

ΕΡΓΑΣΤΗΡΙΑ ΜΑΘΗΜΑΤΟΣ «ΜΕΘΟΔΟΙ ΕΠΙΚΟΙΝΩΝΙΑΚΗΣ ΕΡΕΥΝΑΣ ΙΙ»

Επαναληπτικές ασκήσεις 2024

Ανοίξτε το αρχείο **sleep.sav** που βρίσκεται αναρτημένο στον φάκελο “Εργαστήριο 4” του μενού Έγγραφα του Eclass. Το συγκεκριμένο αρχείο περιλαμβάνει δεδομένα που συγκεντρώθηκαν μέσω κλειστού ερωτηματολογίου από δείγμα 271 εργαζομένων σε πανεπιστήμιο της Μελβούρνης. Η έρευνα μελετά τη συχνότητα, τους προσδιοριστικούς παράγοντες και τις επιπτώσεις των προβλημάτων ύπνου των εργαζομένων. Η βάση δεδομένων περιλαμβάνει 43 μεταβλητές από τις οποίες θα χρησιμοποιήσουμε τις παρακάτω 10:

ID	Ταυτότητα ερωτηματολογίου (271 ερωτηματολόγια)
sex	Φύλο συμμετέχοντα 0=γυναίκα, 1=άντρας
age	Ηλικία συμμετέχοντα
marital	Οικογενειακή κατάσταση συμμετέχοντα 1=Άγαμος, 2=Παντρεμένος/συζύγος ως παντρεμένος, 3=Διαζευγμένος, 4=Χήρος
niteshft	Εργάζεστε βραδινές βάρδιες; 1=Ναι, 2=Όχι
drvsleep	Κοιμηθήκατε ποτέ καθώς οδηγούσατε; 1=Ναι, 2=Όχι
smoke	Καπνίζετε; 1=Ναι, 2=Όχι
smokenum	Αριθμός τσιγάρων την ημέρα
problem	Νιώθετε ότι έχετε πρόβλημα γενικά με τον ύπνο; 1=Ναι, 2=Όχι
stressmo	Πόσο αγχωμένος νιώσατε τον προηγούμενο μήνα; 1=καθόλου έως 10=υπερβολικά αγχωμένος

Από τις ασκήσεις 1 έως 5 καλείστε να επιλέξετε μόνο δύο και στη συνέχεια να κάνετε και την υποχρεωτική άσκηση 6 (δηλ. τρεις ασκήσεις συνολικά).

Άσκηση 1 (2 μονάδες)

i. Συμπληρώστε τον παρακάτω πίνακα με τις απόλυτες και τις σχετικές συχνότητες των εργαζομένων που ανήκουν σε καθεμία από τις τέσσερις κατηγορίες οικογενειακής κατάστασης.

ΟΙΚΟΓΕΝΕΙΑΚΗ ΚΑΤΑΣΤΑΣΗ	Απόλυτες συχνότητες	Σχετικές συχνότητες (έγκυρο %)
Άγαμος	54	19,9%
Έγγαμος / συζύγος ως έγγαμος	188	69,4%
Διαζευγμένος	21	7,7%
Χήρος	8	3%
ΣΥΝΟΛΟ (N)	271	100%

ii. Ποια οικογενειακή κατάσταση εμφανίζεται συχνότερα στο δείγμα; **Έγγαμος / συζύγος ως έγγαμος**

Άσκηση 2 (3 μονάδες)

Για καθεμία από τις μεταβλητές **niteshft** και **drvsleep**, χωριστά, δημιουργήστε έναν απλό πίνακα συχνοτήτων στο SPSS. Κατόπιν απαντήστε στις παρακάτω ερωτήσεις:

- i. Από όσους απάντησαν έγκυρα στη σχετική ερώτηση, τι ποσοστό δουλεύει βράδυ; **4,1%**
- ii. Από όσους απάντησαν έγκυρα στη σχετική ερώτηση, τι ποσοστό έχει κοιμηθεί οδηγώντας; **12,4%**
- iii. Πόσα άτομα δεν απάντησαν στην ερώτηση «*Εργάζεστε βραδινές βάρδιες*»; **2 άτομα**
- iv. Πόσα άτομα δεν απάντησαν στην ερώτηση «*Κοιμηθήκατε ποτέ ενώ οδηγούσατε*»; **22 άτομα**
- v. Φαίνεται να υπάρχει κάποιο λάθος στις τιμές της μεταβλητής **niteshft**; Εξηγήστε.

