

4.5 Έλεγχοι υποθέσεων

Η Επαγωγική Στατιστική ή Στατιστική Συμπερασματολογία (Statistical Inference) μπορεί να οριστεί ως το σύνολο των μεθόδων στις οποίες χρησιμοποιούνται δειγματικά δεδομένα για να ληφθούν αποφάσεις και να προκύψουν συμπεράσματα για ολόκληρο τον πληθυσμό (Γιαλαμάς 2004, σελ. 20). Η στατιστική μεθοδολογία με την οποία θα ασχοληθούμε στα πλαίσια της διπλωματικής εργασίας, περιλαμβάνει τους έλεγχους υποθέσεων (hypotheses tests), την ανάλυση συσχέτισης (correlation analysis), τη γραμμική παλινδρόμηση (linear regression) και την ανάλυση διακύμανσης (analysis of variance). Η σπουδαιότητα της επαγωγικής στατιστικής έγκειται τόσο στην επέκταση των συμπερασμάτων από το δείγμα σε ολόκληρο τον πληθυσμό όσο και στην αντικειμενικότητα των ποσοτικών μεθόδων τις οποίες χρησιμοποιεί, έτσι ώστε οι θεωρητικές προκαταλήψεις του ερευνητή να μην επηρεάζουν τα ευρήματα της έρευνας. Όμως για να υιοθετήσουμε τα αποτελέσματα της Επαγωγικής Στατιστικής θα πρέπει να ακολουθούνται συγκεκριμένες προϋποθέσεις για την κάθε ερευνητική μέθοδο και επιπλέον να δίνεται ιδιαίτερη προσοχή στην εξαγωγή των συμπερασμάτων ώστε να αποφεύγονται παρερμηνείες, στις οποίες αναφερθήκαμε αναλυτικά στην ενότητα 1.4 .

Οι έλεγχοι των στατιστικών υποθέσεων είναι από τα πλέον βασικά θέματα της στατιστικής συμπερασματολογίας. Ως υπόθεση μπορούμε να ορίσουμε μια άποψη ή ένα ερευνητικό ερώτημα, η αξιολόγηση των οποίων χρήζει διερεύνησης. Με τις διαδικασίες του ελέγχου υποθέσεων καθορίζουμε κατά πόσο τέτοιες ερευνητικές υποθέσεις ικανοποιούνται από τα δεδομένα της έρευνας. Οι ερευνητικές υποθέσεις όταν διατυπωθούν με την πιο αυστηρή γλώσσα της στατιστικής οδηγούν στις στατιστικές υποθέσεις που ελέγχονται με μαθηματικές διαδικασίες ως προς την ορθότητά τους.

Οι έλεγχοι υποθέσεων με τους οποίους θα ασχοληθούμε αφορούν τις μέσες τιμές, ενός ή δύο δειγμάτων καθώς και το χ^2 καλής προσαρμογής και ανεξαρτησίας. Η εισαγωγή στην έννοια των στατιστικών υποθέσεων θα γίνει με τη βοήθεια του παρακάτω παραδείγματος.

Παράδειγμα II (Έλεγχος υποθέσεων – z test)

Έχουμε καταγράψει τις βαθμολογίες 10 μαθητών μιας τάξης σε μια γραπτή εξέταση

ενός μαθήματος και μας ενδιαφέρει να εξετάσουμε αν η μέση επίδοση των μαθητών στο συγκεκριμένο μάθημα διαφέρει από το 15, που συνήθως θεωρείται ο μέσος όρος της επίδοσης σε μία τάξη. Η ερευνητική υπόθεση: «ο μέσος όρος της επίδοσης της τάξης είναι διαφορετική από το 15» είναι αποτέλεσμα παρατήρησης από τον ερευνητή-εκπαιδευτικό και θα ελεγχθεί με βάση τα δεδομένα που συλλέχθηκαν από τις βαθμολογίες της γραπτής εξέτασης. Προφανώς η απάντηση δεν μπορεί να δοθεί μόνο από τον υπολογισμό της μέσης τιμής $\bar{x}$ του δείγματος των βαθμολογιών, η οποία υπόκειται σε απρόβλεπτους και τυχαίους παράγοντες. Πρέπει να καθορίσουμε τις προϋποθέσεις και τα κριτήρια με τα οποία θα αποφανθούμε αν η μέση τιμή της επίδοσης είναι ή δεν είναι 15.

■

Η διαδικασία του ελέγχου στατιστικών υποθέσεων θα παρουσιαστεί με τέσσερα βήματα τα οποία θα ακολουθηθούν σε όλα τα παρακάτω παραδείγματα και στην υλοποίηση τους μέσω της R.

1. Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων. Τα ερευνητικά ερωτήματα μπορούν να προκύψουν είτε από ισχυρές πεποιθήσεις του ερευνητή είτε από τη γενική εικόνα που προκύπτει για το δείγμα με τις μεθόδους περιγραφικής στατιστικής. Οι στατιστικές υποθέσεις που προκύπτουν από τα ερευνητικά ερωτήματα σε κάθε έλεγχο υποθέσεων είναι η **μηδενική ή αρχική** υπόθεση (null hypothesis) που συμβολίζεται με H_0 και η **εναλλακτική ή ερευνητική** υπόθεση (alternative ή research hypothesis) που συμβολίζεται με H_a . Ο όρος μηδενική υπόθεση προέκυψε από το γεγονός ότι συχνά οι υποθέσεις σχηματίζονται βάσει κάποιων ερευνητικών ισχυρισμών, π.χ. ότι η μέση επίδοση των μαθητών στο μάθημα του Παραδείγματος 11 διαφέρει από μία συγκεκριμένη τιμή, το 15. Έτσι, η H_0 είναι η υπόθεση που «μηδενίζει», δηλαδή αντιπαρατίθεται στον ισχυρισμό αυτό. Η μηδενική και η εναλλακτική υπόθεση στο Παράδειγμα 11 διατυπώνονται ως εξής:

$$H_0: \text{Η μέση επίδοση της τάξης είναι ίση με 15 ή } H_0: \mu = 15$$

$$H_a: \text{Η μέση επίδοση της τάξης είναι διαφορετική από το 15 ή } H_a: \mu \neq 15.$$

Οι έλεγχοι υποθέσεων όπως ο προηγούμενος λέγονται **αμφίπλευροι έλεγχοι** (two-sided tests). Υπάρχουν επιπλέον και οι **μονόπλευροι έλεγχοι** (one-sided test) οι οποίοι έχουν ίδια μηδενική υπόθεση με τους αμφίπλευρους αλλά διαφορετικές

εναλλακτικές. Στην περίπτωση του Παραδείγματος 11, ο διδάσκων του μαθήματος ενδεχομένως να παρατήρησε ότι το γνωστικό επίπεδο της τάξης στο μάθημά του δεν αξιολογείται απλώς ως διαφορετικό από το 15, αλλά είτε πάνω από το 15, οπότε η εναλλακτική υπόθεση γίνεται:

$$H_a : \text{Η μέση επίδοση της τάξης είναι μεγαλύτερη από το 15 ή } H_a : \mu > 15,$$

είτε κάτω από το 15, οπότε

$$H_a : \text{Η μέση επίδοση της τάξης είναι μικρότερη από το 15 ή } H_a : \mu < 15.$$

2. Προϋποθέσεις. Η πιο συνηθισμένη προϋπόθεση στον έλεγχο υποθέσεων είναι ότι το δείγμα προέρχεται από έναν πληθυσμό που ακολουθεί κανονική κατανομή. Αν υποθέσουμε ότι τα δεδομένα του Παραδείγματος 11 προέρχονται από έναν κανονικό πληθυσμό με μέση τιμή μ και διακύμανση σ^2 αυτό σημαίνει, σύμφωνα με την ενότητα 3.4.1, ότι η καμπύλη συχνοτήτων του δείγματος προσεγγίζεται από την κωδωνοειδή μορφή της κανονικής κατανομής, στην οποία οι περισσότερες βαθμολογίες των μαθητών συγκεντρώνονται γύρω από το μ και με ποσοστό 68% στο διάστημα $\mu \pm \sigma$. Επιπλέον, καθώς οι τιμές της βαθμολογίας απομακρύνονται από το μ αυτές κατανέμονται συμμετρικά γύρω από το μ και με ποσοστά 95% στο διάστημα $\mu \pm 2\sigma$ και 99% των τιμών στο διάστημα $\mu \pm 3\sigma$. Αυτός ο περιγραφικός έλεγχος για την ισχύ της προϋπόθεσης της κανονικότητας δεν είναι αρκετός και απαιτούνται πιο αυστηροί στατιστικοί έλεγχοι, όπως ο στατιστικός έλεγχος των Kolmogorov-Smirnov (Mood et al., 1974, σελ. 508-510 • Stuart, Ord & Arnold, 1999, σελ. 415-416), που όμως ξεφεύγει από το σκοπό της παρούσας μελέτης.

3. Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου. Ένα θεμελιώδες ζήτημα στους ελέγχους υποθέσεων είναι ο καθορισμός του κριτηρίου στατιστικού ελέγχου. Έτσι, για την περίπτωση του μέσου όρου ενός δείγματος, πρέπει να καθορίσουμε πόσο «κοντά» ή «μακριά» πρέπει να βρίσκεται ο δειγματικός μέσος όρος $\bar{x}$ στην καθορισμένη μέση τιμή μ_0 , ώστε να θεωρήσουμε τη διαφορά τους στατιστικώς σημαντική. Συγκεκριμένα, στο Παράδειγμα 11 θέλουμε να εκτιμήσουμε αν η διαφορά των $\bar{x}$ και $\mu_0 = 15$ είναι στατιστικώς σημαντική, για αυτό μας ενδιαφέρει η τιμή της διαφοράς τους $\bar{x} - 15$. Ταυτόχρονα όμως πρέπει να

ληφθεί υπόψη η διασπορά των τιμών του $\bar{x}$ για διαφορετικά δείγματα από τον ίδιο πληθυσμό. Έτσι, εμπειρικά μπορούμε να πούμε ότι η διαφορά $\bar{x} - 15 = 2$ θεωρείται μάλλον σημαντική στην εικοσάβαθμη κλίμακα αξιολόγησης, ενώ θεωρείται μάλλον αμελητέα στην αντίστοιχη εκατοντάβαθμη. Όταν το δείγμα προέρχεται από έναν πληθυσμό με μέση τιμή μ και διακύμανση σ^2 τότε έχει αποδειχθεί (Mood et al. 1974, σελ. 231) ότι η τυπική απόκλιση, που είναι μέτρο διασποράς, των τιμών του $\bar{x}$ ισούται με $\frac{\sigma}{\sqrt{n}}$, που ονομάζεται τυπικό σφάλμα του μέσου όρου (standard error ή συντομογραφικά SE). Λαμβάνοντας υπόψη την παραπάνω παρατήρηση ορίζουμε τη γενική μορφή της στατιστικής συνάρτησης που χρησιμοποιείται για τον έλεγχο υποθέσεων του μέσου όρου ενός δείγματος (Κατσίης 2011, σελ. 84, Εμβαλωτής κ. συν. 2006, σελ. 53 και Γιαλαμάς 2004, σελ. 115):

$$\frac{\text{δειγματικός μέσος όρος} - \text{καθορισμένη τιμή}}{\text{τυπική απόκλιση δειγματικού μέσου όρου}}.$$

Η παραπάνω ποσότητα ονομάζεται στατιστική συνάρτηση ή απλά στατιστικό. Για το Παράδειγμα 11, αν το δείγμα προέρχεται από έναν πληθυσμό με γνωστή διακύμανση σ^2 , το στατιστικό ελέγχου γίνεται:

$$z = \frac{\bar{x} - 15}{\frac{\sigma}{\sqrt{10}}}.$$

4. Κανόνας απόφασης και εξαγωγή συμπερασμάτων. Ο κανόνας απόφασης βασίζεται στην πιθανότητα που υπάρχει για μια στατιστική συνάρτηση (όπως το z) να πάρει τιμές που οφείλονται σε τυχαίους παράγοντες οι οποίοι αλλοιώνουν την πραγματικότητα. Αν η πιθανότητα αυτή, που ονομάζεται τιμή p (p value), είναι μικρή τότε απορρίπτουμε τη μηδενική υπόθεση. Το ανώτατο αποδεκτό όριο για την τιμή p λέγεται **επίπεδο σημαντικότητας** (significance level) και συμβολίζεται με α . Πιο συγκεκριμένα, το α εκφράζει το ανώτατο επίπεδο της πιθανότητας απόρριψης της μηδενικής υπόθεσης όταν στην πραγματικότητα ισχύει. Για αυτό και επιδιώκουμε αυτό το σφάλμα να είναι πολύ μικρό, συνήθως 0.01 ή 0.05.

