας και υπολογισμός της τιμής του, καθώς και του επιπέδου στατιστικής σημαντικότητας (significance) που αντιστοιχεί σε αυτή την τιμή (δηλ. της πιθανότητας το αποτέλεσμα από την περιγραφική μας

ανάλυση να προέκυψε τυχαία και όχι λόγω πραγματικής εξάρτησης μεταξύ των μεταβλητών).¹

6. Σύγκριση του επιπέδου στατιστικής σημαντικότητας που ορίσαμε στο βήμα 3 (δηλαδή του α) με το significance του συγκεκριμένου τεστ (ελέγχου).
ΠΡΟΣΟΧΗ: το 0,05 είναι το ίδιο με το 0,050 και το 0,0500. Άρα, αν το sig ισούται με 0,0499 τότε το 0,0499 είναι μικρότερο του 0,05.
7. Εξαγωγή συμπεράσματος.
 - Αν το significance του τεστ είναι μικρότερο του 0,05 τότε απορρίπτουμε την μηδενική υπόθεση H_0 σε επίπεδο σημαντικότητας $\alpha=0,05$.
 - Αν το significance του τεστ είναι μεγαλύτερο του 0,05 τότε δεν απορρίπτουμε την μηδενική υπόθεση H_0 σε επίπεδο σημαντικότητας $\alpha=0,05$ ή διαφορετικά δεν έχουμε επαρκή δεδομένα για να απορρίψουμε την H_0 .

ΠΑΡΑΔΕΙΓΜΑ

Ανοίγουμε τη βάση δεδομένων Elections_2012.sav που περιλαμβάνει δεδομένα για τους υποψήφιους πολιτικούς στις εκλογές του 2012.

Βήμα 1: Ο ερευνητικός στόχος είναι να διαπιστωθεί αν η εκλογική νίκη των υποψηφίων στις εκλογές του 2012 εξαρτάται από την ύπαρξη προφίλ Facebook. Οι μεταβλητές είναι οι εξής δυο:

- Ύπαρξη προφίλ Facebook που παίρνει τιμές ΝΑΙ/ΟΧΙ: ΝΑΙ αν κάποιος υποψήφιος πολιτικός είχε προφίλ Facebook και ΟΧΙ αν κάποιος υποψήφιος δεν είχε προφίλ Facebook.
- Η εκλογική νίκη που παίρνει τιμές ΝΑΙ/ΟΧΙ: ΝΑΙ αν κάποιος υποψήφιος κέρδισε στις εκλογές και ΟΧΙ αν κάποιος υποψήφιος δεν κέρδισε στις εκλογές.

¹Για κάθε πιθανή τιμή που θα μας δώσει ένα στατιστικό τεστ έχει οριστεί από τους στατιστικολόγους η πιθανότητα εμφάνισης αυτής της τιμής όταν η μηδενική υπόθεση (H_0) είναι αληθινή. Εάν αυτή η πιθανότητα είναι ίση ή μικρότερη του 5%, τότε συνήθως απορρίπτουμε τη μηδενική υπόθεση. Για λόγους ευκολίας, σε αυτά τα εργαστήρια δεν θα παρουσιάσουμε τους μαθηματικούς τύπους βάσει των οποίων υπολογίζονται οι τιμές των διαφόρων στατιστικών ελέγχων (τεστ), ούτε τον τρόπο προσδιορισμού των επιπέδων στατιστικής σημαντικότητας που αντιστοιχούν σε κάθε τιμή, δεδομένου μάλιστα ότι όλα αυτά υπολογίζονται αυτόματα από το πρόγραμμα SPSS. Το μόνο που θα σημειώσουμε είναι ότι η τιμή κάθε στατιστικού κριτηρίου-τεστ αντανακλά το μέγεθος των διαφορών μεταξύ των πραγματικών τιμών του δείγματος και των αναμενόμενων τιμών που θα παίρναμε εάν η μηδενική υπόθεση ήταν αληθινή (ανεξαρτησία μεταβλητών). Όσο μεγαλύτερη είναι η τιμή του κριτηρίου, τόσο περισσότερο αποκλίνουν οι πραγματικές τιμές του δείγματος από τις αναμενόμενες τιμές και επομένως, τόσο πιθανότερο είναι η μηδενική υπόθεση να απορριφθεί και το αποτέλεσμα μας να είναι στατιστικά σημαντικό (ύπαρξη εξάρτησης μεταξύ των μεταβλητών).

Οι δύο παραπάνω μεταβλητές είναι τύπου nominal και μάλιστα διχοτομικές (dichotomous). Οι διχοτομικές μεταβλητές είναι αυτές που μπορούν να πάρουν μόνο δυο τιμές.

Βήμα 2: Διατυπώνουμε τις ερευνητικές υποθέσεις:

H₀: Η εκλογική νίκη δεν εξαρτάται από την ύπαρξη προφίλ Facebook ή δεν υπάρχει εξάρτηση ανάμεσα στην εκλογική νίκη και την ύπαρξη προφίλ Facebook.

