

Μεθοδολογία των επιστημών του Ανθρώπου: Στατιστική Ι

Εργαστήριο 9

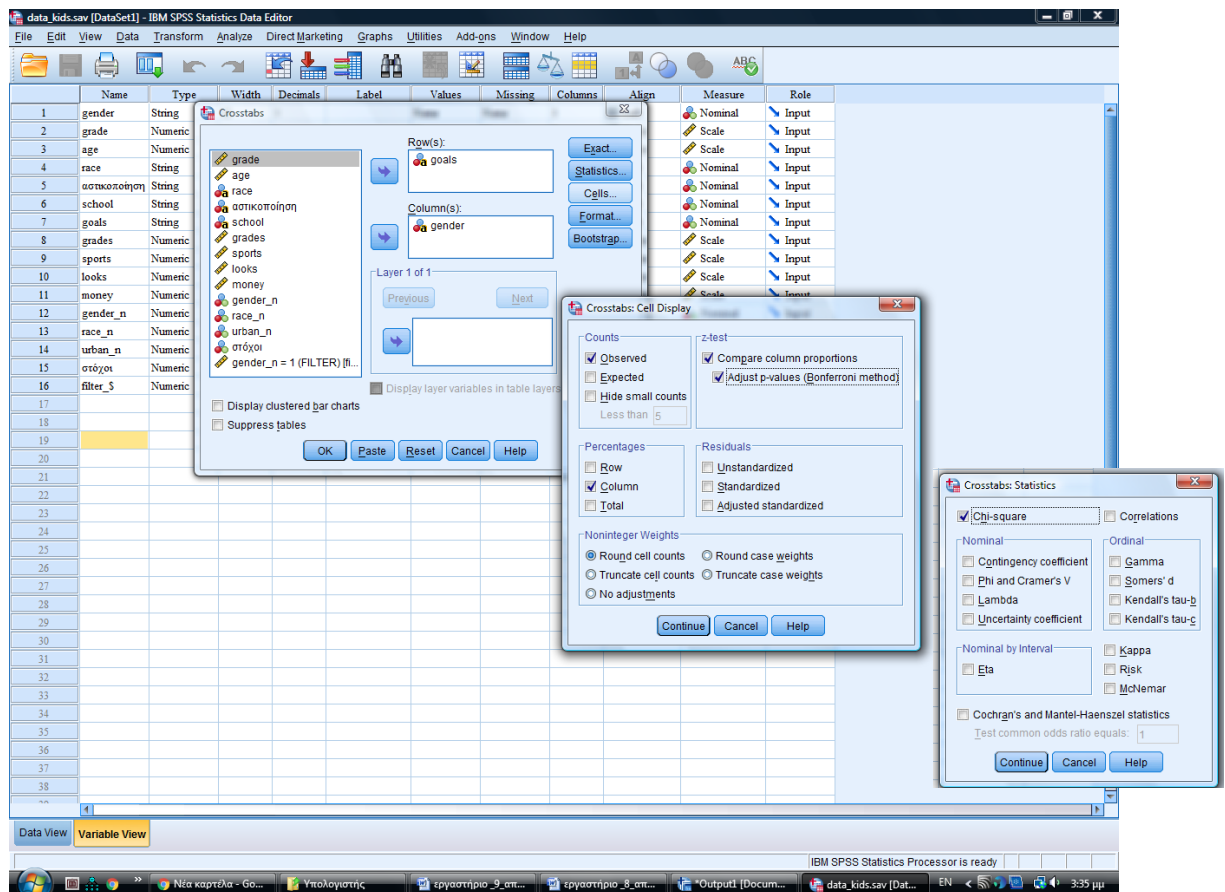
1. Να χρησιμοποιηθεί το αρχείο **data_kids**. Τα δεδομένα του προέρχονται από την έρευνα των Chase και Dummer (1992), μελέτησαν τον ρόλο των σπορ ως καθοριστικό κοινωνικό παράγοντα για τα παιδιά.
 - a. Να διερευνηθεί η σχέση ανάμεσα στο φύλο (μεταβλητή **gender_n**) και την προτεραιότητα στόχων του μαθητή. Για την προτεραιότητα των στόχων (μεταβλητή **goals**), οι μαθητές 10- 12 ετών δήλωσαν ποιο ήταν πιο σημαντικό ανάμεσα στο «καλό βαθμό», «δημοτικότητα» και «σπορ».