Η διαδικασία του κανόνα απόφασης και της εξαγωγής συμπερασμάτων ξεκινά με τον υπολογισμό του κριτηρίου ελέγχου. Οι δυνατές τιμές που μπορεί να πάρει η στατιστική συνάρτηση ελέγχου χωρίζονται σε αυτές που αποτελούν την περιοχή

απόρριψης της μηδενικής υπόθεσης και σε αυτές που σχηματίζουν την περιοχή αποδοχής της μηδενικής υπόθεσης. Ο καθορισμός των περιοχών γίνεται με βάση το επιθυμητό επίπεδο σημαντικότητας. Ειδικότερα, για το Παράδειγμα 11 αποδεικνύεται ότι, κάτω από την υπόθεση της κανονικότητας, η στατιστική συνάρτηση

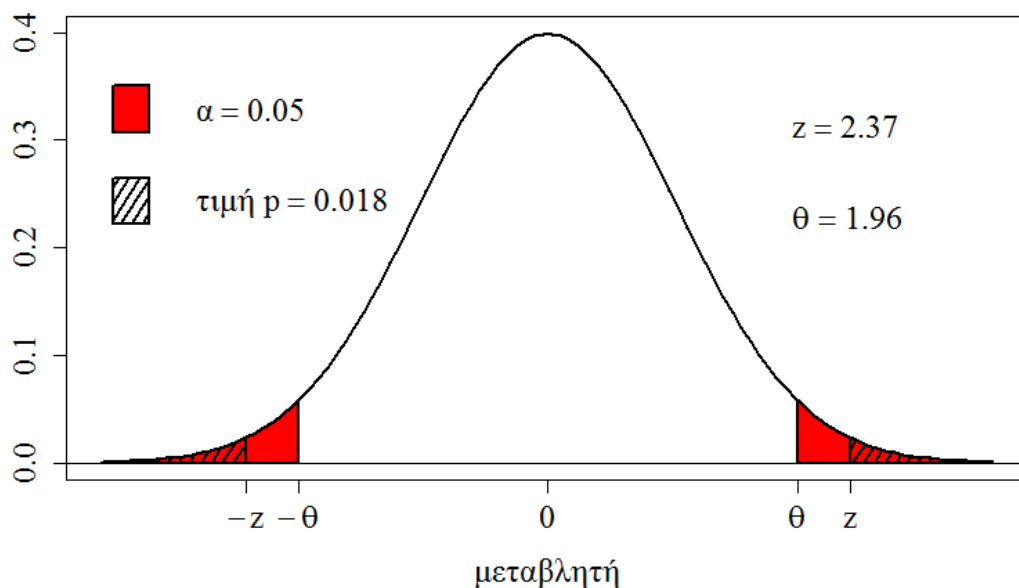
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

ακολουθεί κανονική κατανομή με μέση τιμή 0 και διακύμανση 1 (Παπαϊωάννου & Φερντίνος 1983, σελ. 173-174). Οπότε, όταν $\alpha = 0.05$, η περιοχή απόρριψης καθορίζεται από τη σχέση:

$$z \leq -\theta \text{ ή } z \geq \theta,$$

όπου $\theta = 1.96$ είναι η τιμή για την οποία $\Pr(-\theta \leq z \text{ ή } z \geq \theta) = 0.05$ ή 5%, όπως δείχνει το παρακάτω σχήμα.

Η παραπάνω πιθανότητα αντιστοιχεί στο εμβαδόν του κόκκινου χωρίου του Σχήματος 18. Αν για παράδειγμα, η τιμή της στατιστικής συνάρτησης είναι $z = 2.37$, τότε ισχύει ότι $z \geq \theta$, διότι $2.37 \geq 1.96$, οπότε η μηδενική υπόθεση απορρίπτεται. Στο ίδιο συμπέρασμα μπορούμε να καταλήξουμε εποπτικά, διότι αν θεωρητικά εκτελέσουμε το ίδιο πείραμα πολλές φορές αναμένουμε με βάση το σχήμα, ότι το 95% των τιμών του z θα βρίσκεται στο λευκό κομμάτι του οριζόντιου άξονα, δηλαδή μεταξύ του $-\theta$ και του θ . Έτσι όταν το z βρίσκεται στην κόκκινη περιοχή θεωρούμε ότι είναι «μακριά» από τη μέση τιμή του που είναι το μηδέν και έτσι η μηδενική υπόθεση απορρίπτεται, ενώ αν βρίσκεται στη λευκή περιοχή θεωρούμε ότι είναι «κοντά» στο μηδέν οπότε η μηδενική υπόθεση δεν απορρίπτεται.



Σχήμα 18. Γραφική ερμηνεία του επιπέδου σημαντικότητας α και της τιμής p στο Παράδειγμα 11.

Το σφάλμα που προκύπτει από την εσφαλμένη απόρριψη της μηδενικής υπόθεσης λέγεται **σφάλμα τύπου I**, ενώ το σφάλμα που προκύπτει από την εσφαλμένη αποδοχή της μηδενικής υπόθεσης λέγεται **σφάλμα τύπου II** και η αντίστοιχη πιθανότητα συμβολίζεται με β . Οι δυνατές επιλογές από την εφαρμογή ενός στατιστικού τεστ και τα αντίστοιχα σφάλματα φαίνονται στον παρακάτω πίνακα:

Συμπέρασμα ελέγχου	Πραγματικότητα	
	H_0 είναι αληθινή	H_a είναι αληθινή
Αποδοχή της H_0	Σωστή απόφαση	Λανθασμένη απόφαση - Σφάλμα τύπου II
Απόρριψη της H_0	Λανθασμένη απόφαση – Σφάλμα τύπου I	Σωστή απόφαση

Είναι προφανές ότι θα προτιμήσουμε εκείνο το στατιστικό τεστ με τα μικρότερα α και β ταυτόχρονα. Έχει όμως αποδειχθεί ότι η ταυτόχρονη ελαχιστοποίηση των α και β είναι αδύνατη (Παπαϊωάννου & Φερεντίνος 1983, σελ. 131). Γενικά δεν ασκείται έλεγχος πάνω στο β , είναι όμως γνωστό ότι είναι μεγαλύτερο από το α

(Γιαλαμάς 2004, σελ. 116).

Ένα ισοδύναμος κανόνας απόφασης βασίζεται στον υπολογισμό της τιμής p που αποτελεί το ελάχιστο επίπεδο σημαντικότητας στο οποίο μπορεί να απορριφθεί η μηδενική υπόθεση (Stuart et al., 1999, σελ. 172). Στο παραπάνω παράδειγμα η τιμή του p προκύπτει ως

$$\Pr(z \leq -2.37 \text{ ή } z \geq 2.37) = 0.018$$

και αντιστοιχεί στο εμβαδόν του γραμμοσκιασμένου χωρίου του σχήματος. Σε αυτή την περίπτωση αν το εμβαδόν του χωρίου που αντιστοιχεί στην τιμή p (γραμμοσκιασμένο χωρίο) είναι μικρότερο ή ίσο από το εμβαδόν του χωρίου που αντιστοιχεί στο επίπεδο σημαντικότητας α (κόκκινο χωρίο) τότε απορρίπτουμε τη μηδενική υπόθεση. Δηλαδή, συγκρίνουμε την τιμή p που θα προκύψει από το δείγμα μας με το επίπεδο σημαντικότητας και υιοθετούμε τον ακόλουθο κανόνα απόφασης (Κατσή και συν., 2010, σελ. 80):

- Απορρίπτουμε τη μηδενική υπόθεση αν $p \leq \alpha$
- Δεν απορρίπτουμε τη μηδενική υπόθεση αν $p > \alpha$.

Όπως είδαμε, η στατιστική απόφαση συνίσταται στην απόρριψη ή όχι της μηδενικής υπόθεσης, σύμφωνα με τον κανόνα απόφασης. Ο ερευνητής όμως θα πρέπει να είναι σε θέση να αντλήσει όσο γίνεται περισσότερες πληροφορίες από τα στατιστικά αποτελέσματα. Για να γίνει αυτό θα πρέπει να μετασχηματίσουμε τα στατιστικά αποτελέσματα σε ερευνητικά συμπεράσματα, δηλαδή εκφράσεις και λογικές προτάσεις οι οποίες (King et al., 2000, σελ. 347):

- περιέχουν τα ακριβή αριθμητικά αποτελέσματα που είναι αναγκαία για την εξαγωγή ερευνητικών συμπερασμάτων
- συμπεριλαμβάνουν την έννοια της αβεβαιότητας ή των λαθών στα οποία υπόκεινται οι στατιστικοί υπολογισμοί, όπως είναι τα στατιστικά σφάλματα τύπου I και II, που έχουμε ήδη αναλύσει και
- δεν απαιτούν εξειδικευμένες γνώσεις στατιστικής για την κατανόησή τους.

Επαναφέροντας το Παράδειγμα 11, στην περίπτωση που απορριφθεί η μηδενική υπόθεση το ερευνητικό συμπέρασμα που ικανοποιεί τα παραπάνω κριτήρια, θα μπορούσε να είναι: «Η μέση επίδοση της τάξης στο συγκεκριμένο μάθημα διαφέρει από το 15 με πιθανότητα λάθους 1.8%». Στην περίπτωση που η μηδενική υπόθεση δεν απορρίπτεται η διατύπωση του ερευνητικού συμπεράσματος θα μπορούσε να είναι: «Η μέση επίδοση της τάξης στο συγκεκριμένο μάθημα είναι

15. Επιπλέον, επειδή το δείγμα των 10 μαθητών προέρχεται από τον πληθυσμό όλων των μαθητών οι οποίοι φοιτούν σε αυτή την τάξη που ακολουθεί κανονική κατανομή με γνωστή διακύμανση $\sigma^2=1$, αναμένουμε ότι το 95% των βαθμολογιών σε επόμενη γραπτή εξέταση θα βρίσκονται μεταξύ των τιμών $\mu \pm 3\sigma = 15 \pm 3 \cdot 1$, δηλαδή μεταξύ του 12 και του 18.».

