Έλεγχος υπόθεσης

Εκτός από τις μέσες τιμές, τυπικές αποκλίσεις κλπ, θέλουμε να βρούμε κατά πόσον αυτές οι παρατηρούμενες τάσεις εξαρτώνται από συγκεκριμένες συνθήκες ή προϋποθέσεις. Δηλαδή διατυπώνουμε μια θεωρία ή νόμο που συνοψίζει αυτό που βλέπουμε, υπό μορφή **υπόθεσης**.
Η υπόθεση αυτή πρέπει να ελεγχθεί σε σχέση με άλλες πιθανές εξηγήσεις.

Η πιο απλή εξήγηση που φυσικά πρέπει να αποκλείσουμε πρώτη είναι τα αποτελέσματα να προέκυψαν κατά τύχη.
Άρα, θα διατυπώσουμε μια αρχική "υπόθεση αναφοράς", τη **μηδενική υπόθεση** (συνήθως γράφεται συμβολικά ως H_0) κατά την οποία τα αποτελέσματα που πήραμε ήταν τυχαία.
Ελέγχουμε κατά πόσον, δηλαδή *με ποια πιθανότητα* είναι δυνατό να συμβεί αυτό. Για το σκοπό αυτό, ορίζουμε ένα ποσοστό, κάτι σαν την εμπιστοσύνη που λέγαμε προηγουμένως, αρκετά μικρό, π.χ. 5% ή 1%.
Αν η πιθανότητα ικανοποίησης της μηδενικής υπόθεσης για το δείγμα μας είναι κάτω από αυτό το ποσοστό, τότε τα αποτελέσματα είναι **σημαντικά** (significant) και μπορούμε να πούμε ότι η δική μας **εναλλακτική** υπόθεση, υπερτερεί της υπόθεσης ότι προέκυψαν τυχαία.

Το πώς ακριβώς διατυπώνουμε τη μηδενική υπόθεση εξαρτάται από τη φύση του προβλήματος.
Π.χ. αν έχουμε δύο ενδεχόμενα, τότε αυτά μπορεί να είναι ανεξάρτητα που συνεπάγεται για το καθένα πιθανότητα $P = 50\%$

Παράδειγμα 1:
Έρευνα αν οι μύγες των φρούτων έλκονται από το μέλι ή από το ξύδι.
Μηδενική υπόθεση H_0: έλκονται εξίσου, άρα $P = 50\%$
Δείγμα: n μύγες που πιάνονται και ποσοστό p που πιάστηκαν με το μέλι.
Ορίζουμε εμπιστοσύνη 2% (δηλαδή δεχόμαστε τα αποτελέσματα ως σημαντικά αν είναι κάτω από 2% πιθανότητα να συμβεί το παρατηρούμενο αποτέλεσμα κατά τύχη).
Πάμε στη διωνυμική κατανομή για $p = 50\%$ και μέγεθος δείγματος 15.
Βρίσκουμε ότι για επιτυχίες (να πιαστούν με μέλι) = 12, 13, 14, 15, το άθροισμα των πιθανοτήτων είναι $1.7\% < 2\%$ ενώ για 11 επιτυχίες υπερβαίνει από μόνο του το 5% (είναι 0.059). Άρα, πρέπει να πιαστούν τουλάχιστον 12 μύγες με μέλι για να πούμε ότι είναι κάτω από 2% πιθανό να γίνει κατά τύχη.
Το σύνολο των πιθανών αποτελεσμάτων που είναι σημαντικά αποτελεί την **κρίσιμη περιοχή**
Αν πιάστηκαν π.χ. 13 μύγες τότε είναι μέσα στην κρίσιμη περιοχή και η εναλλακτική υπόθεση δικαιώνεται έναντι της μηδενικής.

Προσοχή! Αν θέλαμε να ελέγξουμε κατά πόσο είναι πιθανό να τις πιάσουμε με το ξύδι, τότε θα θεωρούσαμε ότι θα είχαμε 2 αντί για 13 επιτυχίες. Αλλά το αποτέλεσμα είναι ποσοτικά ίδιο (συμμετρική κατανομή), άρα το μόνο που μπορεί να πει είναι ότι είναι απίθανο να έχει τόσο λίγες επιτυχίες κατά τύχη.

Αυτό που μπορεί να πει σίγουρα είναι μια εναλλακτική υπόθεση κατά την οποία $P \neq 50\%$.

