

ΚΕΦΑΛΑΙΟ 4

ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ

4.1 Βασικές Έννοιες

Σκοπός της ανάλυσης συσχέτισης είναι η εξερεύνηση της σχέσης μεταξύ δύο (ή περισσότερων) μεταβλητών. Το αποτέλεσμα της ανάλυσης συσχέτισης δίνει πληροφορίες στον ερευνητή για την κατεύθυνση και την ένταση της σχέσης μεταξύ των μεταβλητών.

Ενδεικτικά, η ανάλυση συσχέτισης χρησιμοποιείται στην έρευνα αγοράς για να ελέγξει, μεταξύ πολλών άλλων, κατά πόσο συσχετίζονται:

- Το οικογενειακό εισόδημα με τις δαπάνες για ταξίδια αναψυχής.
- Η συχνότητα επίσκεψης στο σουπερμάρκετ με τον αριθμό προϊόντων που αγοράζει ο καταναλωτής.
- Η τιμή του προϊόντος με την αντιλαμβανόμενη από τους καταναλωτές ποιότητά του.
- Οι μηνιαίες πωλήσεις ενός προϊόντος με τη στάση των καταναλωτών απέναντι σ' αυτό.
- Η συχνότητα κατανάλωσης μπίρας με τη συχνότητα διασκέδασης σε νυχτερινά κέντρα.
- Τα μηνιαία κέρδη επιχειρήσεων με τα έτη δραστηριοποίησής τους στην αγορά.
- Η στάση των καταναλωτών απέναντι στη διαφήμιση ενός προϊόντος με τη στάση τους απέναντι σ' αυτό.
- Η εβδομαδιαία κατανάλωση αλκοόλ με την ηλικία των συμμετεχόντων σε μία έρευνα.
- Οι ώρες που οι καταναλωτές χρησιμοποιούν το Internet με τα χρήματα που ξοδεύουν για ηλεκτρονικές (on-line) αγορές.

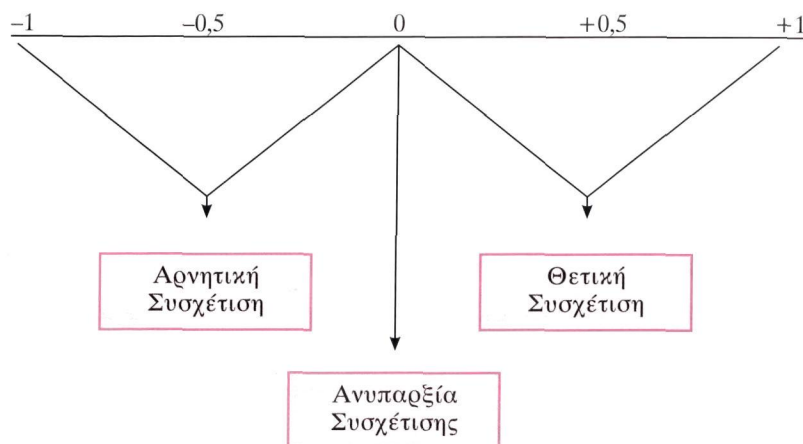
- Ο αριθμός τσιγάρων που καπνίζουν οι ηλικιωμένοι με την ποσότητα αλκοόλ που καταναλώνουν.
- Ο μηνιαίος αριθμός παραγγελιών του προϊόντος μιας επιχείρησης με το μερίδιο αγοράς του.

Οι συντελεστές συσχέτισης είναι αριθμοί που μετρούν το βαθμό εξάρτησης (ένταση της εξάρτησης) της μιας μεταβλητής από την άλλη (magnitude) (π.χ., η στάση απέναντι σε ένα προϊόν συσχετίζεται ισχυρά με τη συχνότητα χρήσης του προϊόντος). Οι συντελεστές συσχέτισης υπολογίζουν επίσης και την «κατεύθυνση» της σχέσης μεταξύ των μεταβλητών, αν δηλαδή η σχέση μεταξύ των μεταβλητών είναι θετική, αρνητική ή ουδέτερη. Για παράδειγμα, η αύξηση των δαπανών για διαφήμιση μπορεί να συνεπάγεται την αύξηση των πωλήσεων ενός προϊόντος (θετική συσχέτιση) ή η αύξηση των δαπανών για προώθηση ενός προϊόντος μπορεί να συνεπάγεται τη μείωση της αντιλαμβανόμενης από τους καταναλωτές ποιότητας του προϊόντος (αρνητική συσχέτιση).

Ο συντελεστής συσχέτισης λαμβάνει τιμή μεταξύ του -1 και του $+1$ (εκτός από ελάχιστες εξαιρέσεις συσχέτισης περισσότερων από δύο μεταβλητών όπου η τιμή του συντελεστή μπορεί να ξεπεράσει τη μονάδα). Εάν η τιμή του συντελεστή βρίσκεται μεταξύ του -1 και του 0 , τότε υπάρχει αρνητική συσχέτιση μεταξύ των δύο μεταβλητών. Αντίθετα, αν ο συντελεστής συσχέτισης λαμβάνει τιμή μεταξύ του 0 και του $+1$, τότε συμπεραίνεται ότι υπάρχει θετική συσχέτιση. Στην περίπτωση που ο συντελεστής συσχέτισης ισούται με μηδέν, ο ερευνητής καταλήγει στο συμπέρασμα ότι δεν υπάρχει γραμμική συσχέτιση μεταξύ των δύο μεταβλητών (Σχήμα 4-1).

Σχήμα 4-1:

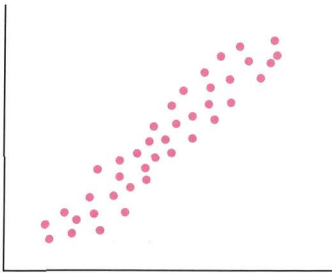
Ο Συντελεστής Συσχέτισης



Για παράδειγμα, μια τιμή του συντελεστή συσχέτισης $+0,87$ ή $+0,92$ παρόμοια με ισχυρή θετική συσχέτιση μεταξύ των μεταβλητών (Σχήμα 4-2). Εάν ο συντελεστής συσχέτισης είναι $-0,83$ ή $-0,91$, τότε συμπεραίνεται ότι υπάρχει ισχυρή αρνητική συσχέτιση μεταξύ των μεταβλητών (Σχήμα 4-4). Τα Σχήματα 4-3 και 4-5 παρουσιάζουν διαγραμματικά την ασθενή θετική και αρνητική συσχέτιση, αντίστοιχα. Τέλος, το Σχήμα 4-6 δείχνει την ανυπαρξία γραμμικής συσχέτισης μεταξύ των μεταβλητών.

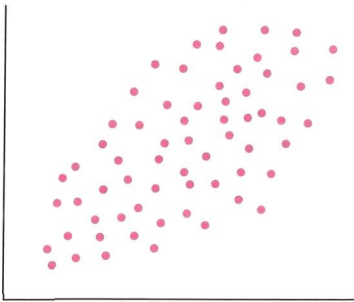
Σχήμα 4-2:

Ισχυρή Θετική Συσχέτιση (συντελεστής συσχέτισης $\approx +0,95$)



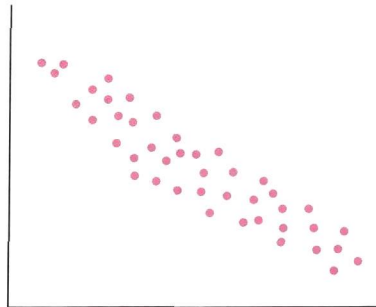
Σχήμα 4-3:

Ασθενής Θετική Συσχέτιση (συντελεστής συσχέτισης $\approx +0,5$)



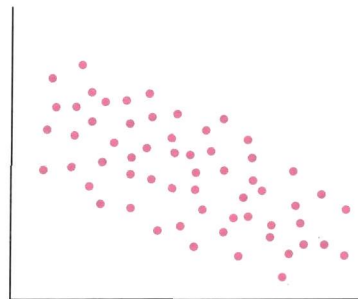
Σχήμα 4-4:

Ισχυρή Αρνητική Συσχέτιση (συντελεστής συσχέτισης $\approx -0,95$)



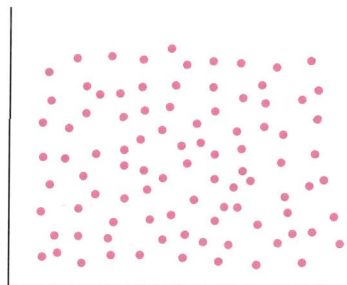
Σχήμα 4-5:

Ασθενής Αρνητική Συσχέτιση (συντελεστής συσχέτισης $\approx -0,5$)



Σχήμα 4-6:

Ανυπαρξία Συσχέτισης (συντελεστής συσχέτισης ≈ 0)



Είναι εξαιρετικά σημαντικό να τονιστεί ότι, αν και η συσχέτιση αντικατοπτρίζει την ένταση της σχέσης μεταξύ των μεταβλητών, δεν εξηγεί τη σχέση αιτίας-αιτιού (δηλαδή, ποια μεταβλητή προκαλεί τη μεταβολή της άλλης).



Ο Πίνακας 4-1 συνοψίζει το εύρος του συντελεστή συσχέτισης που μπορεί να υπάρχει μεταξύ δύο μεταβλητών και τα σχετικά συμπεράσματα που προκύπτουν σχετικά με την ένταση και την κατεύθυνση της συσχέτισης.

Πίνακας 4-1: Εύρος του Συντελεστή Συσχέτισης και Τύπος Σχέσης μεταξύ των Μεταβλητών

Εύρος Συντελεστή Συσχέτισης	Τύπος Σχέσης μεταξύ των Μεταβλητών
Από -1,0 έως -0,7	Ισχυρή αρνητική σχέση
Από -0,7 έως -0,3	Ασθενής αρνητική σχέση
Από -0,3 έως +0,3	Ελάχιστη ή καμία σχέση
Από 0,3 έως +0,7	Ασθενής θετική σχέση
Από +0,7 έως +1,0	Ισχυρή θετική σχέση

4.2 Ο Συντελεστής Συσχέτισης Pearson

Ο συντελεστής συσχέτισης του Pearson χρησιμοποιείται συχνότερα στην έρευνα αγοράς και δείχνει τη γραμμική σχέση μεταξύ δύο μεταβλητών. Απαραίτητη προϋπόθεση του συντελεστή συσχέτισης του Pearson είναι η ύπαρξη μεταβλητών interval ή scale. Η χρήση του συντελεστή συσχέτισης του Pearson θα πρέπει να αποφεύγεται όταν η μία ή και οι δύο εξεταζόμενες μεταβλητές είναι nominal. Η συσχέτιση μεταβλητών nominal δεν έχει νόημα διότι αν, για παράδειγμα, σε μία έρευνα η τιμή 1 μιας μεταβλητής αντιστοιχεί στο απορρυπαντικό Α, η τιμή 2 στο απορρυπαντικό Β και η τιμή 3 αντιστοιχεί στο απορρυπαντικό Γ, είναι ανούσιο να συσχετιστεί η συγκεκριμένη μεταβλητή με την ποσότητα απορρυπαντικού που χρησιμοποιεί μια νοικοκυριά. Δηλαδή, μια ισχυρή συσχέτιση των δύο μεταβλητών (ποσότητας και μάρκας) δε σημαίνει ότι οι χρήστες του απορρυπαντικού Γ καταναλώνουν τριπλάσια ποσότητα σε σχέση με εκείνη που καταναλώνουν οι χρήστες του απορρυπαντικού Γ.

