

A large school of fish swimming in clear blue water. The fish are silvery and appear to be moving in a coordinated pattern. The background is a deep, clear blue, suggesting an underwater environment. The text "FISHER'S EXACT TEST" is overlaid in the center in a bold, white, italicized font.

FISHER'S EXACT TEST



Fisher test

Fisher test : Είναι ένας μη παραμετρικός έλεγχος : $\begin{cases} \text{ανεξαρτησίας} \\ \text{ομοιογένειας} \end{cases}$

Ανάλυση διακριτών δεδομένων όταν τα δύο ανεξάρτητα δείγματα έχουν μικρό μέγεθος. Τα αποτελέσματα αυτών των δειγμάτων αντιστοιχούν στη μια από τις δύο τυχαίες κατηγορίες και κάθε στοιχείο αυτών είναι δυνατόν να έχει το ένα από τα δύο χαρακτηριστικά. Έστω n_1, n_2 δυο ανεξάρτητα τυχαία δείγματα με $N = n_1 + n_2$.

	Χαρ/κο A	Χαρ/κο B	Σύνολο
Ομάδα I	a	b	$n_1 = a + b$
Ομάδα II	c	d	$n_2 = c + d$
Σύνολο	$s_1 = a + c$	$s_2 = b + d$	$N = n_1 + n_2$



Fisher test

Η ακριβής πιθανότητα εμφάνισης του συγκεκριμένου 2x2 πίνακα, δεδομένου ότι τα περιθώρια αθροίσματα n_1 , n_2 , s_1 , s_2 παραμένουν σταθερά, δίνεται από την υπεργεωμετρική κατανομή :

$$P = \frac{\binom{s_1}{a} \binom{s_2}{b}}{\binom{N}{n_1}} = \frac{n_1! n_2! s_1! s_2!}{N! a! b! c! d!} \quad (1)$$

Το εύρος τιμών του a είναι $n_L \leq a \leq n_U$, όπου :

$$n_L = \max\{0, n_1 + s_1 - N\}$$

$$n_U = \min\{n_1, s_1\}$$



Fisher test

Κάποιοι στρατιώτες κάνουν εξάσκηση σαν αλεξιπτωτιστές. Ένα απόγευμα, 55 από αυτούς έλαβαν μέρος σε μια άσκηση κατά την οποία μερικοί από αυτούς προσεδάφιστηκαν στην περιοχή A και κάποιοι άλλοι στην περιοχή B. Από τους 15 που προσεδάφιστηκαν στην περιοχή A, 4 από αυτούς έπαθαν ζημιά στον αστράγαλο, ενώ από τους 40 που προσεδάφιστηκαν στην περιοχή B μόλις 2 από αυτούς υπέστησαν αυτόν τον τραυματισμό. Ο αριθμός των ατυχημάτων στην περιοχή A είναι αρκετά υψηλός, έτσι ο ιατρός που ήταν επικεφαλής αποφάσισε να διεξάγει έρευνα για το αν ο αριθμός των ατυχημάτων είναι τυχαίος ή εξαρτάται σε κάποιο βαθμό από την κάθε περιοχή. Σαν μηδενική υπόθεση έχουμε ότι δεν υπάρχει διαφορά της πιθανότητας τραυματισμού, κατά γενική αναλογία των τραυματισμένων στρατιωτών, για κάθε περιοχή προσεδάφισης.

Θα κάνω έλεγχο, χρησιμοποιώντας fisher test, έχοντας μηδενική υπόθεση $H_0 : p_1 = p_2$ έναντι της εναλλακτικής $H_1 : p_1 \neq p_2$.



Fisher test

Για ευκολία στον υπολογισμό βάζουμε τη μικρότερη τιμή των ομάδων στο πάνω αριστερό κελί.

Πίνακας : Αριθμός στρατιωτών τραυματισμένων και μη κατά την διάρκεια της άσκησης στις δυο περιοχές.

	Τραυμ/νοι	Μη τραυμ/νοι	Σύνολο
Περιοχή Β	2	38	40
Περιοχή Α	4	11	15
Σύνολο	6	49	55



Fisher test

Ο έλεγχος δεν εξετάζει μόνο την παρατηρούμενη περίπτωση, αλλά υπολογίζει την πιθανότητα πιο ακραίων περιπτώσεων με τα ίδια περιθώρια αθροίσματα. Ο αριθμός των ακραίων πιθανών πινάκων με σταθερά τα περιθώρια αθροίσματα είναι ίσος με το μικρότερο περιθώριο άθροισμα + 1. Έτσι :

➤ Εάν, η διαφορά $ad - bc$ είναι αρνητική, τότε οι ακραίες περιπτώσεις λαμβάνονται αν σταδιακά μειώσουμε τα κελιά των a, d και αυξήσουμε τα κελιά των b, c με την ίδια ποσότητα (συνήθως μονάδα).

➤ Εάν, η διαφορά $ad - bc$ είναι θετική, τότε οι ακραίες περιπτώσεις λαμβάνονται αν σταδιακά αυξάνουμε τα κελιά των a, d και μειώνουμε τα κελιά των b, c .