Ναι, υπάρχει λάθος. Η τιμή 3 είναι λάθος καθώς δεν αντιστοιχεί σε κάποιο κωδικό της μεταβλητής niteshft. Η μεταβλητή αυτή μπορεί να πάρει μόνο δύο τιμές (1=Ναι και 2=Όχι).

- vi. Σε ποιο ερωτηματολόγιο (ID) αντιστοιχεί αυτό το λάθος; **Στο ερωτηματολόγιο με ID=41.**

Άσκηση 3 (3 μονάδες)

Η μεταβλητή **age** είναι ίσων διαστημάτων (interval), δηλαδή τύπου scale. Δημιουργήστε μια νέα μεταβλητή ανακωδικοποιώντας τη μεταβλητή **age** ώστε να μετατραπεί σε ιεραρχική (ordinal). Ονομάστε τη νέα μεταβλητή **age_groups** [με Label: Ομαδοποιημένη Ηλικία] και ορίστε τις τιμές της ως εξής: **1: 18-30 ετών, 2: 31-40 ετών, 3: 41-50 ετών, 4: 51-60 ετών, 5: πάνω από 60 ετών**. Στο φύλλο εργασίας Variable View ορίστε το είδος της νέας μεταβλητής (στήλη Measure) και τους κωδικούς της (στήλη Values).

- i. Συμπληρώστε τον πίνακα:

ΗΛΙΚΙΑ	Απόλυτες συχνότητες	Σχετικές συχνότητες (έγκυρο %)
18-30 ετών	44	17,7%
31-40 ετών	52	21%
41-50 ετών	73	29,4%
51-60 ετών	59	23,8%
>60 ετών	20	8,1%
ΣΥΝΟΛΟ (N)	248	100%

- ii. Σε ποια ηλικιακή ομάδα ανήκει το μεγαλύτερο ποσοστό εργαζομένων; **41-50 ετών**
- iii. Σε ποια ηλικιακή ομάδα ανήκει το μικρότερο ποσοστό εργαζομένων; **>60 ετών**

Άσκηση 4 (3 μονάδες)

i. Επιλέξτε την υποομάδα των καπνιστών μέσω της εντολής Select Cases. Στη συνέχεια, κατασκευάστε το ιστόγραμμα της μεταβλητής **smokenum [αριθμός τσιγάρων την ημέρα]** για αυτήν την υποομάδα και εξετάστε εάν οι τιμές της προσεγγίζουν το σχήμα της κανονικής κατανομής. Περιγράψτε τυχόν αποκλίσεις ως προς τη συμμετρία/λοξότητα.

Η κατανομή των τιμών της μεταβλητής **smokenum** **δεν προσεγγίζει** το σχήμα της κανονικής κατανομής. Αποκλίνει σημαντικά από αυτήν. Συγκεκριμένα, παρουσιάζει θετική ασυμμετρία (ή ασυμμετρία προς τα δεξιά) καθώς επιμηκύνεται (έχει μεγάλη ουρά) προς το δεξιό άκρο του οριζόντιου άξονα που αντιστοιχεί στις μεγαλύτερες τιμές της κλίμακας μέτρησης.

ii. Υπάρχουν τιμές που θα θεωρούσατε ακραίες; Αν ναι, ποιες και γιατί; **Ναι, φαίνεται να υπάρχουν ακραίες τιμές που βρίσκονται εντός του διαστήματος 70-80 τσιγάρα την ημέρα.** Μια πιο προσεκτική διερεύνηση δείχνει ότι σε αυτό το διάστημα υπάρχει μια μόνο τιμή που ισούται με 78 τσιγάρα την ημέρα. Η τιμή αυτή είναι ακραία γιατί απέχει πάρα πολύ από όλες τις άλλες τιμές της κατανομής καθώς και από τον μέσο όρο.