Ο συντελεστής συσχέτισης του Pearson προϋποθέτει επίσης ότι οι μεταβλητές κατανέμονται κανονικά στον πληθυσμό. Η υπόθεση αυτή είναι ιδιαίτερα σημαντική όταν το δείγμα είναι μικρό.

Ο παρακάτω πίνακας παρουσιάζει την ποσότητα των προϊόντων που αγόρασαν οι καταναλωτές μέσα σε μία ώρα σε ένα σουπερμάρκετ αυθόρμητα, δίχως δηλαδή να έχουν προγραμματίσει την αγορά των συγκεκριμένων προϊόντων. Επιπλέον, ο παρακάτω πίνακας παρουσιάζει το ποσοστό έκπτωσης για κάθε ένα από τα δέκα προϊόντα της έρευνας.

α' προϋπόθεση για συσχέτισης Pearson

β' προϋπόθεση Pearson

A/A	Έκπτωση (%)	Αυθόρμητες Αγορές
1	16	4
2	29	6
3	36	9
4	43	10
5	20	5
6	32	7
7	20	5
8	3	1
9	4	2
10	6	3

Επιθυμείται να διαπιστωθεί αν υπάρχει συσχέτιση μεταξύ του ποσοστού της έκπτωσης και του αριθμού των αυθόρμητων αγορών που πραγματοποιήθηκαν στον σουπερμάρκετ. Τα δεδομένα του παραδείγματος παρουσιάζονται, εισηγμένα στο SPSS, στο Σχήμα 4-7.

Σχήμα 4-7:

Συντελεστής Συσχέτισης Pearson – Παράδειγμα

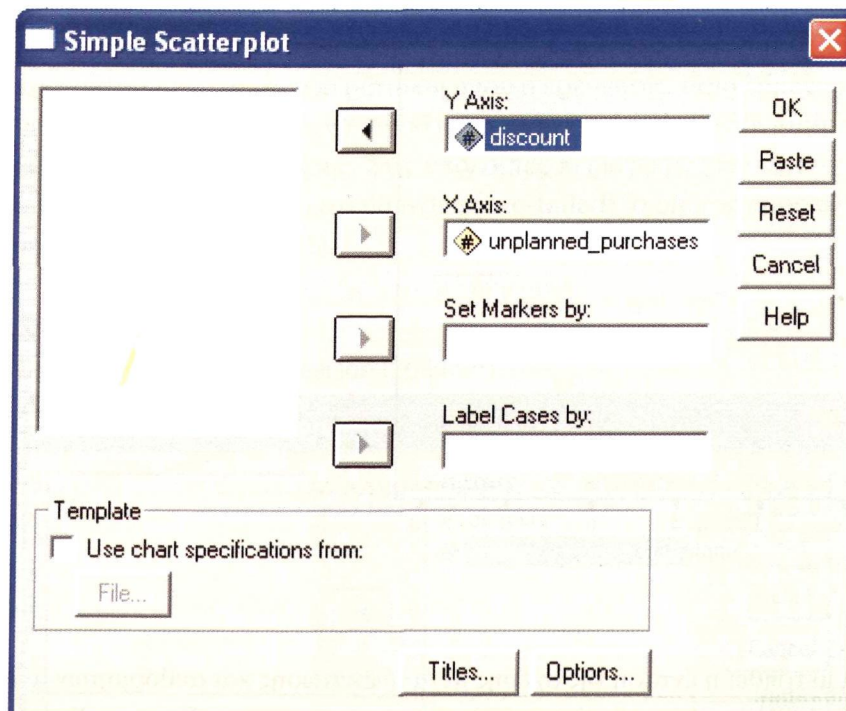
	discount	unplanned purchases
1	16	4
2	29	6
3	36	9
4	43	10
5	20	5
6	32	7
7	20	5
8	3	1
9	4	2
10	6	3
11		

Το πρώτο βήμα για την ανάλυση συσχέτισης μεταξύ των δύο μεταβλητών είναι η κατασκευή ενός scatterplot. Το scatterplot βοηθά στη γραφική αξιολόγηση της σχέσης μεταξύ των δύο μεταβλητών.

Στο πλαίσιο διαλόγου του απλού scatterplot (simple scatterplot), η μεταβλητή “discount” εισάγεται στον κάθετο άξονα (Y Axis), ενώ η μεταβλητή “unplanned purchases” εισάγεται στον οριζόντιο άξονα (X Axis) (Σχήμα 4-8).

Σχήμα 4-8:

Συντελεστής Συσχέτισης Pearson – Scatterplot (1ο βήμα)

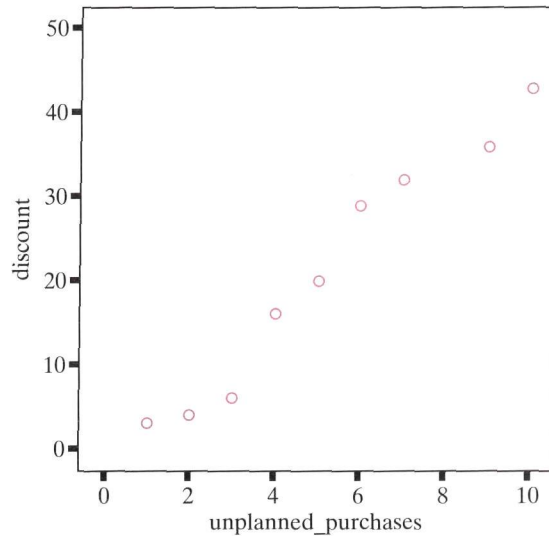


Το scatterplot του παραδείγματος, έτσι όπως προκύπτει από το Output, παρουσιάζεται στο Σχήμα 4-9.

Το γράφημα φανερώνει ότι, καθώς αυξάνεται το ποσοστό της έκπτωσης, οι αγορές που δεν είχαν προγραμματιστεί από τους καταναλωτές επίσης αυξάνονται. Συμπεραίνεται ότι υπάρχει θετική σχέση μεταξύ των δύο μεταβλητών.

Σχήμα 4-9:

Συντελεστής Συσχέτισης Pearson – Scatterplot (2ο βήμα)

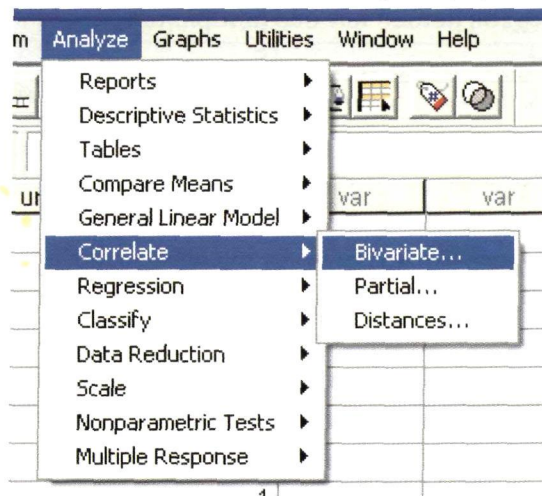


Για να μετρηθεί η ένταση της σχέσης μεταξύ έκπτωσης και αυθόρμητων αγορών, θα πρέπει να υπολογιστεί ο συντελεστής συσχέτισης του Pearson. Για τον υπολογισμό του συντελεστή ακολουθείται η εξής διαδικασία:

Analyze → Correlate → Bivariate (Σχήμα 4-10)

Σχήμα 4-10:

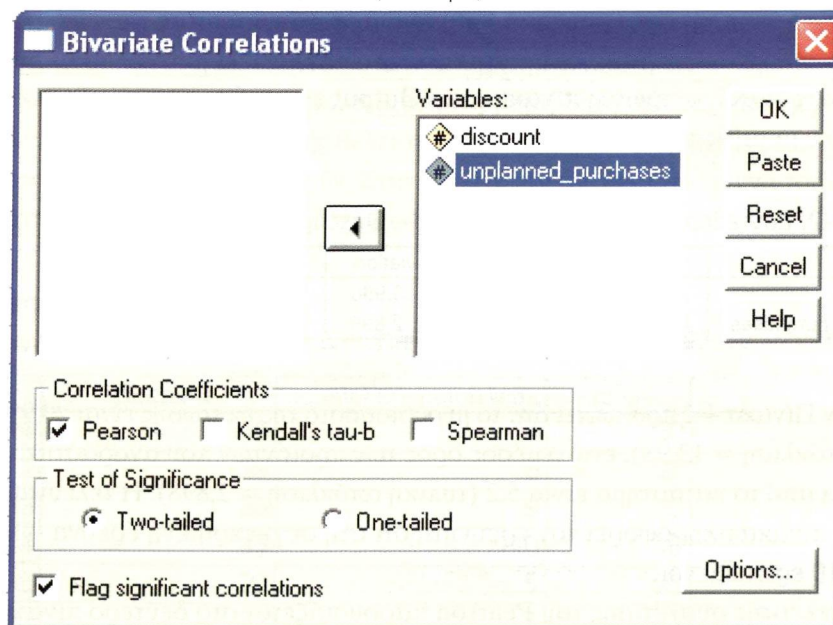
Συντελεστής Συσχέτισης Pearson (1ο Βήμα)



Οι δύο μεταβλητές μεταφέρονται στη λίστα “Variables”. Επιλέγεται ο συντελεστής συσχέτισης του Pearson (Pearson Correlation Coefficient) και το “Two-tailed significance”. Με επιλογή “One-tailed significance” ελέγχεται η μηδενική υπόθεση H_0 : Η συσχέτιση είναι μηδενική. Ο έλεγχος σημαντικότητας μονής κατεύθυνσης προτιμάται μονάχα στην περίπτωση όπου ο ερευνητής, στηριζόμενος σε κάποια θεωρία του MKT, μπορεί να αποκλείσει με βεβαιότητα την αρνητική ή τη θετική κατεύθυνση της συσχέτισης. Στις υπόλοιπες περιπτώσεις επιλέγεται ο έλεγχος σημαντικότητας διπλής κατεύθυνσης (two-tailed), ο οποίος προεπιλέγεται και από το SPSS (Σχήμα 4-11).

