



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

ΦΙΛΟΣΟΦΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΦΙΛΟΣΟΦΙΑΣ-ΠΑΙΔΑΓΩΓΙΚΗΣ-ΨΥΧΟΛΟΓΙΑΣ

ΤΟΜΕΑΣ ΠΑΙΔΑΓΩΓΙΚΗΣ

ΕΡΓΑΣΙΑ ΣΤΟ ΠΛΑΙΣΙΟ ΤΟΥ ΜΑΘΗΜΑΤΟΣ ΜΕΤΡΗΣΗ ΚΑΙ ΒΑΘΜΟΛΟΓΗΣΗ

Καθηγητής Μαθήματος: Αθανάσιος Βέρδης

Ονοματεπώνυμο φοιτήτριας: Καλαμαρά Σταυρούλα

Αριθμός Μητρώου: 1562201400070

Επί πτυχίω

Χειμερινό εξάμηνο του ακαδημαϊκού έτους 2018-2019

Πανεπιστημιούπολη-Αθήνα

Στατιστική με λίγα λόγια από τη Sarah Boslaugh, Paul Andrew Watters

Κεφάλαιο 1. Βασικές έννοιες μέτρησης

Προτού μπορέσετε να χρησιμοποιήσετε στατιστικά στοιχεία για την ανάλυση ενός προβλήματος, πρέπει να μετατρέψετε τα βασικά υλικά του προβλήματος σε δεδομένα. Δηλαδή, πρέπει να καθιερώσετε ή να υιοθετήσετε ένα σύστημα εκχώρησης αξιών, πιο συχνά αριθμών, στα αντικείμενα ή τις έννοιες που είναι κεντρικές στο πρόβλημα που μελετάται. Αυτό δεν είναι μια εσωτερική διαδικασία, αλλά κάτι που κάνετε κάθε μέρα. Για παράδειγμα, όταν αγοράζετε κάτι στο κατάστημα, η τιμή που πληρώνετε είναι μια μέτρηση: αντιστοιχεί έναν αριθμό στο ποσό του νομίσματος που έχετε ανταλλάξει για τα παραληφθέντα αγαθά. Ομοίως, όταν αναβαίνετε στη ζυγαριά του μπάνιου το πρωί, ο αριθμός που βλέπετε είναι μια μέτρηση του σωματικού σας βάρους. Ανάλογα με τον τόπο κατοικίας σας, αυτός ο αριθμός μπορεί να εκφράζεται είτε σε λίρες είτε σε χιλιόγραμμα, αλλά η αρχή της ανάθεσης ενός αριθμού σε μια φυσική ποσότητα (βάρος) ισχύει και στις δύο περιπτώσεις.

Δεν είναι δυνατόν όλα τα δεδομένα να είναι αριθμητικά. Για παράδειγμα, οι κατηγορίες ανδρών και γυναικών που χρησιμοποιούνται συνήθως στην επιστήμη και στην καθημερινή ζωή για να ταξινομήσουν τους ανθρώπους δεν είναι εγγενώς αριθμητικές. Ομοίως, μιλάμε συχνά για τα χρώματα αντικειμένων ευρείας κλάσης, όπως "κόκκινο" ή "μπλε": οι κατηγορίες των οποίων αντιπροσωπεύουν μια μεγάλη απλοποίηση από την άπειρη ποικιλία χρωμάτων που υπάρχουν στον κόσμο. Αυτή είναι μια τόσο συνηθισμένη πρακτική που δύσκολα δίνουμε μια δεύτερη σκέψη.

Πόσο συγκεκριμένοι θέλουμε να είμαστε με αυτές τις κατηγορίες (για παράδειγμα, το "γρανίτης" είναι ένα ξεχωριστό χρώμα από το "κόκκινο"; Πρέπει να μεταβιβάσουν τα ξεχωριστά άτομα σε διαφορετική κατηγορία, πράγμα που εξαρτάται από τον σκοπό: διανοητικές κατηγορίες για το χρώμα από το μέσο άτομο, για παράδειγμα. Ομοίως, το επίπεδο λεπτομέρειας που χρησιμοποιείται στην ταξινόμηση για μια μελέτη εξαρτάται από το σκοπό της μελέτης και τη σημασία της σύλληψης των αποχρώσεων κάθε μεταβλητής.

Μέτρηση

Η μέτρηση είναι η διαδικασία συστηματικής εκχώρησης αριθμών σε αντικείμενα και τις ιδιότητές τους, για να διευκολυνθεί η χρήση των μαθηματικών στη μελέτη και περιγραφή αντικειμένων και των σχέσεών τους. Ορισμένοι τύποι μέτρησης είναι αρκετά συγκεκριμένοι: για παράδειγμα, η μέτρηση του βάρους ενός ατόμου σε λίβρες ή κιλά, ή το ύψος σε πόδια και ίντσες ή σε μέτρα. Σημειώστε ότι το συγκεκριμένο σύστημα μέτρησης που χρησιμοποιείται δεν είναι τόσο σημαντικό

όσο ένα συνεκτικό σύνολο κανόνων: για παράδειγμα, μπορούμε εύκολα να μετατρέψουμε τη μέτρηση από χιλιόγραμμα σε λίβρες. Παρόλο που οποιοδήποτε σύστημα μονάδων μπορεί να φαίνεται αυθαίρετο (προσπαθήστε να υπερασπιστείτε τα πόδια και τις ίντσες σε κάποιον που μεγάλωσε με το μετρικό σύστημα!), Όσο το σύστημα έχει συνεπή σχέση με την ιδιότητα που μετράται, μπορούμε να χρησιμοποιήσουμε τα αποτελέσματα σε υπολογισμούς.

Η μέτρηση δεν περιορίζεται σε φυσικές ιδιότητες όπως το ύψος και το βάρος. Οι δοκιμασίες για τη μέτρηση της αφαίρεσης όπως η νοημοσύνη και η σχολική ικανότητα χρησιμοποιούνται συνήθως στην εκπαίδευση και την ψυχολογία, για παράδειγμα: το πεδίο της ψυχομετρίας ασχολείται σε μεγάλο βαθμό με την ανάπτυξη και την τελειοποίηση των μεθόδων δοκιμής ακριβώς τέτοιων αφηρημένων ιδιοτήτων. Η διαπίστωση ότι μια συγκεκριμένη μέτρηση έχει νόημα είναι πιο δύσκολη όταν δεν μπορεί να παρατηρηθεί άμεσα: ενώ μπορείτε να δοκιμάσετε την ακρίβεια μιας κλίμακας συγκρίνοντας τα αποτελέσματα με τα αποτελέσματα μιας άλλης κλίμακας που είναι γνωστή ότι είναι ακριβής, δεν υπάρχει απλός τρόπος να γνωρίζετε εάν μια δοκιμασία της νοημοσύνης είναι ακριβής επειδή δεν υπάρχει ένας κοινά αποδεκτός τρόπος μέτρησης της αφαίρεσης νοημοσύνης. Για να το θέσουμε με άλλο τρόπο, δεν γνωρίζουμε ποια είναι η πραγματική ευφυΐα κάποιου επειδή δεν υπάρχει κάποιος τρόπος μέτρησης, και μάλιστα δεν μπορούμε να είμαστε σίγουροι τι είναι "η νοημοσύνη", μια κατάσταση εντελώς διαφορετική από αυτή της μέτρησης του ύψους ή του βάρους ενός ατόμου. Αυτά τα θέματα είναι ιδιαίτερα σημαντικά για τις κοινωνικές επιστήμες και την εκπαίδευση, όπου μεγάλη έρευνα επικεντρώνεται σε τέτοιες αφηρημένες έννοιες.

Επίπεδα μέτρησης

Οι στατιστικοί διακρίνουν συνήθως τέσσερις τύπους ή επίπεδα μέτρησης. οι ίδιοι όροι μπορούν επίσης να χρησιμοποιηθούν για να αναφερθούν σε δεδομένα που μετριοούνται σε κάθε επίπεδο. Τα επίπεδα μέτρησης διαφέρουν τόσο ως προς την έννοια των αριθμών όσο και ως προς τους τύπους στατιστικών που είναι κατάλληλα για την ανάλυσή τους.

Ονομαστικά δεδομένα

Με τα ονομαστικά δεδομένα, όπως υπονοεί το όνομα, οι αριθμοί λειτουργούν ως όνομα ή ετικέτα και δεν έχουν αριθμητική σημασία. Για παράδειγμα, μπορείτε να δημιουργήσετε μια μεταβλητή για το φύλο, η οποία παίρνει την τιμή 1 εάν το άτομο είναι αρσενικό και 0 αν το άτομο είναι θηλυκό. Οι αριθμοί 0 και 1 δεν έχουν αριθμητική σημασία αλλά λειτουργούν απλώς ως ετικέτες με τον ίδιο τρόπο που μπορείτε να καταγράψετε τις τιμές ως "M" ή "F." Υπάρχουν δύο κύριοι λόγοι για να επιλέξετε αριθμητικές και όχι τιμές κειμένου σε ονομαστικά δεδομένα κώδικα: είναι πιο εύκολη η επεξεργασία από ορισμένα συστήματα υπολογιστών ως αριθμοί και η

χρήση αριθμών παρακάμπτει ορισμένα ζητήματα στην εισαγωγή δεδομένων, όπως η σύγκρουση μεταξύ κεφαλαίων (σε έναν υπολογιστή, το "M" είναι διαφορετική από "m"). Τα ονομαστικά δεδομένα δεν περιορίζονται σε δύο κατηγορίες: για παράδειγμα, αν μελετούσατε τη σχέση μεταξύ ετών εμπειρίας και μισθού σε παίκτες του μπέιζμπολ, μπορείτε να ταξινομήσετε τους παίκτες σύμφωνα με την αρχική τους θέση χρησιμοποιώντας το παραδοσιακό σύστημα.

Αν δεν μπορείτε να αποφασίσετε αν τα δεδομένα είναι ονομαστικά ή κάποιο άλλο επίπεδο μέτρησης, ρωτήστε τον εαυτό σας αυτή την ερώτηση: οι αριθμοί που αντιστοιχούν σε αυτά τα δεδομένα αντιπροσωπεύουν κάποια ποιότητα τέτοια ώστε μια υψηλότερη τιμή να δείχνει ότι το αντικείμενο έχει μεγαλύτερη ποιότητα από μια χαμηλότερη τιμή ; Για παράδειγμα, υπάρχει κάποιο «φύλο» ποιότητας που οι άνδρες έχουν περισσότερες από τις γυναίκες; Σαφώς όχι και το σχήμα κωδικοποίησης θα λειτουργούσε επίσης αν οι γυναίκες κωδικοποιήθηκαν ως 1 και οι άνδρες ήταν 0. Η ίδια αρχή ισχύει και για το παράδειγμα του μπέιζμπολ: δεν υπάρχει ποιότητα "baseballness" των οποίων οι outfielders έχουν περισσότερα από τους pitchers. Οι αριθμοί είναι απλώς ένας βολικός τρόπος για την επισήμανση των υποκειμένων στη μελέτη και το πιο σημαντικό σημείο είναι ότι κάθε θέση έχει διαφορετική αξία. Ένα άλλο όνομα για τα ονομαστικά δεδομένα είναι τα κατηγορηματικά δεδομένα, αναφερόμενα στο γεγονός ότι οι μετρήσεις τοποθετούν αντικείμενα σε κατηγορίες (αρσενικά ή θηλυκά) αντί να μετρήσουν κάποια εγγενή ποιότητα σε αυτά. Το κεφάλαιο 10 αναλύει τις μεθόδους ανάλυσης που είναι κατάλληλες για αυτόν τον τύπο δεδομένων και πολλές τεχνικές που καλύπτονται στο Κεφάλαιο 11, σχετικά με τις μη παραμετρικές στατιστικές, είναι επίσης κατάλληλες για κατηγορηματικά δεδομένα.

Όταν τα δεδομένα μπορούν να λάβουν μόνο δύο τιμές, όπως στο παράδειγμα ανδρών / γυναικών, μπορεί επίσης να ονομάζονται binary data. Αυτός ο τύπος δεδομένων είναι τόσο συνηθισμένος ώστε να έχουν αναπτυχθεί ειδικές τεχνικές για τη μελέτη του, συμπεριλαμβανομένης της διοικητικής παλινδρόμησης (που συζητείται στο Κεφάλαιο 15), η οποία έχει εφαρμογές σε πολλούς τομείς. Πολλές ιατρικές στατιστικές όπως ο λόγος πιθανότητας και ο λόγος κινδύνου (που συζητήθηκε στο Κεφάλαιο 18) αναπτύχθηκαν για να περιγράψουν τη σχέση μεταξύ δύο δυαδικών μεταβλητών, διότι οι δυαδικές μεταβλητές συμβαίνουν τόσο συχνά στην ιατρική έρευνα.