Οι έλεγχοι υποθέσεων με τους οποίους θα ασχοληθούμε στην παρούσα μελέτη είναι οι εξής:

a. Έλεγχος για τη μέση τιμή ενός πληθυσμού με γνωστή διακύμανση. Η διαδικασία αυτού του ελέγχου υποθέσεων περιλαμβάνει τα εξής στάδια:

1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Για τον αμφίπλευρο έλεγχο έχουμε τις υποθέσεις

$$H_0 : \mu = \mu_0 \text{ έναντι } H_a : \mu \neq \mu_0 ,$$

ενώ οι δύο μονόπλευροι έλεγχοι είναι

$$H_0 : \mu = \mu_0 \text{ έναντι } H_a : \mu < \mu_0 \text{ και } H_0 : \mu = \mu_0 \text{ έναντι } H_a : \mu > \mu_0 .$$

2. **Προϋποθέσεις.** Το δείγμα μεγέθους n προέρχεται από έναν πληθυσμό τα στοιχεία του οποίου ακολουθούν κανονική κατανομή με γνωστή διακύμανση σ^2 .
3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Το κριτήριο ελέγχου είναι η συνάρτηση z με:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} ,$$

η οποία κάτω από τις παραπάνω προϋποθέσεις ακολουθεί κανονική κατανομή με μέση τιμή 0 και διακύμανση 1, όταν ισχύει η υπόθεση H_0 .

4. **Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Ο κανόνας απόφασης για τον αμφίπλευρο έλεγχο βασίζεται στην περιοχή απόρριψης $z \leq -\theta$ ή $z \geq \theta$, όπου θ είναι μια τιμή για την οποία $\Pr(-\theta \leq z \text{ ή } z \geq \theta) = \alpha$. Για τον μονόπλευρο έλεγχο $H_0 : \mu = \mu_0$ έναντι $H_a : \mu < \mu_0$ η περιοχή απόρριψης της μηδενικής υπόθεσης είναι $z \leq -\theta$, όπου θ ικανοποιεί τη σχέση $\Pr(z \leq -\theta) = \alpha$, ενώ για τον άλλο

μονόπλευρο έλεγχο η περιοχή απόρριψης είναι $z \geq \theta$, όπου θ ικανοποιεί τη σχέση $\Pr(z \geq \theta) = \alpha$. Στην πράξη όμως χρησιμοποιείται συνήθως η δεύτερη προσέγγιση για τον κανόνα απόφασης με την τιμή p , η οποία είναι κοινή για όλους τους στατιστικούς ελέγχους. Όπως είδαμε σύμφωνα με αυτήν απορρίπτουμε τη μηδενική υπόθεση αν $p \leq \alpha$ ενώ δεν απορρίπτουμε τη μηδενική υπόθεση αν $p > \alpha$.

Παράδειγμα 11 (*z test* – επεξεργασία με την R)

Οι βαθμολογίες των 10 μαθητών της τάξης είναι: 13, 15, 17, 18, 20, 10, 14, 16, 16, 15. Η εισαγωγή των δεδομένων στην R γίνεται με την εντολή

```
> grades <- c(13, 15, 17, 18, 20, 10, 14, 16, 16, 15)
```

Κατόπιν, ακολουθούμε τη διαδικασία που έχουμε περιγράψει:

1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Για το αμφίπλευρο στατιστικό τεστ έχουμε

H_0 : Η μέση επίδοση της τάξης είναι ίση με 15 ή $H_0: \mu = 15$

H_a : Η μέση επίδοση της τάξης είναι διάφορη από το 15 ή $H_a: \mu \neq 15$.

2. **Προϋποθέσεις.** Υποθέτουμε ότι η δειγματοληψία προέρχεται από κανονική κατανομή με τυπική απόκλιση $\sigma = 1$.

3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Αρχικά ενεργοποιούμε το πακέτο TeachingDemos και ακολούθως καλούμε τη συνάρτηση `z.test()`. Στη συνάρτηση αυτή δίνουμε την προκαθορισμένη τιμή $\mu_0 = 15$ με το όρισμα `mu = 15`, επιλέγουμε τη μορφή της εναλλακτικής υπόθεσης με το όρισμα `alternative` με δυνατές τιμές μία από τις `"two.sided"`, `"less"`, `"greater"`, ορίζουμε την τυπική απόκλιση $\sigma = 1$ με το όρισμα `sd = 1` και τέλος το επίπεδο σημαντικότητας, `conf.level = 0.95`, αν επιθυμούμε $\alpha = 0.05$. Δηλαδή,

```
> z.test(grades, mu=15, sd=1)
```

Επισημαίνουμε ότι δεν χρειάστηκε να δώσουμε τα ορίσματα `alternative` και `conf.level` τα οποία έχουν ως προκαθορισμένες τιμές τις `"two.sided"` και `0.95` αντίστοιχα. Αυτό είναι απαραίτητο να γίνει αν επιλέξουμε για οποιοδήποτε όρισμα διαφορετική από την προκαθορισμένη τιμή. Τα αποτελέσματα που προκύπτουν είναι:

One Sample z-test

```
data: grades
z.mean = 1.2649, n = 10.000, Std. Dev. = 1.000, SE of the sample
mean = 0.316,
p-value = 0.2059
alternative hypothesis: true mean is not equal to 15
95 percent confidence interval:
 14.7802 16.0198
sample estimates:
mean of grades
      15.4
```

2. **Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Από όλα αυτά κυρίως μας ενδιαφέρει η τιμή p , που είναι ίση με 0.2059, η οποία είναι μεγαλύτερη από το $\alpha = 0.05$, οπότε η μηδενική υπόθεση δεν απορρίπτεται. Δηλαδή δεν μπορούμε να απορρίψουμε ότι η μέση επίδοση των μαθητών είναι ίση με 15.

■

b. Έλεγχος για τη μέση τιμή ενός πληθυσμού με άγνωστη διακύμανση. Το Παράδειγμα 11 παρουσιάζει το σημαντικό μειονέκτημα της εκ των προτέρων γνώσης της θεωρητικής διακύμανσης σ^2 , κάτι που συνήθως δε συμβαίνει στην πράξη. Για τις περιπτώσεις αυτές χρησιμοποιείται το t τεστ.

1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Για τον αμφίπλευρο έλεγχο έχουμε τις υποθέσεις

$$H_0 : \mu = \mu_0 \text{ έναντι } H_a : \mu \neq \mu_0,$$

ενώ οι δύο μονόπλευροι έλεγχοι είναι

$$H_0 : \mu = \mu_0 \text{ έναντι } H_a : \mu < \mu_0 \text{ και } H_0 : \mu = \mu_0 \text{ έναντι } H_a : \mu > \mu_0.$$

2. **Προϋποθέσεις.** Το δείγμα μεγέθους n προέρχεται από έναν πληθυσμό τα στοιχεία του οποίου ακολουθούν κανονική κατανομή με άγνωστη διακύμανση.
3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Το κριτήριο ελέγχου αποτελεί η συνάρτηση t με:

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}},$$

όπου s είναι η δειγματική διακύμανση. Η ποσότητα t ακολουθεί κάτω από τις παραπάνω προϋποθέσεις την κατανομή t με $n-1$ βαθμούς

ελευθερίας, όταν ισχύει η υπόθεση H_0 .

4. **Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Απορρίπτουμε τη μηδενική υπόθεση αν $p \leq \alpha$ ενώ δεν απορρίπτουμε τη μηδενική υπόθεση αν $p > \alpha$.

Παράδειγμα 12 (Έλεγχος υποθέσεων – t test με έναν πληθυσμό)

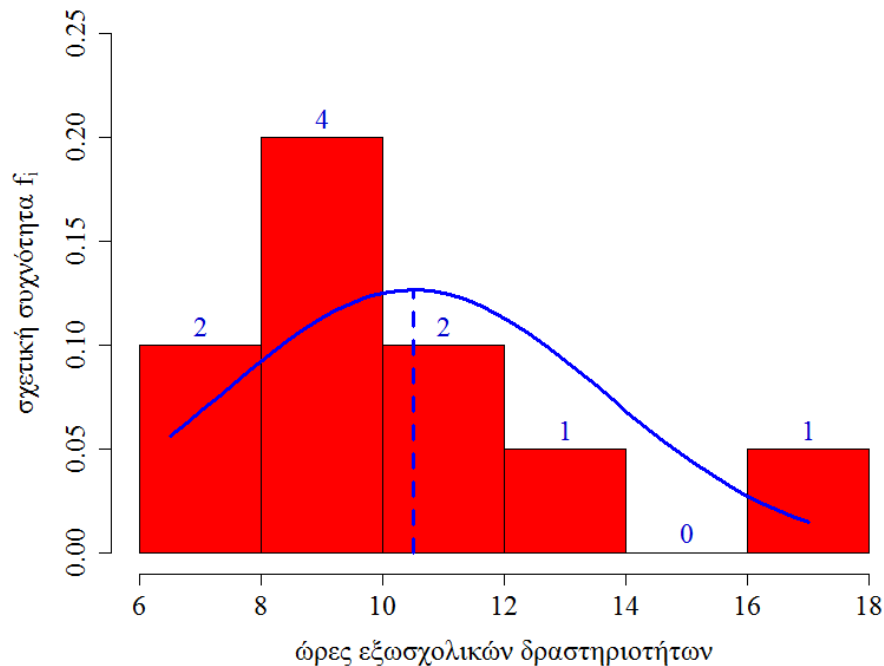
Θεωρούμε ότι οι μαθητές ηλικίας 12 ετών ασχολούνται κατά μέσο όρο 8 ώρες εβδομαδιαίως με εξωσχολικές δραστηριότητες (ξένες γλώσσες, αθλητισμός, κ.λπ). Οι αριθμοί 12, 17, 9.5, 7.5, 6.5, 9.5, 11, 14, 8.5, 9.5 αφορούν τις ώρες που αφιερώνουν την εβδομάδα δέκα τυχαία επιλεγμένοι μαθητές αυτής της ηλικίας για εξωσχολικές δραστηριότητες. Μας ενδιαφέρει να εξετάσουμε εάν ο αρχικός συλλογισμός είναι σωστός. Το παράδειγμα αναφέρεται από τους Κατσή κ. συν. (2010, σελ. 85).

1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Καθορίζουμε τις στατιστικές υποθέσεις:

H_0 : Οι μαθητές απασχολούνται κατά μέσο όρο 8 ώρες την εβδομάδα σε εξωσχολικές δραστηριότητες, δηλαδή $H_0 : \mu = 8$ έναντι

H_a : Οι μαθητές δεν απασχολούνται κατά μέσο όρο 8 ώρες την εβδομάδα σε εξωσχολικές δραστηριότητες, δηλαδή $H_a : \mu \neq 8$.

