Multiple Hypothesis Testing Procedures in Global Test

Chunyang Li

Example Dataset

- Purpose of study: To determine the relationship between aerobic capacity and cardiac gene expression
- Four groups, each group n = 4:LCR trained, LCR sedentary, HCR trained and HCR sedentary.
- number of samples=16; number of genes=31099; annotation=rat2302

Multiple Hypothesis Testing Issues

	Declared non-significant	*Declared significant*	*Total*
True null hypothesis	**U**	**V**	m_0
False null hypothesis	**T**	**S**	$m - m_0$
	$m - \mathbf{R}$	**R**	m

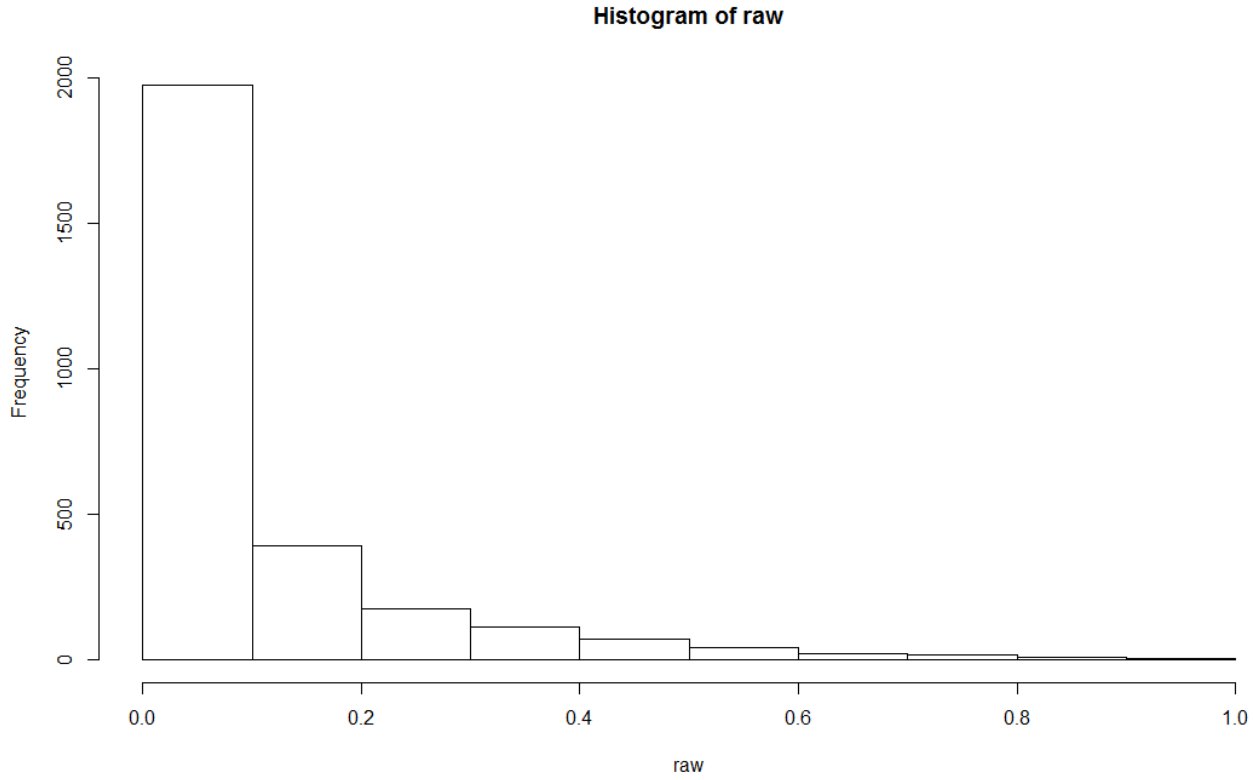
- The Per Comparison Error Rate(PCER): E(V/M)
- The Familywise Error Rate(FWER): P(V>=1) **Bonferroni, Holm and Focus Level**
- The False Discovery Rate(FDR): E(V/R) **Benjamini Hochberg and Benjamini Yekutieli**

Gene Set Testing and the Global Test

- Trying to find out sets of genes that are globally differentially expressed
- Multiple testing problems are reduced, however, still severe
- The testing sets do not have to be the same size

Raw P-value

1588 significant GO terms



Bonferroni Correction

- The Bonferroni Correction rejects all p-values $< \alpha/m$ will control the FWER $< \alpha$.
- Proof: $$FWER = Pr\left\{\bigcup_{I_o}\left(p_i \leq \frac{\alpha}{m}\right)\right\} \leq \sum_{I_o}\left\{Pr\left(p_i \leq \frac{\alpha}{m}\right)\right\} \leq m_0\frac{\alpha}{m} \leq m\frac{\alpha}{m} = \alpha$$
- Advantages:

Strongly controls FWER;

Does not require that the tests be independent.

- Disadvantages:

Power decreases significantly(too conservative) as m increases.