Προκειμένου να διερευνηθεί η σχέση δύο κατηγορικών μεταβλητών κατασκευάζεται ο πίνακας συνάφειας με $n \times k$ αριθμό κελιών, όπου n ο αριθμός των κατηγοριών της μίας μεταβλητής και k ο αριθμός κατηγοριών της άλλης μεταβλητής. Στις στήλες του πίνακα βάζουμε την ανεξάρτητη μεταβλητή ενώ στις γραμμές την εξαρτημένη. Ο πίνακας συνάφειας στο SPSS κατασκευάζεται ως εξής:

Από το μενού Analyze =>Descriptive Statistics => Crosstabs, εισάγουμε στο πλαίσιο Row την μεταβλητή goals και στο πλαίσιο Column την μεταβλητή gender_n. Χρησιμοποιώντας τα κουμπιά επιλογής έχουμε:

Στην επιλογή Cells στο πλαίσιο Percentages τσεκάρουμε Column. Επίσης, στην ίδια επιλογή στο πλαίσιο z-test τσεκάρουμε την επιλογή Compare column proportions και την επιλογή Adjust p-values (Bonferroni method) και πατάμε continue.

Στην επιλογή Statistics τσεκάρουμε Chi-square και πατάμε continue.



Μετά το «κλικ» στο OK του παραθύρου παίρνουμε στο output τους εξής πίνακες:

Πίνακας 1

goals * gender Crosstabulation

		gender		Total
		boy	girl	
Grades	Count	117 _a	130 _a	247
	% within gender_n	51,5%	51,8%	51,7%
	Count	50 _a	91 _b	141
Popular	% within gender_n	22,0%	36,3%	29,5%
	Count	60 _a		90
Sports	% within gender_n	26,4%	12,0%	18,8%
	Count	227	251	478
Total	% within gender_n	100,0%	100,0%	100,0%

Προκειμένου να κατανοήσουμε τη σχέση των δύο μεταβλητών συγκρίνονται σε κάθε γραμμή όλα τα ποσοστά μεταξύ τους (π.χ. το 22% των αγοριών επιλέγουν Popular ενώ ένα πολύ μεγαλύτερο ποσοστό, 36,3% των κοριτσιών επιλέγουν την ίδια κατηγορία.

Each subscript letter denotes a subset of gender_n categories whose column proportions do not differ significantly from each other at the ,05 level.

Πίνακας 1

goals * gender_n Crosstabulation

		gender_n		Total
		boy	girl	
goals	Grades	Count 117 _a	130 _a	247
		% within gender_n 51,5%	51,5%	51,5%
	Popular	Count 50 _a	91 _b	141
		% within gender_n 22,0%	36,3%	29,5%
	Sports	Count 60 _a	30 _b	90
		% within gender_n 26,4%	12,0%	18,8%
Total	Count 227	251	478	
	% within gender_n 100,0%	100,0%	100,0%	

Each subscript letter denotes a subset of gender_n categories whose column proportions do not differ significantly from each other at the ,05 level.

Αφού απορρίφτηκε η μηδενική υπόθεση (βλέπε τον παρακάτω πίνακα) τότε και μόνο τότε διενεργούνται όλες οι δυνατές ανά ζεύγη συγκρίσεις και το αποτέλεσμά τους δίνεται με τη μορφή συμβόλων (a, b, c...) ως δείκτες των απόλυτων συχνοτήτων). Εάν στους δείκτες δύο συχνοτήτων υπάρχει ένα κοινό γράμμα-σύμβολο τότε δεν απορρίπτεται η Η₀ της ισότητας των δύο ποσοστών σε επίπεδο $\alpha = 0,05$. Για παράδειγμα, στην πρώτη γραμμή δεν έχουμε σημαντική διαφορά αφού τόσο για τα αγόρια όσο και για τα κορίτσια ο δείκτης είναι a. Αντιθέτως, στην δεύτερη γραμμή και Τρίτη γραμμή έχουμε σημαντικές διαφορές αφού οι δείκτες είναι διαφορετικοί.

Πίνακας 2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	21,455 ^a	2	,000
Likelihood Ratio	21,769	2	,000
N of Valid Cases	478		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 42,74.

