

Σεμινάριο SPSS

Δραστηριότητα εξάσκησης

Τρίτη, 2 Απριλίου 2024

Κωνσταντίνος Λαβίδας

Ενότητα Περιγραφική Στατιστική

Κατασκευή βάσης ερευνητικών δεδομένων, τα οποία συλλέχθηκαν μέσω εργαλείων παρατήρησης

Για την εισαγωγή ερευνητικών δεδομένων σε υπολογιστικό περιβάλλον ανάλυσης δεδομένων όπως το SPSS, πρέπει να κατασκευαστεί κατάλληλη **βάση δεδομένων**. Αυτή περιλαμβάνει το σύνολο των μεταβλητών (εμφανίζονται στις στήλες του προγράμματος εισαγωγής δεδομένων) λαμβάνοντας υπόψη τα επιμέρους χαρακτηριστικά αυτών.

Αναλυτικότερα, για την κατασκευή της βάσης δεδομένων κατασκευάζουμε τόσες μεταβλητές όσες τουλάχιστον είναι οι **ερωτήσεις ή τα στοιχεία παρατήρησης**. Ειδικότερα στο περιβάλλον του SPSS συνήθως ενημερώνουμε μόνο τα επόμενα χαρακτηριστικά των μεταβλητών:

- Τα **ονόματα των μεταβλητών** «Name» προτείνεται να είναι ο αριθμός της ερώτησης με πρόθεμα Q (π.χ. Q1, Q2 κ.λπ.). Είναι μάλλον εύκολη η αναγνώριση των μεταβλητών κατά την ανάλυση των δεδομένων. *Προσοχή, συνήθως το όνομα μιας μεταβλητής πρέπει να μας υποδεικνύει το περιεχόμενό της.*
- Ο **τύπος** «Type» όλων των μεταβλητών προτείνεται να είναι numeric. Εξαιρούνται οι ερωτήσεις ανοικτού τύπου τις οποίες θα δηλώσουμε αλφαριθμητικού τύπου «string». Για την ανάλυση των δεδομένων από τις ερωτήσεις ανοικτού τύπου ενδεικτικά θα πρέπει να κατασκευαστεί μια ποιοτική μεταβλητή με τιμές που θα προκύψουν από την ανάλυση των απαντήσεων των ερωτώμενων. Προτείνονται ως δυνατές απαντήσεις οι ανόμοιες σημασιολογικά. *Προσοχή, σε πολλές περιπτώσεις οι ερωτήσεις ανοικτού τύπου είναι ερωτήσεις πολλαπλών απαντήσεων που αντιστοιχούν σε δίτιμες μεταβλητές.*
- **Δεκαδικά** «Decimals», δηλώνουμε δεκαδικά, αν η μεταβλητή είναι τέτοια στην οποία είναι απαραίτητα τα δεκαδικά ψηφία (επιθυμούμε μεγαλύτερο επίπεδο ακρίβειας). Για παράδειγμα το ύψος ή μια μεταβλητή που προκύπτει από το γραμμικό συνδυασμό (μετασχηματισμό) άλλων ποσοτικών μεταβλητών όπως ο μέσος όρος βαθμολογίας.
- Στην συνοπτική **περιγραφή** της μεταβλητής «Label» μπορείτε να αντιγράψετε την ερώτηση (copy και μετά paste), ωστόσο, πρέπει να αναδομήσετε το κείμενο κατάλληλα έτσι ώστε να είναι κατάφαση.
- **Λεκτικές περιγραφές** «Values» είναι οι αναλυτικές περιγραφές των τιμών των ποιοτικών μεταβλητών (τις οποίες έχουμε ποσοτικοποιήσει): ονομαστικές ή κατηγορικές «nominal» και διάταξης ή ιεραρχικές «ordinal», οι οποίες αντιστοιχούν στους αριθμούς που «καταχρηστικά» εισάγουμε στην μεταβλητή. Για τις ποσοτικές μεταβλητές (scale στο περιβάλλον του SPSS) αναλογικές ή λόγου «ratio» και ισοδιαστημικές «interval» δεν ασχολούμαστε με το χαρακτηριστικό αυτό.