Fisher test

Πρέπει να βρούμε τις πιθανότητες που συνδέονται με τις ακραίες περιπτώσεις. Το πρόσημο της διαφοράς $ad - cb$ είναι :

$$\begin{aligned} a \cdot d - c \cdot b &= 2 \cdot 11 - 4 \cdot 38 \\ &= -130 < 0 \end{aligned}$$

Θέλουμε $n_L \leq a \leq n_U$ δηλαδή $0 \leq a \leq 6$

Όποτε μειώνουμε τα κελιά a, d σταδιακά και έχουμε :

1	39		40	0	40		40
5	10		15	6	9		15
6	49		55	6	49		55

Στη συνέχεια υπολογίζουμε την πιθανότητα όλων των πιθανών πινάκων από τον τύπο (1) για τα δοσμένα περιθωριακά αθροίσματα.



Fisher test

$$p_2 = \frac{40! 5! 6! 49!}{55! 2! 38! 4! 1! 1!} = 0.0357269$$

$$p_1 = \frac{40! 5! 6! 49!}{55! 1! 39! 5! 1! 0!} = 0.0051794$$

$$p_0 = \frac{40! 5! 6! 49!}{55! 0! 40! 6! 9!} = 0.0001726$$

Το p-value υπολογίζεται με το άθροισμα των πιθανοτήτων όλων των ακραίων περιπτώσεων και του δοσμένου πίνακα. Στο συγκεκριμένο παράδειγμα το p-value θα είναι:

$$\begin{aligned} p - value &= p_0 + p_1 + p_2 = \\ &= 0.0001726 + 0.0051794 + 0.0357269 = \\ &= 0.041 \end{aligned}$$



Fisher test

Στο S-Plus: Η εντολή συντάσσεται ως εξής:

```
fisher.test(x, y=NULL, node.stack.dim=1001,  
            value.stack.dim=10000, hybrid=F)
```

x : είναι είτε πίνακας δεδομένων με διαστάσεις όχι μικρότερες του 2 και μεγαλύτερες του 10 είτε διάνυσμα.

y : είναι διάνυσμα και εισάγεται μόνο στην περίπτωση που ο x είναι διάνυσμα έτσι ώστε να γίνει ο έλεγχος. Έχει την ίδια διάσταση με το x .

`node.stack.dim`, `value.stack.dim`: χρησιμοποιούνται όταν ο x είναι πίνακας μεγαλύτερος από 2×2 .

`hybrid`: αν είναι True χρησιμοποιείται διαφορετική διαδικασία προσέγγισης του p -value που προτάθηκε από τους Mehta και Patel.



Fisher test

Για το παράδειγμα που δόθηκε προηγουμένως:

Κατασκευάζουμε τον πίνακα των δεδομένων:

```
> par_c(4, 11, 2, 38)
> table.par_matrix(par, 2, 2, byrow=T)
> table.par
      [,1] [,2]
[1,]    4   11
[2,]    2   38
```

Ο Fisher-test δίνει:

```
>fisher.test(table.par)
```

```
Fisher's exact test
```

```
data: table.par
```

```
p-value = 0.041
```

```
alternative hypothesis: two.sided
```

Δηλαδή η H_0 απορρίπτεται.



Fisher test

Αν είχαμε κάνει X^2 έλεγχο:

```
> chisq.test(table.par)
```

```
Pearson's chi-square test with Yates' continuity  
correction
```

```
data: table.par
```

```
X-square = 3.2758, df = 1, p-value = 0.0703
```

```
Warning messages:
```

```
Expected counts < 5. Chi-square approximation may not  
be appropriate. in: chisq.test(table.par)
```

Αποδέχεται την H_0 .

Σημ.: Η διόρθωση Yates' χρησιμοποιείται στον X^2 έλεγχο όταν το μέγεθος δείγματος είναι < 50 για να μας δώσει καλύτερη προσέγγιση για το p-value.



Fisher test

Fisher test versus Chi-square test

Σε έλεγχο ανεξαρτησίας για έναν πίνακα ενδεχομένων 2×2 μπορούμε να χρησιμοποιήσουμε είτε τον έλεγχο Fisher είτε τον έλεγχο Chi-square.

Σύγκριση :

- Ο Fisher test είναι ιδανικός στις περιπτώσεις που στον πίνακα ενδεχομένων έχουμε πολύ μικρά μεγέθη (δηλαδή έχω αριθμό μικρότερο του 5).
- Με τη χρήση του Fisher test λαμβάνουμε πιο ακριβές p-value.
- Ο Chi-square test έχει απλούστερους υπολογισμούς, αλλά καταλήγει σε ένα μόνο προσεγγιστικό p-value.
- Ο Fisher test δεν χρησιμοποιείται όταν το μέγεθος του N είναι μεγαλύτερο του 200.
- Όταν οι αριθμοί είναι μεγάλοι τα p-value που προκύπτουν τόσο από τον Fisher test όσο από τον Chi-square test είναι σχεδόν παρόμοια.