Άλλου είδους μηδενική υπόθεση: έστω ότι από την προηγούμενη πείρα μας γνωρίζουμε πως οι μύγες ζούνε γύρω στις 12 μέρες με τυπική απόκλιση πληθυσμού $\sigma = 2$. Αυτή είναι η νέα μηδενική υπόθεση H_0: $\mu = 12$.
εναλλακτική υπόθεση H_1: μέση τιμή $\mu \neq 12$.
Παίρνουμε μεγάλα δείγματα, των 50, και βρίσκουμε $2/\sqrt{50} = 0.28$ (βλ. Κεντρικό Οριακό Θεώρημα).
Θέτουμε επίπεδο σημαντικότητας 1%, δηλαδή εμπιστοσύνη 99% και βρίσκουμε από τον πίνακα τυπικής κανονικής κατανομής $z = \pm 2.58$

Αν από το δείγμα βρίσκουμε 12.5 μέρες αυτό αντιστοιχεί σε z = 1.77 άρα είναι μέσα στο διάστημα εμπιστοσύνης της μηδενικής υπόθεσης και επιβεβαιώνουμε ότι ζούνε γύρω στις 12 μέρες.

Σημαντική σημείωση: απαιτείται προσεκτικός **σχεδιασμός πειράματος** ώστε να αποκλειστούν άλλες πηγές επίδρασης στα αποτελέσματα, π.χ. ο άνεμος έσπρωχνε τις μύγες προς το μέλι.
Βλ. και πειράματα με φάρμακα: όταν δοκιμάζεται ένα νέο φάρμακο σε ομάδα ασθενών, διεξάγεται και το λεγόμενο τυφλό πείραμα (δηλαδή μία ομάδα παίρνει ένα ακίνδυνο ψευδοφάρμακο, "placebo", χωρίς τη δραστική ουσία) για να εξακριβωθεί μέσω της σύγκρισης, ο υποκειμενικός παράγοντας (αυθυποβολή των ασθενών που νομίζουν ότι θεραπεύονται). Αλλά επειδή το ίδιο πρόβλημα υπάρχει και με τους ερευνητές που μπορεί να υποβάλλουν στον εαυτό τους την ιδέα ότι βλέπουν βελτίωση στους ασθενείς που παίρνουν το φάρμακο, πιο σωστή μεθοδολογία είναι το λεγόμενο διπλό τυφλό πείραμα όπου ούτε οι ερευνητές ξέρουν ποια ομάδα παίρνει το αληθινό φάρμακο.

Υπάρχουν δύο είδη σφαλμάτων που θέλουμε να δούμε την πιθανότητά τους:
- Τύπου I (α) = εσφαλμένη θεωρία που επιβεβαιώνουμε κατά τύχη
- Τύπου II (β) = ορθή θεωρία που διαψεύδουμε κατά τύχη.

Όσο μεγαλώνει η πιθανότητα για το ένα μικραίνει για το άλλο (επίπεδο σημαντικότητας, μέγεθος δείγματος, διασπορά πληθυσμού). Επομένως αυτά τα δύο είναι ανταγωνιστικά και δε μπορούμε να τα μειώσουμε συγχρόνως. Θα πρέπει να αποφασίσουμε κατά περίπτωση.
Παραδείγματα:
- Καλύτερα να κρατήσω ένα φάρμακο που δε μου κάνει για περαιτέρω έρευνα παρά να πετάξω ένα χρήσιμο (απαιτείται μείωση του β)
- Καλύτερα να διαψεύσω κατά τύχη μια ορθή θεωρία παρά να επαληθεύσω κατά τύχη μια "πατάτα" (απαιτείται μείωση του α)

Διαφορά Μέσων

Παρόμοια με τον έλεγχο υποθέσεων εργαζόμαστε και για να συγκρίνουμε διαφορές μέσων τιμών και διασπορών από διαφορετικά δείγματα: κατά πόσον αυτές είναι τυχαίες ή αντανακλούν πραγματικές διαφορές;
Για τις διαφορές μέσων τιμών από δύο διαφορετικά δείγματα υπάρχουν δύο δυνατά σενάρια:
- τα δείγματα μπορούν να διαταχθούν σε αντιστοιχία ένα-με-ένα, π.χ. ζεύγη παρατηρήσεων πριν/μετά.
- ανεξάρτητα δείγματα (ίσως και με διαφορετικό μέγεθος) όπου δε γίνεται ή δεν έχει νόημα η ένα με ένα αντιστοίχιση των παρατηρήσεων