- Οι **χαμένες τιμές «Missing values»** αφορούν αφενός τις περιπτώσεις εκείνες που ο ερωτώμενος δεν έχει απαντήσει (ή δεν ήταν δυνατή η όποια καταγραφή των δεδομένων) και αφετέρου τις περιπτώσεις που θέλουμε να τις αποκλείσουμε από την επικείμενη ανάλυση (πιθανά λάθη ή πολύ ακραίες τιμές). Κατά την αρχική εισαγωγή των μεταβλητών δεν ασχολούμαστε με την επιλογή αυτή.
- Η **κλίμακα ή τύπος μέτρησης «measure»** της μεταβλητής μπορεί να είναι: κατηγορική «nominal», διάταξης «ordinal» και ποσοτική «scale» (οι δύο αυτές περιπτώσεις αντιστοιχούν σε ποιοτικές μεταβλητές). Η «scale» αφορά στις ποσοτικές μεταβλητές είτε είναι αναλογική «ratio» είτε ισοδιαστημική «interval».

Δραστηριότητα 1^η

A) Κατασκευής βάσης ερευνητικών δεδομένων

Κάθε στοιχείο παρατήρησης αποτελεί και μεταβλητή στην βάση δεδομένων που θα φτιάχνουμε.

Κάποιες οδηγίες ενδεικτικά:

- Η **πρώτη μεταβλητή που φτιάχνουμε πάντα** είναι ο αύξων αριθμός της καρτέλας παρατήρησης (π.χ ερωτηματολόγιο) και έχει όνομα «α.α.». Η μεταβλητή αυτή είναι ποσοτική και δεν χρησιμοποιείται στην ανάλυση των δεδομένων. Επίσης είναι πολύ χρήσιμη, αφού πολλές φορές χρειάζεται να επιβεβαιώσουμε την εισαγωγή των δεδομένων μας ελέγχοντας το αντίστοιχο ερωτηματολόγιο. *Προσοχή, κατά την εισαγωγή των δεδομένων είναι απαραίτητο να σημειώνουμε στο εκάστοτε ερωτηματολόγιο τον αύξοντα αριθμό του και να εισάγουμε αυτόν στο αντίστοιχο πεδίο.*
- Για τις δίτιμες μεταβλητές χρησιμοποιείτε τους αριθμούς 0 και 1.
- Οι ερωτήσεις πολλαπλών απαντήσεων γίνονται τόσες μεταβλητές (δίτιμες: 0 δεν επιλέχθηκε και 1 επιλέχθηκε) όσες είναι οι δυνατές απαντήσεις. Οι ανοικτού τύπου ερωτήσεις μετά την αρχική ανάλυση που οδηγεί στον προσδιορισμό όλων των δυνατών απαντήσεων, λογίζεται ως ερώτηση πολλαπλών απαντήσεων.
- Οι αριθμοί που αντιστοιχούν στις τιμές (λέξεις) των μεταβλητών διάταξης καλό θα είναι να συνάδουν με την διάταξη των τιμών της μεταβλητής. Έτσι θα είναι πιο εύκολη η ερμηνεία κατά την ανάλυση. Για παράδειγμα μικρότερες τιμές για το μικρότερο βαθμό.
- Σε κάποιες ερωτήσεις περιέχονται περισσότερα του ενός στοιχεία παρατήρησης. Σε αυτή την περίπτωση θα πρέπει να κατασκευάζονται ισάριθμες μεταβλητές.

1.A.1) Να κατασκευαστεί βάση δεδομένων (μεταβλητές και χαρακτηριστικά αυτών) που συλλέχθηκαν με το επόμενο ερωτηματολόγιο (θα μπορούσε να είναι μια οποιαδήποτε καρτέλα παρατήρησης). Οι επόμενες ερωτήσεις είναι κάποιες μόνο από τις ερωτήσεις του ερωτηματολογίου.

Ερωτηματολόγιο ...