Ordinal Data

Τα «κανονικά» δεδομένα αναφέρονται σε δεδομένα που έχουν κάποια σημαντική σειρά, έτσι ώστε οι υψηλότερες τιμές να αντιπροσωπεύουν περισσότερο κάποιο χαρακτηριστικό από τις χαμηλότερες τιμές. Για παράδειγμα, στην ιατρική πρακτική τα εγκαύματα περιγράφονται συνήθως από το βαθμό τους, το οποίο περιγράφει την ποσότητα βλάβης ιστού που προκαλείται από το κάψιμο. Ένα κάψιμο πρώτου

βαθμού χαρακτηρίζεται από ερυθρότητα του δέρματος, μικρό πόνο και βλάβη στην επιδερμίδα μόνο, ενώ ένα έγκαυμα δευτέρου βαθμού περιλαμβάνει φλύκταινες και εμπλέκει το χόριο, ένα κάψιμο τρίτου βαθμού χαρακτηρίζεται από φρύξη του δέρματος και ενδεχομένως να καταστρέφονται οι καταλήξεις των νεύρων. Αυτές οι κατηγορίες μπορούν να ταξινομηθούν με λογική σειρά: τα εγκαύματα πρώτου βαθμού είναι τα λιγότερο σοβαρά από άποψη βλάβης ιστών, τα σοβαρότερα εγκαύματα είναι του τρίτου βαθμού. Ωστόσο, δεν υπάρχει καμία μέτρηση ανάλογη με ένα χάρακα ή κλίμακα για να προσδιοριστεί πόσο μεγάλη είναι η απόσταση μεταξύ των κατηγοριών, ούτε είναι δυνατόν να προσδιοριστεί εάν η διαφορά μεταξύ εγκαυμάτων πρώτου και δευτέρου βαθμού είναι ίδια με τη διαφορά μεταξύ δευτέρου και τρίτου.

Πολλές συγγραφικές κλίμακες περιλαμβάνουν τάξεις: για παράδειγμα, οι υποψήφιοι που υποβάλλουν αίτηση για μια θέση εργασίας μπορούν να ταξινομηθούν από το τμήμα προσωπικού κατά σειρά επιθυμίας ως νέα μίσθωση. Θα μπορούσαμε επίσης να κατατάξουμε τα κράτη των Η.Π.Α. ανάλογα με τον πληθυσμό, τη γεωγραφική περιοχή ή τα ομοσπονδιακά φορολογικά έσοδα. Οι αριθμοί που χρησιμοποιούνται για τη μέτρηση με τα κανονικά δεδομένα έχουν μεγαλύτερη σημασία από εκείνους που χρησιμοποιούνται στα ονομαστικά δεδομένα και έχουν αναπτυχθεί πολλές στατιστικές τεχνικές για την πλήρη χρήση των πληροφοριών που μεταφέρονται κατά την παραγγελία, χωρίς όμως να υποθέσουμε άλλες ιδιότητες των ζυγών. Για παράδειγμα, είναι σωστό να υπολογιστεί η διάμεση (κεντρική τιμή) των σειριακών δεδομένων, αλλά όχι ο μέσος όρος (ο οποίος λαμβάνει δεδομένα διαστήματος). Ορισμένες από αυτές τις τεχνικές συζητούνται αργότερα σε αυτό το κεφάλαιο, ενώ άλλες καλύπτονται στο Κεφάλαιο 11.

Δεδομένα διαστήματος

Τα χρονικά διαστήματα έχουν μια ουσιαστική σειρά και έχουν επίσης την ποιότητα ότι τα ίσα διαστήματα μεταξύ των μετρήσεων αντιπροσωπεύουν ίσες αλλαγές στην ποσότητα όσων μετριοούνται. Το πιο συνηθισμένο παράδειγμα των δεδομένων διαστήματος είναι η κλίμακα θερμοκρασίας Fahrenheit. Αν περιγράψουμε τη θερμοκρασία χρησιμοποιώντας τη κλίμακα Fahrenheit, η διαφορά μεταξύ 10 και 25 βαθμών (διαφορά 15) αντιπροσωπεύει την ίδια ποσότητα αλλαγής θερμοκρασίας με τη διαφορά μεταξύ 60 και 75 βαθμών. Η προσθήκη και η αφαίρεση είναι κατάλληλες με κλίμακες διαστήματος: μια διαφορά 10 βαθμών αντιπροσωπεύει την ίδια ποσότητα σε ολόκληρη τη κλίμακα θερμοκρασίας. Ωστόσο, η κλίμακα Fahrenheit, όπως όλες οι κλίμακες διαστήματος, δεν έχει φυσικό μηδενικό σημείο, επειδή το 0 στην κλίμακα Fahrenheit δεν αντιπροσωπεύει την απουσία θερμοκρασίας αλλά απλά μια θέση σε σχέση με άλλες θερμοκρασίες. Ο πολλαπλασιασμός και ο διαχωρισμός δεν είναι κατάλληλοι με τα δεδομένα

διαστήματος: δεν υπάρχει μαθηματική έννοια στη δήλωση ότι οι 80 βαθμοί είναι διπλάσιοι από τους 40 βαθμούς. Οι κλίμακες διαστήματος είναι μια σπανιότητα: στην πραγματικότητα είναι δύσκολο να σκεφτούμε ένα άλλο κοινό παράδειγμα. Για το λόγο αυτό, ο όρος "δεδομένα διαστήματος" χρησιμοποιείται μερικές φορές για να περιγράψει τόσο τα δεδομένα διαστήματος όσο και την αναλογία (που αναφέρονται στην επόμενη ενότητα).

Δεδομένα αναλογίας

Τα δεδομένα αναλογίας έχουν όλες τις ιδιότητες των δεδομένων διαστήματος (φυσική σειρά, ίσα διαστήματα) συν ένα φυσικό μηδενικό σημείο. Πολλές φυσικές μετρήσεις είναι δεδομένα αναλογίας: για παράδειγμα, το ύψος, το βάρος και η ηλικία, όλα πληρούν τις προϋποθέσεις. Το ίδιο ισχύει και για το εισόδημα: μπορείτε ασφαλώς να κερδίσετε 0 δολάρια σε ένα χρόνο ή να έχετε 0 δολάρια στον τραπεζικό σας λογαριασμό. Τα δεδομένα σε επίπεδο αναφοράς, είναι σκόπιμο να πολλαπλασιάζονται και να διαιρούνται καθώς και να προστίθενται και να αφαιρούνται: είναι λογικό να λέμε ότι κάποιος με \$ 100 έχει διπλάσιο ποσό σε σχέση με κάποιον με \$ 50 ή ότι ένα άτομο ηλικίας 30 ετών είναι 3 φορές μεγαλύτερο από κάποιον που είναι 10 ετών.

Θα πρέπει να σημειωθεί ότι πολύ λίγες ψυχολογικές μετρήσεις (IQ, aptitude, κ.λπ.) είναι πραγματικά διαστήματα, και πολλά είναι στην πραγματικότητα κανονικά (π.χ., αξία στην εκπαίδευση, όπως υποδηλώνεται από μια κλίμακα Likert). Παρόλα αυτά, μερικές φορές θα δείτε τεχνικές διαστήματος ή αναλογίας που εφαρμόζονται σε τέτοια δεδομένα (για παράδειγμα, ο υπολογισμός των μέσων, που περιλαμβάνει τη διαίρεση). Ενώ είναι λανθασμένο από στατιστική άποψη, μερικές φορές πρέπει να πάτε με τις συμβάσεις του τομέα σας, ή τουλάχιστον να τις γνωρίζετε. Για να το θέσουμε με άλλο τρόπο, ένα μέρος των στατιστικών μάθησης μαθαίνει αυτό που είναι κοινώς αποδεκτό στο επιλεγμένο πεδίο προσπάθειάς σας, το οποίο μπορεί να είναι ξεχωριστό θέμα από αυτό που είναι αποδεκτό από καθαρά μαθηματική άποψη.

Συνεχή και διακριτά δεδομένα

Μια άλλη διάκριση που γίνεται συχνά είναι μεταξύ συνεχών και discretedata. Τα συνεχή δεδομένα μπορούν να λάβουν οποιαδήποτε τιμή ή οποιαδήποτε τιμή εντός εύρους. Τα περισσότερα δεδομένα που μετρούνται με κλίμακες διαστήματος και αναλογίας, εκτός από αυτά που βασίζονται στην καταμέτρηση, είναι συνεχή: για παράδειγμα, το βάρος, το ύψος, η απόσταση και το εισόδημα είναι όλα συνεχή.

Κατά την ανάλυση δεδομένων και τη δημιουργία μοντέλων, οι ερευνητές μερικές φορές επανασυνδέουν συνεχόμενα δεδομένα σε κατηγορίες ή μεγαλύτερες μονάδες. Για παράδειγμα, το βάρος μπορεί να καταγραφεί σε λίβρες αλλά

αναλύθηκε σε προσαυξήσεις 10 λιβρών ή ηλικία καταγεγραμμένη σε έτη, αλλά αναλύθηκε με όρους κατηγοριών 0-17, 18-65 και άνω των 65 ετών. Από στατιστική άποψη, δεν υπάρχει απόλυτο σημείο όταν τα δεδομένα γίνονται συνεχή ή διακριτά για τους σκοπούς της χρήσης συγκεκριμένων αναλυτικών τεχνικών: αν καταγράψουμε την ηλικία σε χρόνια, επιβάλλουμε ακόμη διακριτές κατηγορίες σε μια συνεχή μεταβλητή. Για παράδειγμα, ορισμένοι ερευνητές λένε ότι όταν μια μεταβλητή έχει 10 ή περισσότερες κατηγορίες (ή εναλλακτικά, 16 ή περισσότερες κατηγορίες), μπορεί ασφαλώς να αναλυθεί ως συνεχής. Αυτή είναι μια άλλη απόφαση που πρέπει να λαμβάνεται κατά περίπτωση, ενημερωμένη από τα συνήθη πρότυπα και πρακτικές της συγκεκριμένης πειθαρχίας σας και τον τύπο προτεινόμενης ανάλυσης.

Τα διακριτά δεδομένα μπορούν να λάβουν μόνο συγκεκριμένες τιμές και έχουν σαφή όρια. Όπως το παλιό αστείο: μπορείτε να έχετε 2 παιδιά ή 3 παιδιά, αλλά όχι 2.37 παιδιά, οπότε ο "αριθμός παιδιών" είναι μια διακριτή μεταβλητή. Στην πραγματικότητα, οποιαδήποτε μεταβλητή που βασίζεται στην καταμέτρηση είναι διακριτή, είτε μετράτε τον αριθμό των βιβλίων που αγοράσατε σε ένα χρόνο είτε τον αριθμό των επισκέψεων προγεννητικής φροντίδας που έγιναν κατά τη διάρκεια της εγκυμοσύνης. Τα ονομαστικά δεδομένα είναι επίσης διακριτά, όπως και τα δυαδικά και ταξινομημένα δεδομένα.

Λειτουργία

Οι αρχάριοι σε έναν τομέα συχνά πιστεύουν ότι οι δυσκολίες της έρευνας βασίζονται κυρίως στη στατιστική ανάλυση και εστιάζουν τις προσπάθειές τους στην εκμάθηση μαθηματικών τύπων και τεχνικών προγραμματισμού ηλεκτρονικών υπολογιστών για τη διεξαγωγή στατιστικών υπολογισμών. Ωστόσο, ένα σημαντικό πρόβλημα στην έρευνα δεν έχει να κάνει ούτε με τα μαθηματικά ούτε με τα στατιστικά στοιχεία, αλλά με το να γνωρίζετε προσεκτικά το πεδίο σπουδών σας και να σκέφτεστε με πρακτικά προβλήματα. Αυτό είναι το πρόβλημα της <<επιχειρηματοποίησης>>, της διαδικασίας προσδιορισμού του τρόπου με τον οποίο θα οριστεί και θα μετρηθεί μια έννοια. Η επιχειρηματοποίηση αποτελεί ιδιαίτερη μέριμνα στις κοινωνικές επιστήμες και την εκπαίδευση, αλλά ισχύει και για άλλους τομείς.

