

[Αρχική](#) / Τα μαθήματά μου / [HY-390.51 \(Εαρινό 2021\)](#)
/ Διάλεξη 13 & 14: Ανάλυση Μικροσυστοιχιών - Σχεδιασμός statistic για διαφορική έκφραση / [Σειρά ασκήσεων 4](#)

Ξεκίνησε στις	Παρασκευή, 9 Απρίλιος 2021, 4:12 πμ
Κατάσταση	Ολοκληρωμένο
Ολοκληρώθηκε στις	Τρίτη, 20 Απρίλιος 2021, 9:00 μμ
Χρόνος που χρειάστηκε	11 ημέρες 16 ώρες

Ερώτηση **1**
Ολοκλήρωση
Βαθμολογήθηκε
στα 1,00

Θεωρείστε το Position Weight Matrix (PWM) στον παρακάτω σύνδεσμο, καθώς και τις FASTA αλληλουχίες στον παρακάτω σύνδεσμο. Από κάθε FASTA αλληλουχία i , βρείτε την μέγιστη τιμή της εφαρμογής του PWM στην αλληλουχία, έστω ότι την τιμή αυτή την ονομαζουμε S_i . Ποια αλληλουχία δίνει την ελάχιστη τιμή του S_i .

Αν είναι η 1η αλληλουχία του FASTA αρχείου πρέπει να απαντήσετε 1
Αν είναι η δεύτερη αλληλουχία του FASTA πρέπει να απαντήσετε 2, κοκ.

Link για το PWM
http://139.91.162.101/teaching/hy390_2021/pwm_1293992.txt

Link για το αρχείο FASTA
http://139.91.162.101/teaching/hy390_2021/seqFile_1293992.FA

Απάντηση:

Ερώτηση **2**

Ολοκλήρωση

Βαθμολογήθηκε
στα 1,00

Θεωρείστε το Position Weight Matrix (PWM) στον παρακάτω σύνδεσμο, καθώς και τις FASTA αλληλουχίες στον παρακάτω σύνδεσμο. Από κάθε FASTA αλληλουχία i , βρείτε την μέγιστη τιμή της εφαρμογής του PWM στην αλληλουχία, έστω ότι την τιμή αυτή την ονομαζουμε S_i . Ποια αλληλουχία δίνει την MEDIAN (διάμεσος) τιμή του S_i .

Αν είναι η 1η αλληλουχία του FASTA αρχείου πρέπει να απαντήσετε 1

Αν είναι η δεύτερη αλληλουχία του FASTA πρέπει να απαντήσετε 2, κοκ.

Link για το PWM

http://139.91.162.101/teaching/hy390_2021/pwm_1293992.txt

Link για το αρχείο FASTA

http://139.91.162.101/teaching/hy390_2021/seqFile_1293992.FA

Απάντηση:

Ερώτηση **3**

Ολοκλήρωση

Βαθμολογήθηκε
στα 2,00

Θεωρείστε τις εξής περιπτώσεις όπου συγκρίνετε αν δύο δείγματα M, N, από τις κατανομές A και B έχουν στατιστικά ίδιους ή διαφορετικούς μεσους όρους.

Στην πρώτη περίπτωση (περίπτωση 1),

$M = \text{rnorm}(5, 100, 10)$

$N = \text{rnorm}(5, 101, 10)$

Στην δευτερη περίπτωση (περίπτωση 2):

$M = \text{rnorm}(50000, 100, 10)$

$N = \text{rnorm}(50000, 101, 10)$

Σε ποια σύγκριση περιμένετε να βρείτε μικρότερο p-value

Επιλέξτε ένα ή περισσότερα:

- ☒ a. Το p-value θα είναι μικρότερο στην δευτερη περίπτωση περίπτωση
- ☐ b. Το t-test εφαρμόζεται μόνο σε έκφραση γονιδίων, εδώ έχουμε την συνάρτηση rnorm και όχι έκφραση γονιδίων, άρα δεν κάνω t-test.
- ☐ c. Δεν μπορώ να κάνω t-test επειδή δεν ξέρω αν τα δείγματα μου ακολουθούν την κανονική κατανομή
- ☐ d. Το p-value θα είναι μικρότερο στην πρώτη περίπτωση
- ☐ e. Το p-value θα είναι ίσο επειδή οι μέσες τιμές και οι τυπικές αποκλίσεις των κατανομών είναι ίδιες στις περιπτώσεις 1 και 2.

Ερώτηση **4**

Ολοκλήρωση

Βαθμολογήθηκε
στα 1,00

Όταν γράφω κώδικα R και πρέπει να κάνω πράξεις σε vectors, τότε επιλέγω:

Επιλέξτε ένα:

- ☒ a. vectorizations (e.g. $v + w$) επειδή είναι πιο γρήγορα
- ☐ b. Δεν μπορώ να χρησιμοποιήσω vectorizations επειδή η R δεν τα επιτρέπει (πχ δεν γίνεται να κάνω $v+w$ για να προσθέσω στοιχείο-στοιχείο δύο vectors, v και w).
- ☐ c. apply functions, επειδή είναι κομψός κώδικας.
- ☐ d. vectorizations (e.g. $v + w$) επειδή σε αυτήν την περίπτωση δεν υπάρχει numeric overflow.
- ☐ e. for loops, επειδή τα χρησιμοποιεί η C

