

1.

- Remember that  $x=0, \dots, 30$ , don't forget  $x=0$ ! Take  $x=0$  into account in the indexing.
- The alternative hypothesis is two-sided, so the null hypothesis can be rejected with low OR high values!
- The distribution is symmetric.

2.

a)

- Remember to use `set.seed(0511)` so that the results will be comparable to the model solution
- `rnorm` takes standard deviation as a parameter instead of variance
- With replicates we mean study subjects, for example patients. 8 individuals from both groups A and B from which we measure the gene expression levels.
- For each group, generate matrix of size Number of genes x Number of replicates filled with random numbers

b)

- Perform testing for **each gene**, you should end up with 100 p-values
- Use `var.equal=TRUE` with `t.test` function

c)

- The idea in Bonferroni correction: multiply the original p-values with the total number of p-values to be adjusted, find minimum of that and 1
- You can compare the results from your implementation to `p.adjust(..., method="bonferroni")`

d)

- You can compare the results from your implementation with `p.adjust(..., method="BH")`
- Remember to return the p-values in the original order!
- The steps in B-H correction:
  1. order the p-values from smallest to largest
  2. store the original order of the p-values
  3. calculate the  $q_k$  values
  4. make sure that the corrected values do not decrease when  $k$  increases
  5. return the corrected p-values (make sure they are  $\leq 1$ ) in their original order

e)

- 20 rows with  $\mu_a \neq \mu_b$
- 80 rows with  $\mu_a = \mu_b$

f)

- See `?t.test` for how to set the alternative hypothesis
- Use `var.equal=TRUE` with `t.test` function
- Which one of the corrections, Bonferroni or B-H gives less significant p-values? Comment on your findings.

g)

- See 43/45 for FDR
  - 1) Average FDR

- 2) Average FWER
- 3) 0.05

FDR for one gene =  $p \text{ value} * \text{number genes} / ( p \text{ value} * \text{number of genes} + 20 )$

Average FDR =  $\text{sum (FDR for one gene)} / 100$

Average FWER =

- 1) one or more false discovery ( reject the null hypothesis  $p \text{ value} < 0.05$ )
- 2) gene  $p \text{ value} < 0.05$  non differential genes (80) 1  
differential gene (20) 0  
 $p \text{ value} > 0.05$  0  
 $1+1+0+ 1 \dots +1/100$