Άσκηση 5 (3 μονάδες)

i. Υπολογίστε τον μέσο όρο (\bar{X}), την τυπική απόκλιση (s), τη διάμεσο, τη μέγιστη και την ελάχιστη τιμή, καθώς και το εύρος της μεταβλητής **stressmo** για άντρες και γυναίκες χωριστά. Συμπληρώστε τον παρακάτω πίνακα:

ΠΟΣΟ ΑΓΧΩΜΕΝΟΣ/Η ΝΙΩΣΑΤΕ ΤΟΝ ΠΡΟΗΓΟΥΜΕΝΟ ΜΗΝΑ (stressmo)
(1: καθόλου έως 10: υπερβολικά αγχωμένος)

	ΑΝΤΡΑΣ	ΓΥΝΑΙΚΑ
Μέσος όρος	5,29	6,02
Τυπική Απόκλιση	2,318	2,504
Διάμεσος	6	7
Μέγιστη τιμή	10	10
Ελάχιστη τιμή	1	1
Εύρος	9	9
N	121	150

ii. Ποιο φύλο φαίνεται να βιώνει μεγαλύτερα επίπεδα άγχους και από ποιους δείκτες προκύπτει αυτό;

Οι γυναίκες φαίνεται να βιώνουν μεγαλύτερα επίπεδα άγχους από τους άντρες, καθώς ο μέσος όρος της αντίστοιχης μεταβλητής είναι 6,02 για τις γυναίκες και 5,29 για τους άντρες. Επιπλέον, η τιμή της διαμέσου είναι μεγαλύτερη για τις γυναίκες συγκριτικά με τους άντρες (7 και 6, αντίστοιχα).

iii. Παρατηρώντας τους δείκτες διασποράς, φαίνεται ότι κάποια από τις δύο υποομάδες (άντρες ή γυναίκες) είναι πιο ανομοιογενής ως προς τα επίπεδα άγχους από την άλλη; Αιτιολογείστε την απάντησή σας.

Οι γυναίκες παρουσιάζουν ελαφρώς μεγαλύτερη ανομοιογένεια ως προς τα επίπεδα άγχους που βιώνουν συγκριτικά με τους άντρες. Αυτό προκύπτει από τη σύγκριση των δύο τυπικών αποκλίσεων ($s=2,318$ για τους άντρες και $s=2,504$ για τις γυναίκες). **Ωστόσο, αυτή η διαφορά είναι πάρα πολύ μικρή (0,186)** ενώ το εύρος είναι το ίδιο και για τις δύο υποομάδες. Άρα η μεγαλύτερη ανομοιογένεια που παρατηρείται μεταξύ των γυναικών **δεν είναι κάτι το αξιοσημείωτο** και πιθανώς να προέκυψε τυχαία στο συγκεκριμένο δείγμα συμμετεχόντων.

iv. Μπορούμε να γενικεύσουμε τα παραπάνω συμπεράσματα (ερ. ii και iii) στον ευρύτερο πληθυσμό από τον οποίο προέρχεται το δείγμα; Γιατί ή γιατί όχι;

Όχι, δεν μπορούμε να γενικεύσουμε τα παραπάνω συμπεράσματα καθώς δεν εφαρμόσαμε επαγωγικό έλεγχο στατιστικής σημαντικότητας.

Άσκηση 6 (ΥΠΟΧΡΕΩΤΙΚΗ) 4 μονάδες

i. Δημιουργήστε πίνακα διπλής εισόδου (crosstabulation) στο SPSS για να εξετάσετε τη συνάφεια μεταξύ των μεταβλητών **drvsleep** και **problem**. Πριν φτιάξετε τον πίνακα, ορίστε ποια είναι η εξαρτημένη και ποια η ανεξάρτητη μεταβλητή.

Εξαρτημένη μεταβλητή: **drvsleep**

Ανεξάρτητη μεταβλητή: **problem**

ii. Συμπληρώστε τον πίνακα διπλής εισόδου, έτσι ώστε τα ποσοστά στις στήλες να αθροίζουν στο 100%.

ΚΟΙΜΗΘΗΚΑΤΕ ΠΟΤΕ ΚΑΘΩΣ ΟΔΗΓΟΥΣΑΤΕ;	ΝΙΩΘΕΤΕ ΟΤΙ ΕΧΕΤΕ ΠΡΟΒΛΗΜΑ ΓΕΝΙΚΑ ΜΕ ΤΟΝ ΥΠΝΟ;					
	ΝΑΙ		ΟΧΙ		ΣΥΝΟΛΟ	
	Απόλυτη συχνότητα	Σχετική συχνότητα	Απόλυτη συχνότητα	Σχετική συχνότητα	Απόλυτη συχνότητα	Σχετική συχνότητα
ΝΑΙ	20	18,2%	11	8%	31	12,5%
ΟΧΙ	90	81,8%	127	92%	217	87,5%
ΣΥΝΟΛΟ	110	100%	138	100%	248	100%