2. **Προϋποθέσεις.** Αρχικά ελέγχουμε αν τα δεδομένα μας πληρούν την προϋπόθεση της κανονικότητας. Για αυτό στο Σχήμα 19 παραθέτουμε το ιστόγραμμα των δεδομένων μας, τον αριθμό των στοιχείων σε κάθε κλάση και την αντίστοιχη θεωρητική κανονική κατανομή που μπορούμε να προσαρμόσουμε στα δεδομένα (με την μπλε γραμμή). Εποπτικά διαπιστώνουμε ότι επιτυγχάνεται μια ικανοποιητική προσαρμογή της καμπύλης της κανονικής κατανομής στα δεδομένα. Αυτό σημαίνει ότι μπορούμε να εφαρμόσουμε το t test. Για ένα πιο αυστηρό έλεγχο της κανονικότητας υπενθυμίζουμε ότι ενδείκνυται το τεστ των Kolmogorov-Smirnov.



Σχήμα 19. Ιστόγραμμα σχετικών συχνοτήτων των δεδομένων του Παραδείγματος 12. Η συνεχής γραμμή παριστάνει την καμπύλη συχνοτήτων της κανονικής κατανομής που προσαρμόζεται στα δεδομένα και η διακεκομμένη κατακόρυφη γραμμή τη μέση τιμή της κατανομής.

3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Υπολογίζουμε την τιμή του στατιστικού t και την τιμή p για επίπεδο σημαντικότητας $\alpha = 0.05$. Συγκεκριμένα οι υπολογισμοί δίνουν $t = 2.514$ και τιμή $p = 0.033$. Για να υλοποιήσουμε αυτούς τους υπολογισμούς με την R, αρχικά εισάγουμε τα δεδομένα με το διάνυσμα `activities`, ως εξής:

```
> activities <- c(12, 17, 9.5, 7.5, 6.5, 9.5, 11, 14, 8.5, 9.5)
```

Κατόπιν, καλούμε τη συνάρτηση `t.test()`. Συγκεκριμένα, καθορίζουμε ότι το διάνυσμα των δεδομένων `x` αντιστοιχεί στο διάνυσμα `activities` και την προκαθορισμένη τιμή $\mu_0 = 8$ με το όρισμα `mu=8`. Η συνάρτηση αυτή έχει προκαθορισμένες τιμές για το είδος του ελέγχου (αμφίπλευρος) και για το επίπεδο σημαντικότητας (0.05), οπότε δεν χρειάζεται να δοθούν. Έτσι,

```
> t.test(x=activities, mu=8)
```

One Sample t-test

```
data: activities
t = 2.514, df = 9, p-value = 0.03309
alternative hypothesis: true mean is not equal to 8
95 percent confidence interval:
 8.250445 12.749555
```

```
sample estimates:
mean of x
10.5
```

Αν δεν μας ενδιαφέρουν όλες αυτές οι πληροφορίες τότε μπορούμε να επιλέξουμε ορισμένες από αυτές με την εντολή

```
> t.test(x=activities, mu=8)$statistic
t
2.514006
```

η οποία μας δίνει μόνο την τιμή του στατιστικού t ή

```
> t.test(x=activities, mu=8)$p.value
[1] 0.03309254
```

η οποία μας δίνει μόνο την τιμή p .

4. **Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Για $\alpha = 0.05$ παρατηρούμε ότι $p \leq \alpha$, οπότε απορρίπτουμε τη μηδενική υπόθεση. Το συμπέρασμα αυτό συμφωνεί και με το γράφημα του ιστογράμματος, στο ποίο η κατακόρυφη διακεκομμένη γραμμή αντιστοιχεί στη μέση τιμή της κανονικής κατανομής που διαφέρει σημαντικά από το 8. Φαίνεται ότι είναι περίπου 10.5. Το ερευνητικό συμπέρασμα με πιθανότητα λάθους 3.31% είναι ότι οι μαθητές ηλικίας 12 ετών δεν ασχολούνται κατά μέσο όρο 8 ώρες εβδομαδιαίως με εξωσχολικές δραστηριότητες. Με βάση το Σχήμα 19 αν επαναλάβουμε το t test για το ίδιο δείγμα θα πρέπει να αναζητήσουμε την προκαθορισμένη τιμή μ_0 λίγο πάνω από το 10, ίσως 10.5 .

■

- γ. **Έλεγχος για τη σύγκριση των μέσων τιμών δύο πληθυσμών.** Οι έλεγχοι υποθέσεων για τη μέση τιμή ενός πληθυσμού επεκτείνονται σε δείγματα που προέρχονται από δύο πληθυσμούς. Στην κατηγορία των δύο πληθυσμών βάζουμε περιπτώσεις όπου έχουμε πραγματικά δύο διαφορετικούς πληθυσμούς (π.χ. δειγματοληψία από τους μαθητές δύο διαφορετικών τάξεων), αλλά και όταν έχουμε τον ίδιο πληθυσμό από τον οποίο παίρνουμε δείγμα δύο φορές (π.χ. δύο δειγματοληψίες που αφορούν την ίδια μεταβλητή από τους μαθητές της ίδιας τάξης).

Μας ενδιαφέρει να συγκρίνουμε την αποτελεσματικότητα δύο διαφορετικών μεθόδων διδασκαλίας. Για αυτό χωρίζουμε τους μαθητές σε δύο ομάδες, σε κάθε μία από τις οποίες διδάσκεται μία μέθοδος διδασκαλίας. Για να έχουμε αξιόπιστα αποτελέσματα επιδιώκουμε οι δύο ομάδες να αποτελούνται από μαθητές με ισοδύναμες περίπου ικανότητες. Μετά την ολοκλήρωση της εκπαιδευτικής διαδικασίας αξιολογούμε τους ομάδας. Θα ελέγξουμε κατά πόσο διαφέρουν οι μέσοι όροι των βαθμολογιών των δύο ομάδων.

■

Παράδειγμα 14 (Ελεγχος υποθέσεων)

Μας ενδιαφέρει να εξετάσουμε την επίδραση των νέων τεχνολογιών στην εκπαιδευτική διαδικασία. Για αυτό, μετά τον «παραδοσιακό» τρόπο διδασκαλίας μιας διδακτικής ενότητας, διδάσκεται στους μαθητές μιας τάξης η επόμενη διδακτική ενότητα με τη βοήθεια τεχνολογιών πληροφορίας και επικοινωνίας (ΤΠΕ). Οι μαθητές αξιολογούνται πριν και μετά τη διδασκαλία με τις νέες τεχνολογίες. Τα δεδομένα αποτελούνται από ένα ζεύγος βαθμολογιών για κάθε μαθητή. Ο ερευνητής ενδιαφέρεται να εξετάσει αν η επίδοση των μαθητών βελτιώθηκε μετά την χρήση των ΤΠΕ.

■

Στο Παράδειγμα 13 τα δείγματα προέρχονται από μαθητές δύο διαφορετικών ομάδων, οπότε είναι **ανεξάρτητα**, ενώ στο Παράδειγμα 14 προέρχονται από τους ίδιους μαθητές, οπότε είναι **εξαρτημένα**. Σε κάθε περίπτωση, όταν έχουμε δύο δείγματα μεγέθους n_1, n_2 , με δειγματικές μέσες τιμές $\bar{x}, \bar{y}$ και δειγματικές διακυμάνσεις s_1^2, s_2^2 αντίστοιχα, τότε έχουμε δύο ελέγχους υποθέσεων για ανεξάρτητα και εξαρτημένα δείγματα που αναλύονται στη συνέχεια.

I. Ανεξάρτητα δείγματα.

1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** H_0 : Δεν υπάρχει διαφορά στη μέση τιμή των δύο πληθυσμών ή $H_0: \mu_1 = \mu_2$, έναντι

H_a : Υπάρχει διαφορά στη μέση τιμή των δύο πληθυσμών ή $H_a: \mu_1 \neq \mu_2$.

Ανάλογα μπορούν να διατυπωθούν και οι δύο μονόπλευροι έλεγχοι.

2. **Προϋποθέσεις.** Τα δείγματα προέρχονται από δύο πληθυσμούς που ακολουθούν την κανονική κατανομή με θεωρητικές μέσες τιμές μ_1, μ_2 και

διακυμάνσεις σ_1^2, σ_2^2 αντίστοιχα. Οι διακυμάνσεις σ_1^2, σ_2^2 είναι άγνωστες.

3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Το σκεπτικό για το κριτήριο ελέγχου είναι αντίστοιχο με το σκεπτικό που αναπτύχθηκε για το κριτήριο ελέγχου της μέσης τιμής ενός πληθυσμού. Δίνεται από τη σχέση:

$$\frac{\text{διαφορά μέσων όρων}}{\text{συνολική τυπική απόκλιση}}$$

και εκφράζεται από το στατιστικό

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(\nu_1 - 1)s_1^2 + (\nu_2 - 1)s_2^2}{\nu_1 + \nu_2 - 2}} \sqrt{\frac{1}{\nu_1} + \frac{1}{\nu_2}}}$$

όταν οι θεωρητικές διακυμάνσεις σ_1^2, σ_2^2 είναι ίσες και

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{\nu_1} + \frac{s_2^2}{\nu_2}}}$$

όταν οι θεωρητικές διακυμάνσεις σ_1^2, σ_2^2 είναι άνισες. Σε κάθε περίπτωση, η ποσότητα t ακολουθεί κάτω από τις παραπάνω προϋποθέσεις την κατανομή t , όταν ισχύει η υπόθεση H_0 .

4. **Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Απορρίπτουμε τη μηδενική υπόθεση αν $p \leq \alpha$ ενώ δεν απορρίπτουμε τη μηδενική υπόθεση αν $p > \alpha$.