Σχήμα 4-11:

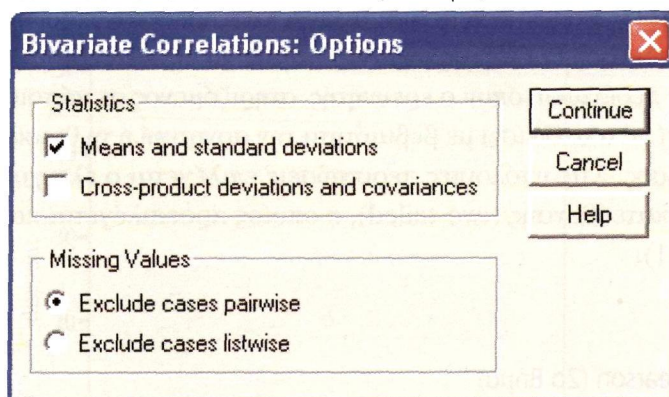
Συντελεστής Συσχέτισης Pearson (2ο Βήμα)



Από την επιλογή “Options” του Σχήματος 4-11 μπορεί ο ερευνητής να μελετήσει στο Output το μέσο και την τυπική απόκλιση της κάθε μεταβλητής καθώς και τον αριθμό των παρατηρήσεων. Για να υπάρχει αυτή η δυνατότητα, στο παράθυρο “Bivariate Correlations: Options” επιλέγεται το “Means and standard deviations” (Σχήμα 4-12).

Σχήμα 4-12:

Συντελεστής Συσχέτισης Pearson (3ο Βήμα)



Πατώντας “OK”, ο πρώτος πίνακας του Output εμφανίζει τα περιγραφικά στοιχεία των μεταβλητών (Πίνακας 4-2).

Πίνακας 4-2: Συντελεστής Συσχέτισης Pearson – Descriptive Statistics

	Mean	Std. Deviation	N
discount	20,90	13,980	10
unplanned_purchases	5,20	2,898	10

Από τον Πίνακα 4-2 προκύπτει ότι, το μέσο ποσοστό της έκπτωσης είναι 20,9% (τυπική απόκλιση = 13,98), ενώ ο μέσος όρος των προϊόντων που αγοράστηκαν αυθόρμητα από το κατάστημα είναι 5,2 (τυπική απόκλιση = 2,898). Η τελευταία στήλη του πίνακα πληροφορεί τον ερευνητή ότι στη συγκεκριμένη έρευνα συμμετείχαν 10 ερωτώμενοι.

Ο συντελεστής συσχέτισης του Pearson παρουσιάζεται στο δεύτερο πίνακα του Output (Πίνακας 4-3).

Πίνακας 4-3: Συντελεστής Συσχέτισης Pearson – Correlations

		discount	unplanned_purchases
discount	Pearson Correlation	1	,982**
	Sig. (2-tailed)	-	,000
	N	10	10
unplanned_purchases	Pearson Correlation	,982**	1
	Sig. (2-tailed)	,000	-
	N	10	10

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

Ο Πίνακας 4-3 παρουσιάζει τη συσχέτιση ανά ζεύγη μεταξύ των μεταβλητών. Ο αριθμός των γραμμών και στηλών του Πίνακα 4-3 αντιπροσωπεύει τον αριθμό των μεταβλητών της ανάλυσης. Εάν στην έρευνα υπήρχαν 3 μεταβλητές και ήταν επιθυμητή η εύρεση του βαθμού συσχέτισης μεταξύ τους, ο Πίνακας “Correlations” του Output θα είχε τρεις γραμμές και τρεις στήλες, οι οποίες θα έδειχναν πώς συσχετίζονται οι μεταβλητές ανά δύο. Στο συγκεκριμένο παράδειγμα, εφόσον υπάρχουν δύο μόνο μεταβλητές η τιμή του συντελεστή είναι η ίδια και στις δύο στήλες. Ο συντελεστής του Pearson είναι 0,982 αντιπροσωπεύοντας μια πολύ ισχυρή συσχέτιση μεταξύ του ποσοστού της έκπτωσης και του αριθμού των αυθόρμητων αγορών. Η συσχέτιση είναι θετική (φαίνεται από το πρόσημο) που σημαίνει ότι καθώς αυξάνεται το ποσοστό της έκπτωσης, αυξάνεται ο αριθμός των προϊόντων που αγοράζονται αυθόρμητα. Ο έλεγχος σημαντικότητας διπλής κατεύθυνσης πιστοποιεί ότι η συσχέτιση μεταξύ των δύο μεταβλητών είναι στατιστικά σημαντική σε επίπεδο στατιστικής σημαντικότητας $\alpha=0,001$.

Ας υποθεθεί ότι υπάρχει και μια τρίτη μεταβλητή στο παράδειγμα (π.χ., τιμή του προϊόντος). Ο ερευνητής θέλει να διαπιστώσει εάν υπάρχει θετική συσχέτιση μεταξύ των τριών μεταβλητών. Στην περίπτωση αυτή, τα δεδομένα του παραδείγματος παρουσιάζονται στο Σχήμα 4-13.

Σχήμα 4-13:

Συντελεστής Συσχέτισης Pearson – Παράδειγμα με Τρεις Μεταβλητές

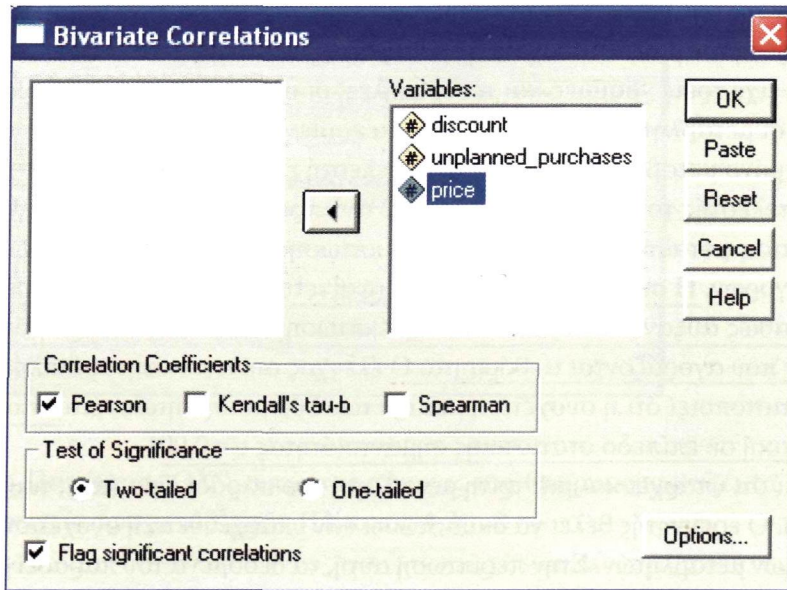
	discount	unplanned purchases	price
1	16	4	30
2	29	6	27
3	36	9	20
4	43	10	18
5	20	5	25
6	32	7	22
7	20	5	35
8	3	1	60
9	4	2	50
10	6	3	40

Για την ανάλυση συσχέτισης με τρεις μεταβλητές, ακολουθείται η ίδια ακριβώς διαδικασία με εκείνη που ακολουθήθηκε για την ανάλυση συσχέτισης με δύο μεταβλητές: Analyze → Correlate → Bivariate (Σχήμα 4-10).

Στο πλαίσιο διαλόγου που εμφανίζεται εισάγονται και οι τρεις μεταβλητές στη λίστα “Variables” (Σχήμα 4-14).

Σχήμα 4-14:

Συντελεστής Συσχέτισης Pearson – Παράδειγμα με Τρεις Μεταβλητές (1ο Βήμα)



Το Output που προκύπτει παρουσιάζεται στους Πίνακες 4-4 και 4-5.

Πίνακας 4-4: Συντελεστής Συσχέτισης Pearson με Τρεις Μεταβλητές – Descriptive Statistics

	Mean	Std. Deviation	N
discount	20,90	13,980	10
unplanned_purchases	5,20	2,898	10
price	32,70	13,720	10

Πίνακας 4-5: Συντελεστής Συσχέτισης Pearson με Τρεις Μεταβλητές – Correlations

		discount	unplanned_purchases	price
discount	Pearson Correlation	1	,982**	-,900**
	Sig. (2-tailed)	-	,000	,000
	N	10	10	10
unplanned_purchases	Pearson Correlation	,982**	1	-,906**
	Sig. (2-tailed)	,000	-	,000
	N	10	10	10
price	Pearson Correlation	-,900**	-,906**	1
	Sig. (2-tailed)	,000	,000	-
	N	10	10	10

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

62

Από τον Πίνακα 4-4 προκύπτει ότι η μέση τιμή των προϊόντων του δείγματος είναι €32,7 (τυπική απόκλιση = 13,72), ενώ οι μέσοι και οι τυπικές αποκλίσεις για τις μεταβλητές “discount” και “unplanned_purchases” είναι οι ίδιοι με αυτές του Πίνακα 4-2.

Από τον Πίνακα 4-5 φαίνεται ότι, ο συντελεστής συσχέτισης μεταξύ των μεταβλητών “discount” και “unplanned_purchases” είναι 0,982 (όπως στον Πίνακα 4-3). Η συσχέτιση μεταξύ αυθόρμητων αγορών και τιμής είναι πολύ ισχυρή (-0,906). Το αρνητικό πρόσημο του συντελεστή πληροφορεί τον ερευνητή για αρνητική συσχέτιση, δηλαδή καθώς αυξάνεται η τιμή των προϊόντων μειώνονται οι μη προγραμματισμένες αγορές. Αρνητική είναι επίσης και η συσχέτιση μεταξύ τιμής και έκπτωσης (όσο αυξάνεται η τιμή των προϊόντων μειώνεται το ποσοστό της έκπτωσης). Η συσχέτιση μεταξύ τιμής και έκπτωσης είναι επίσης αρνητικά ισχυρή (-0,900). Και οι τρεις συσχετίσεις είναι στατιστικά σημαντικές ($p < 0,001$).

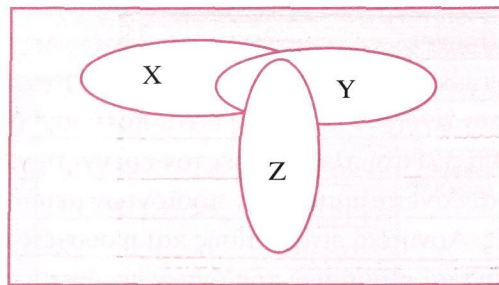
Αν και η ανάλυση με βάση το συντελεστή συσχέτισης του Pearson με τρεις μεταβλητές δείχνει το βαθμό της συσχέτισης των μεταβλητών ανά ζεύγη, αδυνατεί να απομονώσει τη σχέση της τιμής των προϊόντων στη συσχέτιση της έκπτωσης με τις αυθόρμητες αγορές. Το πρόβλημα αυτό μπορεί να λυθεί με το συντελεστή μερικής συσχέτισης, ο οποίος μπορεί να υπολογίσει τη μερική συσχέτιση της έκπτωσης με τις αυθόρμητες αγορές εισάγοντας τη μεταβλητή τιμή ως μεταβλητή ελέγχου (control variable).