Η λειτουργία είναι πάντα απαραίτητη όταν μια ποιότητα ενδιαφέροντος δεν μπορεί να μετρηθεί άμεσα. Ένα προφανές παράδειγμα είναι η νοημοσύνη: δεν υπάρχει κανένας τρόπος για να μετρήσουμε άμεσα τη νοημοσύνη, έτσι στη θέση μιας τέτοιας άμεσης μέτρησης αποδεχόμαστε κάτι που μπορούμε να μετρήσουμε, όπως το σκορ σε ένα τεστ IQ. Ομοίως, δεν υπάρχει άμεσος τρόπος μέτρησης της ετοιμότητας για καταστροφές για μια πόλη, αλλά μπορούμε να υλοποιήσουμε την ιδέα δημιουργώντας μια λίστα ελέγχου των καθηκόντων που πρέπει να εκτελεστούν και δίνοντας σε κάθε πόλη μια βαθμολογία ετοιμότητας για καταστροφή βάσει του

αριθμού των εργασιών που έχουν ολοκληρωθεί και την ποιότητα ή την πληρότητα της ολοκλήρωσης. Για ένα τρίτο παράδειγμα, μπορεί να επιθυμούμε να μετρήσουμε την ποσότητα της φυσικής δραστηριότητας που εκτελείται από τα άτομα σε μια μελέτη: εάν δεν έχουμε την ικανότητα να παρακολουθούμε άμεσα τη συμπεριφορά άσκησής τους, μπορούμε να λειτουργήσουμε με το "ποσό σωματικής άσκησης", ένα αυτοαναφερόμενο ερωτηματολόγιο ή την καταγραφή σε ημερολόγιο.

Επειδή πολλές από τις ιδιότητες που μελετώνται στις κοινωνικές επιστήμες είναι αφηρημένες, η λειτουργικοποίηση είναι ένα κοινό θέμα συζήτησης στους τομείς αυτούς. Ωστόσο, ισχύει και σε πολλά άλλα πεδία. Για παράδειγμα, οι απώτεροι στόχοι του ιατρικού επαγγέλματος είναι η μείωση της θνησιμότητας (θάνατος) και η μείωση του βάρους των ασθενειών και των ταλαιπωριών. Η θνησιμότητα ελέγχεται και ποσοτικοποιείται εύκολα, αλλά είναι συχνά υπερβολικά αμβλεία ως εργαλείο για να είναι χρήσιμη, δεδομένου ότι είναι ευτυχώς σπάνιο αποτέλεσμα για τις περισσότερες ασθένειες. Το "βάρος της νόσου" και "ο πόνος", από την άλλη πλευρά, είναι έννοιες που θα μπορούσαν να χρησιμοποιηθούν για τον καθορισμό των κατάλληλων αποτελεσμάτων για πολλές μελέτες, αλλά που δεν έχουν άμεσο μέσο μέτρησης και πρέπει επομένως να λειτουργούν. Παραδείγματα λειτουργικοποίησης του βάρους της νόσου περιλαμβάνουν τη μέτρηση των επιπέδων ιού στην κυκλοφορία του αίματος για ασθενείς με AIDS και τη μέτρηση του μεγέθους του όγκου για τα άτομα με καρκίνο. Μειωμένα επίπεδα δυστυχίας ή βελτιωμένη ποιότητα ζωής μπορούν να τεθούν σε λειτουργία ως υψηλότερη κατάσταση υγείας που αναφέρθηκε στον εαυτό τους, υψηλότερη βαθμολογία σε ένα εργαλείο έρευνας που σχεδιάστηκε για τη μέτρηση της ποιότητας ζωής, βελτιωμένη κατάσταση διάθεσης όπως μετρήθηκε μέσω προσωπικής συνέντευξης ή μείωση της ποσότητας μορφίνης ζητείται.

Ορισμένοι υποστηρίζουν ότι η μέτρηση ακόμη και των φυσικών μεγεθών, όπως το μήκος, απαιτεί την επιχειρησιακή λειτουργία, επειδή υπάρχουν διάφοροι τρόποι μέτρησης του μήκους (ένας κυβερνήτης μπορεί να είναι το κατάλληλο όργανο σε ορισμένες περιπτώσεις, ένα μικρόμετρο σε άλλους). Ωστόσο, το πρόβλημα της λειτουργικοποίησης είναι πολύ μεγαλύτερο στις ανθρωπιστικές επιστήμες, όταν το αντικείμενο ή οι ιδιότητες ενδιαφέροντος συχνά δεν μπορούν να μετρηθούν άμεσα.

Μέτρηση μεσολάβησης

Ο όρος μέτρηση μεσολάβησης αναφέρεται στη διαδικασία αντικατάστασης μιας μέτρησης. Αν και η λήψη αποφάσεων σχετικά με τις μετρήσεις μεσολάβησης μπορεί να θεωρηθεί ως μια υποκατηγορία επιχειρησιακής λειτουργίας, θα το θεωρήσουμε ως ξεχωριστό θέμα. Η συνηθέστερη χρήση της μέτρησης μεσολάβησης είναι η αντικατάσταση μίας μέτρησης που είναι φθηνή και εύκολα διαθέσιμη από μια διαφορετική μέτρηση που θα ήταν πιο δύσκολη ή δαπανηρή, αν όχι αδύνατη, για τη συλλογή.

Για ένα απλό παράδειγμα μετρικής μέτρησης, εξετάστε μερικές από τις μεθόδους που χρησιμοποιούνται από αστυνομικούς για να αξιολογήσετε την αυτοπεποίθηση των ατόμων ενώ βρίσκεστε στο πεδίο. Έχοντας ένα φορητό ιατρικό εργαστήριο, ένας αξιωματικός δεν μπορεί να μετρήσει άμεσα το περιεχόμενο αλκοόλ στο αίμα για να καθορίσει εάν ένα υποκείμενο είναι νόμιμα μεθυσμένο ή όχι. Επομένως, ο αξιωματικός βασίζεται στην παρατήρηση σημείων που συνδέονται με την μεθυστική κατάσταση, καθώς και σε μερικές απλές δοκιμές πεδίου που πιστεύεται ότι συσχετίζονται καλά με την περιεκτικότητα σε αλκοόλ στο αίμα. Τα συμπτώματα της δηλητηρίασης με οινόπνευμα περιλαμβάνουν την αναπνοή που μυρίζει αλκοόλ, την ομιλία και την έκπλυση του δέρματος. Οι δοκιμές πεδίου που χρησιμοποιούνται για την ταχεία αξιολόγηση της δηλητηρίασης από οινόπνευμα γενικά απαιτούν από τα υποκείμενα να εκτελούν εργασίες όπως στέκονται σε ένα πόδι ή να εντοπίζουν ένα κινούμενο αντικείμενο με τα μάτια τους. Ούτε τα παρατηρούμενα σημάδια ούτε τα μέτρα απόδοσης είναι άμεσα μέτρα μέθης, αλλά είναι γρήγορα και εύκολα στη διαχείριση. Τα άτομα που είναι ύποπτα για μεθυστική κατάσταση όπως αξιολογούνται από αυτά τα μέτρα αντιπροσώπευσης μπορούν στη συνέχεια να υποβληθούν σε πιο ακριβή έλεγχο της περιεκτικότητάς τους σε αλκοόλ στο αίμα.

Μια άλλη κοινή (και μερικές φορές αμφιλεγόμενη) χρήση μέτρησης μεσολάβησης είναι οι διάφορες μέθοδοι που χρησιμοποιούνται συνήθως για την αξιολόγηση της ποιότητας της υγειονομικής περίθαλψης που παρέχεται από νοσοκομεία ή γιατρούς. Θεωρητικά, θα ήταν δυνατή η άμεση μέτρηση της ποιότητας της περίθαλψης, για παράδειγμα μέσω της άμεσης παρατήρησης της παρεχόμενης περίθαλψης και της αξιολόγησής της σε σχέση με τα αποδεκτά πρότυπα (αν και αυτή η διαδικασία θα εξακολουθούσε να είναι μια επιχειρηματοποίηση της αφηρημένης έννοιας "ποιότητα της φροντίδας"). Ωστόσο, η εφαρμογή μιας τέτοιας διαδικασίας θα ήταν απαγορευτικά δαπανηρή καθώς και μια εισβολή στην ιδιωτική ζωή των ασθενών. Μια λύση που υιοθετείται συνήθως είναι η μέτρηση των διαδικασιών που θεωρείται ότι αντανakλούν την υψηλότερη ποιότητα της περίθαλψης: για παράδειγμα, εάν προσφέρθηκε κατά την επίσκεψη γραφείου η παροχή συμβουλών κατά του καπνού ή εάν τα κατάλληλα φάρμακα χορηγήθηκαν αμέσως μετά την εισαγωγή του ασθενούς στο νοσοκομείο.

Οι μετρήσεις μεσολάβησης είναι πιο χρήσιμες εάν, εκτός από την σχετικά εύκολη απόκτηση, είναι καλοί δείκτες της πραγματικής εστίασης ενδιαφέροντος. Για παράδειγμα, εάν η ορθή εκτέλεση των συνταγογραφούμενων διαδικασιών της ιατρικής περίθαλψης για μια συγκεκριμένη θεραπεία είναι στενά συνδεδεμένη με την καλή έκβαση των ασθενών για την εν λόγω κατάσταση, και εάν η ανύπαρκτη εκτέλεση αυτών των διαδικασιών είναι στενά συνδεδεμένη με την κακή έκβαση των ασθενών. Εάν αυτή η στενή σχέση δεν υπάρχει, τότε η χρησιμότητα των μετρήσεων αυτών των διαδικασιών ως υποκατάστατο για την ποιότητα της φροντίδας είναι λιγότερο βέβαιη. Δεν υπάρχει μαθηματική δοκιμασία που θα σας πει εάν ένα μέτρο

είναι καλός πληρεξούσιος για ένα άλλο, παρόλο που υπολογιστικές στατιστικές όπως συσχετισμοί ή chi-squares μεταξύ των μέτρων μπορεί να βοηθήσουν στην αξιολόγηση αυτού του ζητήματος. Όπως πολλά θέματα μέτρησης, η επιλογή καλών μετρήσεων μεσολάβησης είναι θέμα κρίσης που ενημερώνεται από τη γνώση του τομέα, τις συνήθειες πρακτικές στον τομέα και την κοινή λογική.

Αλήθεια και σφάλματα

Μπορούμε να υποθέσουμε με ασφάλεια ότι καμία μέτρηση δεν είναι απολύτως ακριβής. Επειδή η διαδικασία μέτρησης περιλαμβάνει την ανάθεση διακριτών αριθμών σε έναν συνεχή κόσμο, ακόμη και οι μετρήσεις που διενεργούνται από το καλύτερα εκπαιδευμένο προσωπικό χρησιμοποιώντας τα καλύτερα διαθέσιμα επιστημονικά όργανα δεν είναι εντελώς λάθος. Μια ανησυχία της θεωρίας των μετρήσεων είναι η εννοιοποίηση και η ποσοτικοποίηση του βαθμού σφάλματος που υπάρχει σε ένα συγκεκριμένο σύνολο μετρήσεων και η αξιολόγηση των πηγών και των συνεπειών αυτού του σφάλματος.

Η κλασική θεωρία μετρήσεων θεωρεί κάθε μέτρηση ή παρατηρούμενη βαθμολογία ως αποτελούμενη από δύο μέρη: πραγματική βαθμολογία και σφάλμα. Αυτό εκφράζεται στον ακόλουθο τύπο:

$$X = T + E$$

όπου το X είναι η παρατηρούμενη μέτρηση, το T είναι το πραγματικό σκορ και το E είναι το σφάλμα. Για παράδειγμα, η ζυγαριά του μπάνιου μπορεί να μετρήσει το βάρος κάποιου ως 120 κιλά, όταν το πραγματικό βάρος αυτού του ατόμου ήταν 118 κιλά και το σφάλμα των 2 κιλών οφειλόταν στην ανακρίβεια της κλίμακας. Αυτό θα εκφράζεται μαθηματικά ως εξής:

$$120 = 118 + 2$$

που είναι απλώς μια μαθηματική ισότητα που εκφράζει τη σχέση μεταξύ των τριών συστατικών. Ωστόσο, τόσο το T όσο και το E είναι υποθετικά δομήματα: στον πραγματικό κόσμο, δεν γνωρίζουμε ποτέ την ακριβή τιμή του πραγματικού σκορ και επομένως δεν μπορούμε να γνωρίζουμε ούτε την αξία του βαθμού σφάλματος. Μεγάλο μέρος της διαδικασίας μέτρησης περιλαμβάνει την εκτίμηση και των δύο ποσοτήτων και τη μεγιστοποίηση του πραγματικού στοιχείου ενώ ταυτόχρονα ελαχιστοποιείται το σφάλμα. Για παράδειγμα, αν πήραμε αρκετές μετρήσεις σωματικού βάρους σε σύντομο χρονικό διάστημα (έτσι ώστε να θεωρείται ότι το πραγματικό βάρος παρέμεινε σταθερό), χρησιμοποιώντας τις πιο ακριβείς διαθέσιμες κλίμακες, θα μπορούσαμε να αποδεχθούμε τον μέσο όρο όλων των μετρήσεων μια καλή εκτίμηση του πραγματικού βάρους. Στη συνέχεια θα εξετάζαμε τη διακύμανση μεταξύ αυτού του μέσου όρου και κάθε μεμονωμένης μέτρησης ως

το σφάλμα που οφείλεται στη διαδικασία μέτρησης, όπως μικρές ανακρίβειες σε κάθε κλίμακα.