Ερώτηση **5**

Ολοκλήρωση

Βαθμολογήθηκε
στα 1,00

Κάνετε ένα πείραμα στο εργαστήριο όπου μετράτε την τιμή έκφρασης του Γονιδίου MRAXX78 σε τρία κοντρόλ ποντίκια (κατηγορία A) και σε τρία ποντίκια που τους χορηγείτε σουπερμαντολίνη (κατηγορία B). Βρίσκετε ότι ο μέσος όρος έκφρασης του γονιδίου 1 στα ποντίκια της κατηγορίας A είναι κατά 5 μονάδες μεγαλύτερος από ότι στα ποντίκια της κατηγορίας B. Τα επόμενα βήματα που κάνετε είναι τα εξής:

Επιλέξτε ένα:

- ☒ a. Είναι πιθανόν ότι η διακύμανση των τιμών έκφρασης εντός κάθε κατηγορίας (A και B) να είναι μεγάλη και επίσης το δείγμα μας είναι αρκετά μικρό, άρα πρέπει να κάνω στατιστικό τεστ, ώστε να δω αν η διαφορά αυτή είναι στατιστικά σημαντική.
- ☐ b. Δεν χρειάζεται να κάνω στατιστικό τεστ, η διαφορά είναι πολύ μικρή
- ☐ c. Θα κάνω στατιστικό τεστ επειδή μου το ζητάει ο καθηγητής μου
- ☐ d. Θα κάνω πολλά στατιστικά τεστ και θα πάρω όποιο μου δίνει $p\text{-value} < 0.05$
- ☐ e. Don't let p-values destroy your career
- ☐ f. Δεν χρειάζεται να κάνω στατιστικό τεστ, η διαφορά είναι πολύ μεγάλη

Ερώτηση **6**

Ολοκλήρωση

Βαθμολογήθηκε
στα 2,00

Θεωρείστε το Position Weight Matrix (PWM) στον παρακάτω σύνδεσμο, καθώς και τις FASTA αλληλουχίες στον παρακάτω σύνδεσμο. Από κάθε FASTA αλληλουχία i , βρείτε την μέγιστη τιμή της εφαρμογής του PWM στην αλληλουχία, έστω ότι την τιμή αυτή την ονομάζουμε S_i . Ποια αλληλουχία δίνει την μέγιστη τιμή του S_i .

Αν είναι η 1η αλληλουχία του FASTA αρχείου πρέπει να απαντήσετε 1

Αν είναι η δεύτερη αλληλουχία του FASTA πρέπει να απαντήσετε 2, κοκ.

Link για το PWM

http://139.91.162.101/teaching/hy390_2021/pwm_1293992.txt

Link για το αρχείο FASTA

http://139.91.162.101/teaching/hy390_2021/seqFile_1293992.FA

Απάντηση:

Ερώτηση **7**

Ολοκλήρωση

Βαθμολογήθηκε
στα 2,00

Κατεβάστε και φορτώστε το dataset GDS3709 από το GEO του NCBI. Τρέξτε σε κάθε γονίδιο το t.test (με default παραμέτρους) **ΜΟΝΟ ΣΤΑ ΘΗΛΥΚΑ ΑΤΟΜΑ ΚΑΙ ΓΙΑ ΤΙΣ ΚΑΤΗΓΟΡΙΕΣ ΚΑΠΝΙΣΤΡΙΕΣ - ΜΗ_ΚΑΠΝΙΣΤΡΙΕΣ.**

Ποιο γονίδιο έχει το μικρότερο p-value?

(Αν είναι το πρώτο, απαντήστε 1, αν είναι το 2ο, απαντήστε 2 ΚΟΚ).

Απάντηση:

Ερώτηση **8**

Ολοκλήρωση

Βαθμολογήθηκε
στα 1,00

Κάνουμε το t-test όπως το μάθαμε στην διάλεξη σε ένα σύνολο γονιδίων από microarray πείραμα. Όταν το p-value σε ένα τεστ που κάνουμε για να βρούμε αν ένα γονίδιο εκφράζεται διαφορετικά σε 2 συνθήκες είναι 0.00001, τότε αυτό σημαίνει ότι:

Επιλέξτε ένα ή περισσότερα:

- ☐ a. Η πιθανότητα ότι το γονίδιο έχει διαφορετική έκφραση στις δύο κατηγορίες είναι 0.00001, ενώ η πιθανότητα να μην έχει διαφορετική έκφραση στις δύο κατηγορίες είναι 1-0.00001.
- ☐ b. Η πιθανότητα να υπάρχει ασθένεια είναι 0.00001.
- ☒ c. Η πιθανότητα να παρατηρήσουμε τόσο μεγάλη τιμή ή μεγαλύτερη, σε απόλυτη τιμή, του t-statistic είναι 0.00001 δεδομένου ότι η έκφραση του γονιδίου στις δύο κατηγορίες περιγράφεται από την ίδια κατανομή.
- ☐ d. Η πιθανότητα να είναι σωστή η μηδενική υπόθεση είναι 0.00001
- ☐ e. Η πιθανότητα να παρατηρήσουμε τόσο μεγάλη τιμή ή μεγαλύτερη, σε απόλυτη τιμή, του t-statistic είναι 0.00001 δεδομένου ότι η έκφραση του γονιδίου στις δύο κατηγορίες περιγράφεται από διαφορετικές κατανομές.

[◀ Μπόνους Εργασία 1](#)

Μεταπήδηση σε...