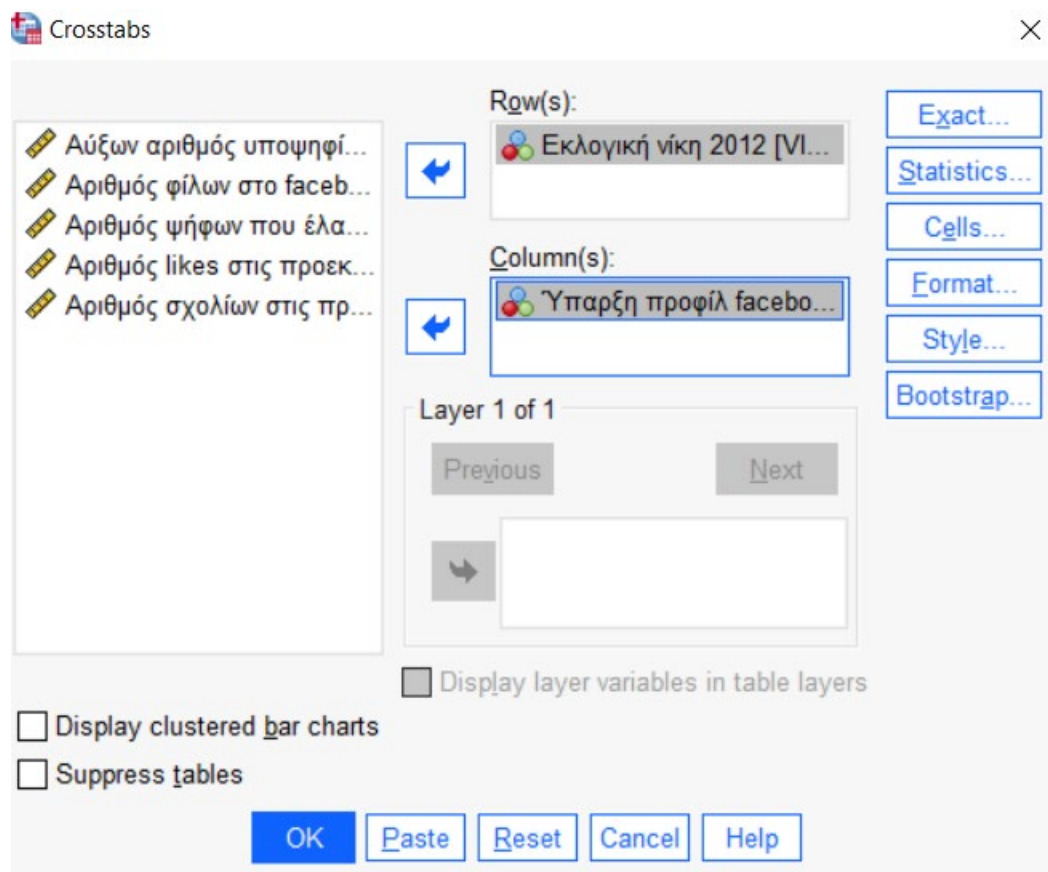
H₁: Η εκλογική νίκη εξαρτάται από την ύπαρξη προφίλ Facebook ή υπάρχει εξάρτηση ανάμεσα στην εκλογική νίκη και την ύπαρξη προφίλ Facebook.

Βήμα 3: Ορίζουμε ως επίπεδο στατιστικής σημαντικότητας το $\alpha=0,05$.

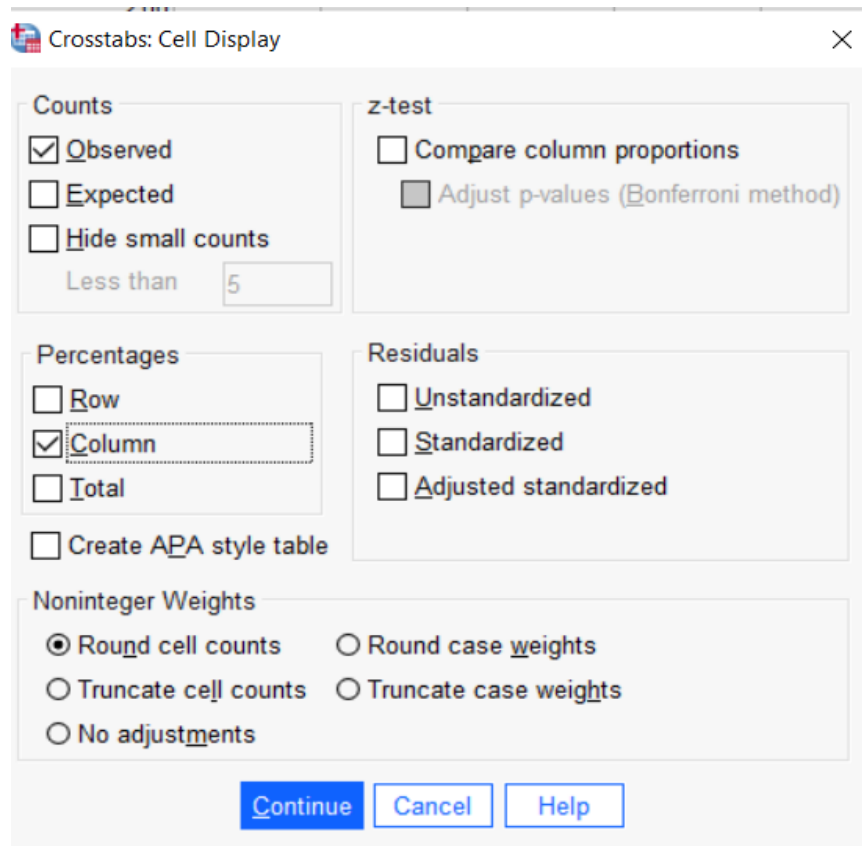
Βήμα 4: Επιλέγουμε το κατάλληλο στατιστικό κριτήριο (τεστ). Αφού έχουμε δυο nominal μεταβλητές, καταλληλότερο κριτήριο είναι ο έλεγχος ανεξαρτησίας χ^2 (**chi-square test**). Στο SPSS επιλέγουμε:

- Analyze
 - Descriptive Statistics
 - Crosstabs

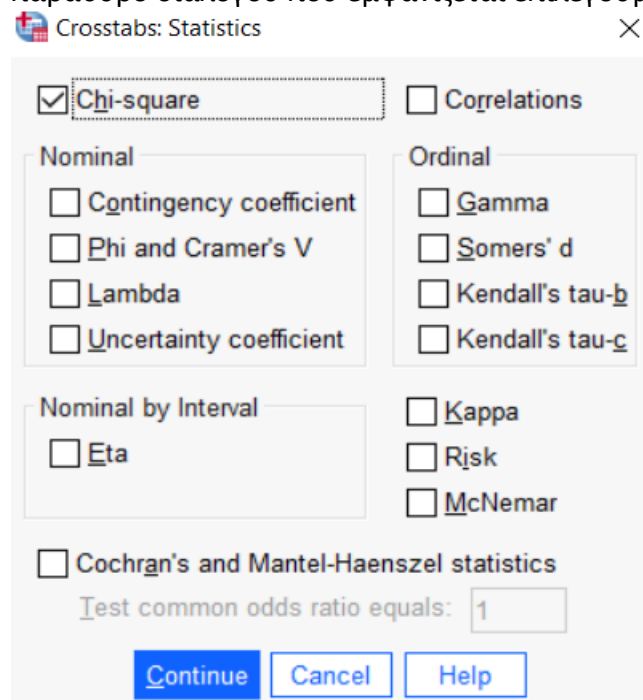
Στο πλαίσιο διαλόγου που εμφανίζεται μεταφέρουμε την εξαρτημένη μεταβλητή «Εκλογική νίκη 2012» στο πλαίσιο Row(s) και την ανεξάρτητη μεταβλητή «Υπαρξη προφίλ facebook» στο πλαίσιο Column(s).



Κατόπιν πατάμε το κουμπί Cells και στο νέο παράθυρο διαλόγου που εμφανίζεται, στην περιοχή Percentages, επιλέγουμε Column. Αυτό μας εξασφαλίζει ότι οι σχετικές συχνότητες της εξαρτημένης μεταβλητής θα υπολογίζονται για κάθε κατηγορία της ανεξάρτητης μεταβλητής ξεχωριστά (και όχι στο σύνολο του δείγματος).



Στη συνέχεια πατάμε Continue. Κατόπιν κάνουμε κλικ στο κουμπί Statistics και στο παράθυρο διαλόγου που εμφανίζεται επιλέγουμε chi-square (πάνω αριστερά).



Πατάμε Continue και μετά OK.

Στο output εμφανίζονται οι πίνακες 1 και 2. Στον Πίνακα 1 που ονομάζεται πίνακας συνάφειας παρουσιάζονται οι συχνότητες (απόλυτες και σχετικές) για κάθε ένα συνδυασμό τιμών των δυο μεταβλητών. Με βάση τα αποτελέσματα, παρατηρείται ότι, από τους 1.134 υποψήφιους που είχαν προφίλ Facebook οι 182 (16%) εκλέχθηκαν, ενώ από τους 1.184 υποψήφιους που δεν είχαν προφίλ Facebook εκλέχθηκαν μόνο οι 92 (8%). Με άλλα λόγια, εκλέχτηκε μεγαλύτερο ποσοστό από εκείνους που είχαν προφίλ Facebook συγκριτικά με εκείνους που δεν είχαν προφίλ Facebook.

ΠΙΝΑΚΑΣ 1
Εκλογική νίκη 2012 * Ύπαρξη προφίλ facebook Crosstabulation

		Ύπαρξη προφίλ facebook		Total	
		ΝΑΙ	ΟΧΙ		
Εκλογική νίκη 2012	ΝΑΙ	Count	182	92	274
		% within Ύπαρξη προφίλ facebook	16,0%	7,8%	11,8%
	ΟΧΙ	Count	952	1092	2044
		% within Ύπαρξη προφίλ facebook	84,0%	92,2%	88,2%
Total	Count	1134	1184	2318	
	% within Ύπαρξη προφίλ facebook	100,0%	100,0%	100,0%	

Βήμα 5: Στον Πίνακα 2 παρουσιάζεται η τιμή του ελέγχου ανεξαρτησίας χ^2 (χ^2 -square=38.090, $df=1$) και το επίπεδο στατιστικής σημαντικότητας που αντιστοιχεί σε αυτήν την τιμή (το significance του ελέγχου). Παρατηρούμε ότι: asymptotic significance (2-sided) < 0,001.² Αυτό είναι το significance του στατιστικού τεστ.

Βήμα 6: Συγκρίνοντας το παραπάνω significance του στατιστικού ελέγχου με το επίπεδο στατιστικής σημαντικότητας $\alpha=0,05$, παρατηρούμε ότι το πρώτο είναι πολύ μικρότερο. Επομένως, μπορούμε να απορρίψουμε τη μηδενική υπόθεση και να υποστηρίξουμε ότι υπάρχει στατιστικά σημαντική εξάρτηση μεταξύ των δυο μεταβλητών.

Βήμα 7: Εξάγουμε, λοιπόν, το συμπέρασμα ότι **οι υποψήφιοι πολιτικοί που διαθέτουν προφίλ Facebook έχουν στατιστικά μεγαλύτερη πιθανότητα να εκλεγούν από ό,τι οι υποψήφιοι πολιτικοί που δεν διαθέτουν προφίλ Facebook.**

²Τις έννοιες των βαθμών ελευθερίας (df , degrees of freedom) και της ασυμπτωματικής σημαντικότητας (asymptotic significance) δεν τις αναλύουμε περισσότερο σε αυτό το εργαστήριο για λόγους ευκολίας.