Τυχαίο και συστηματικό σφάλμα

Επειδή ζούμε στον πραγματικό κόσμο και όχι σε ένα πλατωνικό σύμπαν, υποθέτουμε ότι όλες οι μετρήσεις περιέχουν κάποιο λάθος. Αλλά δεν δημιουργούνται όλα τα σφάλματα ίσα. Το τυχαίο σφάλμα οφείλεται στην τύχη: δεν χρειάζεται ιδιαίτερο σχέδιο και υποτίθεται ότι ακυρώνεται κατά τη διάρκεια επαναλαμβανόμενων μετρήσεων. Για παράδειγμα, οι βαθμολογίες σφάλματος σε διάφορες μετρήσεις του ίδιου αντικειμένου θεωρούνται ότι έχουν μέση τιμή μηδέν. Έτσι, εάν ζυγιστεί κάποιος 10 φορές διαδοχικά στην ίδια κλίμακα, παρατηρούμε μικρές διαφορές στον αριθμό που μας δόθηκε: κάποιιοι θα είναι υψηλότεροι από την πραγματική αξία και κάποιιοι θα είναι χαμηλότεροι. Υποθέτοντας ότι το πραγματικό βάρος είναι 120 λίβρες, ίσως η πρώτη μέτρηση θα επιστρέψει ένα παρατηρούμενο βάρος 119 λιβρών (συμπεριλαμβανομένου ενός λάθους -1 λίρας), το δεύτερο ένα παρατηρούμενο βάρος των 122 κιλών (για σφάλμα +2 κιλά), το τρίτο παρατηρούμενο βάρος 118,5 λίβρες (σφάλμα -1,5 κιλά) και ούτω καθεξής. Εάν η κλίμακα είναι ακριβής και το μοναδικό σφάλμα είναι τυχαίο, το μέσο σφάλμα σε πολλές δοκιμές θα είναι μηδέν και το μέσο παρατηρούμενο βάρος θα είναι 120 λίβρες. Μπορούμε να προσπαθήσουμε να μειώσουμε το ποσό του τυχαίου σφάλματος χρησιμοποιώντας ακριβέστερα όργανα, εκπαιδεύοντας τους τεχνικούς μας για να τα χρησιμοποιήσουμε σωστά κ.ο.κ., αλλά δεν μπορούμε να περιμένουμε να εξαλείψουμε πλήρως τυχαίο λάθος.

Δύο άλλες προϋποθέσεις θεωρούνται ότι ισχύουν για τυχαίο σφάλμα: δεν πρέπει να σχετίζονται με την πραγματική βαθμολογία και η συσχέτιση μεταξύ σφαλμάτων θεωρείται μηδενική. Η πρώτη προϋπόθεση σημαίνει ότι η τιμή του στοιχείου σφάλματος δεν σχετίζεται με την τιμή της πραγματικής βαθμολογίας. Εάν μετρήσαμε τα βάρη ενός αριθμού διαφορετικών ατόμων των οποίων τα πραγματικά βάρη διέφεραν, δεν θα περίμενε κανείς ότι το στοιχείο σφάλματος θα έχει οποιαδήποτε σχέση με το πραγματικό βάρος τους. Για παράδειγμα, το στοιχείο σφάλματος δεν θα πρέπει συστηματικά να είναι μεγαλύτερο όταν το πραγματικό βάρος είναι μεγαλύτερο. Η δεύτερη προϋπόθεση σημαίνει ότι το σφάλμα για κάθε σκορ είναι ανεξάρτητο και δεν σχετίζεται με το σφάλμα για οποιοδήποτε άλλο σκορ: για παράδειγμα, δεν θα πρέπει να υπάρχει ένα πρότυπο με το μέγεθος του σφάλματος να αυξάνεται με την πάροδο του χρόνου (πράγμα που μπορεί να σημαίνει ότι η κλίμακα παρασύρεται από βαθμονόμηση).

Αντίθετα, το συστηματικό σφάλμα είναι ένα παρατηρούμενο πρότυπο, δεν οφείλεται στην τύχη και συχνά έχει μια αιτία ή αιτίες που μπορούν να εντοπιστούν και να αποκατασταθούν. Για παράδειγμα, η κλίμακα μπορεί να βαθμολογηθεί εσφαλμένα για να δείξει ένα αποτέλεσμα που είναι πέντε κιλά πάνω από το

πραγματικό βάρος, οπότε ο μέσος όρος των παραπάνω μετρήσεων θα ήταν 125 κιλά, όχι 120. Το συστηματικό σφάλμα μπορεί επίσης να οφείλεται σε ανθρώπινους παράγοντες: ίσως είμαστε διαβάζοντας την οθόνη της ζυγαριάς σε μια γωνία έτσι ώστε να βλέπουμε τη βελόνα να καταγράφει πέντε κιλά υψηλότερα από ό, τι πραγματικά δείχνει. Μια ζυγαριά που παρασύρεται υψηλότερα (έτσι τα στοιχεία σφάλματος είναι τυχαία στην αρχή του πειράματος, αλλά αργότερα είναι σταθερά υψηλά) είναι ένα άλλο παράδειγμα συστηματικού σφάλματος. Έχουν καταβληθεί πολλές προσπάθειες για τον εντοπισμό πηγών συστηματικού σφάλματος και την επεξεργασία μεθόδων για τον εντοπισμό και την εξάλειψή τους: αυτό συζητείται περαιτέρω στο επικείμενο τμήμα σχετικά με τη μεροληψία μέτρησης.

Αξιοπιστία και εγκυρότητα

Υπάρχουν πολλοί τρόποι με τους οποίους μπορείτε να αντιστοιχίσετε αριθμούς ή κατηγορίες σε δεδομένα και δεν είναι εξίσου χρήσιμα. Δύο πρότυπα που χρησιμοποιούμε για την αξιολόγηση των μετρήσεων είναι η αξιοπιστία και η εγκυρότητα. Στην ιδανική περίπτωση, κάθε μέτρο που χρησιμοποιούμε πρέπει να είναι αξιόπιστο και έγκυρο. Στην πραγματικότητα, αυτές οι ιδιότητες δεν είναι απόλυτες, αλλά είναι θέματα βαθμού και συχνά ειδικών για την περίπτωση: ένα μέτρο που είναι πολύ αξιόπιστο όταν χρησιμοποιείται με μια ομάδα ανθρώπων μπορεί να είναι αναξιόπιστο όταν χρησιμοποιείται με μια διαφορετική ομάδα, για παράδειγμα. Για το λόγο αυτό, είναι πιο χρήσιμο να αξιολογηθεί πόσο έγκυρο και αξιόπιστο είναι ένα μέτρο για έναν συγκεκριμένο σκοπό και αν τα επίπεδα αξιοπιστίας και εγκυρότητας είναι αποδεκτά στο συγκεκριμένο πλαίσιο. Η αξιοπιστία και η εγκυρότητα εξετάζονται επίσης στο Κεφάλαιο 5, στο πλαίσιο του σχεδιασμού της έρευνας και στο Κεφάλαιο 19, στο πλαίσιο των εκπαιδευτικών και ψυχολογικών εξετάσεων.

Αξιοπιστία

Η αξιοπιστία αναφέρεται στο πώς είναι συνεπείς ή επαναλαμβανόμενες οι μετρήσεις. Για παράδειγμα, αν δίνουμε στο ίδιο άτομο την ίδια δοκιμή σε δύο διαφορετικές περιπτώσεις, θα είναι οι βαθμολογίες παρόμοιες και στις δύο περιπτώσεις; Εάν εκπαιδεύσουμε τρία άτομα για να χρησιμοποιήσουμε μια κλίμακα βαθμολόγησης που έχει σχεδιαστεί για να μετράμε την ποιότητα της κοινωνικής αλληλεπίδρασης μεταξύ των ατόμων, τότε έδειξε σε καθένα από αυτά την ίδια ταινία μιας ομάδας ανθρώπων που αλληλεπιδρούν και τους ζήτησε να αξιολογήσουν την κοινωνική αλληλεπίδραση που παρουσιάζεται στην ταινία. Οι αξιολογήσεις θα είναι παρόμοιες; Εάν έχουμε έναν τεχνικό να μετρήσει το ίδιο μέρος 10 φορές, χρησιμοποιώντας το ίδιο όργανο, οι μετρήσεις θα είναι παρόμοιες κάθε φορά; Σε κάθε περίπτωση, αν η απάντηση είναι ναι, μπορούμε να πούμε ότι η δοκιμή, η κλίμακα ή το όργανο είναι αξιόπιστες.

Μεγάλο μέρος της θεωρίας και της πρακτικής της αξιοπιστίας αναπτύχθηκε στον τομέα της εκπαιδευτικής ψυχολογίας και για το λόγο αυτό, τα μέτρα αξιοπιστίας περιγράφονται συχνά όσον αφορά την αξιολόγηση της αξιοπιστίας των εξετάσεων. Ωστόσο, οι εκτιμήσεις αξιοπιστίας δεν περιορίζονται σε εκπαιδευτικούς ελέγχους: οι ίδιες έννοιες ισχύουν για πολλούς άλλους τύπους μετρήσεων, όπως η δημοσκόπηση, οι έρευνες ικανοποίησης και οι αξιολογήσεις συμπεριφοράς.

Η συζήτηση σε αυτό το κεφάλαιο θα διατηρηθεί σε ένα αρκετά βασικό επίπεδο: οι πληροφορίες σχετικά με τον υπολογισμό συγκεκριμένων μέτρων αξιοπιστίας εξετάζονται λεπτομερέστερα στο κεφάλαιο 19, σε σχέση με τη θεωρία των δοκιμών. Επιπλέον, πολλά από τα μέτρα αξιοπιστίας βασίζονται στον συντελεστή συσχέτισης (ο οποίος επίσης ονομάζεται απλά ο συσχετισμός), ο οποίος συζητείται λεπτομερώς στο Κεφάλαιο 9, έτσι ώστε οι στατιστικολόγοι να θέλουν να επικεντρωθούν στη λογική της αξιοπιστίας και της εγκυρότητας και να αφήσουν τις λεπτομέρειες αξιολογώντας τα μέχρι να καταλάβουν την έννοια του συντελεστή συσχέτισης.

Υπάρχουν τρεις βασικές προσεγγίσεις για τη μέτρηση της αξιοπιστίας, καθένα από τα οποία είναι χρήσιμο σε συγκεκριμένα περιβάλλοντα και το καθένα έχει ιδιαίτερα πλεονεκτήματα και μειονεκτήματα:

1. Αξιοπιστία πολλαπλών περιπτώσεων
2. Αξιοπιστία πολλών μορφών
3. Αξιοπιστία εσωτερικής συνοχής

Η αξιοπιστία πολλαπλών περιπτώσεων, μερικές φορές αποκαλείται αξιοπιστία δοκιμής-επανεξέτασης, αναφέρεται στον τρόπο με τον οποίο εκτελεί παρόμοια δοκιμή ή κλίμακα σε επαναλαμβανόμενες δοκιμές. Για το λόγο αυτό αναφέρεται μερικές φορές ως δείκτης χρονικής σταθερότητας, που σημαίνει σταθερότητα με την πάροδο του χρόνου. Για παράδειγμα, ίσως το ίδιο πρόσωπο να κάνει μια ψυχολογική εκτίμηση ενός ασθενούς βάσει μιας βιντεοταινίας, με τις αξιολογήσεις να διενεργούνται δύο εβδομάδες με βάση την ίδια συνέντευξη που έχει εγγραφεί. Για να έχει νόημα αυτό το είδος αξιοπιστίας, πρέπει να υποθέσετε ότι η ποσότητα που μετράται δεν έχει αλλάξει: εξ ου και η χρήση της ίδιας βιντεοταινίας, αντί να διαχωρίσετε ζωντανές συνεντεύξεις με έναν ασθενή. Η αξιοπιστία πολλαπλών περιπτώσεων δεν είναι ένα κατάλληλο μέτρο για πτητικές ιδιότητες, όπως η κατάσταση διάθεσης. Είναι επίσης ακατάλληλο εάν η εστίαση της μέτρησης μπορεί να έχει αλλάξει κατά το χρονικό διάστημα μεταξύ των δοκιμών (για παράδειγμα, εάν ο σπουδαστής μάθει περισσότερα για ένα θέμα μεταξύ των περιόδων δοκιμών) ή μπορεί να αλλάξει ως αποτέλεσμα της πρώτης δοκιμής (για παράδειγμα, αν ο μαθητής θυμάται ποιες ερωτήσεις τέθηκαν σχετικά με την πρώτη δοκιμαστική διοίκηση). Μια κοινή τεχνική για την αξιολόγηση της αξιοπιστίας πολλαπλών περιπτώσεων είναι ο υπολογισμός του συντελεστή συσχέτισης μεταξύ των βαθμών από κάθε περίπτωση δοκιμής: αυτό ονομάζεται συντελεστής σταθερότητας.