4.3 Ο Συντελεστής Μερικής Συσχέτισης (Partial Correlation Coefficient)

Ο συντελεστής μερικής συσχέτισης δείχνει την ένταση και την κατεύθυνση της σχέσης μεταξύ δύο (ή περισσότερων) μεταβλητών, «απομονώνοντας» την αλληλοσυσχέτιση των εξεταζόμενων μεταβλητών με μια τρίτη μεταβλητή (Σχήμα 4-15). Με τη μερική συσχέτιση καθορίζεται ο βαθμός συσχέτισης που θα ίσχυε μεταξύ των δύο μεταβλητών, αν καθεμία από αυτές δε συσχετιζόταν με την τρίτη μεταβλητή. Δηλαδή, η μερική συσχέτιση επιτρέπει τη μέτρηση της συσχέτισης μεταξύ δύο μεταβλητών αν, υποθετικά, η τρίτη μεταβλητή παρέμενε σταθερή.

Σχήμα 4-15:

Διαγραμματική Παρουσίαση της Μερικής Συσχέτισης

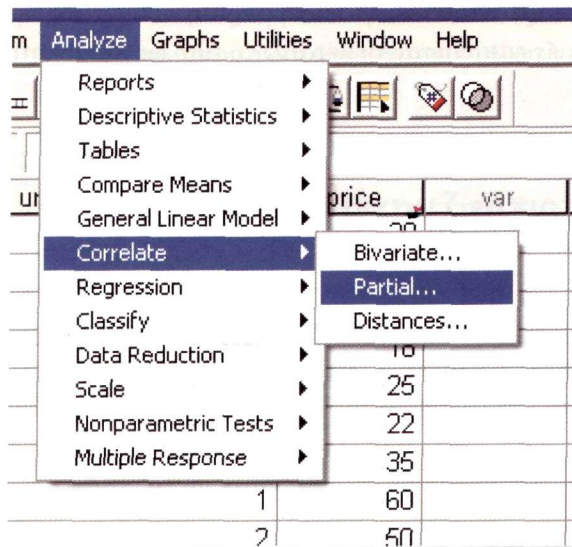


Για να υπολογιστεί ο συντελεστής μερικής συσχέτισης ακολουθείται η παρακάτω διαδικασία:

Analyze → Correlate → Partial (Σχήμα 4-16).

Σχήμα 4-16:

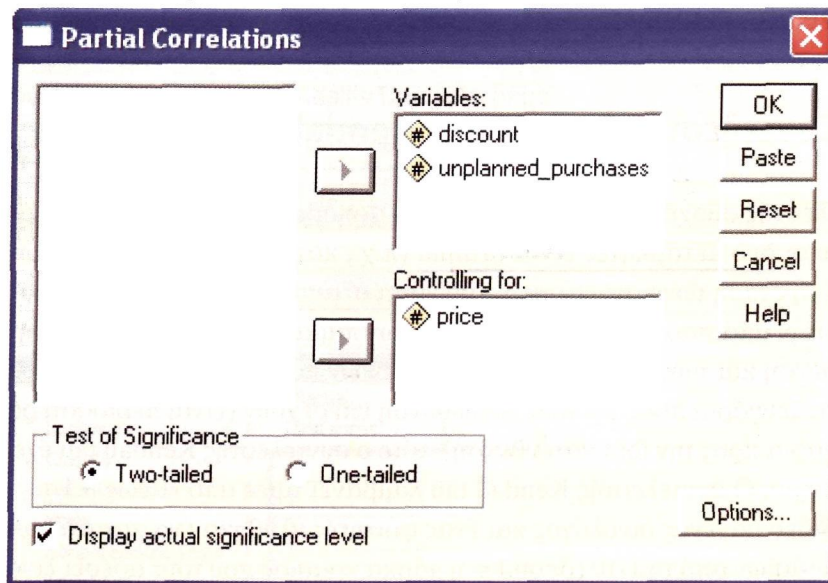
Συντελεστής Μερικής Συσχέτισης (1ο Βήμα)



Οι μεταβλητές “discount” και “unplanned_purchases” εισάγονται στη λίστα “Variables”, ενώ η μεταβλητή “price” μεταφέρεται στο πλαίσιο “Controlling for” (Σχήμα 4-17).

Σχήμα 4-17:

Συντελεστής Μερικής Συσχέτισης (2ο Βήμα)



Πατώντας “OK”, ο πίνακας που εμφανίζεται στο Output δείχνει το συντελεστή μερικής συσχέτισης (Πίνακας 4-6).

Πίνακας 4-6: Συντελεστής Μερικής Συσχέτισης – Correlations

Control Variables			discount	unplanned_purchases
price	discount	Correlation	1,000	,905
		Significance (2-tailed)	-	,001
		df	0	7
unplanned_purchases	unplanned_purchases	Correlation	,905	1,000
		Significance (2-tailed)	,001	-
		df	7	0

Για τις μεταβλητές “unplanned_purchases” και “discount” ο συντελεστής μερικής συσχέτισης είναι 0,905. Η συσχέτιση θεωρείται θετικά ισχυρή, ενώ η συσχέτιση μεταξύ των δύο μεταβλητών είναι στατιστικά σημαντική ($p < 0,005$).

Συγκρίνοντας τον Πίνακα 4-6 με τον Πίνακα 4-5, παρατηρείται ότι μετά την «απομόνωση» της μεταβλητής “price”, ο συντελεστής συσχέτισης μειώθηκε από 0,980 σε 0,905.

Ο συντελεστής μερικής συσχέτισης μπορεί να υπολογιστεί και για περισσότερες από δύο μεταβλητές, χρησιμοποιώντας ως μεταβλητές ελέγχου μία, δύο ή και περισσότερες μεταβλητές.

4.4 Ο Συντελεστής Συσχέτισης Kendall's Tau b

Ο συντελεστής συσχέτισης Kendall's tau (τ) χρησιμοποιείται για στην περίπτωση όπου οι δύο μεταβλητές είναι ordinal (π.χ., κατάταξη πέντε προϊόντων ανάλογα με τη στάση των καταναλωτών απέναντι σ' αυτά). Ο συντελεστής αυτός στηρίζεται στην απόσταση μεταξύ των ζευγών των παρατηρήσεων που έχουν την ίδια κατεύθυνση και των ζευγών των παρατηρήσεων που έχουν αντίθετη κατεύθυνση. Αν τα ζευγάρια προς την ίδια κατεύθυνση υπερέχουν (είναι περισσότερα από τα ζευγάρια προς την ίδια κατεύθυνση), τότε ο συντελεστής Kendall tau έχει θετικό πρόσημο. Ο συντελεστής Kendall tau λαμβάνει τιμές από -1 έως +1.

Για παράδειγμα, ένας οινολόγος και ένας φοιτητής κλήθηκαν να κατατάξουν 10 μάρκες κρασιών από το 1-10 (όπου 1 = η μάρκα κρασιού που τους αρέσει λιγότερο και 10 = η μάρκα κρασιού που τους αρέσει περισσότερο). Τα αποτελέσματα της έρευνας παρουσιάζονται στον παρακάτω πίνακα.

Μάρκες Κρασιών	A	B	Γ	Δ	E	ΣΤ	Z	H	Θ	I
Οινολόγος	1	2	3	4	5	6	7	8	9	10
Φοιτητής	1	3	2	4	6	8	5	7	10	9

Τα δεδομένα του Πίνακα εισηγμένα στο SPSS παρουσιάζονται στο Σχήμα 4-18.

Σχήμα 4-18:

Συντελεστής Συσχέτισης Kendall's Tau b

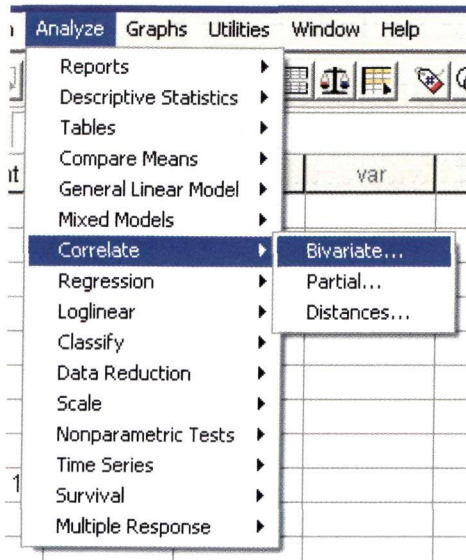
	speciali	student
1	1	1
2	2	3
3	3	2
4	4	4
5	5	6
6	6	8
7	7	5
8	8	7
9	9	10
10	10	9
11		

Non Parametric Correlation

Για να υπολογιστεί ο συντελεστής συσχέτισης Kendall's Tau b ακολουθείται η εξής διαδικασία: Analyze → Correlate → Bivariate (Σχήμα 4-19).

Σχήμα 4-19:

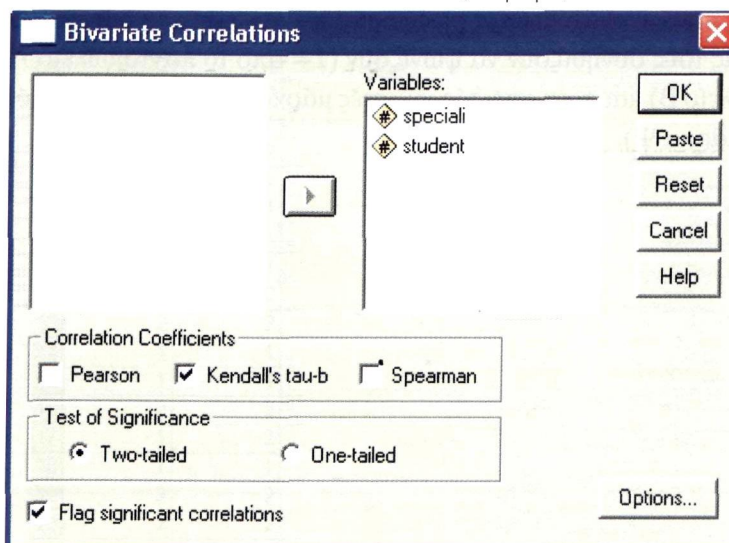
Συντελεστής Συσχέτισης Kendall's Tau b (1ο βήμα)



Στη συνέχεια, εισάγονται οι μεταβλητές “speciali” και “student” στη λίστα “Variables” και επιλέγεται το “Kendall's tau-b” (Σχήμα 4-20).