Παράδειγμα 13 (Συνέχεια – t test με δύο ανεξάρτητους πληθυσμούς)

Ως παράδειγμα θα χρησιμοποιήσουμε τα δεδομένα που έχουν δημοσιευθεί από τον Θεοχαράτου (1999) και παρουσιάζονται από τον Γιαλαμά (2004, σελ. 135-137). Πρόκειται για τη βαθμολογία 17 μαθητών της ομάδας Α και 16 μαθητών της ομάδας Β,

Ομάδα Α: 10, 5.6, 10, 6.2, 7.6, 6.4, 7.4, 5.2, 7.2, 5.6, 5.6, 5.6, 3.6, 7.6, 6.2, 6.6, 10

Ομάδα Β: 2, 2, 6, 3.2, 3.8, 1.8, 1.8, 2.6, 3.2, 2.8, 2.6, 3.4, 2, 2.8, 2.8, 2.8

μέσω των οποίων εξετάζεται η αποτελεσματικότητα δύο μεθόδων διδασκαλίας αρμονίου. Συγκεκριμένα θα εξεταστεί κατά πόσο η πρώτη μέθοδος οδηγεί τους μαθητές σε υψηλότερες επιδόσεις από τη δεύτερη μέθοδο. Ο έλεγχος της υπόθεσής

μας έχει ως εξής:

1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Ο ερευνητής για να έχει μια πρώτη εικόνα των επιδόσεων των μαθητών υπολογίζει τη μέση τιμή της επίδοσης της κάθε ομάδας. Για να εισάγουμε τα δεδομένα στην R πληκτρολογούμε

```
> music.A <- c(10, 5.6, 10, 6.2, 7.6, 6.4, 7.4, 5.2, 7.2, 5.6,
5.6, 5.6, 3.6, 7.6, 6.2, 6.6, 10)
> music.B <- c(2, 2, 6, 3.2, 3.8, 1.8, 1.8, 2.6, 3.2, 2.8, 2.6,
3.4, 2, 2.8, 2.8, 2.8)
```

Για να υπολογίσουμε τις μέσες τιμές τους καλούμε την εντολή **mean**:

```
> mean(music.A)
[1] 6.847059
> mean(music.B)
[1] 2.85
```

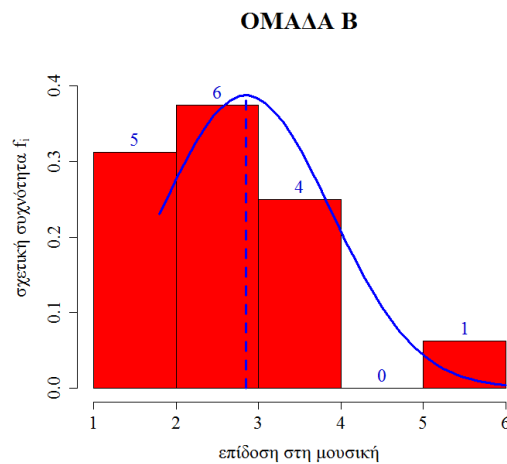
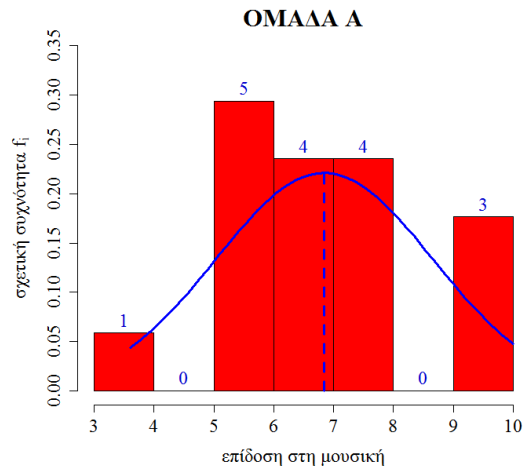
Το ερευνητικό ερώτημα εξετάζει αν οι μαθητές που διδάχθηκαν με τη μέθοδο A έχουν μέση επίδοση μ_1 που είναι μεγαλύτερη από τη μέση επίδοση μ_2 των μαθητών που διδάχθηκαν με τη μέθοδο B, όπως μπορούμε να εικάσουμε από την παρατήρηση των επιδόσεών τους, δηλαδή

$$H_0 : \mu_1 \leq \mu_2 \text{ έναντι } H_a : \mu_1 > \mu_2 .$$

Η ανισότητα της μηδενικής υπόθεσης σημαίνει ότι αυτή αποτελείται από την υπόθεση $\mu_1 = \mu_2$ και άπειρες υποθέσεις της μορφής $\mu_1 < \mu_2$. Ο έλεγχος όμως γίνεται μόνο για $\mu_1 = \mu_2$ διότι αν αυτή απορριφθεί τότε απορρίπτεται και κάθε άλλη περίπτωση στην οποία $\mu_1 < \mu_2$. Επομένως οι στατιστικές υποθέσεις επαναδιατυπώνονται ως εξής:

$$H_0 : \mu_1 = \mu_2 \text{ έναντι } H_a : \mu_1 > \mu_2 .$$

2. **Προϋποθέσεις.** Δεχόμαστε ότι τα δύο δείγματα προέρχονται από πληθυσμούς που ακολουθούν κανονική κατανομή, οι διασπορές των οποίων είναι άγνωστες αλλά ίσες. Τα ιστογράμματα των δεδομένων των δύο δειγμάτων (αριστερά της ομάδας A και δεξιά της ομάδας B) και οι αντίστοιχες καμπύλες της κανονικής κατανομής που προσαρμόζονται σε αυτά, φαίνονται στο παρακάτω σχήμα:



3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Υπολογίζουμε την τιμή του στατιστικού t και την τιμή p . Συγκεκριμένα οι υπολογισμοί δίνουν $t=7.757$ και τιμή $p<0.001$. Για να υλοποιήσουμε την υπολογιστική διαδικασία με την R, καλούμε και πάλι την εντολή `t.test()`, στην οποία δίνουμε τα δεδομένα των δύο δειγμάτων (`x=music.A`, `y=music.B`), τη μορφή της εναλλακτικής υπόθεσης (`alternative="greater"`) και τέλος δηλώνουμε ότι οι άγνωστες διακυμάνσεις των δύο πληθυσμών είναι ίσες (`var.equal=TRUE`).

```
> t.test(x=music.A, y=music.B, alternative="greater",
var.equal=TRUE)$statistic
```

t

7.757083

και

```
> t.test(x=music.A, y=music.B, alternative="greater",
var.equal=TRUE)$p.value
```

[1] 4.716724e-09

4. **Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Για επίπεδο σημαντικότητας $\alpha = 0.05$ παρατηρούμε ότι $p \leq \alpha$, οπότε απορρίπτουμε τη μηδενική υπόθεση, δηλαδή απορρίπτεται ο ισχυρισμός ότι οι μαθητές που διδάχθηκαν με τη μέθοδο Α έχουν μέση επίδοση μ_1 που είναι ίση με τη μέση επίδοση μ_2 των μαθητών που διδάχθηκαν με τη μέθοδο Β. Επομένως, με βάση τα αποτελέσματα δεν μπορούμε να πούμε ότι αποδεικνύεται η εναλλακτική υπόθεση αλλά ότι ενισχύεται ο ισχυρισμός για την ανωτερότητα της μεθόδου Α σε σχέση με τη μέθοδο Β ως προς την αποτελεσματικότητα στη διδασκαλία της μουσικής.

■

II. Εξαρτημένα δείγματα.

Ο έλεγχος αυτός ονομάζεται **έλεγχος συγκρίσεων κατά ζεύγη**, διότι τα δεδομένα προέρχονται από τα ίδια υποκείμενα (π.χ. μαθητές) στα οποία γίνεται καταγραφή των τιμών μιας μεταβλητής (π.χ. επίδοση σε ένα μάθημα) πριν και μετά από μια παρέμβαση ή ένα γεγονός (π.χ. χρήση νέων τεχνολογιών). Ως εκ τούτου τα δεδομένα έχουν τη μορφή (x_i, y_i) , $i = 1, 2, \dots, n$, όπου x_i και y_i είναι οι τιμές της μεταβλητής (πριν και μετά αντίστοιχα) του ίδιου υποκειμένου i . Τότε υπολογίζουμε τις διαφορές μεταξύ των τιμών της μεταβλητής σε κάθε υποκείμενο, δηλαδή

$$d_1 = x_1 - y_1, \dots, d_n = x_n - y_n.$$

Η διαδικασία του ελέγχου υποθέσεων έχει ως εξής:

1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Οι στατιστικές υποθέσεις είναι
 H_0 : Δεν υπάρχει διαφορά στη μέση τιμή των δύο δειγμάτων ή
 $H_0: d = \mu_1 - \mu_2 = 0$
 H_a : Υπάρχει διαφορά στη μέση τιμή των δύο δειγμάτων ή $H_a: d = \mu_1 - \mu_2 \neq 0$
2. **Προϋποθέσεις.** Έστω ότι τα δύο δείγματα προέρχονται από τον ίδιο πληθυσμό και οι διαφορές τους d_i ακολουθούν κανονική κατανομή με άγνωστη διακύμανση.
3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Το κριτήριο ελέγχου δίνεται από τη σχέση

$$\frac{\text{μέσος όρος διαφορών}}{\text{τυπική απόκλιση του μέσου}}$$

και εκφράζεται από το στατιστικό

$$t = \frac{\bar{d}}{s/\sqrt{n}},$$

όπου $\bar{d}$ είναι ο δειγματικός μέσος όρος και s η δειγματική τυπική απόκλιση των $d_1, \dots, d_n$. Η στατιστική συνάρτηση t ακολουθεί κάτω από τις παραπάνω προϋποθέσεις κατανομή t με $n-1$ βαθμούς ελευθερίας, όταν ισχύει η υπόθεση H_0 .

4. **Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Απορρίπτουμε τη μηδενική υπόθεση αν $p \leq \alpha$ ενώ δεν απορρίπτουμε τη μηδενική υπόθεση αν $p > \alpha$.

Παράδειγμα 14 (Συνέχεια – t test με δύο εξηρημένα δείγματα)

Για να συγκρίνουμε την επίδραση των νέων τεχνολογιών στην εκπαιδευτική διαδικασία θα χρησιμοποιήσουμε τα δεδομένα από 10 ζεύγη στα οποία οι αριθμοί εκφράζουν τις δύο αξιολογήσεις των 10 μαθητών της ίδια τάξης, πριν και μετά τη χρήση νέων τεχνολογιών. Η μηδενική υπόθεση είναι ότι η χρήση των νέων τεχνολογιών δεν επιφέρει σημαντική διαφορά στην επίδοση των μαθητών, ενώ η εναλλακτική ότι βελτιώνει την επίδοσή τους. Τα δεδομένα είναι τα εξής:

(15,18), (12,11), (18,19), (12,16), (17,18), (10,11), (14,16), (19,19), (19,20) και (16,18).

Μια πρώτη εικόνα για το ερευνητικό ερώτημα προκύπτει από τις διαφορές των βαθμών κάθε μαθητή, που είναι:

$$d_i = -3, 1, -1, -4, -1, -1, -2, 0, -1, -2.$$

Παρατηρούμε ότι όλες οι διαφορές είναι αρνητικές εκτός από τη δεύτερη που είναι θετική κατά μία μόνο μονάδα και την όγδοη η οποία είναι μηδενική. Επίσης με την ακολουθία των εντολών

```
> method.A <- c(15, 12, 18, 12, 17, 10, 14, 19, 19, 16)
> method.B <- c(18, 11, 19, 16, 18, 11, 16, 19, 20, 18)
> d <- method.A - method.B
> mean(d)
```

[1] -1.4

αφενός εισάγουμε στην R τα δεδομένα αφετέρου υπολογίζουμε τη μέση τιμή των διαφορών d_i , η οποία είναι αρνητική. Αυτές οι μετρήσεις αποτελούν μια ένδειξη ότι η χρήση νέων τεχνολογιών βελτιώνει την επίδοση των μαθητών, κάτι που επιβεβαιώνεται και από την εφαρμογή του t test.

