Έλεγχοι Υποθέσεων

Καθηγητής Χρήστος Γ. Μασούρος

Ο έλεγχος υποθέσεων αναφέρεται στις ιδιότητες μιας άγνωστης παραμέτρου του πληθυσμού:

- Ο πληθυσμός της πόλης έχει προσβληθεί από τον ιό της γρίπης
- Η μέση τιμή των χρημάτων των μαθητών της Γ Λυκείου κατά την διάρκεια της εξαήμερης εκδρομής τους είναι $\mu = 120€$

Κάθε υπόθεση συνοδεύεται από μια εναλλακτική:

- Ο πληθυσμός της πόλης δεν έχει προσβληθεί από τον ιό της γρίπης
- Η μέση τιμή των χρημάτων των μαθητών της Γ Λυκείου κατά την διάρκεια της εξαήμερης εκδρομής τους είναι $\mu \neq 120€$

- Μια υπόθεση μπορεί να είναι αληθής ή ψευδής
- Η απόρριψη ή όχι μιας υπόθεσης μπορεί να είναι σωστή ή λανθασμένη

Η αρχική υπόθεση, συμβολίζεται με H_0, είναι μια υπόθεση για μία ή περισσότερες παραμέτρους του πληθυσμού. Η υπόθεση αυτή θεωρούμε ότι ισχύει μέχρις ότου έχουμε επαρκή στατιστικά ευρήματα για να αποφασίσουμε να την απορρίψουμε. Π.χ. $\mu = 120$€

Η εναλλακτική υπόθεση, συμβολίζεται με H_1, είναι η υπόθεση που καλύπτει όλες τις άλλες περιπτώσεις που δεν συμπεριλαμβάνονται στην αρχική. Π.χ. $\mu \neq 120$€

Οι H_0 και H_1 είναι:

Αμοιβαία αποκλειόμενες

- Μόνο μία μπορεί να είναι αληθής.

Συμπληρωματικές

- Ο συνδυασμός τους καλύπτει όλα τα πιθανά ενδεχόμενα και επομένως είτε η μία είτε η άλλη θα είναι αληθής

Ο στατιστικός έλεγχος υποθέσεων (hypothesis testing) είναι μια συμπερασματική διαδικασία/μεθοδολογία που ανέπτυξε η Στατιστική και βρίσκει εφαρμογή σε στοχαστικά προβλήματα απόφασης μεταξύ δύο εναλλακτικών υποθέσεων.

Η μία υπόθεση έχει επικρατήσει να συμβολίζεται με H_0 και ονομάζεται **μηδενική υπόθεση** (null hypothesis), και η άλλη με H_1 και ονομάζεται **εναλλακτική υπόθεση** (alternative hypothesis).

Η γενική ιδέα της διαδικασίας στατιστικού ελέγχου υποθέσεων είναι η εξής:

θέτουμε ως μηδενική υπόθεση H_0 αυτή για την οποία αμφιβάλουμε, αυτή που αμφισβητείται, και εξετάζουμε αν ένα τυχαίο δείγμα που παίρνουμε από τον πληθυσμό συνηγορεί υπέρ της αποδοχής της ή της απόρριψής της, έναντι της εναλλακτικής H_1.

Δηλαδή, η H_0, απορρίπτεται ή γίνεται αποδεκτή με βάση το τι παρατηρείται στο τυχαίο δείγμα που πήραμε από τον πληθυσμό.

Πιο συγκεκριμένα:

Υποθέτοντας ότι η H_0 είναι αληθής, αν αυτό που παρατηρείται στο δείγμα είναι ακραίο, δηλαδή, αν έχει πολύ μικρή πιθανότητα να συμβεί, τότε απορρίπτουμε την H_0.

Στην αντίθετη περίπτωση, δηλαδή, αν αυτό που παρατηρείται στο δείγμα δεν είναι ακραίο-σπάνιο (όταν είναι αληθής η H_0) τότε το δείγμα που πήραμε δε μας δίνει αρκετές ενδείξεις για την απόρριψη της H_0 και «αποτυγχάνουμε να την απορρίψουμε».

Η διαδικασία όμως αυτή περιέχει «αβεβαιότητα». Αν κάτι έχει πολύ μικρή πιθανότητα να συμβεί δεν έπεται ότι είναι αδύνατο να συμβεί.

