

Μέθοδοι Στατιστικής Επεξεργασίας των βαθμολογιών των υποψηφίων που έλαβαν μέρος στις Γραπτές Εξετάσεις 2021 για Εγγραφή και Κατάταξη στους Πίνακες Διορισίμων

1. Εισαγωγή

Το 2021 πραγματοποιήθηκαν Εξετάσεις για Εγγραφή και Κατάταξη στους Πίνακες Διορισίμων για 27 ειδικότητες. Όλες οι ειδικότητες που εξετάστηκαν το 2021 είχαν, επίσης, εξεταστεί το 2019, ενώ 20 ειδικότητες είχαν εξεταστεί και το 2017.

Σύμφωνα με τις πρόνοιες της σχετικής νομοθεσίας, οι βαθμολογίες των υποψηφίων έτυχαν στατιστικής επεξεργασίας, η οποία περιγράφεται πιο κάτω:

2. Περιγραφή της μεθοδολογίας για τη συγκρισιμότητα των βαθμολογιών των εξετάσεων των περιόδων 2017, 2019 και 2021

Προκειμένου να διασφαλιστεί η συγκρισιμότητα των βαθμολογιών των εξεταστικών περιόδων 2017, 2019 και 2021, για κάθε ειδικότητα εφαρμόστηκε στατιστική επεξεργασία των αντίστοιχων βαθμολογιών σε δύο στάδια.

(α) Στο πρώτο στάδιο, για κάθε ειδικότητα, υπολογίστηκε ο βαθμός δυσκολίας της εξέτασης του 2021 σε σχέση με τον βαθμό δυσκολίας της αντίστοιχης εξέτασης του 2019, ο οποίος είναι ήδη σταθμισμένος σε σχέση με το βαθμό δυσκολίας του μαθήματος στην εξέταση του 2017 που αποτελεί και το σημείο αναφοράς. Αυτό το στάδιο ήταν απαραίτητο, προκειμένου να εξευρεθεί η βαθμολογία της εξέτασης του 2021 που αντιστοιχεί με το 50% της βαθμολογίας του 2017, που αποτελούσε και το βασικό κριτήριο επιτυχίας/αποτυχίας στην εξέταση. Στην περίπτωση ειδικοτήτων που εξετάστηκαν μόνο το 2021 και το 2019, ο βαθμός δυσκολίας της εξέτασης του 2019 θεωρείται το βασικό κριτήριο επιτυχίας/αποτυχίας στην εξέταση.

(β) Στο δεύτερο στάδιο, οι βαθμολογίες μετασηματίστηκαν σύμφωνα με τη μέθοδο που χρησιμοποιήθηκε το 2017, προκειμένου ο αριθμός μορίων των υποψηφίων του 2017, του 2019 και του 2021 να είναι άμεσα συγκρίσιμος.

Στις πιο κάτω παραγράφους, περιγράφονται πιο αναλυτικά τα δύο στάδια στατιστικής επεξεργασίας.

Στάδιο 1: Για κάθε ειδικότητα υπολογίστηκε η διαφορά του βαθμού δυσκολίας των εξετάσεων των εξεταστικών περιόδων 2019 και 2021. Για κάθε ειδικότητα υπολογίστηκε ένας δείκτης D_{21} όπου:

Στάδιο 1^α:

Αναπροσαρμοσμένος βαθμός του 2021 = Αρχική βαθμολογία εξέτασης του 2021 + D_{21} .

Νοείται ότι το D_{21} έχει θετική τιμή αν η εξέταση το 2021 ήταν πιο δύσκολη από αυτήν του 2019. Το D_{21} παίρνει αρνητική τιμή, αν η εξέταση το 2021 ήταν ευκολότερη από αυτήν του 2019. Στην περίπτωση, βέβαια, που οι εξετάσεις των δύο περιόδων ήταν της ίδιας δυσκολίας, η τιμή του D_{21} είναι 0.