ΠΙΝΑΚΑΣ 2 Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	38,090 ^a	1	<,001		
Continuity Correction ^b	37,300	1	<,001		
Likelihood Ratio	38,636	1	<,001		
Fisher's Exact Test				<,001	<,001
Linear-by-Linear Association	38,074	1	<,001		
N of Valid Cases	2318				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 134,04.

b. Computed only for a 2x2 table

Τέλος, χρειάζεται να πούμε ότι απαραίτητη προϋπόθεση για τη διεξαγωγή του ελέγχου ανεξαρτησίας χ^2 είναι το ποσοστό των κελιών του πίνακα συνάφειας που έχει αναμενόμενη συχνότητα³ μικρότερη του 5 να μην ξεπερνά το 20% (βλέπε την υποσημείωση α στον Πίνακα 2).

Πώς γράφουμε σε κείμενο τα αποτελέσματα της παραπάνω ανάλυσης;

Για να διαπιστωθεί η ύπαρξη εξάρτησης μεταξύ των μεταβλητών «προφίλ Facebook» και «εκλογική νίκη» διεξήχθη έλεγχος ανεξαρτησίας χ^2 . Βάσει των αποτελεσμάτων εντοπίστηκε στατιστικά σημαντική εξάρτηση σε επίπεδο σημαντικότητας $\alpha=0,05$ μεταξύ της ύπαρξης προφίλ Facebook και της εκλογικής νίκης ενός υποψηφίου ($\chi^2=38,090$, $df=1$, $sig<0,001$). Επομένως, μπορούμε να απορρίψουμε τη μηδενική υπόθεση και να υποστηρίξουμε ότι η ύπαρξη προφίλ Facebook αυξάνει στατιστικά τις πιθανότητες ενός υποψήφιου πολιτικού να πετύχει εκλογική νίκη.

2. Pearson's r για μεταβλητές scale

Όταν θέλουμε να διαπιστώσουμε τη συσχέτιση μεταξύ δυο μεταβλητών τύπου scale (ποσοτικές μεταβλητές χωρίς κλάσεις) τότε χρησιμοποιούμε το συντελεστή συσχέτισης Pearson product-moment ή Pearson r που είναι ένα παραμετρικό τεστ. Βασική προϋπόθεση για την εφαρμογή του εν λόγω τεστ είναι οι δύο μεταβλητές να έχουν μετρηθεί στο ίδιο δείγμα ατόμων (εξαρτημένα δείγματα).

³Όπως εξηγήσαμε και στην υποσημείωση 1, η αναμενόμενη συχνότητα σε ένα κελί είναι εκείνη η συχνότητα που θα περίμενε κανείς εάν η μηδενική υπόθεση H_0 ήταν αληθινή, δηλαδή εάν οι δύο υπό μελέτη μεταβλητές ήταν ανεξάρτητες μεταξύ τους.

ΠΑΡΑΔΕΙΓΜΑ

Έστω ότι θέλουμε να διαπιστώσουμε αν οι ψήφοι που συγκέντρωσε κάθε υποψήφιος στις εκλογές συσχετίζονται με τον αριθμό των φίλων που είχε στο προφίλ Facebook που διέθετε. Το δείγμα αποτελείται από 917 υποψηφίους βουλευτές που είχαν Facebook profile και το οποίο δεν ήταν κλειστό (δεν ήταν private) λογαριασμός. Η μεταβλητή VOTES (αριθμός ψήφων κάθε υποψήφιου βουλευτή) είναι scale, όπως και η μεταβλητή FRIENDS (αριθμός φίλων στο προφίλ κάθε υποψηφίου). Επίσης, τόσο οι αριθμοί των ψήφων, όσο και οι αριθμοί των φίλων στο προφίλ Facebook, προέρχονται από το ίδιο δείγμα ατόμων (εξαρτημένα δείγματα). Επομένως, εξασφαλίζονται οι δύο βασικές προϋποθέσεις για την εφαρμογή του δείκτη συσχέτισης Pearson r .

Βήμα 1: Διατύπωση Υποθέσεων

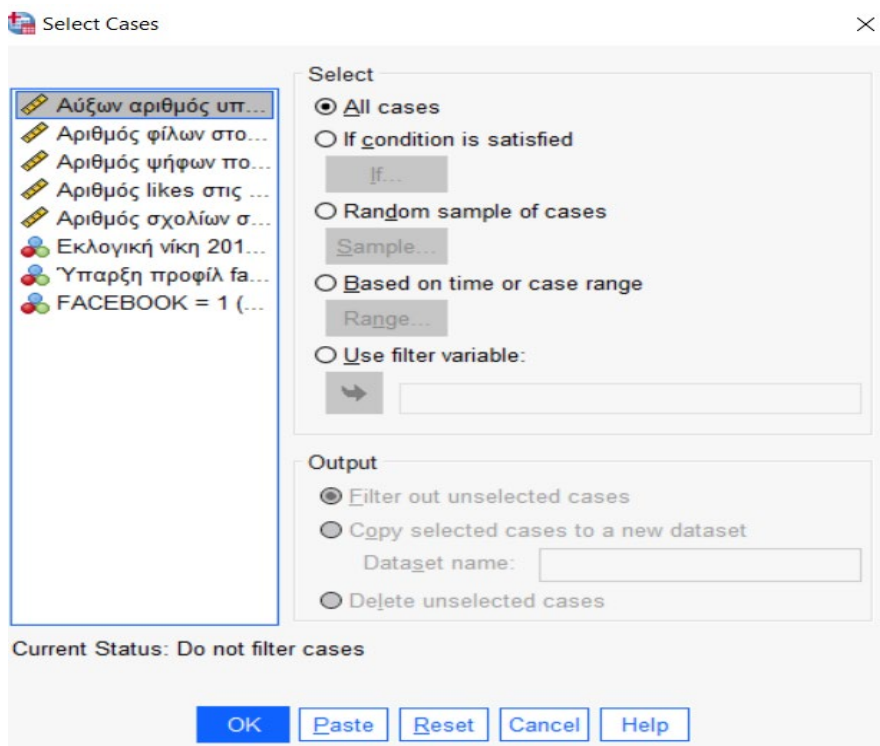
H_0 : **Δεν υπάρχει συσχέτιση** μεταξύ του αριθμού των ψήφων που θα λάβει ο υποψήφιος πολιτικός και του αριθμού των φίλων που θα συγκεντρώσει στο προφίλ Facebook που διαθέτει.