Σχήμα 4-20:

Συντελεστής Συσχέτισης Kendall's Tau b (2ο βήμα)



Πατώντας “OK”, εμφανίζεται ο μοναδικός Πίνακας του Output (Πίνακας 4-7).

Πίνακας 4-7: Συντελεστής Συσχέτισης Kendall's Tau b – Correlations

			SPECIALI	STUDENT
Kendall's tau_b	SPECIALI	Correlation Coefficient	1,000	,778**
		Sig. (2-tailed)	-	,002
		N	10	10
	STUDENT	Correlation Coefficient	,778**	1,000
		Sig. (2-tailed)	,002	-
		N	10	10

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

Η συσχέτιση της προτίμησης μαρκών κρασιών του οινολόγου και του φοιτητή βρέθηκε να είναι θετικά ισχυρή ($\tau = 0,778$). Η συσχέτιση είναι στατιστικά σημαντική σε επίπεδο στατιστικής σημαντικότητας $\alpha = 0,005$.

4.5 Ο Συντελεστής Συσχέτισης Phi (Phi Coefficient)

Ο συντελεστής συσχέτισης Phi (ϕ) χρησιμοποιείται στην περίπτωση όπου οι δύο εξεταζόμενες μεταβλητές της έρευνας είναι διχοτομικές (π.χ., Αθήνα/ Περιφέρεια, κατανάλωση μπίρας στο σπίτι/ κατανάλωση μπίρας έξω από το σπίτι). Ο συντελεστής ϕ εξετάζει την κατεύθυνση και την ένταση της συσχέτισης μεταξύ των μεταβλητών.

Για παράδειγμα, είκοσι καταναλωτές ρωτήθηκαν σε ποιο από τα δύο παντοπωλεία της γειτονιάς τους συνηθίζουν να ψωνίζουν (1= από το παντοπωλείο Α, 2= από το παντοπωλείο Β) και ποια από δύο γνωστές μάρκες γάλακτος προτιμούν (1= μάρκα Χ, 2= μάρκα Ψ).

A/A	Παντοπωλείο	Μάρκα Γάλακτος
1	1	1
2	2	2
3	2	2
4	2	1
5	1	2
6	1	1
7	2	2
8	1	1
9	2	1
10	1	1
11	1	1
12	1	2
13	2	2
14	1	1
15	1	1
16	2	2
17	1	2
18	1	1
19	2	2
20	2	2

Τα στοιχεία του παραδείγματος παρουσιάζονται εισηγμένα στο SPSS στο Σχήμα 4-21.

Σχήμα 4-21:

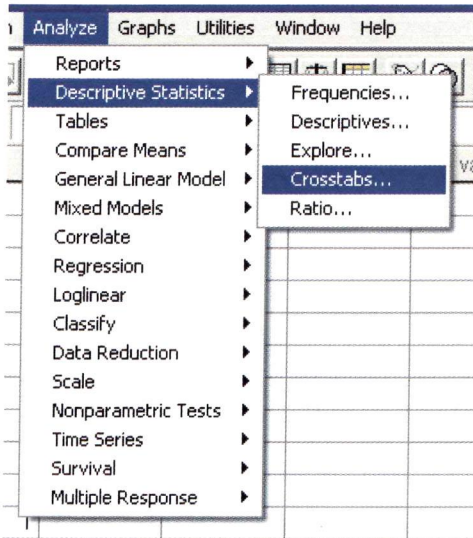
Ο Συντελεστής Συσχέτισης Phi (1ο βήμα)

	shop	milk
1	1	1
2	2	2
3	2	2
4	2	1
5	1	2
6	1	1
7	2	2
8	1	1
9	2	1
10	1	1
11	1	1
12	1	2
13	2	2
14	1	1
15	1	1
16	2	2
17	1	2
18	1	1
19	2	2
20	2	2

Για τον υπολογισμό του Phi ακολουθείται η εξής διαδικασία:
Analyze → Descriptive Statistics → Crosstabs (Σχήμα 4-22).

Σχήμα 4-22:

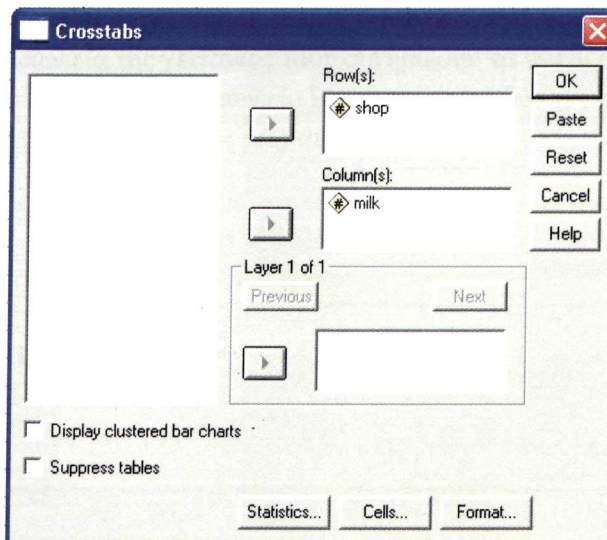
Ο Συντελεστής Συσχέτισης Phi (2ο βήμα)



Η μεταβλητή “shop” μεταφέρεται στο πλαίσιο “Row(s)” και η μεταβλητή “milk” στο πλαίσιο “Column(s)” (Σχήμα 4-23).

Σχήμα 4-23:

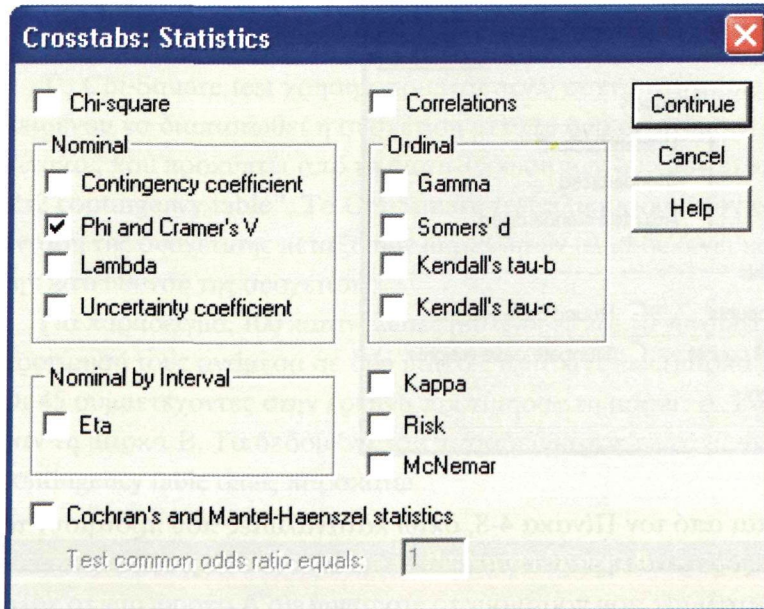
Ο Συντελεστής Συσχέτισης Phi (3ο βήμα)



Από την επιλογή “Statistics” του Σχήματος 4-23, εμφανίζεται το πλαίσιο διαλόγου “Statistics” στο οποίο ο ερευνητής θα πρέπει να επιλέξει το “Phi and Cramer’s V” (Σχήμα 4-24).

Σχήμα 4-24:

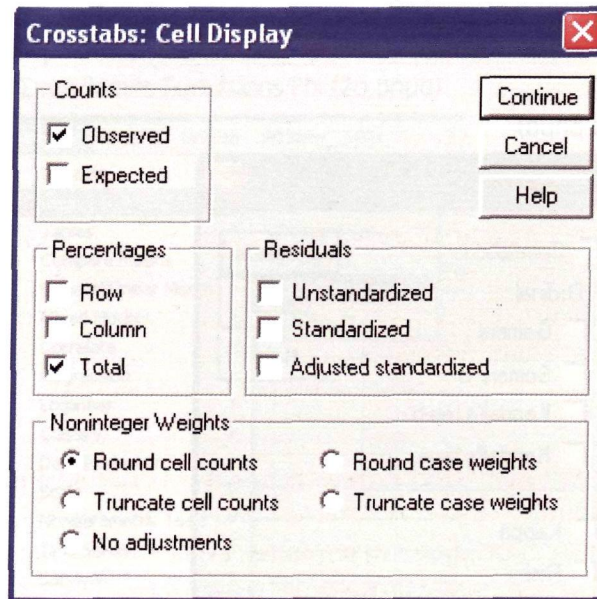
Ο Συντελεστής Συσχέτισης Phi (4ο βήμα)



Ο πρώτος Πίνακας του Output παρουσιάζει τις πινακοειδείς διασταυρώσεις (cross tabs) μεταξύ των δύο μεταβλητών. Για να παρουσιάσει το Output και τα ποσοστά των διασταυρώσεων μεταξύ των δύο μεταβλητών, από την επιλογή “Cells” του Σχήματος 4-23 ζητείται η εμφάνιση των ποσοστών “Total” από τα “Percentages” (Σχήμα 4-25).

Σχήμα 4-25:

Ο Συντελεστής Συσχέτισης Phi (5ο βήμα)



Όπως φαίνεται από τον Πίνακα 4-8, οκτώ καταναλωτές που προτιμούν το παντοπωλείο Α αγοράζουν τη μάρκα γάλακτος Χ (ποσοστό 40%), ενώ 3 από τους 11 συνολικά καταναλωτές που προτιμούν το παντοπωλείο Α αγοράζουν το γάλα Ψ (ποσοστό 15%). Αντίθετα, 2 από τους 9 συνολικά καταναλωτές που φωνίζουν από το παντοπωλείο Β αγοράζουν το γάλα Χ (ποσοστό 10%) και 7 από αυτούς το γάλα Ψ (ποσοστό 35%).

Πίνακας 4-8: Ο Συντελεστής Συσχέτισης Phi - Shop*Milk Cross Tabulation

			milk		Total
			1	2	
shop 1	Count	8	3	11	
	% of Total	40,0%	15,0%	55,0%	
2	Count	2	7	9	
	% of Total	10,0%	35,0%	45,0%	
Total	Count	10	10	20	
	% of Total	50,0%	50,0%	100,0%	

Ο Πίνακας 4-9 παρουσιάζει το συντελεστή συσχέτισης Phi (0,503). Η συσχέτιση μεταξύ της επιλογής του παντοπωλείου και της επιλογής γάλακτος είναι ασθενώς θετική. Από τη στήλη του Πίνακα “Approx.Sig.” προκύπτει ότι η συσχέτιση μεταξύ των δύο μεταβλητών είναι στατιστικά σημαντική ($p < 0,05$).