Αν κρίνουμε ότι αυτό που παρατηρείται στο τυχαίο δείγμα είναι ακραίο και <u>απορρίψουμε</u> την H_0, τότε ακριβώς ένα από τα παρακάτω μπορεί να συνέβη:

(α) είτε η H_0 πράγματι **δεν είναι αληθής**, όποτε αποφασίσαμε **σωστά**,

(β) είτε η H_0 **είναι αληθής** και το ακραίο οφείλεται στην τύχη, δηλαδή, συνέβη κάτι σπάνιο (εμφανίσθηκε ένα δείγμα που σπάνια εμφανίζεται). Στην περίπτωση αυτή, απορρίψαμε **λανθασμένα** την H_0

Αυτό το σφάλμα ονομάζεται ***σφάλμα τύπου Ι***

Ανάλογα, είναι δυνατόν, λανθασμένα να <u>μην απορρίψουμε</u> την H_0. Δηλαδή, να αποτύχουμε να απορρίψουμε την H_0, ενώ είναι αληθής η H_1. Αυτό το σφάλμα ονομάζεται ***σφάλμα τύπου ΙΙ***

Αναφύεται το ερώτημα:

Πώς υπολογίζονται οι πιθανότητες

- σφάλματος τύπου Ι (λανθασμένη απόρριψη της H_0),

  P(*σφάλμα τύπου I*) = P(*απόρριψη της H_0 | αληθής η H_0*)

και

- σφάλματος τύπου ΙΙ (λανθασμένη μη απόρριψη της H_0),

  P(*σφάλμα τύπου II*) = P(*μη απόρριψη της H_0 | αληθής η H_1*).

Θα απαντήσουμε μέσα από ένα παράδειγμα

Παράδειγμα 1

Οι λαμπτήρες με κεραμικό καυστήρα που κατασκευάζει μια εταιρεία για τον φωτισμό πόλεων, έχουν μέσο χρόνο ζωής 15.000 ώρες με τυπική απόκλιση 1.750 ώρες. Η κατασκευάστρια εταιρεία, προκειμένου να συμμετάσχει στην αναβάθμισης της υπάρχουσας εγκατάστασης φωτισμού μιας πόλης, μέσω της αντικατάστασης των ήδη υπαρχόντων παλαιών λαμπτήρων νατρίου με τους νέους, ισχυρίζεται ότι βελτίωσε τον κεραμικό καυστήρα που χρησιμοποιεί και πέτυχε έτσι ακόμη μεγαλύτερο ωφέλιμο χρόνο ζωής των λαμπτήρων της.

Παράδειγμα 1

Για να ελεγχθεί ο ισχυρισμός της κατασκευάστριας εταιρείας, **ως μηδενική υπόθεση θέτουμε την H_0: μ=15000 ώρες**, δηλαδή, αυτή η οποία αμφισβητείται από τον ισχυρισμό που ελέγχουμε.

Γενικά, η H_0 δηλώνει ότι στον πληθυσμό η κατάσταση παραμένει αμετάβλητη. Για αυτό έχει επικρατήσει να λέγεται μηδενική υπόθεση, διότι αυτή διαλαμβάνει τη μηδενική μεταβολή στην τιμή της ελεγχόμενης παραμέτρου. Ή διαφορετικά, ότι η ανεξάρτητη μεταβλητή (βελτίωση του κεραμικού καυστήρα του λαμπτήρα στο παράδειγμά μας) δεν έχει επίδραση στην εξαρτημένη μεταβλητή (αύξηση του ωφέλιμου χρόνου ζωής των λαμπτήρων του παραδείγματος).

Παράδειγμα 1

Ως εναλλακτική υπόθεση θέτουμε την H_1: $\mu > 15000$ ώρες, δηλαδή, η H_1 δηλώνει ότι η βελτίωση του κεραμικού καυστήρα επηρεάζει, και ειδικότερα αυξάνει, τον ωφέλιμο χρόνο ζωής των λαμπτήρων.

Γενικά, η H_1 δηλώνει ότι στον πληθυσμό υπάρχει μεταβολή ή αλλιώς, ότι η ανεξάρτητη μεταβλητή έχει επίδραση στην εξαρτημένη μεταβλητή για τον πληθυσμό.