Ο έλεγχος χ^2 εμφανίζεται στην πρώτη γραμμή του πίνακα. Απορρίπτεται η μηδενική υπόθεση της ανεξαρτησίας των δύο μεταβλητών αφού $p < 0.001$

Ικανοποιούνται οι προϋποθέσεις για την εκτέλεση του ελέγχου χ^2 αφού το ποσοστό των κελιών με αναμενόμενη συχνότητα μικρότερη του 5 είναι 0% και η ελάχιστη αναμενόμενη συχνότητα είναι 42,74.

Πριν αξιολογηθεί το αποτέλεσμα του ελέγχου χ^2 πρέπει να διαπιστωθεί κατά πόσο ο έλεγχος είναι έγκυρος. Η εγκυρότητα του εξαρτάται από την ικανοποίηση δύο προϋποθέσεων:

α. το ποσοστό των κελιών με αναμενόμενη συχνότητα μικρότερη του 5, δεν πρέπει να ξεπερνά το 20%

β. Τη ελάχιστη αναμενόμενη συχνότητα πρέπει να είναι μεγαλύτερη από 1.

Η ικανοποίηση των παραπάνω προϋποθέσεων φαίνονται στο κάτω μέρος του πίνακα 2.

Το παραπάνω αποτέλεσμα μπορεί να γραφτεί στα πλαίσια μιας εργασίας ως εξής: Ελέγξαμε την υπόθεση της ανεξαρτησίας ανάμεσα στο φύλο και τους στόχους που θέτει ο μαθητής με την βοήθεια του ελέγχου χ^2 ανεξαρτησίας. Από τα αποτελέσματα του ελέγχου προκύπτει ότι υπάρχει σημαντική σχέση ανάμεσα στο φύλο και τους στόχους του μαθητή [$\chi^2 (2) = 21.455, p < 0.001$]. Από τις επιμέρους ανά ζεύγη συγκρίσεις σε κάθε γραμμή με τον έλεγχο z προέκυψε ότι τα κορίτσια προτιμούν σημαντικά περισσότερο από τα αγόρια την δημοτικότητα (36,3% έναντι 22%), ενώ αντιθέτως τα αγόρια προτιμούν σε σημαντικά υψηλότερο ποσοστό τα σπορ σε σχέση με τα κορίτσια (26,4% έναντι 12%).

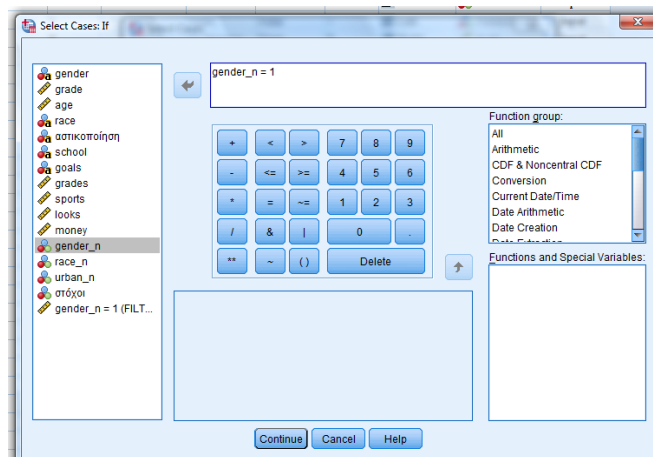
- b. Μπορεί να υποστηριχτεί ότι τα αγόρια δίνουν σημαντικά περισσότερο προτεραιότητα σε ορισμένους στόχους; Αν ναι, σε ποιους;

Προκειμένου να διερευνηθεί εάν στον πληθυσμό των αγοριών υπάρχει κάποιος στόχος που προτιμάται περισσότερο από τους υπόλοιπους θα εκτελεστεί ο έλεγχος χ^2 καλής προσαρμογής με το δείγμα των αγοριών με μηδενική υπόθεση την παρακάτω:

H_0 : grades = popular = sports = 33,3%

H_A : τουλάχιστον μία κατηγορία πρέπει να διαφέρει σε ποσοστό από τις υπόλοιπες

Για να εκτελεστεί ο έλεγχος στο SPSS πρέπει να επιλέξουμε από το αρχείο το δείγμα των αγοριών ως εξής: Data=> Select Cases=> If condition is satisfied. Στην συνθήκη γράφουμε `gender_n = 1`.