H_1 : **Υπάρχει συσχέτιση** μεταξύ του αριθμού των ψήφων που θα λάβει ο υποψήφιος πολιτικός και του αριθμού των φίλων που θα συγκεντρώσει στο προφίλ Facebook που διαθέτει.

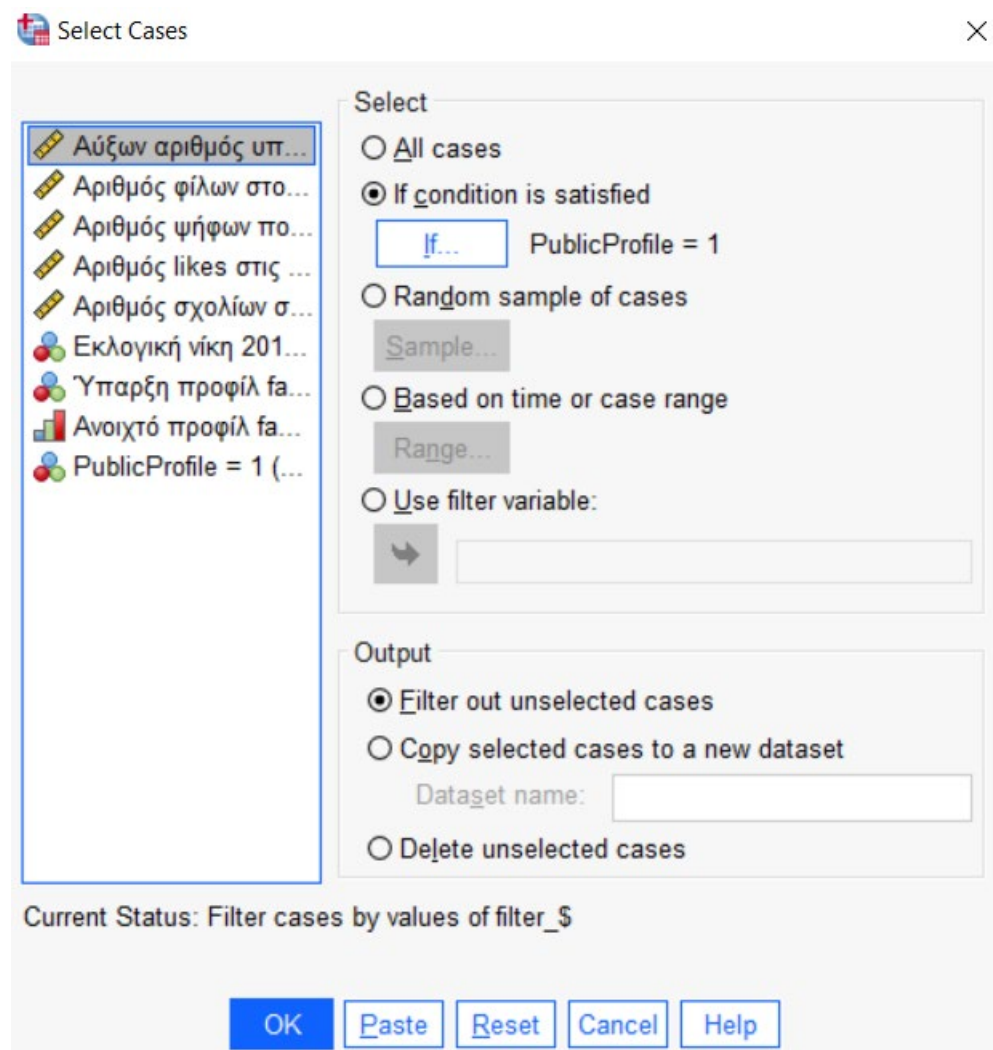
Βήμα 2: Ορίζουμε το επίπεδο στατιστικής σημαντικότητας σε $\alpha=0,05$.

Βήμα 3: Επιλέγουμε το κατάλληλο στατιστικό test. Από τη στιγμή που έχουμε δυο scale μεταβλητές και δύο εξαρτημένα δείγματα, το καταλληλότερο είναι να υπολογιστεί ο συντελεστής συσχέτισης Pearson r .

Αρχικά, επιλέγουμε όλους τους υποψηφίους πολιτικούς που έχουν ανοιχτό προφίλ Facebook: **Data → Select Cases**.



Στο παράθυρο διαλόγου που εμφανίζεται επιλέγουμε «if condition is satisfied» και στη συνέχεια, ορίζουμε τη μεταβλητή PublicProfile = 1.