Παράδειγμα 1

Ο έλεγχος που μόλις διατυπώσαμε, είναι ένας μονόπλευρος και ειδικότερα δεξιόπλευρος έλεγχος. Γενικότερα, οι έλεγχοι,

$$H_0: \mu \le \mu_0 \qquad\qquad H_0: \mu \ge \mu_0$$
$$H_1: \mu > \mu_0 \quad \text{και} \quad H_1: \mu < \mu_0$$

ονομάζονται **μονόπλευροι** (one-tailed) έλεγχοι (δεξιόπλευρος και αριστερόπλευρος αντίστοιχα) και ο έλεγχος,

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

ονομάζεται **αμφίπλευρος** (two-tailed).

Παράδειγμα 1

Αφού διατυπώσαμε την υπόθεση ότι η άγνωστη μέση τιμή του πληθυσμού του ωφέλιμου χρόνου ζωής των λαμπτήρων μετά τη βελτίωση του κεραμικού καυστήρα είναι 15.000 ώρες (H_0 : μ = 15.000 ώρες), παίρνουμε ένα τυχαίο δείγμα λαμπτήρων από το σύνολο της παραγωγής του εργοστασίου και μετράμε τον ωφέλιμο χρόνου ζωής κάθε λαμπτήρα του δείγματος. Έστω ότι λαμβάνουμε το τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ μεγέθους n=50, το οποίο μας έδωσε τις μετρήσεις $x_1, x_2, \ldots, x_{50}$ των οποίων η μέση τιμή είναι $\bar{x}$**=15.500** ώρες.

Αυτό που παρατηρείται στο δείγμα, συμφωνεί άραγε με την υπόθεση

H_0 : μ = 15.000 ώρες

ή είναι εναντίον της H_0 και υπέρ της H_1;

Παράδειγμα 1

Για να απαντήσουμε σε αυτό το ερώτημα, πρέπει να ορίσουμε μια κατάλληλη στατιστική συνάρτηση, η οποία, όταν ισχύει η H_0, να ακολουθεί γνωστή κατανομή (χωρίς άγνωστες παραμέτρους) ώστε να μπορούμε να υπολογίσουμε τις απαιτούμενες για τον έλεγχο πιθανότητες.

Στο παράδειγμά μας, που αφορά στον έλεγχο της μέσης τιμής μ, του πληθυσμού, είναι λογικό να επιλέξουμε ως στατιστική συνάρτηση, τη δειγματική μέση τιμή

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Παράδειγμα 1

Η κατανομή της $\bar{X}$ είναι γνωστή:

Σύμφωνα με το Κ.Ο.Θ. αφού το μέγεθος του δείγματος που πήραμε είναι αρκετά μεγάλο έχουμε:

$$\bar{X} \sim N\left(15000, \frac{1750^2}{50}\right)$$

Ισοδύναμα, τυποποιώντας, μπορούμε να επιλέξουμε ως στατιστική συνάρτηση, τη

$$Z = \frac{\bar{X} - 15000}{1750/\sqrt{50}} = \frac{(\bar{X} - 15000)\sqrt{50}}{1750} \sim N(0,1)$$

Παράδειγμα 1

Έτσι «*Αυτό που παρατηρείται στο δείγμα*» στο παράδειγμά μας εκφράζεται από τη στατιστική συνάρτηση $\bar{X}$ με τιμή, στο συγκεκριμένο δείγμα που πήραμε, 15.500 ώρες ή, ισοδύναμα, από την

$$Z = \frac{(\bar{X} - 15000)\sqrt{50}}{1750}$$

με τιμή, στο συγκεκριμένο δείγμα που πήραμε:

$$Z = \frac{(15500 - 15000)\sqrt{50}}{1750} = 2{,}02$$

Ας δούμε τώρα πως – με ποιο κριτήριο – αποδεχόμαστε ή απορρίπτουμε την H_0

Παράδειγμα 1

1ος τρόπος: Ορίζουμε το ανεκτό μέγεθος σφάλματος τύπου Ι

Αν η H_0: μ=15000 ώρες είναι αληθής, είναι λογικό να αναμένουμε ότι η τιμή της στατιστικής συνάρτησης $\bar{X}$ στο δείγμα που πήραμε (δηλαδή, η μέση τιμή του δείγματος) θα είναι «κοντά» στην τιμή 15000 ώρες. Αντιθέτως, αν η H_0: μ=15000 ώρες δεν είναι αληθής, αναμένουμε η μέση τιμή του δείγματος να είναι «μακριά» από το 15000 (μεγαλύτερη σύμφωνα με την H_1). Πρέπει λοιπόν να καθορίσουμε το «κοντά» και το «μακριά». Έτσι ορίζουμε μια τιμή ***c*** με βάση την οποία θα κρίνεται αν η δειγματική μέση τιμή $\bar{X}$ βρίσκεται «μακριά» από την μ = 15000 ώρες. Έτσι, αν επιλέξουμε c=15250 ώρες, τότε επειδή 15500>15250, η H_0 απορρίπτεται, ενώ αν ορίσουμε c=15650 ώρες, τότε επειδή 15500<15650, η H_0 δεν απορρίπτεται.