Στη συνέχεια από το μενού Analyze=> Nonparametric Tests=> Legacy Dialogs=> Chi-square εισάγουμε στο πλαίσιο Test Variable List την μεταβλητή στόχοι. Στο πλαίσιο Expected Values αφήνουμε επιλεγμένη την All categories equal, αφού αυτό δηλώνεται στη μηδενική υπόθεση που προαναφέραμε. Αν η μηδενική μας υπόθεση ήταν διαφορετική, π.χ.

H_0 : grades = 40% popular = 40% sports = 20%,

θα έπρεπε πατώντας Values να δώσουμε μία προς μία τις τιμές 40, 40, 20.

Μετά το «κλικ» στο OK του παραθύρου παίρνουμε στο output τους εξής πίνακες:

Πίνακας 3

	στόχοι		
	Observed N	Expected N	Residual
1 Grades	117	75,7	41,3
2 Popular	50	75,7	-25,7
3 Sports	60	75,7	-15,7
Total	227		

Αφού απορριφθεί η μηδενική υπόθεση (βλέπε τον παρακάτω πίνακα 4) τότε και μόνο τότε εξετάζουμε τις παρατηρούμενες συχνότητες προκειμένου να διατυπώσουμε την ισχύουσα εναλλακτική υπόθεση. Από τις παρατηρούμενες συχνότητες προκύπτει σημαντικά μεγαλύτερη προτίμηση στους βαθμούς (117/227) σε σχέση με τους άλλους δύο στόχους.

Πίνακας 4

Test Statistics	
	στόχοι
Chi-Square	34,529 ^a
df	2
Asymp. Sig.	,000

a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 75,7.

Απορρίπτεται η H_0 αφού $p < 0,001$. Άρα τουλάχιστον μία κατηγορία διαφέρει σημαντικά σε ποσοστό από τις άλλες.

Ικανοποιούνται οι προϋποθέσεις για την εκτέλεση του ελέγχου χ^2 αφού το ποσοστό των κελιών με αναμενόμενη συχνότητα μικρότερη του 5 είναι 0% και η ελάχιστη αναμενόμενη συχνότητα είναι 75,7.

Το παραπάνω αποτέλεσμα μπορεί να γραφτεί στα πλαίσια μιας εργασίας ως εξής: Ελέγξαμε την υπόθεση ότι τα αγόρια δεν έχουν προτίμηση σε κάποιον από τους τρεις στόχους με την βοήθεια του ελέγχου χ^2 καλής προσαρμογής. Από τα αποτελέσματα του ελέγχου προκύπτει ότι υπάρχει σημαντική προτίμηση σε τουλάχιστον έναν στόχο σε σχέση με τους υπόλοιπους [$\chi^2(2) = 34.53, p < 0.001$]. Τα αγόρια προτιμούν σημαντικά περισσότερο τους βαθμούς (grades: $117/227 = 51,5\%$) σε σχέση με τους δύο άλλους στόχους (popular = $50/227 = 22\%$, sports = $60/227 = 26,4\%$).

- c. Να διερευνηθεί η σχέση ανάμεσα στο βαθμό αστικοποίησης της περιοχής που ζει ο μαθητής και την προτεραιότητα στους στόχων του. Επιλέχθηκαν τυχαία δείγματα μαθητών από τρεις περιοχές με διαφορετικό βαθμό αστικοποίησης: «Αγροτική», «Ημιαστική», «Αστική» (μεταβλητή **αστικοποίηση**).
- d. Το ερώτημα b να επαναληφτεί για την περίπτωση των κοριτσιών. Έχει νόημα η αναζήτηση αυτή να γίνεται κατά φύλο ή θα μπορούσαμε αναλύσουμε και τα δυο δείγματα (αγόρια, κορίτσια) μαζί;

(Τα θέματα c και d δίνονται ως δραστηριότητα στις φοιτήτριες/τες).

Η ύλη του ελέγχου χ^2 βρίσκεται στο αρχείο διαφανειών «Επαγωγική Στατιστική_9.pptx» στο φάκελο «Παρουσιάσεις».

Στο video με τίτλο « χ^2 ανεξαρτησίας.swf» θα βρείτε την εκτέλεση του ελέγχου χ^2 ανεξαρτησίας στο SPSS.