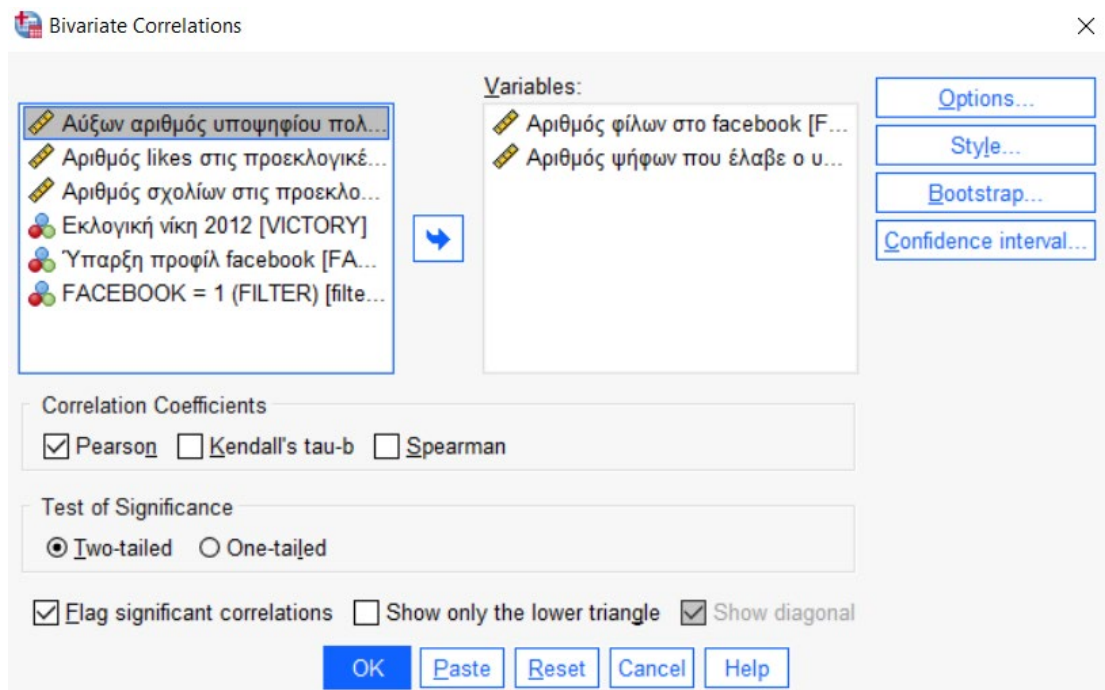


Πατάμε Continue και OK.

Προχωρούμε στο test:

- Analyze
 - Correlate
 - Bivariate

Στο παράθυρο διαλόγου που ανοίγει μεταφέρουμε τις μεταβλητές «Αριθμός ψήφων» και «Αριθμός φίλων» στο πλαίσιο Variables. Δεν έχει σημασία με ποια σειρά θα τοποθετηθούν οι μεταβλητές καθώς ο συντελεστής συσχέτισης δεν υπολογίζει ποια μεταβλητή είναι εξαρτημένη και ποια ανεξάρτητη. Κατόπιν επιλέγουμε Pearson στην περιοχή Correlation Coefficients και μετά πατάμε OK.



Στο Output εμφανίζονται τα αποτελέσματα της ανάλυσης συσχέτισης (Πίνακας 3).

ΠΙΝΑΚΑΣ 3 Correlations

		Αριθμός φίλων στο facebook	Αριθμός ψήφων που έλαβε ο υποψήφιος στις εκλογές
Αριθμός φίλων στο facebook	Pearson Correlation	1	,470**
	Sig. (2-tailed)		<,001
	N	925	917
Αριθμός ψήφων που έλαβε ο υποψήφιος στις εκλογές	Pearson Correlation	,470**	1
	Sig. (2-tailed)	<,001	
	N	917	917

** . Correlation is significant at the 0.01 level (2-tailed).

Βήμα 4: Στον πίνακα 3 παρουσιάζεται η τιμή Pearson $r = 0,470$ και το επίπεδο στατιστικής σημαντικότητας που αντιστοιχεί σε αυτήν την τιμή ($p < 0,001$). Η τιμή $p < 0,001$ είναι το significance του στατιστικού τεστ.

Βήμα 5: Συγκρίνοντας την τιμή p (significance) του συγκεκριμένου τεστ με το επίπεδο στατιστικής σημαντικότητας $\alpha = 0,05$, παρατηρούμε ότι το πρώτο είναι πολύ μικρότερο. Επομένως, μπορούμε να απορρίψουμε τη μηδενική υπόθεση και να

υποστηρίζουμε ότι υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ των δυο μεταβλητών σε επίπεδο σημαντικότητας $\alpha=0,05$.

Αφού εντοπίστηκε στατιστικά σημαντική σχέση, το επόμενο βήμα είναι να διαπιστωθεί η ένταση αυτής της σχέσης (ελάχιστη, ασθενής, μέτρια, ισχυρή) και η κατεύθυνση της σχέσης (θετική ή αρνητική). **Θετική συσχέτιση** υπάρχει στην περίπτωση που όταν αυξάνεται η μία μεταβλητή, αυξάνεται και η άλλη. **Αρνητική συσχέτιση** υπάρχει στην περίπτωση που όταν αυξάνεται η μία μεταβλητή, μειώνεται η άλλη.