Παράδειγμα 1

Τίθεται, επομένως, το ερώτημα: *πώς επιλέγουμε την τιμή της σταθεράς c;*

Για οποιοδήποτε c, έχουμε:

$$P(\bar{X} \geq c) = P\left(\frac{\bar{X} - 15000}{1750/\sqrt{50}} \geq \frac{c - 15000}{1750/\sqrt{50}}\right)$$

$$= P\left(Z \geq \frac{(c - 15000)\sqrt{50}}{1750}\right) = 1 - \Phi\left(\frac{(c - 15000)\sqrt{50}}{1750}\right)$$

Παράδειγμα 1

Από τα παραπάνω, είναι φανερό ότι η τιμή της σταθεράς **c** καθορίζει την πιθανότητα σφάλματος τύπου I που κάνουμε. Έτσι, μπορούμε να επιλέξουμε την τιμή της **c** ως εξής:

Ορίζουμε ένα *μέγιστο ανεκτό μέγεθος σφάλματος τύπου I* και με βάση αυτό υπολογίζουμε την τιμή της **c**.

Το ανεκτό επίπεδο σφάλματος τύπου I που προκαθορίζουμε, συμβολίζεται με ***α*** και ονομάζεται *επίπεδο σημαντικότητας* (*level of significance*) του ελέγχου. Συνήθως το επίπεδο σημαντικότητας ***α***, ορίζεται ίσο με **0,01** ή **0,05**

Παράδειγμα 1

Ας ολοκληρώσουμε τον έλεγχο, στο παράδειγμά μας, θέτοντας επίπεδο σημαντικότητας α=0,05. Πρέπει να επιλέξουμε το c έτσι ώστε:

$$P(\bar{X} \geq c) \leq 0{,}05 \Leftrightarrow P\left(\frac{\bar{X} - 15000}{1750/\sqrt{50}} \geq \frac{c - 15000}{1750/\sqrt{50}}\right) \leq 0{,}05$$

$$\Leftrightarrow P\left(Z \geq \frac{(c - 15000)\sqrt{50}}{1750}\right) \leq 0{,}05 \Leftrightarrow$$

Παράδειγμα 1

$$\Leftrightarrow 1 - \Phi\left(\frac{(c - 15000)\sqrt{50}}{1750}\right) \leq 0{,}05 \Leftrightarrow$$

$$\Leftrightarrow \Phi\left(\frac{(c - 15000)\sqrt{50}}{1750}\right) \geq 0{,}95 \Leftrightarrow$$

$$\Leftrightarrow \frac{(c - 15000)\sqrt{50}}{1750} \geq z_{0,05} = 1{,}645 \Leftrightarrow$$

$$\Leftrightarrow c \geq 15000 + 1{,}645\frac{1750}{\sqrt{50}} \Leftrightarrow c \geq 15.407$$

Παράδειγμα 1

Έτσι, επιλέγοντας c=15.407 έχουμε ότι το $\bar{X}$ το οποίο είναι 15.500 είναι μεγαλύτερο από το c και επομένως απορρίπτουμε την H_0 με πιθανότητα εσφαλμένης απόρριψης το πολύ 0,05.

Αν όμως επιβάλλουμε πιο αυστηρό έλεγχο στον ισχυρισμό του εργοστασίου και κάνουμε τον έλεγχο με μικρότερη ανοχή σε εσφαλμένη απόρριψη της H_0, π.χ. με α=0,01 τότε έχουμε:

Παράδειγμα 1

$$P(\bar{X} \geq c) \leq 0{,}01 \Leftrightarrow P\left(\frac{\bar{X} - 15000}{1750/\sqrt{50}} \geq \frac{c - 15000}{1750/\sqrt{50}}\right) \leq 0{,}01$$

$$\Leftrightarrow P\left(Z \geq \frac{(c - 15000)\sqrt{50}}{1750}\right) \leq 0{,}01 \Leftrightarrow$$