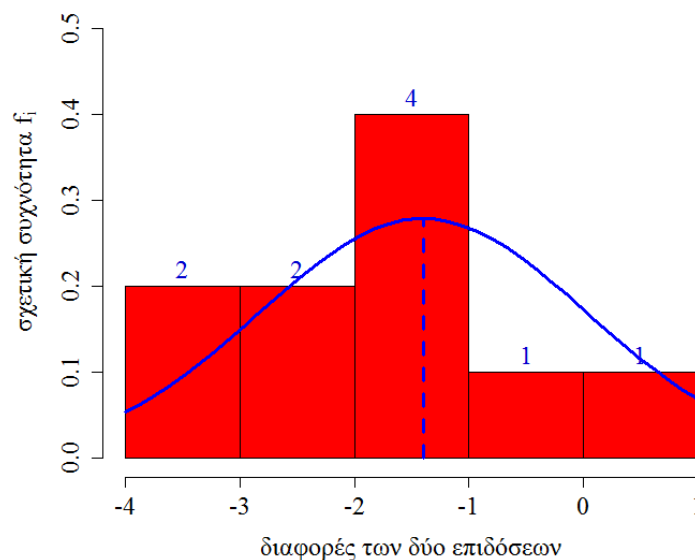
1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Καθορίζουμε τις στατιστικές υποθέσεις.

H_0 : Η χρήση των νέων τεχνολογιών δεν επιφέρει σημαντική διαφορά στην επίδοση των μαθητών ή $H_0: d = \mu_1 - \mu_2 = 0$, έναντι

H_a : Η χρήση των νέων τεχνολογιών βελτιώνει την επίδοση των μαθητών ή

$H_a: d = \mu_1 - \mu_2 < 0 \Leftrightarrow \mu_1 < \mu_2$.

2. **Προϋποθέσεις.** Η προϋπόθεση για να εφαρμόσουμε το t test συγκρίσεων ζευγών είναι οι τιμές $d_1, \dots, d_n$ να ακολουθούν κανονική κατανομή κάτι που επιβεβαιώνεται μέσω του εποπτικού ελέγχου στο παρακάτω σχήμα:



3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Υπολογίζουμε την τιμή του στατιστικού t και την τιμή p . Συγκεκριμένα οι υπολογισμοί δίνουν $t = -3.096$ και τιμή $p = 0.0064$. Για να υλοποιήσουμε την υπολογιστική διαδικασία με την R, καλούμε και πάλι την εντολή `t.test()`, στην οποία δίνουμε τα δεδομένα των δύο δειγμάτων (`x=method.A`, `y=method.B`), τη

μορφή της εναλλακτικής υπόθεσης (`alternative="less"`) και τέλος δηλώνουμε ότι πρόκειται για έλεγχο συγκρίσεως κατά ζεύγη (`paired=TRUE`).

```
> t.test(x=method.A, y=method.B, alternative="less",  
paired=TRUE)$statistic  
t  
-3.096281  
> t.test(x=method.A, y=method.B, alternative="less",  
paired=TRUE)$p.value  
[1] 0.006399521
```

4. **Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Με επίπεδο σημαντικότητας $\alpha = 0.05$ παρατηρούμε ότι $p \leq \alpha$, οπότε απορρίπτουμε τη μηδενική υπόθεση. Δηλαδή με αμελητέα πιθανότητα λάθους, ίση με 0.64%, απορρίπτεται ο ισχυρισμός ότι η χρήση των νέων τεχνολογιών δεν επιφέρει σημαντική διαφορά στην επίδοση των μαθητών. Επομένως, τα αποτελέσματά μας συνηγορούν υπέρ της άποψης ότι η χρήση νέων τεχνολογιών επιφέρει στατιστικώς σημαντική βελτίωση στην επίδοση των μαθητών σε σχέση με τον «παραδοσιακό» τρόπο διδασκαλίας.

■

- d. **Έλεγχοι υποθέσεων που βασίζονται στο κριτήριο του χ^2 .** Το κριτήριο του χ^2 είναι κατάλληλο όταν εργαζόμαστε με ποιοτικές μεταβλητές. Αυτό σημαίνει ότι δεν επεξεργαζόμαστε αριθμητικά δεδομένα με τα οποία συνήθως μετράμε ένα χαρακτηριστικό των συμμετεχόντων αλλά τους εντάσσουμε σε κατηγορίες. Για παράδειγμα, δεν μας ενδιαφέρει ο γενικός μέσος όρος ενός ελέγχου επίδοσης, αλλά η ταξινόμηση του μαθητή σε μια από τις κατηγορίες με τους χαρακτηρισμούς επίδοσης «μέτρια», «καλά», «πολύ καλά», «άριστα».

Το χ^2 εφαρμόζεται είτε για μια ποιοτική μεταβλητή ως έλεγχος καλής προσαρμογής είτε για δύο ποιοτικές μεταβλητές ως έλεγχος ανεξαρτησίας.

I. Έλεγχος καλής προσαρμογής.

Ο έλεγχος καλής προσαρμογής για μια ποιοτική μεταβλητή εξετάζει αν υπάρχουν στατιστικώς σημαντικές διαφορές μεταξύ των κατηγοριών του δείγματος, συγκρίνοντας τις συχνότητες εμφάνισης των κατηγοριών του δείγματος (πραγματικές συχνότητες) με τις συχνότητες που αναμένουμε (αναμενόμενες συχνότητες) όταν ισχύει η μηδενική υπόθεση, δηλαδή όταν υποθέτουμε ότι όλες οι

κατηγορίες έχουν την ίδια συχνότητα εμφάνισης. Για την κατανόηση της χρήσης και ωφελιμότητας του ελέγχου παραθέτουμε ένα παράδειγμα από το βιβλίο των Ρούσσου & Τσαούση (2011, σελ. 469).

Παράδειγμα 15 (Ελεγχος υποθέσεων – χ^2 test καλής προσαρμογής)

Σε μία έρευνα έχουν καταγραφεί οι τρόποι μελέτης 120 φοιτητών διαφόρων σχολών, στις κατηγορίες μεθοδική (καθημερινή μελέτη), ακανόνιστη (χωρίς κανένα προγραμματισμό) και συνδυαστική (συνδυασμός των δύο προηγούμενων). Οι αντίστοιχες συχνότητες ν_i , $i=1,2,3$ των συμμετεχόντων στις τρεις κατηγορίες δίνονται στον πίνακα που ακολουθεί:

Τρόπος Μελέτης			
Μεθοδική	Ακανόνιστη	Συνδυαστική	Σύνολο
51	27	42	120

Το ερώτημα που απασχολεί τον ερευνητή είναι αν τα δεδομένα αντικατοπτρίζουν στατιστικώς σημαντικές διαφορές στον τρόπο μελέτης των φοιτητών ή αν πρόκειται για τυχαίες διαφορές μεταξύ των τριών μεθόδων.

■

Η διαδικασία του ελέγχου καλής προσαρμογής είναι:

- Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Αν υποθέσουμε ότι εντάσσουμε τον πληθυσμό σε k κατηγορίες, οι στατιστικές υποθέσεις του ελέγχου είναι η μηδενική H_0 : οι συχνότητες εμφάνισης των k κατηγοριών του πληθυσμού δεν είναι διαφορετικές μεταξύ τους ή $H_0: \nu_1 = \nu_2 = \dots = \nu_k$, έναντι της εναλλακτικής H_a : οι συχνότητες εμφάνισης των k κατηγοριών του πληθυσμού είναι διαφορετικές μεταξύ τους ή H_a : δεν ισχύει ότι $\nu_1 = \nu_2 = \dots = \nu_k$.
- Προϋποθέσεις.** Ένα άλλο σημαντικό χαρακτηριστικό των ελέγχων υποθέσεων που βασίζονται στο κριτήριο του χ^2 είναι ότι δεν είναι παραμετρικοί έλεγχοι υποθέσεων, δηλαδή δεν απαιτείται καμία προϋπόθεση για την ακριβή μορφή της κατανομής του πληθυσμού. Παρόλα αυτά υπόκειται σε κάποιους περιορισμούς (Ρούσσος & Τσαούσης, 2011, σελ. 484-486):

- πρέπει οι συμμετέχοντες να εμφανίζονται μόνο μία φορά στον πίνακα συχνοτήτων
- ο συνολικός αριθμός των συμμετεχόντων να είναι τουλάχιστον 50
- η πραγματική συχνότητα του κάθε κελιού να είναι τουλάχιστον 5.

3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Με το κριτήριο του χ^2 συγκρίνεται η πραγματική συχνότητα Π του κάθε κελιού του πίνακα με την αντίστοιχη αναμενόμενη συχνότητα A . Όσο μεγαλύτερες είναι οι διαφορές $\Pi - A$, τόσο πιθανότερο είναι να απορριφθεί η μηδενική υπόθεση, δηλαδή να υπάρχουν στατιστικώς σημαντικές διαφορές μεταξύ των κατηγοριών του δείγματος. Το χ^2 αντανακλά το μέγεθος των διαφορών μεταξύ των πραγματικών και των αναμενόμενων συχνοτήτων $\Pi - A$ μέσω του τύπου:

$$\chi^2 = \sum \frac{(\Pi - A)^2}{A}.$$

Ο αριθμητής στον παραπάνω τύπο υπολογίζει και αθροίζει όλες τις διαφορές $\Pi - A$, οι οποίες υψώνονται στο τετράγωνο για να είναι πάντα θετικές. Ο παρονομαστής είναι αναγκαίος ώστε να συγκρίνουμε την διαφορά $(\Pi - A)^2$ με το πλήθος των περιπτώσεων της κατηγορίας, που εκφράζεται από το A . Για παράδειγμα, αν $\Pi = 36$, $A = 21$, τότε $\Pi - A = 15$, όπως και στην περίπτωση $\Pi = 336$, $A = 321$, $\Pi - A = 15$. Όμως η συνεισφορά του όρου $\frac{(\Pi - A)^2}{A}$, στο συνολικό άθροισμα είναι πολύ μικρότερη στη δεύτερη περίπτωση εξαιτίας του μεγαλύτερου A σε σχέση με την πρώτη περίπτωση. Συγκεκριμένα, είναι

$$\frac{(\Pi - A)^2}{A} = \frac{(36 - 21)^2}{21} = 10.71 \text{ και } \frac{(\Pi - A)^2}{A} = \frac{(336 - 321)^2}{321} = 0.70.$$

Το κριτήριο ελέγχου χ^2 ακολουθεί την κατανομή του χ^2 με $k - 1$ βαθμούς ελευθερίας, όταν ισχύει η υπόθεση H_0 .

4. **Κανόνες απόφασης και εξαγωγή συμπερασμάτων.** Απορρίπτουμε τη μηδενική υπόθεση αν $p \leq \alpha$ ενώ δεν απορρίπτουμε τη μηδενική υπόθεση αν $p > \alpha$.

Παράδειγμα 15 (Συνέχεια – χ^2 test καλής προσαρμογής)

Ο έλεγχος της υπόθεσής μας έχει ως εξής:

1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Από τα δεδομένα αναδύεται το ερευνητικό ερώτημα της προτίμησης των φοιτητών στη μεθοδική μελέτη σε βάρος των άλλων δύο κατηγοριών (ακανόνιστη και συνδυαστική). Συγκεκριμένα οι στατιστικές υποθέσεις είναι:

H_0 : οι συχνότητες των τριών τύπων μελέτης είναι ίσες μεταξύ τους ή

$H_0: \nu_1 = \nu_2 = \nu_3$, έναντι

H_a : οι συχνότητες των τριών τύπων μελέτης δεν είναι ίσες μεταξύ τους ή H_a :

δεν ισχύει $\nu_1 = \nu_2 = \nu_3$.