Η αξιοπιστία πολλαπλών τύπων (που ονομάζεται επίσης αξιοπιστία παράλληλων μορφών) αναφέρεται στο πώς οι διαφορετικές εκδοχές μιας δοκιμής ή ερωτηματολογίου εκτελούν τη μέτρηση της ίδιας οντότητας. Ένας κοινός τύπος αξιοπιστίας πολλαπλών τύπων είναι η διακεκομμένη αξιοπιστία, στην οποία δημιουργείται μια ομάδα αντικειμένων που πιστεύεται ότι είναι ομοιογενής και τα μισά στοιχεία διατίθενται για το σχήμα Α και το ήμισυ για το Β. Εάν οι δύο (ή περισσότερες) μορφές του δοκιμής χορηγούνται στους ίδιους ανθρώπους με την ίδια ευκαιρία, ο συσχετισμός μεταξύ των βαθμολογιών που λαμβάνονται σε κάθε έντυπο είναι μια εκτίμηση της αξιοπιστίας πολλών φορμών. Αυτή η συσχέτιση ονομάζεται μερικές φορές συντελεστής ισοδυναμίας. Η αξιοπιστία πολλαπλών τύπων είναι σημαντική για τυποποιημένες δοκιμές που υπάρχουν σε πολλαπλές εκδόσεις: για παράδειγμα, οι διαφορετικές μορφές του SAT (Scholastic Aptitude Test, που χρησιμοποιούνται για τη μέτρηση της ακαδημαϊκής ικανότητας των φοιτητών που εφαρμόζονται σε Αμερικανικά κολέγια και πανεπιστήμια) βαθμονομούνται, ανεξάρτητα από τη μορφή που χρησιμοποιείται.

Η εσωτερική αξιοπιστία συνάφειας αναφέρεται στο πόσο καλά τα στοιχεία που συνθέτουν ένα τεστ αντανακλούν την ίδια δομή. Για να το θέσουμε με άλλο τρόπο, η αξιοπιστία της εσωτερικής συνοχής μετρά πόσο τα στοιχεία μιας δοκιμής μετρούν το ίδιο πράγμα. Αυτός ο τύπος αξιοπιστίας μπορεί να εκτιμηθεί με τη χορήγηση μιας μόνο δοκιμής σε μία μόνο περίπτωση. Η εσωτερική αξιοπιστία της συνέπειας είναι μια πιο πολύπλοκη ποσότητα για μέτρηση από την αξιοπιστία πολλαπλών περιπτώσεων ή παράλληλων μορφών και έχουν αναπτυχθεί αρκετές διαφορετικές μέθοδοι για την αξιολόγηση της: αυτές εξετάζονται περαιτέρω στο Κεφάλαιο 19. Ωστόσο, όλα εξαρτώνται κυρίως από τη συσχέτιση μεταξύ στοιχείων, δηλαδή τη συσχέτιση κάθε στοιχείου στην κλίμακα με κάθε άλλο στοιχείο. Αν οι συσχετισμοί αυτοί είναι υψηλοί, αυτό ερμηνεύεται ως απόδειξη ότι τα αντικείμενα μετρούν το ίδιο πράγμα και ότι τα διάφορα στατιστικά στοιχεία που χρησιμοποιούνται για τη μέτρηση της εσωτερικής αξιοπιστίας συνάφειας θα είναι όλα υψηλά. Εάν οι συσχετίσεις μεταξύ στοιχείων είναι χαμηλές ή ασυνεπείς, τα στατιστικά στοιχεία αξιοπιστίας εσωτερικής συνοχής θα είναι χαμηλά και αυτό ερμηνεύεται ως απόδειξη ότι τα στοιχεία δεν μετρούν το ίδιο πράγμα.

Δύο απλά μέτρα εσωτερικής συνέπειας που είναι πολύ χρήσιμα για δοκιμές που αποτελούνται από πολλαπλά αντικείμενα που καλύπτουν το ίδιο θέμα, παρόμοιας δυσκολίας και που θα βαθμολογηθούν ως σύνθετα, είναι η μέση συσχέτιση και συσχέτιση μεταξύ στοιχείου και συνολικού στοιχείου. Για να υπολογίσουμε τη μέση συσχέτιση μεταξύ στοιχείων, βρίσκουμε τη σχέση μεταξύ κάθε ζεύγους στοιχείων και λαμβάνουμε το μέσο όρο όλων των συσχετίσεων. Για να υπολογίσουμε τη μέση σχέση μεταξύ στοιχείου και ολικής συσχέτισης, δημιουργούμε ένα συνολικό σκορ προσθέτοντας βαθμολογίες σε κάθε μεμονωμένο στοιχείο της κλίμακας και στη συνέχεια υπολογίζουμε τη συσχέτιση κάθε στοιχείου με το σύνολο. Ο μέσος όρος

συσχέτισης μεταξύ στοιχείου και μέσου όρου είναι ο μέσος όρος αυτών των μεμονωμένων στοιχείων-συνολικών συσχετισμών.

Η αξιοπιστία κατά το ήμισυ, που περιγράφεται παραπάνω, είναι μια άλλη μέθοδος προσδιορισμού της εσωτερικής συνέπειας. Αυτή η μέθοδος έχει το μειονέκτημα ότι, αν τα στοιχεία δεν είναι πραγματικά ομοιογενή, διαφορετικούς διαχωρισμούς θα δημιουργήσουν οι μορφές ανισότητας δυσκολίας και ο συντελεστής αξιοπιστίας θα είναι διαφορετικός για κάθε ζεύγος μορφών. Μια μέθοδος που υπερνικά αυτή τη δυσκολία είναι το άλφα του Cronbach (συντελεστής άλφα), το οποίο είναι ισοδύναμο με το μέσο όρο όλων των δυνατών υποδιαιρέσεων. Για περισσότερες πληροφορίες σχετικά με το άλφα του Cronbach, συμπεριλαμβανομένης μιας επίδειξης για τον τρόπο υπολογισμού του, βλ. Κεφάλαιο 19.

Μέτρα συμφωνίας

Οι τύποι αξιοπιστίας που περιγράφονται παραπάνω είναι χρήσιμοι κυρίως για συνεχείς μετρήσεις. Όταν ένα πρόβλημα μέτρησης αφορά κατηγορηματικές κρίσεις, για παράδειγμα την κατάταξη των εξαρτημάτων μηχανών ως αποδεκτών ή ελαττωματικών, οι μετρήσεις συμφωνίας είναι καταλληλότερες. Για παράδειγμα, θα θέλαμε να αξιολογήσουμε τη συνοχή των αποτελεσμάτων από δύο διαφορετικές διαγνωστικές εξετάσεις για την παρουσία ή την απουσία ασθένειας. Ή θα θέλαμε να αξιολογήσουμε τη συνοχή των αποτελεσμάτων από τρεις κριτές που κατατάσσουν τη συμπεριφορά στην τάξη ως αποδεκτή ή απαράδεκτη. Σε κάθε περίπτωση, κάθε κριτής αποδίδει ένα μοναδικό αποτέλεσμα από ένα περιορισμένο σύνολο επιλογών και μας ενδιαφέρει το πόσο καλά συμφωνούν αυτά τα αποτελέσματα σε όλες τις δοκιμές ή τους κριτές.

Η ποσοστιαία συμφωνία είναι το απλούστερο μέτρο συμφωνίας: υπολογίζεται διαιρώντας τον αριθμό των περιπτώσεων στις οποίες οι κριτές συμφωνούν με τον συνολικό αριθμό αξιολογήσεων. Στο παρακάτω παράδειγμα, το ποσοστό συμφωνίας είναι $(50 + 30) / 100$ ή 0.80. Ένα σημαντικό μειονέκτημα της απλής ποσοστιαίας συμφωνίας είναι ότι μπορεί να επιτευχθεί υψηλός βαθμός συμφωνίας απλά τυχαία και έτσι είναι αδύνατο να συγκριθεί η ποσοστιαία συμφωνία μεταξύ διαφορετικών καταστάσεων όπου διαφέρει η κατανομή των δεδομένων.

Αυτό το κενό μπορεί να ξεπεραστεί χρησιμοποιώντας ένα άλλο κοινό μέτρο συμφωνίας που ονομάζεται Cohen's kappa ή απλά κάπα, το οποίο σχεδιάστηκε αρχικά για να συγκρίνει δύο δοκιμές. Το κάπα είναι προτιμότερο σε σχέση με το ποσοστό συμφωνίας, διότι διορθώνεται για συμφωνία λόγω τυχαίας (αν και οι στατιστικολόγοι διαφωνούν για το πόσο επιτυχής είναι αυτή η διόρθωση (δείτε την πλαϊνή γραμμή παρακάτω για μια σύντομη εισαγωγή στα θέματα). Το κάπα υπολογίζεται εύκολα με ταξινόμηση των αποκρίσεων σε συμμετρικό πλέγμα και πραγματοποίηση υπολογισμών όπως υποδεικνύεται στον Πίνακα 1-1. Αυτό το

υποθετικό παράδειγμα αφορά δύο εξετάσεις για την παρουσία (D +) ή την απουσία (D-) ασθένειας.

Πίνακας 1-1. Συμφωνία δύο ποσοστών για ένα διχοτόμο αποτέλεσμα

		Test 2		
		+	-	
Test 1	+	50	10	60
	-	10	30	40
		60	40	100

Τα τέσσερα στοιχεία που περιέχουν δεδομένα ταυτολογούνται συνήθως ως εξής:

	+	-
+	<i>A</i>	<i>b</i>
-	<i>C</i>	<i>d</i>

Τα κύτταρα *a* και *d* αντιπροσωπεύουν μια συμφωνία (*a* περιέχει τις περιπτώσεις που ταξινομούνται ως έχουσες την ασθένεια και από τις δύο δοκιμασίες, περιγράφει τις περιπτώσεις που ταξινομούνται ως μη ασθένειες και από τις δύο δοκιμές), ενώ τα κύτταρα *b* και *c* αντιπροσωπεύουν διαφωνία.

Ο τύπος για το κάπα είναι:

όπου p_o = παρατηρούμενη συμφωνία και p_e = αναμενόμενη συμφωνία.

$p_o = (a + d) / (a + b + c + d)$, δηλ. ο αριθμός των περιπτώσεων σε συμφωνία διαιρεμένος με τον συνολικό αριθμό των περιπτώσεων.

p_e = η αναμενόμενη συμφωνία, η οποία μπορεί να υπολογιστεί σε δύο στάδια. Πρώτα, για τα κελιά *a* και *d*, βρείτε τον αναμενόμενο αριθμό περιπτώσεων σε κάθε κελί, πολλαπλασιάζοντας τα σύνολα σειρών και σειρών και διαιρώντας με το

συνολικό αριθμό περιπτώσεων. Για ένα, αυτό είναι $(60 \times 60) / 100$ ή 36. για d είναι $(40 \times 40) / 100$ ή 16. Δεύτερον, βρείτε την αναμενόμενη συμφωνία προσθέτοντας τον αναμενόμενο αριθμό περιπτώσεων σε αυτά τα δύο κελιά και διαιρώντας με το συνολικό αριθμό των περιπτώσεων. Επομένως, η αναμενόμενη συμφωνία είναι:

$$pe = (36 + 16) / 100 = 0,52$$

Ως εκ τούτου, το κάπα μπορεί να υπολογιστεί ως:

$$\kappa = \frac{0.8 - 0.52}{1 - 0.52} = 0.583$$

Το κάπα έχει ένα εύρος -1-1: η τιμή θα είναι 0 αν παρατηρηθεί συμφωνία, ήταν η ίδια με συμφωνία τυχαίας και 1 αν όλες οι περιπτώσεις συμφωνούσαν. Δεν υπάρχουν απόλυτα πρότυπα βάσει των οποίων να κρίνεται μια συγκεκριμένη τιμή κάπα ως υψηλή ή χαμηλή. Ωστόσο, πολλοί ερευνητές χρησιμοποιούν τις οδηγίες που δημοσιεύονται από τους Landis και Koch (1977):

<0

Ελάχιστος

0-0,20

Μικρός

0,21-0,40

Δίκαιος

0,41-0,60

Μέτριος

0,61-0,81

Ουσιώδης

0,81-1,0

Σχεδόν τέλειο

Σημειώστε ότι η κάπα είναι πάντα μικρότερη ή ίση με τη συμφωνία ποσοστού, διότι διορθώνεται για τυχαία συμφωνία.