Πίνακας 4-9: Ο Συντελεστής Συσχέτισης Phi -Symmetric Measures

		Value	Approx. Sig.
Nominal by	Phi	,503	,025
Nominal	Cramer's V	,503	,025
N of Valid Cases		20	

4.6 Contingency Tables - Chi-Square (χ^2)

Το Chi-Square test χρησιμοποιείται πολύ συχνά στην έρευνα αγοράς προκειμένου να διαπιστωθεί η συσχέτιση μεταξύ δύο διχοτομικών μεταβλητών. Ο πίνακας που προκύπτει από τη διασταύρωση των δύο μεταβλητών ονομάζεται “2x2 contingency table”. Το Chi-Square test πληροφορεί τον ερευνητή για την ένταση της συσχέτισης μεταξύ των μεταβλητών αλλά δε δίνει καμία ένδειξη για την κατεύθυνση της συσχέτισης.

Για παράδειγμα, 100 καταναλωτές (50 άνδρες και 50 γυναίκες) εξέφρασαν την προτίμησή τους ανάμεσα σε δύο μάρκες αυτοκινήτων (μάρκα A και μάρκα B). Οι 45 συμμετέχοντες στην έρευνα προτίμησαν τη μάρκα A, ενώ οι 55 προτίμησαν τη μάρκα B. Τα δεδομένα του παραδείγματος εμφανίζονται σε έναν 2x2 contingency table όπως παρακάτω.

	Προτίμηση Μάρκας A	Προτίμηση Μάρκας B	Σύνολο
Άνδρες	30	20	50
Γυναίκες	15	35	50
Σύνολο	45	55	100

Τα δεδομένα του πίνακα παρουσιάζονται στο Σχήμα 4-26 εισηγμένα στο SPSS.

Σχήμα 4-26:

Chi-Square – Παράδειγμα

	group	choice	count
1	men	brandA	30
2	men	brandB	20
3	women	brandA	15
4	women	brandB	35
5			

Στο συγκεκριμένο παράδειγμα υπάρχουν δύο κατηγοριοποιημένες μεταβλητές: «φύλο» (άνδρας/ γυναίκα) και «προτιμητέα μάρκα» (μάρκα A/ μάρκα B). Η μηδενική υπόθεση είναι H_0 : Δεν υπάρχει σχέση μεταξύ των δύο μεταβλητών.

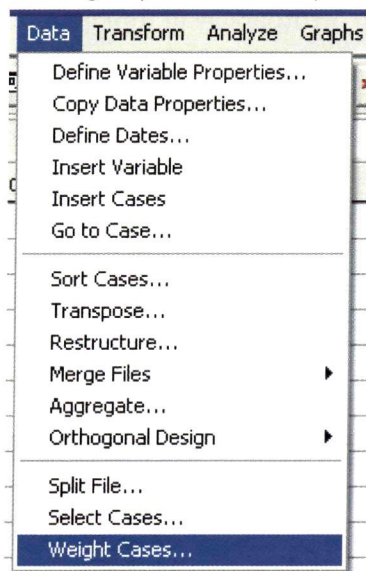
Όμως, εφόσον οι αριθμοί της στήλης “count” αντιπροσωπεύουν συχνότητες και όχι απλές τιμές, η εντολή “Weight Cases” πρέπει να προηγηθεί της ανάλυσης, έτσι ώστε το πρόγραμμα να προσαρμόσει την ανάλυση Chi-Square ανάλογα.

Για να δηλωθούν ότι οι τιμές της μεταβλητής “count” αντιπροσωπεύουν συχνότητες ακολουθείται η παρακάτω διαδικασία:

Data → Weight Cases (Σχήμα 4-27).

Σχήμα 4-27:

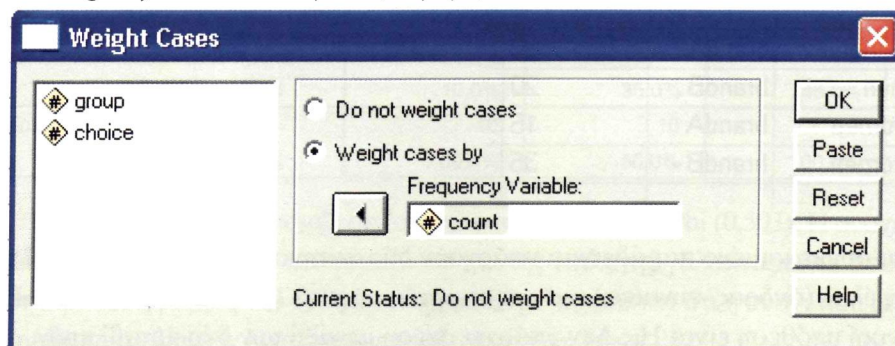
Contingency Table - Chi-Square (1ο βήμα)



Στη συνέχεια, επιλέγεται το “Weight cases by” και η μεταβλητή “count” μεταφέρεται στο πλαίσιο “Frequency Variable” (Σχήμα 4-28).

Σχήμα 4-28:

Contingency Table - Chi-Square (2ο βήμα)

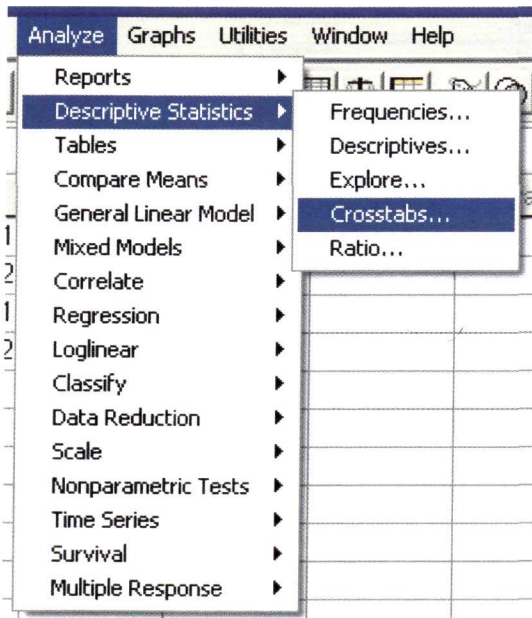


Πατώντας “OK”, η εντολή έχει γίνει δεκτή και ο ερευνητής μπορεί να προχωρήσει στην ανάλυση chi-square. Για τον υπολογισμό του συντελεστή chi-square ακολουθείται η εξής διαδικασία:

Analyze → Descriptive Statistics → Crosstabs (Σχήμα 4-29).

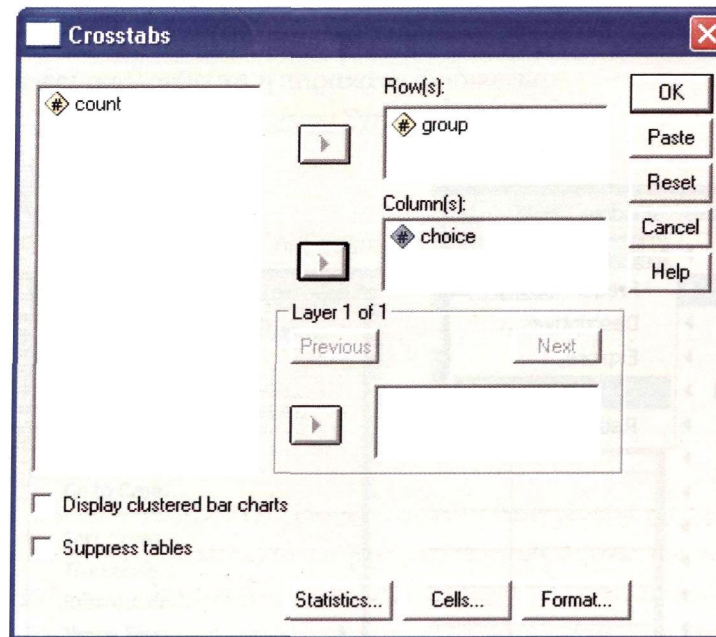
Σχήμα 4-29:

Chi-Square (3ο βήμα)



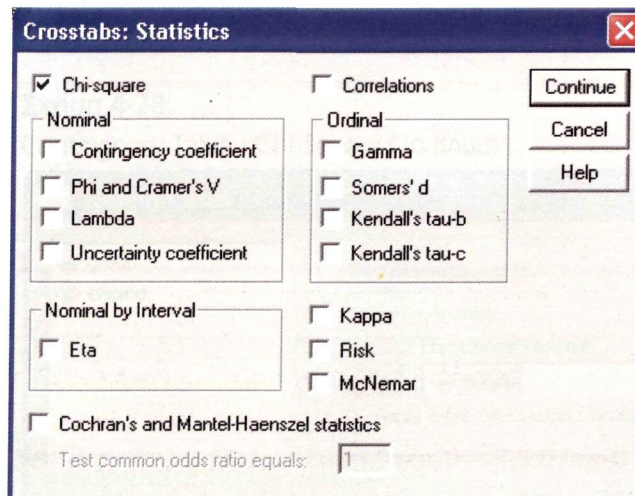
Στη συνέχεια, στο πλαίσιο διαλόγου “Crosstabs” εισάγεται η μεταβλητή “group” στο πλαίσιο “Row(s)” και η μεταβλητή “choice” στο πλαίσιο “Column(s)” (Σχήμα 4-30).

Σχήμα 4-30:
Chi-Square (4ο βήμα)



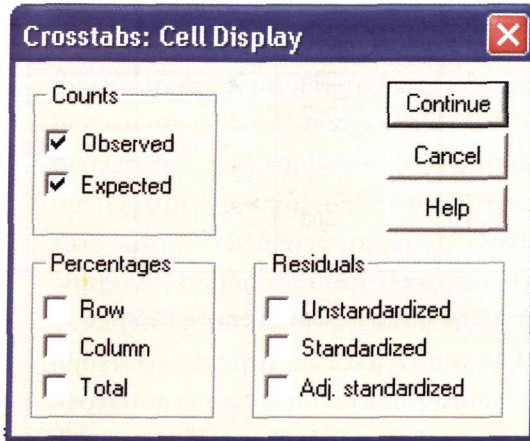
Από το “Statistics” επιλέγεται το “Chi-square” (Σχήμα 4-31) και από το “Cells” επιλέγεται το “Expected Counts” (Σχήμα 4-32).

Σχήμα 4-31:
Chi-Square (5ο βήμα)



Σχήμα 4-32:

Chi-Square (6ο βήμα)



Για τους 2x2 contingency tables, για τον υπολογισμό του chi-square είναι επιθυμητό ο αριθμός όλων των αναμενόμενων συχνοτήτων (expected frequencies) να είναι μεγαλύτερος ή ίσος με 5. Σε μεγαλύτερους πίνακες απαιτείται όλες οι αναμενόμενες συχνότητες να είναι μεγαλύτερες ή ίσες με 1 και το 20% των αναμενόμενων συχνοτήτων να είναι μεγαλύτερες από 5. Η επιλογή “Expected Counts” επιτρέπει στον ερευνητή να ελέγξει τις παραπάνω προϋποθέσεις.