Κάποιες από τις ερωτήσεις ενός ερωτηματολογίου:

1. Ποιο είναι το έτος γέννησης σας;
2. Ποια ήταν η κατεύθυνση των σπουδών σας στο Λύκειο; (1. Θετική, 2. Τεχνολογική, 3. Θεωρητική)
3. Ποιο είναι το αντικείμενο των βασικών σας (προπτυχιακό επίπεδο) σπουδών;
4. Ποιες από τις επόμενες κατηγορίες υπολογιστικών εργαλείων (λογισμικών) χρησιμοποιείτε;

	...Καθόλου	...Λίγο	...Αρκετά	...Πολύ	...Πάρα Πολύ
Προγράμματα επεξεργασίας κειμένου (όπως το word)					
Προγράμματα στατιστικής επεξεργασίας δεδομένων (όπως το spss)					

5. Ποια από τα επόμενα έχετε διδαχθεί; (μπορείτε να επιλέξετε περισσότερα από ένα)
 - Περιγραφική Στατιστική
 - Επαγωγική Στατιστική
 - Πολυμεταβλητή Ανάλυση
6. Έχετε ποτέ εμπλακεί σε ερευνητική διαδικασία και αν ναι, ποιο ήταν το αντικείμενο της έρευνας;

Δραστηριότητα 2^η

Α) Ανάλυση των δεδομένων που συλλέχθηκαν με το ερωτηματολόγιο

2.A.1) Ανοίξτε το αρχείο «responses MEE.PP test.sav»

Στο αρχείο αυτό υπάρχουν δεδομένα των ερωτηματολογίων που έχουν καταχωρήσει Μεταπτυχιακοί Φοιτητές (ΜΦ) του ακαδημαϊκού έτους 20..-20... Τα δεδομένα αυτά που βρίσκονται σε αυτό το αρχείο υποθέτουμε ότι τα έχετε εσείς καταχωρήσει, αφού πρώτα κατασκευάσατε τη βάση δεδομένων (ένα μέρος έγινε από εσάς στη προηγούμενη δραστηριότητα). Προσοχή, τα στοιχεία που θα δείτε στο εν λόγω αρχείο δεν ανταποκρίνονται στην πραγματικότητα.

Γενικά τα ερευνητικά ερωτήματα θα μπορούσαν να έχουν την παρακάτω μορφή:

1. **Παρουσίαση του φαινομένου που μας ενδιαφέρει να μελετήσουμε:** Ποιες είναι οι προσδοκίες των μεταπτυχιακών φοιτητών σχετικά με το πρόγραμμα;
2. **Εξήγηση της διακύμανσης του φαινομένου που μας ενδιαφέρει να μελετήσουμε (αναζήτηση σχέσεων μεταξύ μεταβλητών):** Διαφοροποιούνται (σχετίζονται) και πως οι προσδοκίες ανάλογα με τα α, β, γ, δ και ε χαρακτηριστικά των ΜΦ;

Β) Ανάλυση

Πριν ξεκινήσουμε καλό θα είναι να δημιουργήσουμε αλλά και να ελέγξουμε τις μεταβλητές μας. Ας επικεντρωθούμε ενδεικτικά στις επόμενες:

2.B.1) Μας ενδιαφέρει να μάθουμε ποια είναι η ηλικία των ΜΦ. Δημιουργήστε μια μεταβλητή **ηλικία** που θα προκύπτει από την μεταβλητή... Πως;

2.B.2) Επίσης ελέγξτε την κατανομή της μεταβλητής ηλικία μήπως κάποιες τιμές είναι **ακραίες (extreme values)** και πρέπει να εξαιρεθούν από την ανάλυση.

Ο προσδιορισμός των **ακραίων τιμών** είναι πολύ σημαντικός αφού η μέση τιμή που είναι ένας βασικός εκτιμητής των τιμών των ποσοτικών μεταβλητών επηρεάζεται έντονα από τις ακραίες τιμές. Ενδεικτικό παράδειγμα, είναι η λανθασμένη χρήση της μέσης τιμής για τον προσδιορισμό του μέσου εισοδήματος των πολιτών μιας χώρας. Στην περίπτωση δηλαδή, που τα υψηλά εισοδήματα είναι έντονα (θετική ασυμμετρία) «τραβούν τον μέσο όρο ψηλά» και παρουσιάζουν έτσι μια λανθασμένη εικόνα του εισοδήματος των πολιτών μιας χώρας.