Εύρος Συντελεστή	Τύπος Εξάρτησης / Συσχέτισης
Από -1,0 έως -0,7	Ισχυρή Αρνητική
Από -0,7 έως -0,5	Μέτρια Αρνητική
Από -0,5 έως -0,3	Ασθενής Αρνητική
Από -0,3 έως +0,3	Ελάχιστη ή Καμία
Από 0,3 έως 0,5	Ασθενής Θετική
Από 0,5 έως 0,7	Μέτρια Θετική
Από 0,7 έως 1,0	Ισχυρή Θετική

Στον πίνακα 3, παρατηρούμε εκ νέου την τιμή του συντελεστή Pearson r . Η τιμή του συντελεστή Pearson ισούται με 0,470. Επομένως, υπάρχει μια θετική (όσο αυξάνει η μία μεταβλητή αυξάνει και η άλλη γραμμικά) και ταυτόχρονα ασθενής προς μέτρια συσχέτιση ανάμεσα στον αριθμό των Facebook φίλων ενός υποψήφιου πολιτικού και τον αριθμό των ψήφων που λαμβάνει στις εκλογές. Με άλλα λόγια, **καθώς αυξάνει ο αριθμός των φίλων που έχει ένας υποψήφιος πολιτικός στο προφίλ Facebook που διαθέτει, αυξάνει (με ασθενή προς μέτρια ένταση) και ο αριθμός των ψήφων που θα πάρει στις εκλογές.**

Πως γράφουμε σε κείμενο τα αποτελέσματα της παραπάνω ανάλυσης;

Για να διαπιστωθεί η ύπαρξη συσχέτισης μεταξύ του αριθμού των ψήφων που έλαβε ένας υποψήφιος βουλευτής και του αριθμού των Facebook φίλων του, διεξήχθη ανάλυση συσχέτισης και υπολογίστηκε ο συντελεστής συσχέτισης Pearson r . Βάσει των αποτελεσμάτων, εντοπίστηκε στατιστικά σημαντική συσχέτιση σε επίπεδο σημαντικότητας $\alpha=0,05$ μεταξύ του αριθμού των ψήφων που έλαβε ο υποψήφιος και του αριθμού των Facebook φίλων που συγκέντρωσε στο προφίλ του. Ο συντελεστής συσχέτισης Pearson Correlation Coefficient ισούται με 0,470 και είναι στατιστικά σημαντικός σε επίπεδο $\alpha=0,05$ ($\text{sig}<0,001$). Συνεπώς, υπάρχει μια μέτρια θετική συσχέτιση μεταξύ του αριθμού των ψήφων που έλαβε ένας υποψήφιος βουλευτής και του αριθμού των Facebook φίλων που συγκέντρωσε στο προφίλ του. Επομένως, όσο αυξάνεται ο αριθμός των φίλων στο Facebook profile ενός υποψηφίου τόσο αυξάνονται (με μέτρια ένταση) οι ψήφοι που αυτός συγκεντρώνει στις εκλογές. Συνεπώς, μπορούμε να απορρίψουμε τη μηδενική υπόθεση H_0 .

3. Spearman Rho για μεταβλητές ordinal

Όταν θέλουμε να εξετάσουμε αν υπάρχει συσχέτιση μεταξύ δυο μεταβλητών τύπου ordinal (ιεραρχικών μεταβλητών) τότε ο δείκτης Pearson r δεν είναι κατάλληλος. Σε μια τέτοια περίπτωση μπορούμε να χρησιμοποιήσουμε τον δείκτη Spearman Rho που συνιστά ένα μη παραμετρικό τεστ.

ΠΑΡΑΔΕΙΓΜΑ

Ας ανοίξουμε το αρχείο Store_spss_lab.sav που περιλαμβάνει δεδομένα από μια έρευνα ικανοποίησης πελατών 4 καταστημάτων κινητής τηλεφωνίας. Έστω ότι θέλουμε να διαπιστώσουμε αν η ηλικία επηρεάζει τη συχνότητα επίσκεψης των καταναλωτών στα καταστήματα κινητής τηλεφωνίας. Οι υπό μελέτη μεταβλητές είναι οι εξής:

- Ηλικία: [18-30 έτη], [31-40 έτη], [41-50 έτη], [51-60 έτη] και [61 και άνω].
- Συχνότητα επίσκεψης: [1 φορά], [2-3 φορές], [Περισσότερες από 3 φορές]

Παρατηρούμε ότι, οι μεταβλητές είναι ordinal αφού έχουν κλάσεις με διαβάθμιση.

Βήμα 1: Διατύπωση των ερευνητικών υποθέσεων:

H_0 : Δεν υπάρχει συσχέτιση ανάμεσα στην ηλικία του ερωτώμενου και τη συχνότητα επίσκεψης σε κατάστημα κινητής τηλεφωνίας.