Για μια εναλλακτική άποψη του κάπα (που προορίζεται για πιο εξελιγμένους στατιστικούς), δείτε την πλαϊνή γραμμή παρακάτω.

ΑΝΤΙΜΕΤΩΠΙΣΗ ΣΤΟ ΚΑΠΠΑ

Η κάπα του Cohen είναι μια κοινά διδαχθείσα και ευρέως χρησιμοποιούμενη στατιστική, αλλά η εφαρμογή της δεν είναι χωρίς αμφιβολία. Ο Κάπα ορίζεται συνήθως ότι αντιπροσωπεύει συμφωνία πέρα από εκείνη που αναμένεται τυχαία ή απλά συμφωνία διορθωμένη για τυχαία. Έχει δύο χρήσεις: ως στατιστική δοκιμασία για να καθορίσει εάν δύο ομάδες αξιολογήσεων συμφωνούν συχνότερα από ό, τι θα περίμενε τυχαία (που είναι μια απόφαση ναι / όχι) και ως μέτρο του επιπέδου συμφωνίας (το οποίο εκφράζεται ως αριθμός μεταξύ 0 και 1).

Ενώ οι περισσότεροι ερευνητές δεν έχουν κανένα πρόβλημα με την πρώτη χρήση του κάπα, κάποιοι αντιτίθενται στη δεύτερη. Το πρόβλημα είναι ότι ο υπολογισμός συμφωνίας που αναμένεται τυχαία μεταξύ οποιωνδήποτε δύο οντοτήτων, όπως οι βαθμολογητές, βασίζεται στην υπόθεση ότι οι αξιολογήσεις είναι ανεξάρτητες, μια κατάσταση που συνήθως δεν πληρούται στην πράξη. Επειδή η κάπα χρησιμοποιείται συχνά για την ποσοτικοποίηση της συμφωνίας για πολλαπλά άτομα που βαθμολογούν την ίδια περίπτωση, είτε πρόκειται για συμπεριφορά ενός παιδιού στην τάξη είτε για ακτινογραφία θώρακα από άτομο που μπορεί να έχει φυματίωση, δεν υπάρχει κανένας λόγος να υποθέσουμε ότι οι αξιολογήσεις είναι ανεξάρτητες. Στην πραγματικότητα το αντίθετο. Οι επικρίσεις του κάπα, συμπεριλαμβανομένης και μιας μακράς βιβλιογραφίας σχετικών άρθρων, βρίσκονται στην ιστοσελίδα του John Uebersax, Ph.D., στη διεύθυνση <http://ourworld.compuserve.com/homepages/jsuebersax/kappa.htm>.

Εγκυρότητα

Η ισχύς αναφέρεται στο βαθμό στον οποίο μια κλίμακα ελέγχου ή βαθμολόγησης μετράει τι υποτίθεται ότι πρέπει να μετρήσει. Μερικοί ερευνητές ορίζουν την επικύρωση ως τη διαδικασία συγκέντρωσης αποδεικτικών στοιχείων που υποστηρίζουν τους τύπους συμπερασμάτων που πρόκειται να αντληθούν από τις εν λόγω μετρήσεις. Οι ερευνητές διαφωνούν σχετικά με το πόσες μορφές εγκυρότητας υπάρχουν και η επιστημονική συναίνεση ποικίλει με την πάροδο των ετών καθώς οι διαφορετικοί τύποι εγκυρότητας υπάγονται σε ένα μόνο τίτλο ένα χρόνο και στη συνέχεια διαχωρίζονται και αντιμετωπίζονται ως διακεκριμένοι. Για λόγους απλούστευσης, θα ακολουθήσουμε μια ευρέως αποδεκτή κατηγοριοποίηση της εγκυρότητας που αναγνωρίζει τέσσερις τύπους: την εγκυρότητα του περιεχομένου, την εγκυρότητα της δομής, την παράλληλη εγκυρότητα και την εγκυρότητα της πρόβλεψης, με την προσθήκη της εγκυρότητας του προσώπου, η οποία σχετίζεται στενά με την εγκυρότητα του περιεχομένου. Αυτά τα είδη εγκυρότητας συζητούνται περαιτέρω στο πλαίσιο του σχεδιασμού της έρευνας στο Κεφάλαιο 5.

Η εγκυρότητα του περιεχομένου αναφέρεται στο πόσο καλά η διαδικασία μέτρησης αντανακλά το σημαντικό περιεχόμενο του τομέα ενδιαφέροντος. Είναι ιδιαίτερα

σημαντικό όταν ο σκοπός της μέτρησης είναι να συνάγει συμπεράσματα σχετικά με ένα μεγαλύτερο πεδίο ενδιαφέροντος. Για παράδειγμα, οι δυνητικοί εργαζόμενοι που αναζητούν εργασία ως προγραμματιστές υπολογιστών μπορεί να κληθούν να ολοκληρώσουν μια εξέταση που απαιτεί από αυτούς να γράψουν και να ερμηνεύσουν προγράμματα στις γλώσσες που θα χρησιμοποιήσουν. Μόνο περιορισμένες ικανότητες περιεχομένου και προγραμματισμού μπορεί να συμπεριληφθούν σε μια τέτοια εξέταση, σε σχέση με το τι μπορεί πραγματικά να απαιτείται να είναι ο επαγγελματίας προγραμματιστής. Ωστόσο, εάν το υποσύνολο του περιεχομένου και των ικανοτήτων είναι καλά επιλεγμένο, το αποτέλεσμα μιας τέτοιας εξέτασης μπορεί να είναι μια καλή ένδειξη της ικανότητας του ατόμου να συνεισφέρει στην επιχείρηση ως προγραμματιστής.

Μια στενά συνδεδεμένη αντίληψη για την εγκυρότητα του περιεχομένου είναι γνωστή ως ισχύουσα. Ένα μέτρο με καλή εγκυρότητα του προσώπου εμφανίζεται, σε ένα μέλος του ευρύτερου κοινού ή σε ένα τυπικό πρόσωπο που μπορεί να αξιολογηθεί, να είναι μια δίκαιη αξιολόγηση των ποιοτικών μελετών. Για παράδειγμα, εάν οι μαθητές που κάνουν μια δοκιμή άλγεβρας στην τάξη αισθάνονται ότι οι ερωτήσεις αντανakλούν αυτό που έχουν σπουδάσει στην τάξη, τότε η δοκιμή έχει καλή εγκυρότητα του προσώπου. Η εγκυρότητα του προσώπου είναι σημαντική επειδή εάν τα άτομα που δοκιμάζουν αισθάνονται ότι ένα μέσο μέτρησης δεν είναι δίκαιο ή δεν μετρά αυτό που ισχυρίζεται ότι μετράει, μπορεί να μην επιθυμούν να συνεργαστούν και να καταβάλουν τις καλύτερες προσπάθειές τους και οι απαντήσεις τους μπορεί να μην αντικατοπτρίζουν πραγματικά τις απόψεις τους ή ικανότητες.

Η ταυτόχρονη εγκυρότητα αναφέρεται στο πόσο καλά μπορούν να χρησιμοποιηθούν τα συμπεράσματα που προκύπτουν από μία μέτρηση για την πρόβλεψη κάποιας άλλης συμπεριφοράς ή απόδοσης που μετράται ταυτόχρονα. Η προβλεπτική εγκυρότητα είναι παρόμοια, αλλά αφορά την ικανότητα να συνάγει συμπεράσματα σχετικά με κάποιο γεγονός στο μέλλον. Για παράδειγμα, αν η βαθμολογία του επιτεύγματος επιτυχίας σχετίζεται σε μεγάλο βαθμό με τις σύγχρονες σχολικές επιδόσεις ή με βαθμολογίες σε άλλες δοκιμές που χορηγούνται ταυτόχρονα, έχει υψηλή ταυτόχρονη εγκυρότητα. Αν σχετίζεται σε μεγάλο βαθμό με τις επιδόσεις του σχολείου ή με άλλες δοκιμές για αρκετά χρόνια στο μέλλον, έχει υψηλή εγκυρότητα.

Τριγωνισμός

Επειδή κάθε σύστημα μέτρησης έχει τα ελαττώματά του, οι ερευνητές συχνά χρησιμοποιούν αρκετές διαφορετικές μεθόδους για να μετρήσουν το ίδιο πράγμα. Για παράδειγμα, τα κολλέγια συνήθως χρησιμοποιούν πολλαπλούς τύπους πληροφοριών για να αξιολογήσουν τις σχολικές ικανότητες. Οι μετρήσεις που χρησιμοποιούνται για το σκοπό αυτό περιλαμβάνουν βαθμολογίες στα μαθήματα

SAT, δοκιμασίες, προσωπική δήλωση ή δοκίμιο και συστάσεις από δασκάλους. Με παρόμοιο τρόπο, οι αποφάσεις πρόσληψης σε μια εταιρεία γίνονται συνήθως μετά από εξέταση πολλών τύπων πληροφοριών, συμπεριλαμβανομένης αξιολόγησης της εργασιακής εμπειρίας κάθε αιτούντος, της εκπαίδευσης, της εντύπωσης κατά τη διάρκεια μιας συνέντευξης και ενδεχομένως ενός δείγματος εργασίας και μιας ή περισσότερων ικανοτήτων ή εξετάσεις προσωπικότητας.

Αυτή η διαδικασία συνδυασμού πληροφοριών από πολλαπλές πηγές για να φθάσουμε σε μια "αληθινή" ή τουλάχιστον ακριβέστερη τιμή ονομάζεται θωράκιση, μια χαλαρή αναλογία με τη διαδικασία στη γεωμετρία της εύρεσης της θέσης ενός σημείου μετρώντας τις γωνίες και τις πλευρές του τριγώνου που σχηματίζεται από το άγνωστο σημείο και δύο άλλες γνωστές θέσεις. Η λειτουργική ιδέα στην τριγωνισμό είναι ότι μία μόνο μέτρηση μιας έννοιας μπορεί να περιέχει υπερβολικά μεγάλο σφάλμα (γνωστών ή άγνωστων τύπων) είτε αξιόπιστο είτε έγκυρο από μόνο του, αλλά συνδυάζοντας πληροφορίες από διάφορους τύπους μπορούμε να καταλήξουμε σε μια αποδεκτή μέτρηση της άγνωστης ποσότητας. Αναμένουμε ότι κάθε μέτρηση περιέχει σφάλμα, αλλά ελπίζουμε ότι δεν πρόκειται για τον ίδιο τύπο σφάλματος, έτσι ώστε μέσω πολλαπλών μετρήσεων να έχουμε μια εύλογη εκτίμηση της ποσότητας που είναι η εστίασή μας.

Η θέσπιση μιας μεθόδου τριγωνισμού δεν είναι απλή υπόθεση. Μία ιστορική προσπάθεια για να γίνει αυτό είναι η μήτρα πολλαπλών τάσεων (MTMM) που αναπτύχθηκε από τους Campbell και Fiske (1959). Η ιδιαίτερη ανησυχία τους ήταν να χωριστεί το τμήμα της μέτρησης λόγω της ποιότητας του ενδιαφέροντος από το μέρος αυτό, λόγω της χρησιμοποιούμενης μεθόδου μέτρησης. Αν και η συγκεκριμένη μεθοδολογία τους χρησιμοποιείται σήμερα λιγότερο και η πλήρης συζήτηση της τεχνικής MTMM είναι πέρα από το πεδίο ενός αρχικού κειμένου, η έννοια παραμένει χρήσιμη ως παράδειγμα ενός τρόπου να σκεφτούμε το σφάλμα μέτρησης και την εγκυρότητα.

Το MTMM είναι ένας πίνακας συσχετισμών μεταξύ των μετρήσεων πολλών εννοιών (τα "γνωρίσματα") που μετρούνται με διάφορους τρόπους ("μέθοδοι"). Ιδανικά, θα χρησιμοποιηθούν οι ίδιες αρκετές μέθοδοι για κάθε χαρακτηριστικό. Εντός αυτού του πίνακα, αναμένουμε ότι διαφορετικά μέτρα του ίδιου χαρακτηριστικού συσχετίζονται σε μεγάλο βαθμό: για παράδειγμα, οι βαθμολογίες που μετρούν τη νοημοσύνη με διαφορετικές μεθόδους, όπως μια δοκιμασία με μολύβι και χαρτί, πρακτική επίλυση προβλημάτων και μια δομημένη συνέντευξη, πρέπει να συσχετιστούν. Με την ίδια λογική, τα αποτελέσματα που αντανακλούν διαφορετικά κατασκευάσματα, τα οποία μετρούνται με τον ίδιο τρόπο δεν πρέπει να σχετίζονται ιδιαίτερα: για παράδειγμα, η νοημοσύνη, η νοημοσύνη και η κοινωνικότητα, όπως μετράται με έρευνα σε μολύβι και χαρτί, δεν πρέπει να συσχετίζονται σε μεγάλο βαθμό.