Υπό αυτή την έννοια, οι ακραίες τιμές μπορεί να διαστρεβλώσουν τα αποτελέσματα της έρευνας και πολλές φορές πρέπει να εξαιρούνται από την ανάλυση. Στην περίπτωση που κάποιες ακραίες τιμές δεν μπορούν να εξαιρεθούν από την ανάλυση συμπληρωματικά θα μπορούσε να χρησιμοποιηθεί η διάμεσος ως εκτιμητής των τιμών της μεταβλητής.

Ο προσδιορισμός των ακραίων τιμών, θα πρέπει να γίνεται κυρίως σε όλες τις ποσοτικές μεταβλητές (τύπου scale). Τιμές που βρίσκονται εντός του διαστήματος: $[Q1-1,5IQR, Q3+1,5IQR]$ είναι συνήθως αποδεκτές. Όπου: Q1, Q3, το πρώτο και το τρίτο τεταρτημόριο και IQR το ενδοτεταρτημοριακό εύρος της κατανομής των τιμών.

Για εντοπίσουμε τις ακραίες τιμές:

1^{ος} τρόπος, με την βοήθεια του Θηκογράμματος (boxplot). (Εντολές: **Graphs- Legacy Dialogs- boxplot – Simple -summaries of separate variables**)

2^{ος} τρόπος, μέσω της εντολής διερεύνησης της κατανομής «ποσοτικών» μεταβλητών, διαδικασία explore. (Εντολές: **Analyze- Descriptive Statistics- Explore- Statistics... επιλέγω outliers**)

3^{ος} τρόπος, υπολογίζουμε τις **τυπικές τιμές (z-values)** των μεταβλητών. Δημιουργώ καινούργιες μεταβλητές με τις τυπικές τιμές των μεταβλητών. Οι τυπικές τιμές συνήθως εκτός του ορίου (-3, 3) υποδεικνύουν ακραίες τιμές που μπορεί να δημιουργήσουν πρόβλημα στην εκτίμηση των αποτελεσμάτων. (Εντολές: **Analyze-**

Descriptive Statistics- Descriptives επιλέγω κάτω και αριστερά “Save standardized values as variables”

2.B.3) Υπάρχουν ακραίες τιμές ηλικίας και τι προτείνετε να κάνουμε; Έχουμε μήπως στη διάθεσή μας τα ερωτηματολόγια; Και αν τα έχουμε πως θα λύσουμε το πρόβλημα που παρουσιάζεται;

Γ) Κατασκευή μεταβλητής «σύνοψης» δεδομένων

2.Γ.1) Πρέπει επίσης να κατασκευάσουμε μια ποσοτική μεταβλητή που να αναδεικνύει σύντομα το βαθμό χρήσης των βασικών υπολογιστικών εργαλείων (**ομάδα ερωτήσεων**) (αντίστοιχη μεταβλητή q12.1, q12.2, q12.3) που είναι απαραίτητα για τις σπουδές των ΜΦ στο ΜΠ.

Πολλές φορές ο καλύτερος τρόπος να ομαδοποιήσεις τις ερωτήσεις (με στόχο να τις αντικαταστήσεις με μια καινούργια μεταβλητή) είναι να ελέγξεις αν έχουν σημασιολογικές ομοιότητες. Προσοχή, για να αντιπροσωπεύεται ικανοποιητικά η διακύμανση της ομάδας των δηλώσεων από τη καινούργια μεταβλητή, θα πρέπει τουλάχιστον να διέπεται η ομάδα των δηλώσεων από ικανοποιητική εσωτερική συνοχή (**internal consistency**). Ο δείκτης **Cronbach’s Alpha** μας μετράει την εσωτερική συνοχή μιας ομάδας δηλώσεων και θεωρείται ικανοποιητικός αν είναι τουλάχιστον 0,7. **Ο δείκτης αυτός αυξάνει όσο αυξάνει το πλήθος των δηλώσεων αλλά και η μέση συσχέτιση τους.**

(Εντολές: Analyze-scale-reliability analysis- model:Alpha).