Πατώντας “OK”, ο πρώτος πίνακας του Output παρουσιάζει τις πινακοειδείς διασταυρώσεις (cross tabulation) των συχνοτήτων και των αναμενόμενων συχνοτήτων των δύο μεταβλητών (Πίνακας 4-10).

Πίνακας 4-10: Chi-Square – Crosstabulation

			CHOICE		Total
			brandA	brandB	
GROUP	men	Count	30	20	50
		Expected Count	22,2	27,5	50,0
	women	Count	15	35	50
		Expected Count	22,5	27,5	50,0
Total		Count	45	55	100
		Expected Count	45,0	55,0	100,0

Παρατηρείται ότι όλες οι συχνότητες του Πίνακα 4-10 είναι μεγαλύτερες από 5 και συνεπώς είναι εφικτός ο υπολογισμός του chi-square.

Ο Πίνακας 4-11 παρουσιάζει το συντελεστή chi-square και τη σημαντικότητά του.

Πίνακας 4-11: Chi-square tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9,091 ^b	1	,003		
Continuity Correction ^a	7,919	1	,005		
Likelihood Ratio	9,240	1	,002		
Fisher's Exact Test				,005	,002
Linear-by-Linear Association	9,000	1	,003		
N of Valid Cases	100				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 22,50.

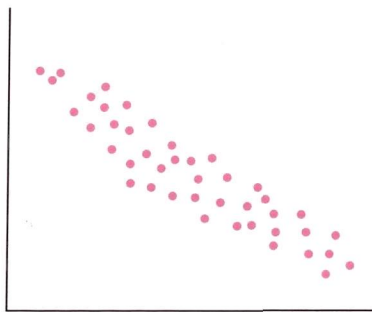
Από τον Πίνακα 4-11 συνάγεται ότι, υπάρχει ισχυρή συσχέτιση μεταξύ του φύλου και της προτίμησης μάρκας αυτοκινήτου ($\chi^2 = 9,091$, $df = 1$). Η συσχέτιση είναι στατιστικά σημαντική σε επίπεδο στατιστικής σημαντικότητας $\alpha = 0,005$. Η υποσημείωση "b", που φαίνεται στο κάτω μέρος του πίνακα, πληροφορεί τον ερευνητή πόσα «κελιά» του contingency table έχουν αναμενόμενη συχνότητα μικρότερη από 5. Αν κάποια κελιά είχαν αναμενόμενες συχνότητες μικρότερες από 5 (στη συγκεκριμένη περίπτωση δεν υπάρχουν), τότε το "Fisher's Exact Test" θεωρείται καταλληλότερο.

Περίληψη

Η ανάλυση συσχέτισης δίνει πληροφορίες στον ερευνητή για την κατεύθυνση και την ένταση της σχέσης μεταξύ των μεταβλητών της έρευνας. Συνήθως, οι συντελεστές συσχέτισης (π.χ., του Pearson, ο συντελεστής μερικής συσχέτισης, κ.ά.) λαμβάνουν τιμές μεταξύ του -1 και του +1. Εάν η τιμή του συντελεστή βρίσκεται μεταξύ του -1 και του 0, τότε υπάρχει αρνητική συσχέτιση μεταξύ των δύο μεταβλητών. Αντίθετα, αν ο συντελεστής συσχέτισης λαμβάνει τιμή μεταξύ του 0 και του +1, τότε συμπεραίνεται ότι υπάρχει θετική συσχέτιση. Στην περίπτωση που ο συντελεστής συσχέτισης ισούται με μηδέν, ο ερευνητής καταλήγει στο συμπέρασμα ότι δεν υπάρχει γραμμική συσχέτιση μεταξύ των δύο μεταβλητών. Τιμές του συντελεστή συσχέτισης από +1 έως +0,7 (ή από -1 έως -0,7) παραπέμπουν σε ισχυρή συσχέτιση μεταξύ των μεταβλητών. Εάν ο συντελεστής συσχέτισης είναι μεταξύ του +0,7 και +0,3 (ή -0,7 έως -0,3), τότε συμπεραίνεται ότι υπάρχει ασθενής συσχέτιση μεταξύ των μεταβλητών. Υπενθυμίζεται ότι, αν και η ανάλυση συσχέτισης αντικατοπτρίζει την ένταση της σχέσης μεταξύ των μεταβλητών, δεν εξηγεί τη σχέση αιτίας-αιτιατού.

Σωστό ή Λάθος;

1. Ο συντελεστής συσχέτισης Pearson υπολογίζει και την «κατεύθυνση» της σχέσης μεταξύ των μεταβλητών.
 Σωστό
 Λάθος
2. Εάν η τιμή του συντελεστή συσχέτισης βρίσκεται μεταξύ του -1 και του 0, τότε υπάρχει πάντοτε αρνητική ισχυρή συσχέτιση μεταξύ των δύο μεταβλητών.
 Σωστό
 Λάθος
3. Αν η τιμή του συντελεστή συσχέτισης είναι +0,934, τότε συμπεραίνεται ότι υπάρχει ισχυρή θετική συσχέτιση μεταξύ των μεταβλητών.
 Σωστό
 Λάθος
4. Το παρακάτω διάγραμμα απεικονίζει ασθενή αρνητική συσχέτιση μεταξύ των μεταβλητών.



- Σωστό
 Λάθος
5. Ο συντελεστής συσχέτισης του Pearson μπορεί να υπολογιστεί και για nominal μεταβλητές, αρκεί αυτές να μην είναι διχοτομικές.
 Σωστό
 Λάθος

6. Στο scatterplot απεικονίζεται και η σχέση αιτίας-αιτιατού (δηλ. ποια μεταβλητή προκαλεί μεταβολή της άλλης) μεταξύ των μεταβλητών.
- Σωστό
 Λάθος
7. Στο συντελεστή Pearson με την επιλογή “One-tailed significance” ελέγχεται η μηδενική υπόθεση « H_0 : Η συσχέτιση είναι μηδενική».
- Σωστό
 Λάθος
8. Για την ανάλυση συσχέτισης με τρεις μεταβλητές ακολουθείται η διαδικασία: Analyze → Correlate → Multivariate.
- Σωστό
 Λάθος
9. Στην ανάλυση συσχέτισης με τρεις μεταβλητές παρουσιάζεται στο Output ο βαθμός της συσχέτισης των μεταβλητών (συντελεστής συσχέτισης του Pearson) ανά ζεύγη.
- Σωστό
 Λάθος
10. Για να υπολογιστεί ο συντελεστής μερικής συσχέτισης, ακολουθείται η διαδικασία: Analyze → Correlate → Partial.
- Σωστό
 Λάθος
11. Ο συντελεστής συσχέτισης Kendall's tau (τ) χρησιμοποιείται σε περιπτώσεις όπου οι δύο μεταβλητές είναι διχοτομικές.
- Σωστό
 Λάθος
12. Ο συντελεστής συσχέτισης Phi (ϕ) χρησιμοποιείται στην περίπτωση όπου οι δύο εξεταζόμενες μεταβλητές της έρευνας είναι ordinal.
- Σωστό
 Λάθος
13. Η μηδενική υπόθεση στην ανάλυση chi-square είναι « H_0 : Δεν υπάρχει σχέση μεταξύ των δύο μεταβλητών».

- Σωστό
 Λάθος
14. Πριν από τον υπολογισμό του συντελεστή συσχέτισης Pearson θα πρέπει πάντα να προηγείται η διαδικασία του “Weight Cases”.
- Σωστό
 Λάθος
15. Στους 2x2 contingency tables, για τον υπολογισμό του chi-square είναι επιθυμητό ο αριθμός όλων των αναμενόμενων συχνοτήτων να είναι μεγαλύτερος ή ίσος με 5.
- Σωστό
 Λάθος
16. Το chi-square δείχνει και την κατεύθυνση της σχέσης μεταξύ των μεταβλητών.
- Σωστό
 Λάθος
17. Το Fisher’s Exact Test θεωρείται καταλληλότερο από το Chi-Square στις περιπτώσεις όπου υπάρχουν περισσότερες από δύο μεταβλητές.
- Σωστό
 Λάθος
18. Στο Chi-Square, οι αναμενόμενες συχνότητες παρουσιάζονται στον πίνακα “Chi-square Tests”.
- Σωστό
 Λάθος
19. Από την εντολή “Weight Cases” εμφανίζεται μία νέα μεταβλητή στη Data View, η οποία αντιπροσωπεύει τις συχνότητες των μεταβλητών.
- Σωστό
 Λάθος
20. Ο πίνακας που προκύπτει από τη διασταύρωση των δύο διχοτομικών μεταβλητών ονομάζεται 2x2 contingency table.
- Σωστό
 Λάθος

Ασκήσεις

Άσκηση 1

Σε μία έρευνα κλήθηκαν 30 καταναλωτές (10 έγγαμοι και 10 άγαμοι) να ανακαλέσουν στη μνήμη τους τη διαφήμιση μιας μάρκας παιδικού παιχνιδιού γνωστής εταιρείας. Τα αποτελέσματα της έρευνας παρουσιάζονται στον παρακάτω Πίνακα.

A/A	Οικογ. Κατάσταση	Ανάκληση Διαφήμισης	A/A	Οικογ. Κατάσταση	Ανάκληση Διαφήμισης
1	Έγγαμος	ΝΑΙ	16	Άγαμος	ΟΧΙ
2	Έγγαμος	ΝΑΙ	17	Άγαμος	ΟΧΙ
3	Έγγαμος	ΝΑΙ	18	Άγαμος	ΝΑΙ
4	Έγγαμος	ΝΑΙ	19	Άγαμος	ΟΧΙ
5	Έγγαμος	ΟΧΙ	20	Άγαμος	ΟΧΙ
6	Έγγαμος	ΝΑΙ	21	Άγαμος	ΟΧΙ
7	Έγγαμος	ΝΑΙ	22	Άγαμος	ΟΧΙ
8	Έγγαμος	ΟΧΙ	23	Άγαμος	ΟΧΙ
9	Έγγαμος	ΝΑΙ	24	Άγαμος	ΟΧΙ
10	Έγγαμος	ΝΑΙ	25	Άγαμος	ΝΑΙ
11	Έγγαμος	ΝΑΙ	26	Άγαμος	ΝΑΙ
12	Έγγαμος	ΝΑΙ	27	Άγαμος	ΟΧΙ
13	Έγγαμος	ΝΑΙ	28	Άγαμος	ΟΧΙ
14	Έγγαμος	ΝΑΙ	29	Άγαμος	ΝΑΙ
15	Έγγαμος	ΝΑΙ	30	Άγαμος	ΝΑΙ

Ζητούμενα

1. Εισάγετε τα στοιχεία του Πίνακα στο SPSS.
2. Πόσοι καταναλωτές συνολικά ανακάλεσαν τη διαφήμιση του παιδικού παιχνιδιού;
3. Πόσοι άγαμοι καταναλωτές κατάφεραν να ανακαλέσουν στη μνήμη τους τη διαφήμιση του παιχνιδιού;
4. Τι ποσοστό των έγγαμων καταναλωτών κατάφεραν να ανακαλέσουν στη μνήμη τους τη διαφήμιση του παιχνιδιού;
5. Υπολογίστε το συντελεστή συσχέτισης Phi. Υπάρχει ισχυρή συσχέτιση μεταξύ της οικογενειακής κατάστασης των ερωτώμενων και της ικανότητας ανάκλησης της διαφήμισης; Είναι η συσχέτιση στατιστικά σημαντική;
6. Κατασκευάστε ένα 2x2 contingency table για τα δεδομένα του πίνακα.
7. Υπολογίστε το chi-square αφού πρώτα διατυπώσετε τις ερευνητικές υποθέσεις H_0 και H_1 .