Πρώτη περίπτωση: δείγματα διατάξιμα κατά ζεύγη.
Παράδειγμα 2:
Δοκιμή συγκεκριμένης δίαιτας σε n=49 άτομα.
Παίρνω διαφορές ΔΒ = Βάρος (μετά) – Βάρος (πριν)
Αν δεν είχε αποτέλεσμα περιμένω μ(ΔΒ) = 0 = *μηδενική υπόθεση*
Εναλλακτική: έχασαν βάρος, άρα μ(ΔΒ) < 0
Αν ισχύει η μηδενική υπόθεση περιμένω κανονική κατανομή των ΔB_i . Αλλά σ = άγνωστο άρα χρησιμοποιώ $s = \sigma_{\bar{x}} = s/\sqrt{n}$ με τη βοήθεια του οποίου βρίσκω το z (ορθό αφού δείγμα > 30)
Έστω επίπεδο εμπιστοσύνης α = 0.05
Βρίσκω κρίσιμη περιοχή z <= -1.65
Έστω ότι n = 49, $\bar{x}$ = -5, s = 3.5 επομένως $\sigma_{\bar{x}} = 3.5/7 = 0.5$ => z = (-5 -0) / 0.5 = -10 σαφώς στην κρίσιμη περιοχή, άρα η δίαιτα έφερε αποτέλεσμα.

Αν το δείγμα ήταν κάτω των 30 ατόμων θα χρησιμοποιούσαμε κατανομή του t υποθέτοντας κανονική

κατανομή του πληθυσμού.

Παράδειγμα 3: σφάλμα τύπου ΙΙ.
Επειδή τα επίπεδα εμπιστοσύνης είναι λίγο πολύ αυθαίρετα, υπάρχει ο κίνδυνος να απορριφθούν παρατηρήσεις που ισχύουν, ως τυχαίες. Χρειάζεται προσοχή και διεύρυνση των δειγμάτων αν υπάρχει σχετική υποψία. Π.χ. παρατηρήσεις σε ζεύγη διδύμων που έχουν ανατραφεί χωριστά (σε οικογένεια και ίδρυμα) για την επίπτωση στο δείκτη ευφυίας iq έδωσε:

Ζεύγη	Οικογένεια	Ίδρυμα	Διαφορά
1	105	95	-10
2	95	83	-12
3	103	103	0
4	98	96	-2
5	103	97	-6

Μηδενική υπόθεση: μ = 0
Εναλλακτική: μ ≠ 0
Εμπιστοσύνη α = 5%
Υποθέτοντας κανονική κατανομή πληθυσμού, η κατανομή του t δίνει κρίσιμη περιοχή |t| >= 2.78
$\bar{x} = -6, s = 5.1 \Rightarrow t = (\bar{x} - \mu)/s\sqrt{n} = -6/2.3 = 2.61$ άρα μη σημαντικό αποτέλεσμα. (df = 5 – 1 = 4 και μοιράζοντας το 0.05 στις δύο ουρές της κατανομής βρίσκω P(4; 0.025) = 2.776)
ΑΛΛΑ το δείγμα είναι *πολύ* μικρό και η απόρριψη είναι *οριακή*. Αν πάρουμε εμπιστοσύνη α = 10% βρίσκουμε |t| >= 2.13 που περιλαμβάνει το δείγμα. Με την πιο αυστηρή εμπιστοσύνη υπάρχει κίνδυνος να απορρίψουμε μια σωστή θεωρία!

Δεύτερη περίπτωση: ανεξάρτητα δείγματα.
Κλασσικό παράδειγμα: δείγμα συγκρινόμενο με τυφλό δείγμα.
Εδώ δεν έχει νόημα να κάνουμε αντιστοιχία και να πάρουμε διαφορές γιατί δεν υπάρχει συγκεκριμένη αντιστοιχία.
Πώς εργαζόμαστε:
Οι παρατηρήσεις δίνουν δύο μέσες τιμές μ_1, μ_2
Διατυπώνουμε τη μηδενική υπόθεση H_0: $\mu_1 = \mu_2$
Εναλλακτική υπόθεση H_1: $\mu_1 \neq \mu_2$ ή $\mu_1 > \mu_2$ ή $\mu_1 < \mu_2$, ανάλογα με τη θεωρία μας.
Εμπιστοσύνη α = 1% ή 5% (ή όσο θέλουμε).
Χρησιμοποιούμε z ή t, ανάλογα με το μέγεθος του δείγματος και την "κανονικότητα" του πληθυσμού.

Αλλά, ορίζουμε $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ ή $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_1}}}$ όπου $s = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

Παράδειγμα 4: ύψος φυτών που καλλιεργούνται με και χωρίς φάρμακο.

	n	$\bar{x}$	s
1	100	27	5
2	100	29	4

Βρίσκουμε z = 3.13. Αν α = 1%, τοτε |z| <= 2.33 και αν η εναλλακτική είναι μ1 < μ2, τότε βρισκόμαστε στην κρίσιμη περιοχή.