2. **Προϋποθέσεις.** Ικανοποιούνται οι προϋποθέσεις εφαρμογής του ελέγχου, διότι το μέγεθος του δείγματος είναι 120, επομένως κατά πολύ μεγαλύτερο από το 50 και οι συχνότητες κάθε κελιού 51, 27 και 42 είναι μεγαλύτερες από το 5.

3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Η διαδικασία υπολογισμού του χ^2 είναι η εξής:

- Υπολογίζουμε τις αναμενόμενες συχνότητες βασιζόμενοι στη μηδενική υπόθεση $\nu_1 = \nu_2 = \nu_3$. Επειδή επιπλέον ισχύει $\nu_1 + \nu_2 + \nu_3 = 120$ προκύπτει ότι $A = 40$.
- Αντικαθιστούμε τις τιμές των πραγματικών και αναμενόμενων συχνοτήτων στον τύπο του χ^2 :

$$\chi^2 = \frac{(51-40)^2}{40} + \frac{(27-40)^2}{40} + \frac{(42-40)^2}{40} = 3.025 + 4.225 + 0.100 = 7.35$$

Η υλοποίηση των παραπάνω υπολογισμών με τη βοήθεια της R, γίνεται ως εξής:

Καταρχήν, εισάγουμε στο RStudio ή στη βασική έκδοση της R, τα δεδομένα με το διάνυσμα

```
> method.study <- c(51, 27, 42)
```

Για επιβεβαίωση πληκτρολογούμε

```
> method.study
```

```
[1] 51 27 42
```

Για τον υπολογισμό του χ^2 χρησιμοποιούμε την εντολή

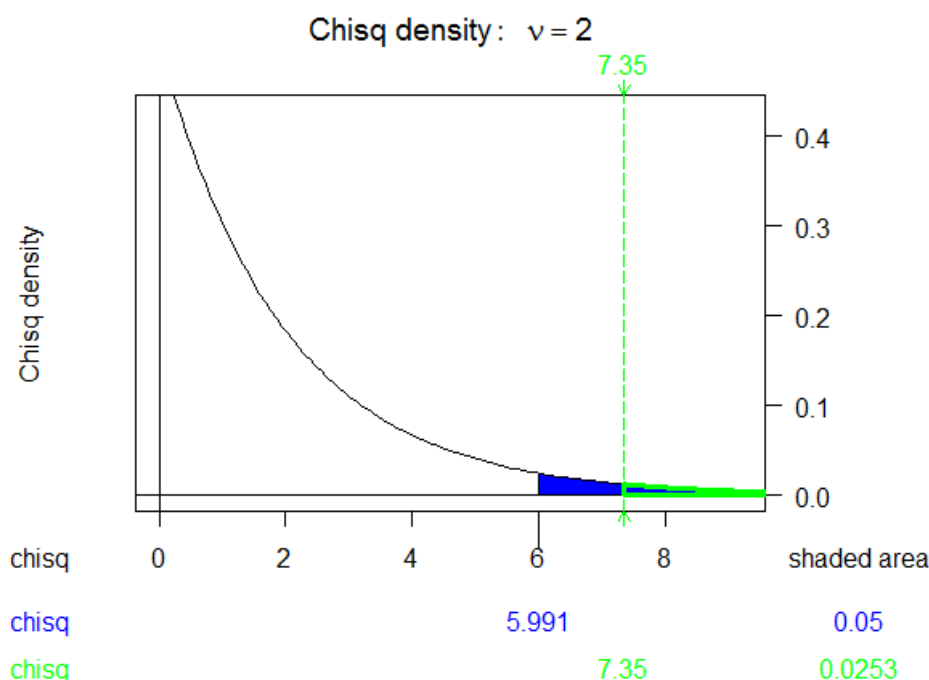

```
> chisq.test(method.study)
```

Chi-squared test for given probabilities

```
data: method.study
```

```
X-squared = 7.35, df = 2, p-value = 0.02535
```

από την οποία επιβεβαιώνουμε την τιμή του χ^2 , τους βαθμούς ελευθερίας της κατανομής που είναι $k-1=2$ και βρίσκουμε την τιμή p που ισούται με 0.02535.



Σχήμα 20. Η καμπύλη συχνοτήτων της κατανομής χ^2 που αντιστοιχεί στο Παράδειγμα 15 και είναι το αποτέλεσμα της εντολής `prob.chisq.pvalue()`.

4. **Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Συγκρίνουμε την τιμή $p = 0.02535$ με το επίπεδο σημαντικότητας της επιλογής μας (π.χ. $\alpha = 0.05$), από όπου προκύπτει ότι πρέπει να απορρίψουμε τη μηδενική υπόθεση. Η στατιστική απόφαση επιβεβαιώνεται από το Σχήμα 20 στο οποίο το μπλε τμήμα του οριζόντιου άξονα αντιστοιχεί στην κρίσιμη περιοχή, δηλαδή στην περιοχή απόρριψης της μηδενικής υπόθεσης. Αυτό σημαίνει ότι αν η τιμή του στατιστικού χ^2 , υπερβαίνει το 5.991 τότε απορρίπτουμε τη μηδενική υπόθεση, διότι η πιθανότητα να συμβεί κάτι τέτοιο είναι πολύ μικρή και μικρότερη ή ίση με 5% (εμβαδόν μπλε χωρίου). Αυτό ακριβώς συμβαίνει και στο Παράδειγμα

15, στο οποίο το $\chi^2 = 7.35 > 5.991$. Η πιθανότητα λανθασμένης απόφασης ισούται με την τιμή p που είναι 2.53% (εμβαδόν μπλε χωρίου που περικλείεται από την πράσινη γραμμή).

Επομένως, τα παραπάνω ευρήματα είναι επαρκή για να υποστηρίξουν την υπόθεση ότι οι φοιτητές των διάφορων σχολών χρησιμοποιούν ένα τρόπο μελέτης περισσότερο από τους άλλους.

■

II. Έλεγχος ανεξαρτησίας.

Εκτός από τον έλεγχο καλής προσαρμογής, το στατιστικό κριτήριο χ^2 εξετάζει την ύπαρξη ανεξαρτησίας μεταξύ δύο ποιοτικών μεταβλητών. Για να επιτευχθεί αυτός ο στόχος χρησιμοποιούνται **πίνακες συνάφειας** (contingency tables), όπως στο παρακάτω παράδειγμα.

Παράδειγμα 16 (Έλεγχος υποθέσεων – χ^2 έλεγχος ανεξαρτησίας)

Ο παρακάτω πίνακας συνάφειας:

	Αποφοίτησε	Δεν αποφοίτησε	Σύνολο
Με ενισχυτική διδασκαλία	73	12	85
Χωρίς ενισχυτική διδασκαλία	43	39	82
Σύνολο	116	51	167

αναπαριστά τις συχνότητες 167 ατόμων, για τα οποία εξετάζουμε αν αποφοίτησαν από τη δευτεροβάθμια εκπαίδευση (ποιοτική μεταβλητή X , με τιμές ΝΑΙ και ΟΧΙ) και αν παρακολούθησαν κατά τη διάρκεια των σπουδών τους ενισχυτική διδασκαλία επιπροσθέτως του βασικού προγράμματος σπουδών (ποιοτική μεταβλητή Y , με τιμές ΝΑΙ και ΟΧΙ). Στον πίνακα αυτό οι αριθμοί παριστάνουν συχνότητες εμφάνισης. Για παράδειγμα το κελί με τον αριθμό 73 σημαίνει ότι 73 μαθητές που παρακολούθησαν πρόγραμμα ενισχυτικής διδασκαλίας αποφοίτησαν από τη δευτεροβάθμια εκπαίδευση, ενώ το κελί με συχνότητα 43 αναφέρεται σε αυτούς που αποφοίτησαν χωρίς να παρακολουθήσουν πρόγραμμα ενισχυτικής διδασκαλίας επί συνόλου 167

ατόμων που συμμετείχαν στην έρευνα.

Η πρώτη εικόνα του πίνακα συνάφειας δείχνει ότι το πρόγραμμα ενισχυτικής διδασκαλίας συνετέλεσε ώστε οι μαθητές που το παρακολούθησαν να έχουν μεγαλύτερο ποσοστό αποφοίτησης από τη δευτεροβάθμια εκπαίδευση σε σχέση με τους μαθητές που δεν το παρακολούθησαν. Με τη βοήθεια του ελέγχου ανεξαρτησίας του χ^2 θα εξετάσουμε αν τα δεδομένα μας οδηγούν στο συμπέρασμα ότι υπάρχει στατιστικώς σημαντική εξάρτηση μεταξύ της παρακολούθησης των προγραμμάτων ενισχυτικής διδασκαλίας της ολοκλήρωσης των σπουδών στη δευτεροβάθμια εκπαίδευση.

■

Η διαδικασία του ελέγχου ανεξαρτησίας χ^2 έχει ως εξής:

- 1. Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Αν υποθέσουμε ότι έχουμε δύο ποιοτικές μεταβλητές με k και r κατηγορίες αντίστοιχα, οι στατιστικές υποθέσεις του ελέγχου είναι η μηδενική H_0 : οι μεταβλητές είναι ανεξάρτητες, έναντι της εναλλακτικής H_a : οι μεταβλητές δεν είναι ανεξάρτητες.
- 2. Προϋποθέσεις.** Οι προϋποθέσεις εφαρμογής του ελέγχου ανεξαρτησίας χ^2 είναι ακριβώς ίδιες με αυτές του ελέγχου καλής προσαρμογής.
- 3. Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Αν και το κριτήριο ελέγχου παραμένει το ίδιο με αυτό του χ^2 καλής προσαρμογής, δηλαδή είναι

$$\chi^2 = \sum \frac{(\Pi - A)^2}{A},$$

είναι σημαντικό να κατανοήσουμε τη λογική που διέπει τις αναμενόμενες συχνότητες και τον τρόπο υπολογισμού τους. Για αυτό θα χρησιμοποιήσουμε τα δεδομένα του Παραδείγματος 16. Παρατηρούμε ότι από τους ερωτηθέντες οι μισοί περίπου παρακολούθησαν πρόγραμμα ενισχυτικής διδασκαλίας και οι άλλοι μισοί όχι (85 και 82 άτομα αντίστοιχα). Θα περιμέναμε λοιπόν, αν δεν υπήρχε σχέση μεταξύ του προγράμματος σπουδών και της σχολικής διαρροής (δηλαδή του ποσοστού των μαθητών που δεν αποφοιτούν), να εμφανιζόταν το ίδιο περίπου ποσοστό, δηλαδή περίπου 50%, μεταξύ αυτών που αποφοίτησαν και αυτών που δεν αποφοίτησαν. Έτσι, για το κελί με συχνότητα 73, η

αναμενόμενη συχνότητα προκύπτει ως εξής:

$$A = \left(\frac{\text{ποσοστό όσων παρακολούθησαν}}{\text{ενισχυτική διδασκαλία}} \right) \cdot (\text{αριθμός ατόμων που αποφοίτησαν}),$$

που ισούται με $A = \frac{85}{167} \cdot 116 = 59.042$. Ομοίως, για το κελί με συχνότητα 43, η

αναμενόμενη συχνότητα προκύπτει ως εξής:

$$A = \left(\frac{\text{ποσοστό όσων δεν παρακολούθησαν}}{\text{ενισχυτική διδασκαλία}} \right) \cdot (\text{αριθμός ατόμων που αποφοίτησαν})$$

που ισούται με $A = \frac{82}{167} \cdot 116 = 56.958$. Ομοίως μπορούν να υπολογιστούν οι

αναμενόμενες συχνότητες των υπολοίπων κελιών.