2.Γ.2) Υπολογίστε τον συντελεστή **Cronbach’s Alpha**. Με ποιο τρόπο θα κατασκευάσω αυτή τη μεταβλητή;

2.Γ.3) Να υπολογίσεις τα περιγραφικά στατιστικά (min, max, mean και standard deviation) της μεταβλητής που έφτιαξες στις διάφορες ομάδες: α) άντρες, γυναίκες και β) ομάδες κατεύθυνσης των σπουδών στο Λύκειο. Να φτιάξετε δύο αντίστοιχους και να σχολιάσετε τα αποτελέσματα.

Κάποιες επιπλέον σκέψεις

Στην μεταβλητή «q18», καταγράψαμε την αρχική Αξιολόγηση - Εκτίμηση του ΜΠ σύμφωνα με την όποια αρχική ενημέρωση είχε ο φοιτητής. Η έκφραση αυτή έχει αποτυπωθεί με μια κλίμακα πέντε σημείων (0, 1, 2, 3, 4). Θέλω να ελέγξω αν η αρχική εκτίμηση διαφοροποιείται από την τελική εκτίμηση. Υπενθυμίζω ότι η τελική

εκτίμηση (μεταβλητή «q19») έχει καταγραφεί με μια κλίμακα 4 σημείων (0, 1, 2, 3).

Πως θα γίνει αυτός ο έλεγχος;

2.Γ.5) Φτιάξτε μια (ποιοτική μεταβλητή) δύο σημείων η οποία να συνοψίζει την εκτίμηση μου για το ΜΠ, ως εξής: αν η αρχική αξιολόγηση είναι μικρότερη ή ίση από την τελική αξιολόγηση να θεωρείται ικανοποιητικό το Πρόγραμμα Σπουδών (ΠΣ) διαφορετικά μη Ικανοποιητικό ΠΣ. Να την ονομάσετε Brief_estimate (Συνοπτική Εκτίμηση)

2.Γ.6) Ελέγξτε και παρουσιάστε πως διαφοροποιείται η συνοπτική εκτίμηση των ΜΦ για το πρόγραμμα των μεταπτυχιακών τους σπουδών ανάλογα με το φύλο τους;

Δ) Διαχείριση ερωτήσεων πολλαπλών απαντήσεων (multiple responses)

Η ερώτηση (αντίστοιχη μεταβλητή q13.1, q13.2, q13.3), σχετικά με το αν έχουν διδαχθεί οι ΜΦ περιγραφική, επαγωγική και πολυμεταβλητή ανάλυση, είναι πολλαπλών απαντήσεων.

2.Δ.1) Τι ποσοστό των ερωτώμενων έχει διδαχθεί στο παρελθόν καθένα από τα παραπάνω;

Στην περίπτωση που επιθυμούμε να μάθουμε τι ποσοστό των ερωτώμενων έχει διδαχθεί στο παρελθόν τα παραπάνω, τότε πίνακες συχνοτήτων για κάθε μια μεταβλητή ξεχωριστά.

2.Δ.2) Τι ποσοστό από το καθένα από τα παραπάνω έχουν διδαχθεί **όσοι απάντησαν στην ερώτηση**. (υπάρχουν και ΜΦ που δεν απάντησαν στην ερώτηση αυτή)

Για την συνολική εκτίμηση των απαντήσεων μιας ερώτησης πολλαπλών απαντήσεων κατασκευάζουμε μια καινούργια (εικονική) μεταβλητή πολλαπλών απαντήσεων: (Εντολές: **Analyze – tables – multiple response set- dichotomies – counted value: 1-set name – add- ok**). Η μεταβλητή αυτή είναι μόνιμη (αρκεί να σώσεις το αρχείο) αλλά δεν φαίνεται στη βάση δεδομένων (είναι βοηθητική). Η όποια διαχείρισή της γίνεται με το **analyze – tables – custom tables**. Η μεταβλητή πολλαπλών απαντήσεων φαίνεται αριστερά στις μεταβλητές και στο τέλος. (Εντολές: **analyze – tables – custom tables.... τοποθετούμε τη μεταβλητή πολλαπλών απαντήσεων στη γραμμή και επιλέγουμε:**

Column N % ή Column Response % (Base: Count), Βασισμένο στον αριθμό των ερωτώμενων που απάντησε (επέλεξε) τουλάχιστον μια από τις δυνατές περιπτώσεις (απαντήσεις). Αντιστοιχεί στο ποσοστό των απαντώντων που έχουν επιλέξει την κάθε απάντηση (συχνότητα επιλογής της απάντησης προς το σύνολο των απαντώντων)

Column Total N %, Βασισμένο στο σύνολο των συμμετεχόντων. Αντιστοιχεί στο ποσοστό των συμμετεχόντων που έχουν επιλέξει την κάθε απάντηση.