Υπενθυμίζεται, ωστόσο, ότι για κάθε ειδικότητα έχει ήδη υπολογιστεί η διαφορά του βαθμού δυσκολίας των εξετάσεων των εξεταστικών περιόδων 2017 και 2019. Για κάθε ειδικότητα, υπολογίστηκε ένα δείκτης D_{19} (δείτε το σχετικό αρχείο «Μέθοδοι στατιστικής

επεξεργασίας» του 2019: <https://archeia.moec.gov.cy/ed/231/nsde-2019-statistiki-epexergasia-vathmologion.pdf>) όπου:

Αναπροσαρμοσμένος βαθμός 2019 = Αρχική βαθμολογία εξέτασης του 2019 + D_{19} .

Ως εκ τούτου, προκειμένου να σταθμίσουμε «αλυσιδωτά» τις βαθμολογίες του 2021 με αυτές του 2017, μέσω των εξετάσεων του 2019, ο τελικός αναπροσαρμοσμένος βαθμός του 2021, για το Στάδιο 1, υπολογίζεται ως εξής:

Στάδιο 1^β:

Αναπροσαρμοσμένος βαθμός του 2021 = Αρχική βαθμολογία εξέτασης του 2021 + D_{21} + D_{19} .

Στάδιο 2: Για κάθε ειδικότητα οι αναπροσαρμοσμένες βαθμολογίες που προέκυψαν από το Στάδιο 1 μετασχηματίστηκαν με τη μέθοδο που είχε χρησιμοποιηθεί το 2017. Η μέθοδος περιγράφεται αναλυτικά στο σχετικό κείμενο που είχε δημοσιευτεί στις 30.3.2018 και το οποίο είναι διαθέσιμο στον ακόλουθο ιστότοπο της Υπηρεσίας Εξετάσεων:

http://archeia.moec.gov.cy/ed/117/2018_03_30_statistiki_epexergasia_vathmologion.pdf

3. Μέθοδοι υπολογισμού του δείκτη D_{21} και D_{19}

Ο υπολογισμός του βαθμού δυσκολίας μιας εξέτασης σε σχέση με το βαθμό δυσκολίας μιας άλλης εξέτασης που προηγήθηκε χρονικά είναι διαδικασία πολύπλοκη. Συνήθως, η διεθνής βιβλιογραφία αλλά και η διεθνής πρακτική απαιτούν την αξιοποίηση «κοινών ερωτήσεων». Αυτό σημαίνει ότι κάποιες από τις ερωτήσεις των εξετάσεων του 2017 ή/και του 2019 θα έπρεπε να επαναχρησιμοποιηθούν και στις εξετάσεις του 2021. Η διαδοχική στάθμιση της δυσκολίας εξετάσεων σειριακά από χρόνο σε χρόνο (π.χ. του 2021 με το 2019 και του 2019 με το 2017) είναι γνωστή στη βιβλιογραφία ως «αλυσιδωτή» («chain») και ο αναγνώστης παραπέμπεται στον Guo (2010) για περισσότερες λεπτομέρειες. Αν και σε κάποιες χώρες τέτοιου είδους σταθμίσεις με κοινές ερωτήσεις είναι υπόθεση ρουτίνας, η αξιοποίηση «κοινών ερωτήσεων» σε δημόσιες εξετάσεις είναι εκτός συζήτησης στην Κυπριακή πραγματικότητα. Για περισσότερες τεχνικές λεπτομέρειες για τη μέθοδο των «κοινών ερωτήσεων» παραπέμπουμε στους Kolen and Brennan (2014).