Ένδειξη μέτρησης

Η εξέταση της μεροληψίας μέτρησης είναι σημαντική σε κάθε τομέα, αλλά αποτελεί ιδιαίτερη μέριμνα στις ανθρωπιστικές επιστήμες. Έχουν εντοπιστεί και οριστεί πολλοί ειδικοί τύποι προκατάληψης: δεν θα προσπαθήσουμε να τα ονομάσουμε όλα εδώ, αλλά θα συζητήσουμε μερικούς κοινούς τύπους. Τα περισσότερα εγχειρίδια σχεδιασμού της έρευνας αντιμετωπίζουν αυτό το θέμα με μεγάλη λεπτομέρεια και μπορούν να συμβουλεύουν για περαιτέρω συζήτηση αυτού του θέματος. Το πιο σημαντικό σημείο είναι ότι ο ερευνητής πρέπει να είναι σε εγρήγορση για τη δυνατότητα μεροληψίας στη μελέτη του, επειδή η αποτυχία να εξετάσει και να ασχοληθεί με ζητήματα που σχετίζονται με την προκατάληψη μπορεί να ακυρώσει τα αποτελέσματα μιας κατά τα άλλα υποδειγματικής μελέτης.

Η Bias μπορεί να εισέλθει σε μελέτες με δύο βασικούς τρόπους: κατά την επιλογή και τη διατήρηση των αντικειμένων μελέτης, ή στον τρόπο συλλογής πληροφοριών σχετικά με τα αντικείμενα. Και στις δύο περιπτώσεις, το οριστικό χαρακτηριστικό της μεροληψίας είναι ότι αποτελεί πηγή συστηματικού και όχι τυχαίου σφάλματος. Το αποτέλεσμα της μεροληψίας είναι ότι οι πληροφορίες που αναλύονται σε μια μελέτη είναι λανθασμένες με συστηματικό τρόπο, γεγονός που μπορεί να οδηγήσει σε ψευδή συμπεράσματα παρά την εφαρμογή ορθών στατιστικών διαδικασιών και τεχνικών. Τα επόμενα δύο τμήματα συζητούν μερικά από τα πιο κοινά είδη μεροληψίας, τα οποία οργανώνονται σε δύο μεγάλες κατηγορίες: μεροληψία στην επιλογή και διατήρηση δείγματος και προκατάληψη που προκύπτει από τη συλλογή ή την καταγραφή πληροφοριών διαφορετικά για διαφορετικά θέματα.

Εξέλιξη της επιλογής και διατήρησης του δείγματος

Οι περισσότερες μελέτες λαμβάνουν χώρα σε δείγματα ατόμων, είτε πρόκειται για ασθενείς με λευχαιμία είτε για μικροσκοπικά είδη που παράγονται από τοπικό εργοστάσιο, διότι θα ήταν απαγορευτικά δαπανηρό αν όχι αδύνατο να μελετηθεί ολόκληρος ο πληθυσμός που μας ενδιαφέρει. Το δείγμα πρέπει να είναι μια καλή αναπαράσταση του πληθυσμού μελέτης (του πληθυσμού στον οποίο πρόκειται να εφαρμοστούν τα αποτελέσματα), προκειμένου ο ερευνητής να είναι άνετος χρησιμοποιώντας τα αποτελέσματα του δείγματος για να περιγράψει τον πληθυσμό. Αν το δείγμα είναι προκατειλημμένο, δηλαδή με κάποιο συστηματικό τρόπο δεν είναι αντιπροσωπευτικό του πληθυσμού της μελέτης, τα συμπεράσματα που προκύπτουν από το δείγμα της μελέτης ενδέχεται να μην ισχύουν για τον πληθυσμό της μελέτης.

Η μεροληψία επιλογής υπάρχει εάν ορισμένα άτομα είναι πιθανότερο να επιλέξουν το δείγμα της μελέτης. Αυτός ο όρος συνήθως δεσμεύεται για προκατάληψη που συμβαίνει λόγω της διαδικασίας δειγματοληψίας. Για παράδειγμα, οι τηλεφωνικές έρευνες που πραγματοποιήθηκαν με τη χρήση αριθμών από τους δημοσιευμένους

καταλόγους, αφαιρούν ακούσια από την ομάδα δυνητικών ερωτηθέντων άτομα με αδημοσίευτους αριθμούς ή που έχουν αλλάξει τηλεφωνικούς αριθμούς από τότε που δημοσιεύθηκε ο κατάλογος. Οι τεχνικές τυχαίας ψηφιοποίησης (RDD) ξεπερνούν αυτά τα προβλήματα, αλλά εξακολουθούν να μην περιλαμβάνουν τους ανθρώπους που ζουν σε νοικοκυριά χωρίς τηλέφωνα ή που έχουν μόνο ένα κινητό τηλέφωνο. Αυτό είναι ένα πρόβλημα για μια ερευνητική μελέτη, εάν οι αποκλεισμένοι άνθρωποι διαφέρουν συστηματικά σε ένα χαρακτηριστικό ενδιαφέροντος και επειδή είναι πολύ πιθανό να διαφέρουν, αυτό το ζήτημα πρέπει να αντιμετωπιστεί από οποιονδήποτε πραγματοποιεί τηλεφωνικές έρευνες. Για παράδειγμα, οι άνθρωποι που ζουν σε νοικοκυριά που δεν διαθέτουν τηλεφωνική υπηρεσία τείνουν να είναι φτωχότεροι από όσους έχουν τηλέφωνο και οι άνθρωποι που έχουν μόνο ένα κινητό τηλέφωνο (δηλαδή όχι "χερσαία γραμμή") τείνουν να είναι νεότεροι από όσους έχουν συμβατική τηλεφωνική υπηρεσία .

Η εθελοντική προκατάληψη αναφέρεται στο γεγονός ότι οι άνθρωποι που θέλουν να συμμετέχουν συνήθως δεν είναι αντιπροσωπευτικοί του πληθυσμού στο σύνολό του. Για το λόγο αυτό, τα αποτελέσματα δειγμάτων, όπως οι τηλεφωνικές δημοσκοπήσεις που εμφανίζονται σε ορισμένα τηλεοπτικά προγράμματα, δεν είναι χρήσιμα για επιστημονικούς σκοπούς, εκτός εάν ο πληθυσμός που ενδιαφέρει είναι άτομα που εθελοντικά συμμετέχουν σε αυτές τις δημοσκοπήσεις (και όχι στο ευρύ κοινό). Μπορούν να λειτουργούν πολλαπλά στρώματα: για να ανταποκριθεί, το άτομο πρέπει να παρακολουθεί το εν λόγω τηλεοπτικό πρόγραμμα, το οποίο σημαίνει πιθανώς ότι είναι στο σπίτι όταν απαντά (ως εκ τούτου οι απαντήσεις στις δημοσκοπήσεις που διεξάγονται κατά τη διάρκεια της κανονικής εργάσιμης ημέρας μπορούν να προσελκύσουν κοινό κυρίως οι συνταξιούχοι, οι νοικοκυρές και οι άνεργοι), έχουν εύκολη πρόσβαση σε ένα τηλέφωνο και έχουν τα χαρακτηριστικά γνωρίσματα της προσωπικότητας που θα τους επηρεάσουν για να πάρουν το τηλέφωνό τους και να καλέσουν έναν αριθμό που βλέπουν στην οθόνη της τηλεόρασης.

Η απόκλιση μη ανταπόκρισης αναφέρεται στην αλληλεπικαλυπτόμενη πλευρά της εθελοντικής προκατάληψης, όπως οι άνθρωποι που εθελοντικά συμμετέχουν σε μια μελέτη, είναι πιθανό να διαφέρουν συστηματικά από εκείνους που δεν προσφέρονται εθελοντικά. Πιθανότατα γνωρίζετε τους ανθρώπους που αρνούνται να συμμετάσχουν σε κάθε είδους τηλεφωνική έρευνα. Φαίνεται ότι είναι μια τυχαία επιλογή από τον γενικό πληθυσμό; Πιθανόν όχι. Η έρευνα της Υγείας διαπίστωσε όχι μόνο διαφορετικά ποσοστά ανταπόκρισης για τους καναδούς έναντι των Αμερικανών, αλλά επίσης βρήκε πόλωση μη ανταπόκρισης για σχεδόν όλες τις σημαντικές καταστάσεις υγείας και μέτρα πρόσβασης στην υγειονομική περίθαλψη (τα αποτελέσματα συνοψίζονται στο http://www.allacademic.com/meta/p_mla_apa_research_citation/0/1/6/8/4/p16845_index.html).

Η απώλεια παρακολούθησης μπορεί να προκαλέσει προκατάληψη σε οποιαδήποτε διαχρονική μελέτη (μια μελέτη όπου τα δεδομένα συλλέγονται σε μια χρονική περίοδο). Η απώλεια θεμάτων κατά τη διάρκεια μιας μακροχρόνιας μελέτης είναι σχεδόν αναπόφευκτη, αλλά το πραγματικό πρόβλημα έρχεται όταν τα υποκείμενα δεν αποχωρούν τυχαία αλλά για λόγους που σχετίζονται με το σκοπό της μελέτης. Ας υποθέσουμε ότι συγκρίνουμε δύο ιατρικές θεραπείες για μια χρόνια ασθένεια με τη διεξαγωγή μιας κλινικής δοκιμής στην οποία τα άτομα κατανέμονται τυχαία σε μία από τις διάφορες ομάδες θεραπείας και ακολουθούνται για πέντε χρόνια για να δουν πώς εξελίσσεται η ασθένειά τους. Χάρη στη χρήση ενός τυχαίου σχεδιασμού, ξεκινάμε με μια τέλεια ισορροπημένη ομάδα θεμάτων. Ωστόσο, με την πάροδο του χρόνου τα άτομα για τα οποία η καθορισμένη θεραπεία δεν αποδεικνύεται αποτελεσματική θα είναι πιο πιθανό να εγκαταλείψουν τη μελέτη, ενδεχομένως να αναζητήσουν θεραπεία αλλού, οδηγώντας σε μεροληψία. Το τελικό δείγμα των αντικειμένων που θα αναλύσουμε θα αποτελείται από εκείνους που παραμένουν στη δοκιμή μέχρι το συμπέρασμά του και εάν η απώλεια για παρακολούθηση δεν ήταν τυχαία, το δείγμα που θα αναλύσουμε δεν θα είναι πλέον το τυχαίο δείγμα που ξεκινήσαμε. Αντίθετα, εάν η απόρριψη σχετίζεται με την αναποτελεσματικότητα της θεραπείας, η τελική ομάδα ασθενών θα είναι μεροληπτική υπέρ αυτών που ανταποκρίθηκαν αποτελεσματικά στην θεραπεία που τους αποδόθηκε.

Πληροφορίες Bias

Ακόμη και αν το τέλει δείγμα επιλεγεί και διατηρηθεί, η προκατάληψη μπορεί να εισέλθει στη μελέτη μέσω των μεθόδων που χρησιμοποιούνται για τη συλλογή και την καταγραφή δεδομένων. Αυτός ο τύπος μεροληψίας συχνά αποκαλείται *biasbecause* επηρεάζει την εγκυρότητα των πληροφοριών επί των οποίων βασίζεται η μελέτη, η οποία με τη σειρά της μπορεί να ακυρώσει τα αποτελέσματα της μελέτης.

Όταν τα δεδομένα συλλέγονται με προσωπικές ή τηλεφωνικές συνεντεύξεις, υπάρχει κοινωνική σχέση μεταξύ του ερευνητή και του θέματος για την πορεία της συνέντευξης. Αυτή η σχέση μπορεί να επηρεάσει δυσμενώς την ποιότητα των δεδομένων που συλλέγονται. Όταν η προκατάληψη εισάγεται στα δεδομένα που συλλέγονται λόγω της στάσης ή της συμπεριφοράς του ερωτηθέντος, αυτό είναι γνωστό ως προκατάληψη από τον συνεντευξιαζόμενο. Αυτός ο τύπος μεροληψίας μπορεί να δημιουργηθεί ακούσια όταν ο ερευνητής γνωρίζει τον σκοπό της μελέτης ή το καθεστώς των ατόμων που έχουν ερωτηθεί: για παράδειγμα, οι ερευνητές μπορούν να ζητήσουν περισσότερες ερωτήσεις για να ενθαρρύνουν το υποκείμενο να ανακαλεί τοξικές χημικές εκθέσεις αν γνωρίζουν ότι το θέμα είναι που πάσχουν από σπάνιο τύπο καρκίνου που σχετίζεται με χημική έκθεση. Μπορεί επίσης να δημιουργηθεί προκατάληψη από τους ερωτηθέντες εάν οι συνεντευξιαζόμενοι

εμφανίζουν προσωπικές συμπεριφορές ή απόψεις που υποδηλώνουν ότι δεν αποδέχονται τις συμπεριφορές που μελετώνται, όπως είναι η αταξία ή η χρήση ναρκωτικών, καθιστώντας τα άτομα λιγότερο πιθανό να αναφέρουν αυτές τις συμπεριφορές.