8. Υπάρχουν αναμενόμενες συχνότητες που να είναι μικρότερες του 5;
9. Θεωρείτε ότι το “Fisher’s Exact Test” θα ήταν καταλληλότερο; Γιατί;
10. Είναι το chi-square στατιστικά σημαντικό;

Άσκηση 2

Είκοσι καταναλωτές κλήθηκαν να αξιολογήσουν (από το 1 έως το 10) ένα προϊόν και να δηλώσουν πόσο τους αρέσει το διαφημιστικό μήνυμα (από το 1 έως το 10) του προϊόντος που προβάλλεται στην τηλεόραση.

Τα αποτελέσματα της έρευνας παρουσιάζονται στον Πίνακα που ακολουθεί.

A/A	Προϊόν	Διαφήμιση
1	5	5
2	5	6
3	6	7
4	3	3
5	10	6
6	4	2
7	7	5
8	6	6
9	7	5
10	9	9
11	8	8
12	5	3
13	10	9
14	10	8
15	5	4
16	5	5
17	4	5
18	5	6
19	3	3
20	4	2

Ζητούμενα

1. Εισάγετε τα δεδομένα του παραπάνω Πίνακα στο SPSS.
2. Κατασκευάστε ένα scatterplot για τα δεδομένα της άσκησης. Τι παρατηρείτε; Υπάρχει θετική ή αρνητική συσχέτιση μεταξύ των μεταβλητών;
3. Υπολογίστε το συντελεστή συσχέτισης Pearson.
4. Ποια είναι η μέση στάση (αξιολόγηση) για το προϊόν; Ποια είναι η τυπική απόκλιση;

5. Ποια είναι η μέση στάση (αξιολόγηση) για τη διαφήμιση του προϊόντος; Ποια είναι η τυπική απόκλιση;
6. Είναι η συσχέτιση μεταξύ των δύο μεταβλητών ισχυρή;
7. Είναι ο συντελεστής συσχέτισης Pearson στατιστικά σημαντικός;

Άσκηση 3

Στον πίνακα που ακολουθεί καταγράφονται ο αριθμός των καταναλωτών (ελεύθεροι επαγγελματίες και μισθωτοί) που προτιμούν το κατάστημα Χ ή το κατάστημα Ψ.

	Προτίμηση Καταστήματος Χ	Προτίμηση Καταστήματος Ψ	Σύνολο
Ελεύθεροι Επαγγελματίες	20	30	50
Μισθωτοί	45	5	50
Σύνολο	65	35	100

Ζητούμενα

1. Εισάγετε τα στοιχεία του Πίνακα στο SPSS.
2. Εφαρμόστε την εντολή “Weight cases” έτσι ώστε να είναι εφικτός ο υπολογισμός του chi-square test.
3. Υπολογίστε το chi-square αφού πρώτα διατυπώσετε τις ερευνητικές υποθέσεις H_0 και H_1 .
4. Τι ποσοστό των ελεύθερων επαγγελματιών προτιμούν το κατάστημα Χ; Τι ποσοστό των μισθωτών προτιμούν το κατάστημα Ψ;
5. Υπάρχουν αναμενόμενες συχνότητες που να είναι μικρότερες του 5;
6. Θεωρείτε ότι το “Fisher’s Exact Test” θα ήταν καταλληλότερο; Γιατί;
7. Είναι το chi-square στατιστικά σημαντικό;

Mini Projects

Mini Project A

Ανοίξτε το αρχείο “Kefaleo4_project_1” από το CD-ROM. Τα δεδομένα του αρχείου προέρχονται από μία έρευνα σχετικά με τα χρήματα που δαπανούν οι μαθητές στην Αγγλία σε τσιγάρα (μεταβλητή *moneycig*) και αλκοόλ (*moneyalc*) την εβδομάδα (σε £). Οι μαθητές επίσης ρωτήθηκαν για το χαρτζιλίκι που λαμβάνουν το μήνα (μεταβλητή *allmoney*).

Ζητούμενα

1. Υπολογίστε το συντελεστή συσχέτισης Pearson για τις τρεις μεταβλητές.
2. Πόσα χρήματα δαπανούν κατά μέσο όρο την εβδομάδα οι μαθητές για τσιγάρα; Πόσα χρήματα δαπανούν κατά μέσο όρο την εβδομάδα οι μαθητές για αλκοόλ;
3. Είναι η συσχέτιση μεταξύ των χρημάτων που λαμβάνουν οι μαθητές την εβδομάδα και των χρημάτων που ξοδεύουν για τσιγάρα ισχυρή; Είναι θετική ή αρνητική;
4. Είναι η συσχέτιση μεταξύ των χρημάτων που λαμβάνουν οι μαθητές την εβδομάδα και των χρημάτων που ξοδεύουν για αλκοόλ ισχυρή; Είναι θετική ή αρνητική;
5. Είναι οι δύο συσχετίσεις στατιστικά σημαντικές;
6. Ποιος είναι ο συντελεστής μερικής συσχέτισης για τις μεταβλητές “*moneycig*” και “*moneyalc*”, αν χρησιμοποιηθεί η μεταβλητή “*allmoney*” ως μεταβλητή ελέγχου;
7. Συγκρίνετε το συντελεστή μερικής συσχέτισης με το συντελεστή συσχέτισης Pearson για τις αντίστοιχες μεταβλητές. Τι παρατηρείτε;

Mini Project B

Ανοίξτε το αρχείο “kefaleo4_project_2”. Το αρχείο περιλαμβάνει στοιχεία από μία έρευνα σχετικά με τη χρήση του Internet. Για τους σκοπούς της έρευνας, 300 καταναλωτές [135 απόφοιτοι ΑΕΙ/ΤΕΙ (τιμή: 1) και 165 απόφοιτοι γυμνασίου/λυκείου (τιμή: 0)] ρωτήθηκαν αν χρησιμοποιούν το Internet (0= όχι, 1=Ναι).

Ζητούμενα

1. Πόσοι καταναλωτές συνολικά χρησιμοποιούν το Internet;
2. Πόσοι απόφοιτοι ΑΕΙ/ΤΕΙ χρησιμοποιούν το Internet;
3. Τι ποσοστό των αποφοίτων ΑΕΙ/ΤΕΙ δε χρησιμοποιούν το Internet;

4. Υπολογίστε το συντελεστή συσχέτισης Phi. Υπάρχει ισχυρή συσχέτιση μεταξύ του μορφωτικού επιπέδου και της χρήσης του Internet;
5. Υπολογίστε το chi-square, αφού πρώτα διατυπώσετε τις ερευνητικές υποθέσεις H_0 και H_1 .
6. Υπάρχουν αναμενόμενες συχνότητες που να είναι μικρότερες του 5;
7. Είναι το chi-square στατιστικά σημαντικό σε επίπεδο στατιστικής σημαντικότητας $\alpha=0,001$;

Βιβλιογραφία Κεφαλαίου 4

- Bobko, P. (2001), *Correlation and Regression*, 2nd edition, Thousand Oaks, CA: Sage Publications.
- Chen, P.Y. and P.M. Popovich (2002), *Correlation: Parametric and Nonparametric Measures*, Thousand Oaks, CA: Sage Publications.
- Cochran, W.G. (1952), "The χ^2 Test of Goodness of Fit," *Annals of Mathematical Statistics*, Vol. 23, pp. 315-345.
- Cohen, J. and P.Cohen (1983), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2nd Edition. Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Fineberg, S.E. (1977), *The Analysis of Cross-Classified Categorical Data*, Boston: MIT Press.
- Goodman, L.A. and W.H. Kruskal (1954), "Measures of Association for Cross Classification," *Journal of the American Statistical Association*, Vol. 49 (December), pp. 732-764.
- Goodman, L.A. and W.H. Kruskal (1954), "Measures of Association for Cross Classification: Appropriate Sampling Theory," *Journal of the American Statistical Association*, Vol. 88 (June), pp. 310-254.
- Gorsuch, R.L. (1973), "Data Analysis of Correlated Independent Variables," *Multivariate Behavioral Research*, Vol. 8 (January), pp. 89-107.
- Hellevik, O. (1984), "Statistical Tests and Substantive Significance," *The American Sociologist*, Vol. 4 (February).
- Kamakura, W.A. and M. Wedel (1977), "Statistical Data Fusion for Cross Tabulations," *Journal of Marketing Research*, Vol. 34, No. 2 (November), pp. 485-498.
- Kendall, M. and J. Dickinson-Gibbons (1990), *Rank Correlation Methods*, 5th Edition, NY: Oxford Univ Press.
- Martin, W.S. (1978), "Effects of Scaling on the Correlation Coefficient: Additional Considerations," *Journal of Marketing Research*, Vol. 15 (May), pp. 304-308.
- Liebetrau, A.M. (1983), *Measures of Association*, Beverly Hills, Thousand Oaks, CA: Sage Publications.

- Olsson, U. (1980), "Measuring Correlation in Ordered Two-Way Contingency Tables," *Journal of Marketing Research*, Vol. 17 (August), pp. 391-394.
- Peterson, R.A. and S. Sharma (1977), "Adjusting Correlation Coefficients for the Effects of Scaling," *Proceedings*, Business and Economic Statistics Section, American Statistical Association, pp. 784-786.
- Reynolds, H.T. (1977), *The Analysis of Cross-Classifications*, New York: The Free Press.
- Sharma, S. and R.A. Peterson (1982), "Effects of Scaling on the Correlation Coefficient: Issues and Outlooks," presented at the *Annual Conference of the American Institute for Decision Sciences*.
- Selvin, H.C. and A. Stuart (1966), "Data-Dredging Procedures in Survey Analysis," *The American Statistician*, June, pp. 20-23.