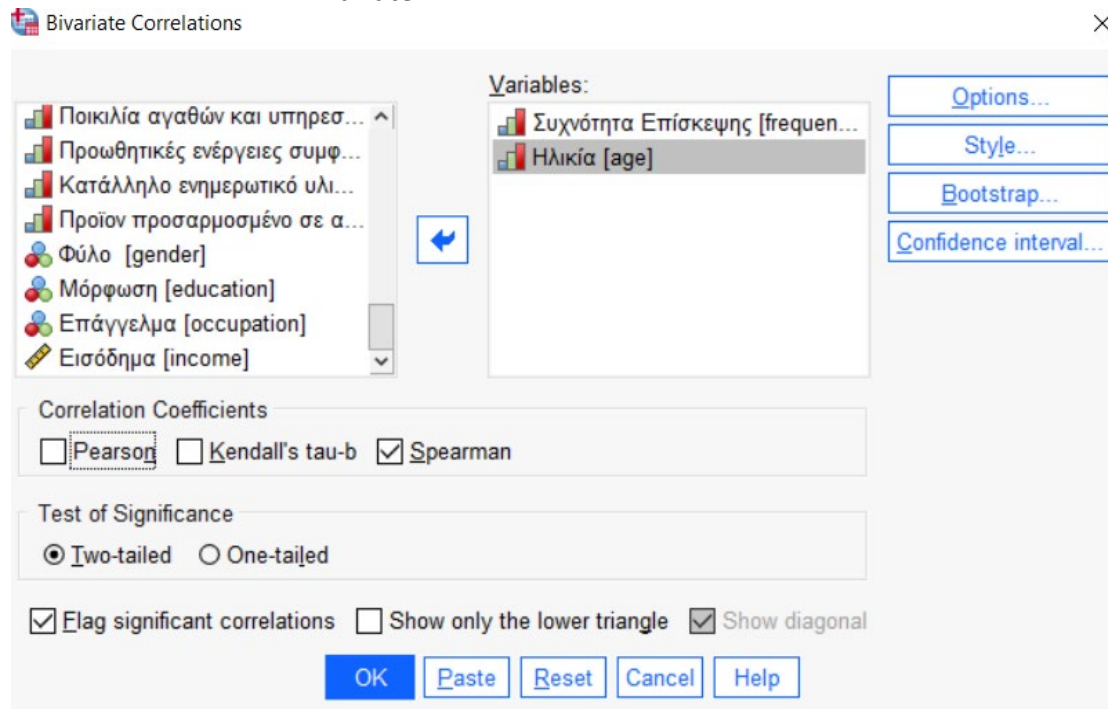
H_1 : Υπάρχει συσχέτιση ανάμεσα στην ηλικία του ερωτώμενου και τη συχνότητα επίσκεψης σε κατάστημα κινητής τηλεφωνίας.

Βήμα 2: Ορίζουμε το επίπεδο στατιστικής σημαντικότητας ως 0.05.

Βήμα 3: Επιλέγουμε το κατάλληλο στατιστικό τεστ.

Από τη στιγμή που οι δυο μεταβλητές είναι ordinal, τότε το κατάλληλο τεστ για να υπολογίσουμε τη συσχέτιση ανάμεσα τους είναι το μη παραμετρικό τεστ Spearman Rho.

- Analyze
 - Correlate
 - Bivariate



Επιλέγουμε τις μεταβλητές που θέλουμε να ελέγξουμε ως προς τη συσχέτισή τους και τις μεταφέρουμε στο κουτάκι στα δεξιά. Και εδώ δεν έχει σημασία με ποια σειρά θα μπουν οι μεταβλητές. Στην περιοχή Correlation Coefficients επιλέγουμε το Spearman. Πατάμε OK. Ο Πίνακας 4 παρουσιάζει τα αποτελέσματα της ανάλυσης συσχέτισης.

ΠΙΝΑΚΑΣ 4
Correlations

			Συχνότητα Επίσκεψης	Ηλικία
Spearman's rho	Συχνότητα Επίσκεψης	Correlation Coefficient	1,000	,229**
		Sig. (2-tailed)	.	,001
		N	200	200
	Ηλικία	Correlation Coefficient	,229**	1,000
		Sig. (2-tailed)	,001	.
		N	200	200

** . Correlation is significant at the 0.01 level (2-tailed).

Βήμα 4: Η τιμή του δείκτη Spearman Rho είναι ίση με 0,229 και το επίπεδο στατιστικής σημαντικότητας που αντιστοιχεί σε αυτήν την τιμή είναι το $p=0,001$. Αυτό είναι το significance του στατιστικού τεστ.

Βήμα 5: Συγκρίνοντας την τιμή p (significance) του συγκεκριμένου τεστ με το επίπεδο στατιστικής σημαντικότητας $\alpha=0,05$, παρατηρούμε ότι το πρώτο είναι πολύ μικρότερο. Επομένως, μπορούμε να απορρίψουμε τη μηδενική υπόθεση και να υποστηρίξουμε ότι υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ των δυο μεταβλητών σε επίπεδο σημαντικότητας $\alpha=0,05$.

Αφού εντοπίστηκε στατιστικά σημαντική συσχέτιση, το επόμενο βήμα είναι να διαπιστωθεί η ένταση και η κατεύθυνση της. Γι' αυτό το λόγο, παρατηρούμε εκ νέου στον Πίνακα 4 την τιμή του δείκτη Spearman Rho. Αυτή ισούται με 0,229. Επομένως, υπάρχει μια θετική ελάχιστη συσχέτιση ανάμεσα στην ηλικία και στη συχνότητα επίσκεψης στο κατάστημα.

Πως γράφουμε σε κείμενο τα αποτελέσματα της παραπάνω ανάλυσης;

Για να διαπιστωθεί η ύπαρξη συσχέτισης μεταξύ της ηλικίας του καταναλωτή και της συχνότητας επίσκεψης στο κατάστημα υπολογίστηκε ο δείκτης Spearman Rho. Βάσει των αποτελεσμάτων, εντοπίστηκε στατιστικά σημαντική συσχέτιση σε επίπεδο σημαντικότητας $\alpha=0,05$ μεταξύ της ηλικίας του καταναλωτή και της συχνότητας επίσκεψης στο κατάστημα. Ο δείκτης συσχέτισης Spearman Rho ισούται με 0,229 και είναι στατιστικά σημαντικός σε επίπεδο $\alpha=0,05$ ($p=0,001<0,05$). Συνεπώς, υπάρχει μια ελάχιστη θετική συσχέτιση μεταξύ της ηλικίας και της συχνότητας επίσκεψης στο κατάστημα. Επομένως, όσο η ηλικία του ερωτώμενου αυξάνεται τόσο αυξάνεται (με ελάχιστη ένταση) η συχνότητα επίσκεψης στο κατάστημα. Συνεπώς, μπορούμε να απορρίψουμε τη μηδενική υπόθεση H_0 .