R5 homework

March 9, 2011

For this homework, you will use data collected from breast cancer tissues regarding the gene methylation status of 5 genes (among other variables). The genes are Hin-1, Twist, Cyclin D2, RAR-beta, RASSF1A (all genes that are known to be associated with breast cancer). The data for these genes are in the 8th-12th columns of the dataset.

1. Read the methylationdata.csv file into R.
2. Create a function for estimating an odds ratio between two binary variables and its 95% confidence interval. Write the function so that the OR estimate and its 95% confidence limits are returned as a vector with length 3.
3. Apply the function you created in 2 to estimate the ORs between the Hin-1 and Twist.
4. Generate a new variable called ‘highgrade’ which takes value 1 if grade =3 and value 0 if grade=1 or2.
5. Create a loop to do the following:
   1. Estimate the OR and 95% CI between highgrade and each of the genes using your function
   2. Save the results in a matrix.

The resulting matrix should have dimension 5 by 3.

1. Label the rows and columns of the results matrix meaningful labels (hint: convert your results matrix into a data.frame and use the row.names() and names() commands).
2. Round your results to 2 decimal places.
3. Print our your matrix. It should look similar to this:

OR 95% CI.L 95% CI.U

hin1 0.43 0.17 1.05

twist 0.57 0.24 1.30

Cd2 0.68 0.30 1.53

rar-B 0.56 0.22 1.39

rassf1a 0.24 0.07 0.69