Τυποποιώντας την παραπάνω διαδικασία, καταλήγουμε στον τύπο με τον οποίο υπολογίζονται οι αναμενόμενες συχνότητες:

$$A = \frac{(\text{συχνότητα γραμμής}) \cdot (\text{συχνότητα στήλης})}{\text{σύνολο συχνοτήτων}}.$$

Επομένως, ο πίνακας συνάφειας με τις αντίστοιχες αναμενόμενες συχνότητες σε παρένθεση γίνεται:

	Αποφοίτησε	Δεν αποφοίτησε	Σύνολο
Με ενισχυτική διδασκαλία	73 (59.042)	12 (25.958)	85
Χωρίς ενισχυτική διδασκαλία	43 (56.958)	39 (25.042)	82
Σύνολο	116	51	167

Το κριτήριο του χ^2 στο έλεγχο ανεξαρτησίας ακολουθεί κατανομή χ^2 με $(k-1) \cdot (r-1)$ βαθμούς ελευθερίας, όταν ισχύει η υπόθεση H_0 . Στη συνήθη περίπτωση των πινάκων συνάφειας με 2 γραμμές ($r = 2$) και 2 στήλες ($k = 2$) η κατανομή είναι χ^2 με 1 βαθμό ελευθερίας.

- Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Απορρίπτουμε τη μηδενική υπόθεση αν $p \leq \alpha$ ενώ δεν απορρίπτουμε τη μηδενική υπόθεση αν $p > \alpha$.

Παράδειγμα 16 (Συνέχεια – χ^2 έλεγχος ανεξαρτησίας)

Ο έλεγχος της υπόθεσής μας έχει ως εξής:

1. **Διατύπωση ερευνητικών ερωτημάτων και καθορισμός των στατιστικών υποθέσεων.** Οι στατιστικές υποθέσεις είναι:

H_0 : Η παρακολούθηση προγράμματος ενισχυτικής διδασκαλίας και η σχολική διαρροή είναι ανεξάρτητες

H_a : Η παρακολούθηση προγράμματος ενισχυτικής διδασκαλίας και η σχολική διαρροή δεν είναι ανεξάρτητες.

2. **Προϋποθέσεις.** Οι προϋποθέσεις εφαρμογής του ελέγχου ανεξαρτησίας ικανοποιούνται διότι οι συμμετέχοντες εμφανίζονται μόνο μία φορά στον πίνακα συνάφειας, ο συνολικός αριθμός των συμμετεχόντων είναι τουλάχιστον 50 και η πραγματική συχνότητα κάθε κελιού είναι τουλάχιστον 5.

3. **Κατασκευή και υπολογισμός κριτηρίου στατιστικού ελέγχου.** Το στατιστικό χ^2 ισούται με 22.01. Πράγματι,

$$\chi^2 = \frac{(73-59.042)^2}{59.042} + \frac{(12-25.958)^2}{25.958} + \frac{(43-56.958)^2}{56.958} + \frac{(39-25.042)^2}{25.042} = 22.01$$

Η τιμή p του ελέγχου είναι μικρότερη από 0.0001 .

Η υλοποίηση των παραπάνω υπολογισμών με τη βοήθεια της R, γίνεται ως εξής:

Καταρχήν, εισάγουμε στο RStudio ή στη βασική έκδοση της R, τα δεδομένα με δύο διανύσματα που αντιστοιχούν στις γραμμές του πίνακα συνάφειας.

```
> curriculum.yes <- c(73, 12)
```

```
> curriculum.no <- c(43, 39)
```

Κατόπιν συνδυάζουμε τα δύο διανύσματα για να δημιουργήσουμε τον πίνακα συνάφειας.

```
> curriculum.graduation <- rbind(curriculum.yes, curriculum.no)
```

```
> curriculum.graduation
```

```
      [,1] [,2]
```

```
curriculum.yes    73    12
```

```
curriculum.no     43    39
```

Αν θέλουμε να προσθέσουμε ονόματα στις γραμμές και στήλες του πίνακα συνάφειας, τότε:

```
> rownames(curriculum.graduation) <- c("Με ενισχυτική
διδασκαλία", "Χωρίς ενισχυτική διδασκαλία")
> colnames(curriculum.graduation) <- c("Αποφοίτησε", "Δεν
αποφοίτησε")
```

Ο πίνακας συνάφειας παίρνει τότε την πιο γνώριμη μορφή

```
> curriculum.graduation
```

	Αποφοίτησε	Δεν αποφοίτησε
Με ενισχυτική διδασκαλία	73	12
Χωρίς ενισχυτική διδασκαλία	43	39

Τέλος, για τον υπολογισμό του χ^2 χρησιμοποιούμε την εντολή

```
> chisq.test(curriculum.graduation, correct=FALSE)
```

Pearson's Chi-squared test

```
data: curriculum.graduation
X-squared = 22.0059, df = 1, p-value = 2.718e-06
```

Η μεταβλητή `correct=FALSE`, αποτρέπει την R να διορθώσει την αριθμητική τιμή του χ^2 . Η δυνατότητα διόρθωσης προσφέρεται από την R και βασίζεται στην άποψη ότι το κριτήριο χ^2 δεν είναι πολύ ασφαλές στην περίπτωση που έχουμε 1 βαθμό ελευθερίας (Ρούσσο & Τσαούσης, 2011, σελ. 486). Για να επιβεβαιώσουμε ότι η αριθμητικές τιμές των X-squared και p-value είναι ακριβείς, αλλά κυρίως για να εμπεδώσουμε τον τρόπο που οι στατιστικές ψηφιακές εφαρμογές δημιουργούν συναρτήσεις όπως η `chisq.test`, μπορούμε να κάνουμε τα εξής:

```
> Expected <- chisq.test(curriculum.graduation)$expected
> Expected
```

	Αποφοίτησε	Δεν αποφοίτησε
Με ενισχυτική διδασκαλία	59.04192	25.95808
Χωρίς ενισχυτική διδασκαλία	56.95808	25.04192

που υπολογίζει τις αναμενόμενες τιμές του πίνακα συνάφειας. Ακολουθώντας με τον πίνακα `Observed` παίρνουμε τις πραγματικές συχνότητες εμφάνισης.

```
> Observed <- chisq.test(curriculum.graduation)$observed
> Observed
```

	Αποφοίτησε	Δεν αποφοίτησε
Με ενισχυτική διδασκαλία	73	12
Χωρίς ενισχυτική διδασκαλία	43	39

και με τον τύπο

```
> chi.sq <- (Observed - Expected)^2 / Expected
> chi.sq
```

	Αποφοίτησε	Δεν αποφοίτησε
Με ενισχυτική διδασκαλία	3.299827	7.505489
Χωρίς ενισχυτική διδασκαλία	3.420552	7.780080

όλους τους όρους της μορφής $\frac{(\Pi - A)^2}{A}$. Η τιμή του χ^2 προκύπτει από το

άθροισμα όλων των όρων του πίνακα `chi.sq`.

```
> sum(chi.sq)
```

```
[1] 22.00595
```

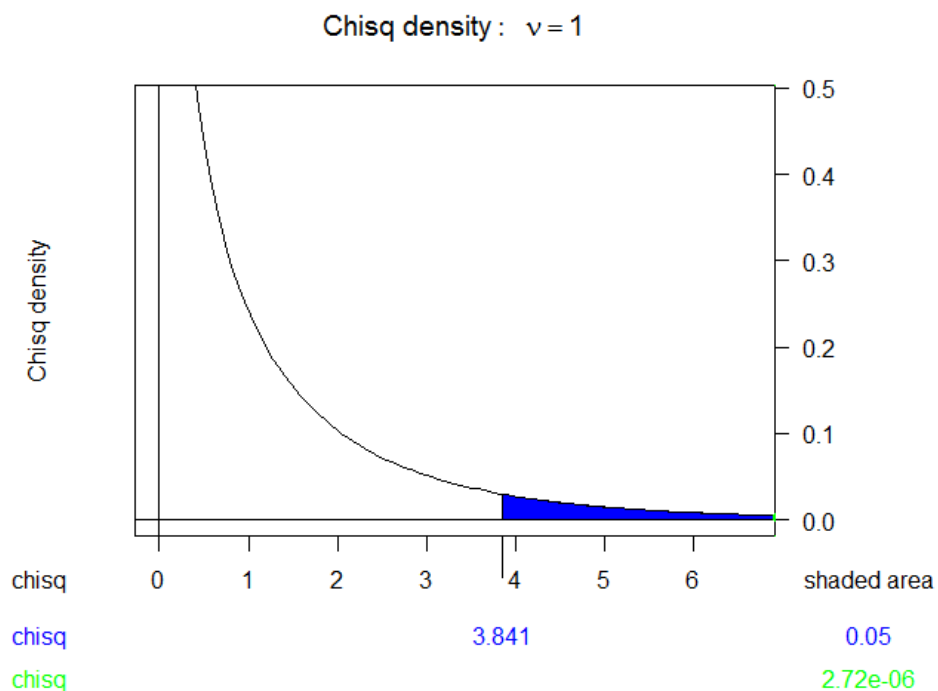
Η τιμή p του ελέγχου προκύπτει μέσω της εντολής

```
> 1-pchisq(22.0059, df=1)
```

```
[1] 2.718136e-06
```

Τα αποτελέσματα αυτά επιβεβαιώνονται και από το Σχήμα 21, το οποίο δημιουργήθηκε με την εντολή

```
> prob.chisq.pvalue(df=1, chisq.obs=22.0059)
```



Σχήμα 21. Η καμπύλη συχνοτήτων της κατανομής χ^2 που αντιστοιχεί στο Παράδειγμα 16.

- Κανόνας απόφασης και εξαγωγή συμπερασμάτων.** Επειδή η τιμή $p < 0.0001$, η μηδενική υπόθεση με επίπεδο σημαντικότητας 0.05 απορρίπτεται. Επειδή η τιμή p είναι πολύ μικρή μπορούμε με σχετική ασφάλεια (πιθανότητα λάθους μικρότερη από 0.01%) να απορρίψουμε τη μηδενική υπόθεση της ανεξαρτησίας. Καταλήγουμε λοιπόν στο συμπέρασμα ότι η ενισχυτική

διδασκαλία και η σχολική διαρροή δεν είναι ανεξάρτητες, επομένως έχει νόημα να εξετάσουμε επιπλέον την ωφελιμότητα της ενισχυτικής διδασκαλίας στον περιορισμό της σχολικής διαρροής.