Παράδειγμα 5: Έστω ότι προμηθευόμαστε ανταλλακτικά από διαφορετικούς προμηθευτές για λογαριασμό μιας βιομηχανικής μονάδας όπου εργαζόμαστε. Θέλουμε να δούμε αν υπάρχουν σημαντικές διαφορές απόδοσης, ποιότητας κλπ για να αποφασίσουμε με ποιον θα συνεχίσουμε να

συνεργαζόμαστε στο μέλλον.

	n	$\bar{x}$	s
1	50	150	10
2	100	153	5

Μηδενική υπόθεση H_0: $\mu_1 = \mu_2$
Εναλλακτική H_1: $\mu_1 \neq \mu_2$
Επίπεδο σημαντικότητας $\alpha = 1\%$
Κρίσιμη περιοχή $|z| >= 2.58$
Από τα δεδομένα βρίσκουμε $z = 2.0$ που δεν είναι στην κρίσιμη περιοχή άρα η διαφορά κρίνεται μη σημαντική.
Παραδοχές για την παραπάνω μεθοδολογία για ανεξάρτητα δείγματα:

- κανονική κατανομή *αμφότερων πληθυσμών*
- ίση τυπική απόκλιση $\sigma_1 = \sigma_2 = \sigma$ αμφότερων πληθυσμών

Φυσικά, δε γνωρίζω τα χαρακτηριστικά των πληθυσμών με ακρίβεια γιατί αν τα γνώριζα δε θα χρειαζόμουν τα δείγματα. Οπότε καταφεύγω σε εκτίμηση, π.χ. κάνω ιστόγραμμα των δειγμάτων. Αν το ένα είναι ασύμμετρο ή $s_1 >> s_2$ τότε πιθανότατα παραβιάζονται οι παραδοχές.

Παρόμοια εργαζόμαστε και *για t-test*: προϋποτίθενται κανονικά δείγματα, $\sigma_1 = \sigma_2$. Στην πράξη, θέλουμε δείγματα παρόμοιου μεγέθους χωρίς μεγάλη ασυμμετρία στην κατανομή τους.

Διαφορές Διασπορών

Πολλές φορές θέλουμε να ξέρουμε αν οι διακυμάνσεις διαφέρουν σημαντικά από δείγμα σε δείγμα επειδή π.χ. οι μεγάλες διακυμάνσεις είναι ανεπιθύμητες. Για παράδειγμα, εργαζόμαστε σε μια βιομηχανία που παράγει κάποιο υλικό και θέλουμε να προμηθευτούμε πρώτη ύλη της οποίας η σύσταση έχει διακυμάνσεις που μπορούμε να ανεχθούμε μέχρι ένα όριο πάνω από το οποίο θα προκύψουν ανεπιθύμητες αντιδράσεις που θα αλλοιώσουν το τελικό προϊόν. Για το σκοπό αυτό, συγκρίνουμε δείγματα διαφορετικών προμηθευτών για να αποφανθούμε για τις διαφορές των διασπορών τους.
Η γενική μεθοδολογία για τα προβλήματα αυτού του είδους συνίσταται στο να ορίσω το λόγο

$$F = \frac{s_1^2}{s_{2,}^2} \quad \text{όπου } s_1 > s_{2.}$$

Αν η κατανομή του πληθυσμού είναι κανομική, αποδεικνύεται ότι ο λόγος F ακολουθεί τη λεγόμενη “κατανομή του F” που . Αυτή δίνεται σε πίνακες με γραμμές και στήλες ανάλογα με τους βαθμούς ελευθερίας $df_i = n_i - 1$ όπου σε κάθε entry δίνεται η τιμή της ουράς (δηλαδή ανάποδα απ' ο,τι για την κανονική κατανομή) για επίπεδο σημαντικότητας 5% και 1%.

Σε περίπτωση που δεν ενδιαφερόμαστε ειδικά για $\sigma_1 > \sigma_2$ (μία ουρά) αλλά γενικά για $\sigma_1 \neq \sigma_2$ (δύο ουρές) τότε θα χρησιμοποιήσουμε το διπλάσιο του ποσοστού εμπιστοσύνης ή σημαντικότητας για να πάμε στον πίνακα, π.χ. αν θέσουμε $\alpha = 1\%$ θα πάμε στην αντίστοιχη τιμή για τα δεδομένα df_1, df_2 αλλά θα θεωρήσουμε ότι το αποτέλεσμα δίνεται με σημαντικότητα 2% να προκύψει τυχαία.