iii. Από όσους νιώθουν ότι έχουν πρόβλημα ύπνου, τι ποσοστό έχει κοιμηθεί οδηγώντας; _____

iv. Από όσους δεν νιώθουν ότι έχουν πρόβλημα ύπνου, τι ποσοστό έχει κοιμηθεί οδηγώντας; _____

v. Συγκρίνετε τα ποσοστά όσων έχουν κοιμηθεί οδηγώντας μεταξύ αυτών που έχουν και αυτών που δεν έχουν πρόβλημα ύπνου. Παρατηρείτε κάποια σχέση μεταξύ προβλημάτων ύπνου και επικινδυνότητας της οδήγησης; Εξηγήστε (αιτιολογήστε) ολοκληρωμένα αυτό που συμπεραίνετε.

Φαίνεται να υπάρχει συνάφεια μεταξύ προβλημάτων ύπνου και επικινδυνότητας της οδήγησης στο δείγμα των 248 ατόμων. Το ποσοστό των συμμετεχόντων που έχουν κοιμηθεί στην οδήγηση είναι μεγαλύτερο μεταξύ αυτών που νιώθουν ότι έχουν προβλήματα ύπνου από ό,τι μεταξύ αυτών που δεν νιώθουν ότι έχουν προβλήματα ύπνου (18,2% και 8%, αντίστοιχα). Όσοι νιώθουν ότι έχουν προβλήματα με τον ύπνο έχουν υπερδιπλάσια πιθανότητα να κοιμηθούν την ώρα που οδηγούν συγκριτικά με εκείνους που δεν νιώθουν ότι έχουν τέτοια προβλήματα. Ωστόσο, για να γενικεύσουμε την παρατηρούμενη συνάφεια στον ευρύτερο πληθυσμό απαιτείται η διεξαγωγή του στατιστικού κριτηρίου χ^2 για ανεξάρτητα δείγματα (βλ. παρακάτω).

vi. Εφαρμόστε το στατιστικό κριτήριο χ^2 (chi-square test) για να εξετάσετε εάν η παρατηρούμενη σχέση μεταξύ των μεταβλητών **drvsleep** και **problem** είναι στατιστικά σημαντική ($\alpha=0,05$). Ελέγξτε, δηλαδή, εάν μπορεί να γενικευτεί στον ευρύτερο πληθυσμό από τον οποίο προέρχεται το δείγμα.

Διατυπώστε τη μηδενική και την εναλλακτική υπόθεση.

H_0 : Δεν υπάρχει στατιστικά σημαντική σχέση μεταξύ των μεταβλητών **drvsleep** και **problem**.

H_1 : Υπάρχει στατιστικά σημαντική σχέση μεταξύ των μεταβλητών **drvsleep** και **problem**.

Καταγράψτε την τιμή του κριτηρίου χ^2 και την τιμή της πιθανότητας p (asympt. sig.). Τι συμπεραίνετε;

$\chi^2=5,835$
 $p=0,016$

Συμπεραίνουμε ότι υπάρχει στατιστικά σημαντική σχέση μεταξύ των μεταβλητών **drvsleep** και **problem** καθώς $p=0,016 < 0,05$ (επίπεδο σημαντικότητας $\alpha=0,05$).

Με άλλα λόγια, η πιθανότητα τα αποτελέσματα που περιγράφονται στον πίνακα του υποερωτήματος ii να έχουν προκύψει τυχαία είναι μικρότερη του 5%. Επομένως, απορρίπτουμε τη μηδενική υπόθεση και δεχόμαστε την εναλλακτική.

ΒΑΣΙΚΑ ΒΗΜΑΤΑ ΓΙΑ ΤΗΝ ΕΠΙΛΥΣΗ ΤΩΝ ΑΣΚΗΣΕΩΝ

ΑΣΚΗΣΗ 1: Analyze→Descriptive Statistics→Frequencies. Μεταφέρω τη μεταβλητή **marital** [οικογενειακή κατάσταση συμμετέχοντα] στο πλαίσιο Variable(s) δεξιά. Πατώ OK.

ΑΣΚΗΣΗ 2: Analyze→Descriptive Statistics→Frequencies. Αφαιρώ τη μεταβλητή **marital** από το πλαίσιο Variable(s) μεταφέροντας την αριστερά. Μεταφέρω τις μεταβλητές **niteshft** και **drvsleep** στο πλαίσιο Variable(s) δεξιά. Πατώ OK.