Μια άλλη συνηθισμένη πρακτική στο εξωτερικό, προκειμένου να υπολογιστεί ο βαθμός δυσκολίας μιας εξέτασης σε σχέση με το βαθμό δυσκολίας μιας άλλης εξέτασης που προηγήθηκε είναι η αξιοποίηση των αποτελεσμάτων των «κοινών υποψηφίων». Αυτό πρακτικά σημαίνει ότι για τον υπολογισμό του D_{21} θα πρέπει να συγκριθούν τα αποτελέσματα των υποψηφίων που παρακάθισαν στην εξέταση του 2019 με την εξέταση του 2021 (και βεβαίως για τον υπολογισμό του D_{19} θα πρέπει να συγκριθούν τα αποτελέσματα των υποψηφίων που παρακάθισαν στην εξέταση του 2017 με την εξέταση του 2019). Επειδή ο αριθμός των υποψηφίων που εξετάζονται από χρόνο σε χρόνο είναι μεγάλος (π.χ. υπήρχαν πέραν των 1700 υποψηφίων που εξετάστηκαν στη Γνώση Ελληνικής Γλώσσας και το 2017 και το 2019), η μέθοδος των «κοινών υποψηφίων» προκρίθηκε ως μια από τις μεθόδους που αξιοποιήθηκαν για τον υπολογισμό των δεικτών D .

Αν και η μέθοδος των «κοινών υποψηφίων» είναι πολύ χρήσιμη, αποφασίστηκε να διασταυρωθούν τα αποτελέσματά της με αυτά άλλων τεχνικών προκειμένου να επιτευχθεί ελαχιστοποίηση του στατιστικού σφάλματος. Μελετώντας τη διεθνή πρακτική και τη διεθνή βιβλιογραφία προκρίθηκε η επιπρόσθετη χρήση των ακόλουθων τεχνικών: (α) propensity

score matching, (β) pseudo anchor method, (γ) 3DC standard setting και (δ) Angoff standard setting. Αξίζει να αναφερθεί ότι οι μέθοδοι αυτές αξιοποιούνται διεθνώς σε εξετάσεις υψηλού διακυβεύματος. Για παράδειγμα, χρησιμοποιούνται για να διατηρηθούν σταθερά από χρόνο σε χρόνο τα επίπεδα δυσκολίας των γνωστών εξετάσεων GCE A' Level. Επίσης χρησιμοποιούνται σε διάφορες χώρες (π.χ. Αυστραλία, Αγγλία) για να διατηρείται από χρόνο σε χρόνο η συγκρισιμότητα των επιπέδων των εξετάσεων για πιστοποίηση άσκησης του ιατρικού επαγγέλματος (οι γνωστές «medical licensing exams»). Για πιο πολλές πληροφορίες παραπέμπουμε στους Ward, Chiavaroli, Fraser et al. (2018).

Για κάθε ειδικότητα αξιοποιήθηκε αριθμός από αυτές τις τεχνικές για τον υπολογισμό των δεικτών D. Για κάθε μέθοδο υπολογίστηκε το D, αλλά και το τυπικό σφάλμα εκτίμησης του D. Για κάθε ειδικότητα οι διάφορες εκτιμήσεις του D συνυπολογίστηκαν προκειμένου να υπολογιστεί ένα συνολικό D με το ελάχιστο δυνατό σφάλμα.

Για σύντομες περιγραφές της κάθε μεθόδου, ο αναγνώστης παραπέμπεται στο σχετικό αρχείο για τη Στατιστική Επεξεργασία των βαθμολογιών των εξετάσεων του 2019.

4. Σύνθεση των δεικτών D από διαφορετικές μεθόδους

Δεν ήταν πρακτικά δυνατόν να αξιοποιηθούν όλες οι μέθοδοι στάθμισης για όλες τις ειδικότητες λόγω πρακτικών δυσκολιών. Για παράδειγμα, κάποιες ειδικότητες είχαν πολύ μικρό αριθμό υποψηφίων, άρα κάποιες μέθοδοι δεν μπορούσαν να χρησιμοποιηθούν. Ωστόσο, έγινε κάθε προσπάθεια να συνδυαστούν όσο το δυνατόν περισσότερες μέθοδοι για κάθε ειδικότητα, ώστε να έχουμε πολλαπλές πηγές πληροφόρησης και να πετύχουμε ελαχιστοποίηση του σφάλματος μέσω της τριγωνοποίησης των αποτελεσμάτων.