Η υπενθύμιση αναφέρεται στο γεγονός ότι άτομα με εμπειρίες στη ζωή, όπως σοβαρές ασθένειες ή τραυματισμοί, είναι πιο πιθανό να θυμούνται γεγονότα που πιστεύουν ότι σχετίζονται με την εμπειρία. Για παράδειγμα, οι γυναίκες που έχουν υποστεί αποβολή μπορεί να έχουν περάσει πολύ χρόνο εξετάζοντας τις αναμνήσεις τους για εκθέσεις ή περιστατικά που πιστεύουν ότι θα μπορούσαν να προκαλέσουν την αποβολή. Οι γυναίκες που είχαν κανονική γέννα ίσως είχαν παρόμοια έκθεση, αλλά δεν τους έδωσαν περαιτέρω σκέψη και συνεπώς δεν θα τις ανακαλέσουν όταν τους ζητηθεί σε μια έρευνα.

Η προκατάληψη ανίχνευσης αναφέρεται στο γεγονός ότι ορισμένα χαρακτηριστικά μπορεί να είναι πιο πιθανό να ανιχνευθούν ή να αναφερθούν σε μερικούς ανθρώπους απ' ό,τι σε άλλα. Για παράδειγμα, οι αθλητές σε ορισμένα αθλήματα υπόκεινται σε τακτικούς ελέγχους για φάρμακα που βελτιώνουν την απόδοση και τα αποτελέσματα των δοκιμών δημοσιοποιούνται. Οι κολυμβητές παγκόσμιας κλάσης εξετάζονται τακτικά για αναβολικά στεροειδή, για παράδειγμα, και οι θετικές δοκιμές καταγράφονται επίσημα και συχνά απελευθερώνονται και στα μέσα ενημέρωσης. Οι αθλητές που ανταγωνίζονται σε χαμηλότερο επίπεδο ή σε άλλα αθλήματα μπορεί να χρησιμοποιούν τα ίδια φάρμακα, αλλά επειδή δεν ελέγχονται τακτικά, ή επειδή τα αποτελέσματα των δοκιμών δεν έχουν αναφερθεί δημοσίως, δεν υπάρχουν στοιχεία για τη χρήση ναρκωτικών ουσιών. Θα ήταν λανθασμένο να υποθέσουμε, για παράδειγμα, ότι επειδή η χρήση αναβολικών στεροειδών είναι υψηλότερη στην κολύμβηση από ό,τι στο μπέιζμπολ, το πραγματικό ποσοστό χρήσης στεροειδών είναι υψηλότερο στη κολύμβηση από ό,τι στο μπέιζμπολ. Η φαινομενική διαφορά στα αποτελέσματα μπορεί να οφείλεται σε πιο επιθετικές δοκιμές εκ μέρους των υπαλλήλων κολύμβησης και σε μεγαλύτερη δημοσιότητα των αποτελεσμάτων των δοκιμών.

Η προκατάληψη της κοινωνικής επιθυμίας προκαλείται από την επιθυμία των ανθρώπων να παρουσιαστούν σε ευνοϊκό κλίμα. Αυτό συχνά τους παρακινεί να δώσουν απαντήσεις που πιστεύουν ότι θα ευχαριστήσουν το άτομο που θέτει την ερώτηση. Αυτός ο τύπος μεροληψίας μπορεί να λειτουργήσει ακόμα και αν ο ερωτώμενος δεν υπάρχει στην πραγματικότητα, για παράδειγμα όταν τα υποκείμενα ολοκληρώνουν μια έρευνα με μολύβι και χαρτί. Αυτό είναι ένα ιδιαίτερο πρόβλημα στις έρευνες που ρωτούν για συμπεριφορές ή συμπεριφορές που υπόκεινται σε κοινωνική απόρριψη, όπως η εγκληματική συμπεριφορά που θεωρείται ενοχλητική. Η κοινωνική επιθυμητή προκατάληψη μπορεί επίσης να

επηρεάσει τις απαντήσεις σε έρευνες όπου οι ερωτήσεις τίθενται με τέτοιο τρόπο ώστε να υποδηλώνουν ποια είναι η "σωστή" απάντηση.

Εργασία

Ακολουθεί μια επισκόπηση των θεμάτων που καλύπτονται σε αυτό το κεφάλαιο.

Πρόβλημα: Δεδομένης της κατανομής των δεδομένων στον παρακάτω πίνακα, υπολογίστε το ποσοστό συμφωνίας, τις αναμενόμενες τιμές για τα κύτταρα a και d και το κάπα για τον κριτή 1 και τον κριτή 2.

Ritter 2

+ -

Ritter 1 + 70 15 85

- 30 25 55

100 40 140

Λύση

Συμφωνία ποσοστού = $(70 + 25) / 140 = 0,679$

Αναμενόμενες τιμές:

60.7

15.7

a: $(85 \times 100) / 140 = 60,7$

d: $(55 \times 40) / 140 = 15,7$

ρ_o = παρατηρούμενη συμφωνία = $(70 + 25) / 140 = 0.679$

ρ_e = αναμενόμενη συμφωνία = $(60,7 + 15,7) / 140 = 0,546$

$$\kappa = \frac{0.679 - 0.546}{1 - 0.546} = \frac{0.133}{0.454} = 0.293$$

Η ΚΛΙΜΑΚΑ ΤΟΥ ΛΙΚΕΡΤ

Η κλίμακα Likert μπορεί να είναι ο πιο συνηθισμένος τύπος κλίμακας αξιολόγησης που χρησιμοποιείται στην έρευνα για τα ανθρώπινα υποκείμενα. Αυτός ο τύπος κλίμακας περιγράφηκε για πρώτη φορά το 1932 από τον Rensis Likert (1903-1981), οργανωτικό ψυχολόγο ο οποίος διετέλεσε διευθυντής του Ινστιτούτου Κοινωνικών

Ερευνών του Πανεπιστημίου του Μίτσιγκαν από το 1946 έως το 1970. Οι ερωτήσεις που χρησιμοποιούν την κλίμακα Likert παρουσιάζουν συνήθως μια δήλωση και θέματα που καλούνται να επιλέξουν την ανταπόκρισή τους σε μια ταξινομημένη, μονόδρομη σειρά επιλογών (συνήθως πέντε, αλλά μερικές φορές επτά ή εννέα). Παρακάτω είναι ένα παράδειγμα.

Οι Ηνωμένες Πολιτείες θα πρέπει να υιοθετήσουν ένα εθνικό σύστημα ασφάλισης υγείας.

1. Συμφωνώ απόλυτα
2. Συμφωνώ
3. Ούτε συμφωνείτε ούτε διαφωνείτε
4. Διαφωνώ
5. Διαφωνώ έντονα

Μερικές φορές παρέχεται ένας άρτος αριθμός απαντήσεων, έτσι ώστε να μην υπάρχει ουδέτερη μεσαία επιλογή: αυτή ονομάζεται μέθοδος "αναγκαστικής επιλογής" επειδή ο ερωτώμενος αναγκάζεται να κάνει την επιλογή να συμφωνήσει ή να διαφωνήσει με τη δήλωση. Συχνά η σειρά των απαντήσεων αλλάζει μέσα σε ένα ερωτηματολόγιο έτσι 1 = Σοβαρά διαφωνώ και 5 = Συμφωνώ απόλυτα, για να ανιχνεύσουμε αν οι άνθρωποι επιλέγουν αυτόματα την πρώτη ή την τελευταία επιλογή χωρίς να διαβάσουν τα στοιχεία.

Τα στοιχεία που συλλέγονται από τα στοιχεία του Likert παρόλο που είναι τακτικά δεν υπάρχει λόγος να πιστεύουμε ότι υπάρχουν ίσα διαστήματα μεταξύ τους. Για παράδειγμα, δεν έχουμε τρόπο να γνωρίζουμε εάν η απόσταση μεταξύ των "Συμφωνώ απόλυτα" και "Συμφωνώ" είναι ίδια με την απόσταση μεταξύ του "Συμφωνώ" και "Ούτε συμφωνώ ούτε διαφωνώ".

DEWEY DEFEATS TRUMAN

Ορισμένες προεδρικές εκλογές των Ηνωμένων Πολιτειών παρουσίασαν ανακριβείς προβλέψεις βάσει προκατειλημμένων δειγμάτων. Είναι πάντα χιουμοριστικό να βλέπεις μια αξιολογία δημοσίευση ή μια οργάνωση να είναι εντελώς λανθασμένη, αλλά τα περιστατικά αυτά χρησιμεύουν επίσης ως προειδοποιητική ιστορία για το τι μπορεί να συμβεί όταν οι στατιστικές που διεξάγονται σε μη αντιπροσωπευτικό δείγμα υποτίθεται ότι ισχύουν για τον γενικό πληθυσμό.

Το 1936, το περιοδικό Literary Digest, το οποίο είχε προβλέψει σωστά τον νικητή των προεδρικών εκλογών το 1916, 1920, 1924, 1928 και 1932, πρόβλεψε ότι ο Δημοκρατικός Alf Landon θα νικήσει τον Δημοκρατικό Φράνκλιν Ρούσβελτ. Ωστόσο, η ιστορία δείχνει ότι ο Ρούσβελτ κέρδισε τις εκλογές του 1936. Το πρόβλημα με την

πρόβλεψη του Literary Digest ήταν ότι αν και βασίστηκε σε ένα μεγάλο δείγμα (πάνω από 2.3 εκατομμύρια ερωτηθέντες από τα 10 εκατομμύρια που κλήθηκαν να συμμετάσχουν), το δείγμα ήταν προκατειλημμένο επειδή αποτελούταν από άτομα που είχαν αυτοκίνητα ή τηλέφωνα ή είχαν εγγραφεί στο λογοτεχνικό περιοδικό. Το 1936, τέτοια άτομα τείνουν να είναι πλουσιότερα από τον γενικό πληθυσμό, και επίσης πιο πιθανό να είναι Ρεπουμπλικάνοι. Επειδή ήταν απαραίτητο να επιστρέψει μια κάρτα για να συμμετάσχει στην ψηφοφορία, το δείγμα Literary Digest υποβλήθηκε και σε εθελοντική προκατάληψη.

Το 1948, κάθε μεγάλη δημοσκόπηση πρόβλεπε ότι ο Ρεπουμπλικάνος Τόμας Ντέιβιθα νικήσει τον Δημοκρατικό Χάρι Σ. Τρούμαν. Αν και οι τεχνικές ψηφοφορίας βελτιώθηκαν από το 1936, αρκετές πηγές προκατάληψης εξακολουθούσαν να υπάρχουν στις δημοσκοπήσεις, γεγονός που οδήγησε σε αυτήν την ανακριβή πρόβλεψη. Ένα πρόβλημα ήταν ότι οι τηλεφωνικές έρευνες χρησιμοποιήθηκαν χωρίς στατιστική διόρθωση για το γεγονός ότι η τηλεφωνική ιδιοκτησία ήταν πολύ πιο κοινή μεταξύ των εύπορων, οι οποίοι ήταν επίσης πιο πιθανό να υποστηρίξουν τον Dewey. Ένας άλλος παράγοντας ήταν ότι υπήρχαν μεγάλοι αριθμοί αναποφάσιστων ψηφοφόρων στις ημέρες που προηγήθηκαν της εκλογής και καμία από τις δημοσκοπήσεις δεν είχε μια καλή μέθοδο για να προβλέψει ποιον θα ψηφίσουν τελικά αυτά τα άτομα. Ένα τρίτο πρόβλημα που αφορούσε άμεσα το φιάσκο του Chicago Tribune ήταν ότι η υποστήριξη του Dewey ήταν ισχυρότερη στην Ανατολή και λόγω των διαφορών στις ζώνες ώρας τα αποτελέσματα των εκλογών αναφέρθηκαν πρώτα. Το Tribune αποφάσισε να εκτυπώσει έγγραφα βάσει αυτών των πρώιμων αποτελεσμάτων, τα οποία βασίστηκαν σε ένα προκατειλημμένο δείγμα αποτελεσμάτων από τα ανατολικά κράτη. Αυτό που δεν προσδοκούσε το Tribune ήταν ότι ο Τρούμαν θα υποστηριζόταν από πολλά δυτικά κράτη, συμπεριλαμβανομένης της Καλιφόρνιας, συγκεντρώνοντας έτσι επαρκείς εκλογικές ψήφους για να κερδίσει τις εκλογές.