Για τον εντοπισμό του λάθους, επιλέγω **Data→Sort cases**. Μεταφέρω τη μεταβλητή **niteshft** στο πλαίσιο Sort by δεξιά. Στην περιοχή Sort Order, επιλέγω Descending. Πατώ OK. Στο φύλλο εργασίας Data View, η λάθος τιμή της μεταβλητής **niteshft** εμφανίζεται πάνω πάνω (στην πρώτη σειρά). Βρίσκω το ID του αντίστοιχου ερωτηματολογίου.

ΑΣΚΗΣΗ 3: Transform→Recode into Different Variables. Μεταφέρω τη μεταβλητή **age** στο πλαίσιο Numeric Variable→Output Variable. Στην περιοχή Name ορίζω το όνομα της νέας μεταβλητής ως **age_groups**. Στην περιοχή Label την περιγράφω ως **Ομαδοποιημένη Ηλικία**. Πατώ **Change**.

Πατώ **Old and New Values**. Επιλέγω **Range** αριστερά και δίνω τιμές **'18 through 30'**. Στο πλαίσιο **New Value→Value** δεξιά δίνω **τιμή 1** και πατώ **Add**. Ξαναεπιλέγω **Range** αριστερά και δίνω τιμές **'31 through 40'**. Στο πλαίσιο **New Value→Value** δεξιά δίνω τιμή 2 και πατώ **Add**. Επιλέγω ξανά **Range** δίνοντας τιμές **'41 through 50'**, ενώ στο πλαίσιο **New Value→Value** δίνω τιμή 3 και πατώ **Add**. Επιλέγω και πάλι **Range** και δίνω τιμές **'51 through 60'**, ενώ στο πλαίσιο **New Value→Value** δίνω τιμή 4 και πατώ **Add**. Τέλος, επιλέγω **Range, Value through highest** και δίνω τιμή 61. Στο πλαίσιο **New Value→Value** δίνω τιμή 5 και πατώ **Add**. Πατώ **Continue** και **OK**.

Στο φύλλο εργασίας **Variable View** ορίζω τα χαρακτηριστικά της νέας μεταβλητής στις στήλες **Values** και **Measure**.

Δημιουργώ απλό πίνακα συχνότητας της νέας μεταβλητής **age_groups** μέσω των εντολών **Analyze→Descriptive Statistics→Frequencies**.

ΑΣΚΗΣΗ 4: Data →Select Cases → If condition is satisfied → If. Επιλέγω τη μεταβλητή **smoke**, την μεταφέρω στο πλαίσιο δεξιά και σημειώνω **smoke=1** (η τιμή 1 αντιστοιχεί στους καπνιστές). Πατώ **Continue** και **OK**.

Analyze→Descriptive Statistics→Frequencies. Μεταφέρω τη μεταβλητή **smokenum** στο πλαίσιο Variable(s) δεξιά και αναιρώ την επιλογή **Display frequency tables**. Πατώ το κουμπί **Charts** και επιλέγω **Histograms** και **Show normal curve on histogram**. Πατώ **Continue** και **OK**.

ΑΣΚΗΣΗ 5: Data →Select Cases → All cases → OK.

Analyze→Compare Means and Proportions→Means. Μεταφέρω τη μεταβλητή **stressmo** στο πλαίσιο **Dependent List** δεξιά και τη μεταβλητή **sex** στο πλαίσιο **Independent List**. Πατώ το κουμπί **Options** και στο νέο παράθυρο μεταφέρω τους δείκτες **Mean, Number of Cases, Standard Deviation, Median, Minimum, Maximum** και **Range** στο πλαίσιο **Cell Statistics** δεξιά. Πατώ **Continue** και **OK**.

ΑΣΚΗΣΗ 6: Analyze→Descriptive Statistics→Crosstabs. Μεταφέρω τη μεταβλητή **drvsleep** στο πλαίσιο **Row(s)** δεξιά και τη μεταβλητή **problem** στο πλαίσιο **Column(s)**. Πατάω το κουμπί **Cells** και στην περιοχή **Percentages** επιλέγω **Column**. Πατώ **Continue** → Πατώ το κουμπί **Statistics** και στο νέο παράθυρο επιλέγω **Chi-square**. Πατώ **Continue** και **OK**.