Παράδειγμα 1

$$\Leftrightarrow 1-\Phi\left(\frac{(c-15000)\sqrt{50}}{1750}\right) \leq 0{,}01 \Leftrightarrow$$

$$\Leftrightarrow \Phi\left(\frac{(c-15000)\sqrt{50}}{1750}\right) \geq 0{,}99 \Leftrightarrow$$

$$\Leftrightarrow \frac{(c-15000)\sqrt{50}}{1750} \geq z_{0,01} = 2{,}33 \Leftrightarrow$$

$$\Leftrightarrow c \geq 15000 + 2{,}33\frac{1750}{\sqrt{50}} \Leftrightarrow c \geq 15.577$$

Παράδειγμα 1

Έτσι, επιλέγοντας c=15.577 έχουμε ότι το $\bar{X}$ το οποίο είναι 15.500 είναι μικρότερο από το c και επομένως δεν απορρίπτουμε την H_0.

Η σταθερά ***c*** ονομάζεται **κρίσιμη τιμή** ή **όριο απόρριψης** (critical value, rejection limit)

Παράδειγμα 1

2^{ος} τρόπος: Υπολογίζουμε την P-Τιμή (P-Value) του δείγματος

Με δεδομένο ότι η H_0: μ=15000 ώρες είναι αληθής, υπολογίζουμε την πιθανότητα να εμφανισθεί η τιμή $\bar{x} = 15.500$ που εμφανίσθηκε στο δείγμα που πήραμε ή κάποια μεγαλύτερή της (δηλαδή, προς την κατεύθυνση της H_1). Ζητάμε δηλαδή την πιθανότητα $P(\bar{X} \geq 15.500)$. Έχουμε:

$$P(\bar{X} \geq 15.500) = \ P\left(\frac{\bar{X} - 15.000}{1.750/\sqrt{50}} \geq \frac{15.500 - 15.000}{1.750/\sqrt{50}}\right) =$$

$$= \ P(Z \geq 2{,}02) = 1 - \Phi(2{,}02) = 0{,}0217$$

Παράδειγμα 1

Αυτή η πιθανότητα ονομάζεται P-Τιμή (P-Value) του δείγματος ή κρίσιμο επίπεδο (critical level) και είναι η πιθανότητα να εμφανισθεί η τιμή της στατιστικής συνάρτησης ελέγχου που εμφανίσθηκε στο παράδειγμά μας ($\bar{x} = 15.500\ ώρες$) ή κάποια πιο μακριά ακόμη, προς την κατεύθυνση της H_1. Έτσι, υπολογίζοντας την P-τιμή του δείγματος, γνωρίζουμε πόσο πιθανή ήταν η εμφάνιση του δείγματος που πήραμε με την υπόθεση ότι η H_0 είναι αληθής. Επομένως, όσο πιο μικρή είναι η P-Τιμή τόσο ισχυρότερες ενδείξεις εναντίον της H_0 προκύπτουν από το συγκεκριμένο τυχαίο δείγμα.

Παράδειγμα 1

Στο παράδειγμά μας, υπολογίσαμε ότι η P-Τιμή του δείγματος που πήραμε, είναι ίση με 0.0217 ή 2.17%. Επομένως, αν κάνουμε τον έλεγχο σε επίπεδο σημαντικότητας α=0,01=1%, δηλαδή, αν θέλουμε πιο «ισχυρές αποδείξεις» εναντίον της H_0 από αυτές που παρατηρούνται στο δείγμα, τότε **δεν την απορρίπτουμε**, ενώ αν κάνουμε τον έλεγχο σε επίπεδο σημαντικότητας α=0,05=5% τότε **την απορρίπτουμε** (γιατί στην περίπτωση αυτή, απαιτούμε «ασθενέστερες αποδείξεις» εναντίον της H_0).

Συμπερασματικά:

Η P-Τιμή είναι η ελάχιστη τιμή του επιπέδου σημαντικότητας για την οποία απορρίπτεται η H_0.

Ο κανόνας απόφασης διαμορφώνεται ως εξής:

- αν $\alpha \geq$ P-Τιμή, τότε, σε επίπεδο σημαντικότητας α, η H_0 **απορρίπτεται**.
- αν $\alpha <$ P-Τιμή, τότε, σε επίπεδο σημαντικότητας α, η H_0 **δεν απορρίπτεται.**