Κάθε μέθοδος αποδίδει μια εκτίμηση για το D και ένα τυπικό σφάλμα για αυτή την εκτίμηση. Προκειμένου να αποδώσουμε σε κάθε δείκτη D μια «δίκαιη» βαρύτητα, αξιοποιήσαμε το μέγεθος του τυπικού σφάλματος. Σε πιο επισφαλείς εκτιμήσεις (με μεγάλο σφάλμα) δόθηκε μικρότερη βαρύτητα. Ο μέσος όρος (weighted average), για κάθε ειδικότητα, υπολογίστηκε ως:

$$\text{Συνθετικός μέσος όρος D} = \frac{\sum m \cdot w_m \cdot D_m}{\sum m \cdot w_m}$$

όπου

$$w_m = 1 / (SE_m \cdot SE_m)$$

και w_m υποδηλώνει τη βαρύτητα του δείκτη για τη μέθοδο m (για όσες μεθόδους αξιοποιήθηκαν για τη στάθμιση των αποτελεσμάτων κάθε ειδικότητας).

5. Συμπερασματικά

Η διεθνής βιβλιογραφία και η διεθνής πρακτική προτείνουν πολλές μεθόδους για τη διαχρονική στάθμιση εξεταστικών δοκιμών. Προκειμένου να μεγιστοποιηθεί η ακρίβεια στους υπολογισμούς και προκειμένου να πετύχουμε την πιο δίκαιη διαχρονική στάθμιση, αξιοποιήθηκε μια μεγάλη συλλογή από μεθόδους, κατά τα πρότυπα άλλων Ευρωπαϊκών χωρών.

Κάθε μία από αυτές τις μεθόδους που αξιοποιήθηκαν έχει πλεονεκτήματα και μειονεκτήματα. Σε κάθε περίπτωση, η ομάδα των ειδικών που έφερε σε πέρας τη διαδικασία της στάθμισης διερεύνησε διεξοδικά τα μειονεκτήματα και τα πλεονεκτήματα της κάθε μεθόδου και αξιοποίησε μόνο τις πιο κατάλληλες για κάθε ειδικότητα.



Για περισσότερες πληροφορίες, ενδιαφερόμενοι με εξειδικευμένες γνώσεις παραπέμπονται σε σχετική βιβλιογραφία που παρατίθεται στο τέλος αυτού του κειμένου.

Αναφορές και βιβλιογραφία

Bramley, T. & Vidal Rodeiro, C.L. (2014). Using statistical equating for standard maintaining in GCSEs and A levels. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Bramley, T. (2018, November). Evaluating the 'similar items method' for standard maintaining. Paper presented at the 19th annual conference of the Association for Educational Assessment in Europe, Arnhem-Nijmegen, The Netherlands.

Brandon, P. R. (2004). Conclusions about frequently studied modified Angoff standard-setting topics. *Applied Measurement in Education*, 17(1), 59-88.

Guo, H. (2010). Accumulative equating error after a chain of linear equatings. *Psychometrika*, 75(3), 438-453.

Hambleton, R. K., & Plake, B. S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education*, 8(1), 41-55.

Keuning, J., Straat, J. H., & Feskens, R. C. (2017). The Data-Driven Direct Consensus (3DC) Procedure: New Approach to Standard Setting. In *Standard Setting in Education* (pp. 263-278). Springer, Cham.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices*. Springer Science & Business Media.

Livingston, S. A., Dorans, N. J., & Wright, N. K. (1990). What combination of sampling and equating methods works best? *Applied Measurement in Education*, 3(1), 73-95.

Ward, H., Chiavaroli, N., Fraser, J. et al. (2018). Standard setting in Australian medical schools. *BMC Medical Education*, 18, 80. <https://doi.org/10.1186/s12909-018-1190-6>