Column Responses %, Βασισμένο στο σύνολο των απαντήσεων. Αντιστοιχεί στο ποσοστό των απαντήσεων που έχουν επιλέξει την κάθε απάντηση (συχνότητα επιλογής της απάντησης προς το σύνολο των απαντήσεων)

2.Δ.3) Ποια είναι η ενότητα της στατιστικής που έχουν παρακολουθήσει περισσότερο οι ΜΦ.

(Εντολές: **analyze – tables – custom tables.... τοποθετούμε τη μεταβλητή πολλαπλών απαντήσεων στη γραμμή και επιλέγουμε «Column Responses %»**).

2.Δ.4) Να μεταφέρετε τον πίνακα (copy – paste) του προηγούμενου ερωτήματος στο excel και να φτιάξετε ένα οριζόντιο ραβδόγραμμα.

Δραστηριότητες για το σπίτι

1^η Να κατασκευάσετε μια βάση για την εισαγωγή δεδομένων που θα συλλέξουμε με τις επόμενες ερωτήσεις. Το αρχείο, τύπου sav που θα φτιάξεις, να το ανεβάσεις στην ομάδα εργασιών αυτής της εβδομάδας.

1. Είσαι: Αγόρι, Κορίτσι, Άλλο
2. Πόσων ετών είσαι;
3. Ποιο είναι το οικογενειακό σας εισόδημα (σε χιλιάδες €)
4. Η θρησκεία μου είναι: Χριστιανός/ή, Μουσουλμάνος/α, Ινδουιστής/ια, Εβραίος/α, Καμία, Άλλη

5. Κατά τη διάρκεια αυτής της σχολικής χρονιάς	Ποτέ	Σπάνια	Μερικές φορές	Συχνά	Σχεδόν πάντα
1. Παίζω συχνά ψηφιακά παιχνίδια					
2. Συμβουλευόμαι τους φίλους για τα καινούργια ψηφιακά παιχνίδια					

6. Ποιο ήταν το τελευταίο ψηφιακό παιχνίδι που έπαιξες; Γράψε τον τίτλο του.
7. Σε ποιες συσκευές παίζεις συνήθως ψηφιακά παιχνίδια; (διάλεξε όλες όσες χρησιμοποιείς): Κονσόλα, Υπολογιστή PC/Mac, Κινητό τηλέφωνο/Tablet, Φορητή Κονσόλα, άλλο

2^η Στο αρχείο δεδομένων “viomatikes.sav” παρουσιάζονται οι απαντήσεις των εκπαιδευτικών για την αποτελεσματικότητα των βιωματικών τεχνικών στην μάθηση (μεταβλητές q1, q2, q3 και q4).

- I. Να ελέγξετε την αξιοπιστία αυτής της κλίμακας και να κατασκευάσετε μια μεταβλητή από αυτές (Q_apotelesmatikohta), που να συνοψίζει τις απαντήσεις των εκπαιδευτικών σε αυτές.
- II. Να υπολογίστε τα περιγραφικά στατιστικά της καινούργιας μεταβλητής (Q_apotelesmatikohta) για κάθε ομάδα εκπαιδευτικών σύμφωνα με τα δημογραφικά χαρακτηριστικά των εκπαιδευτικών (φύλο, έτη υπηρεσίας και βαθμίδα εκπαίδευσης).
- III. Να φτιάξετε αντίστοιχους πίνακες και να σχολιάσετε τα αποτελέσματα. Γράψτε μια αναφορά για τα παραπάνω και ανεβάστε το αρχείο word στην ομάδα εργασιών αυτής της εβδομάδας. Στο αρχείο αυτό μετά τον σχολιασμό να υπάρχουν και οι αντίστοιχοι συνοδευτικοί πίνακες.