Holm's Correction

Sequential Bonferonni

Procedures:

1. Sort p-values $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$;
2. Compare P(i) to $\alpha/(m - i + 1)$, beginning with the smallest p-value
3. Reject the corresponding null hypothesis and repeat step 2 until the p-value is no longer significant

Advantages:

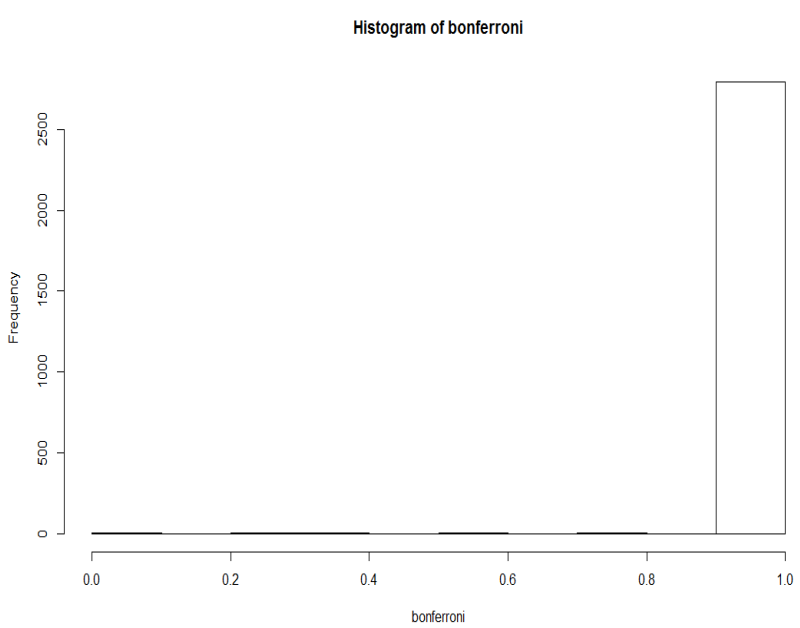
Strongly controls FWER; More powerful than Bonferroni

Disadvantages:

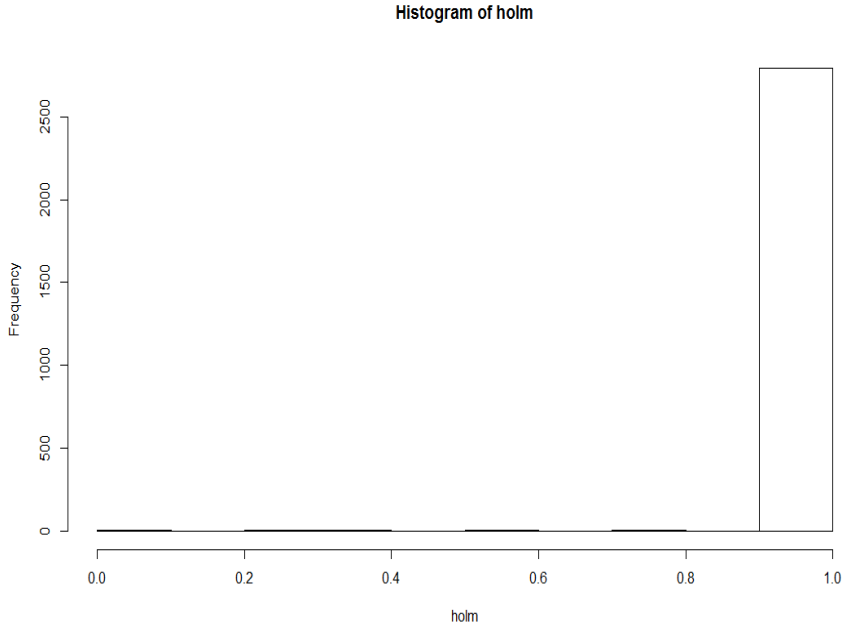
Power is still low with large m

Holm vs Bonferroni

Bon:1 significant GO terms



Holm:1 significant GO terms



Benjamini Hochberg's Correction

Procedures:

1. Sort p-values $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$;
2. Compare P(i) to (i/m)α, beginning with the largest p-value
3. Do not reject the corresponding null hypothesis and repeat step 2 until the p-value is significant

Advantages:

Controls FDR; More powerful than Holm's method

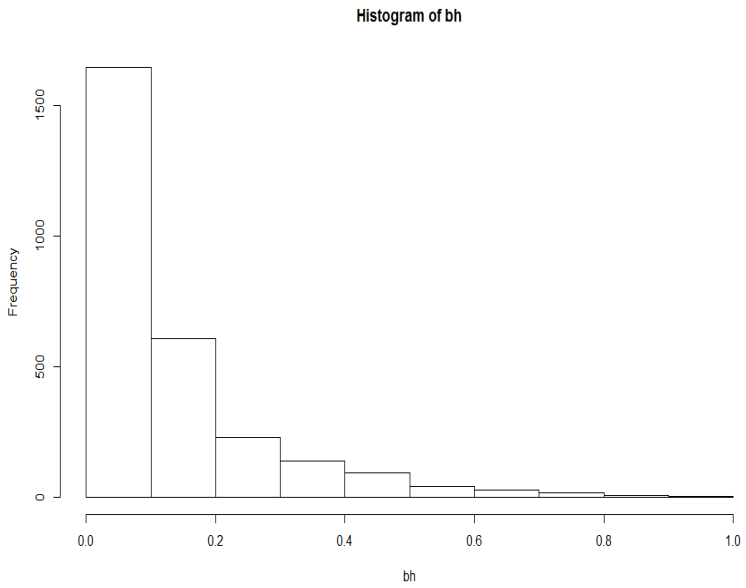
Disadvantages:

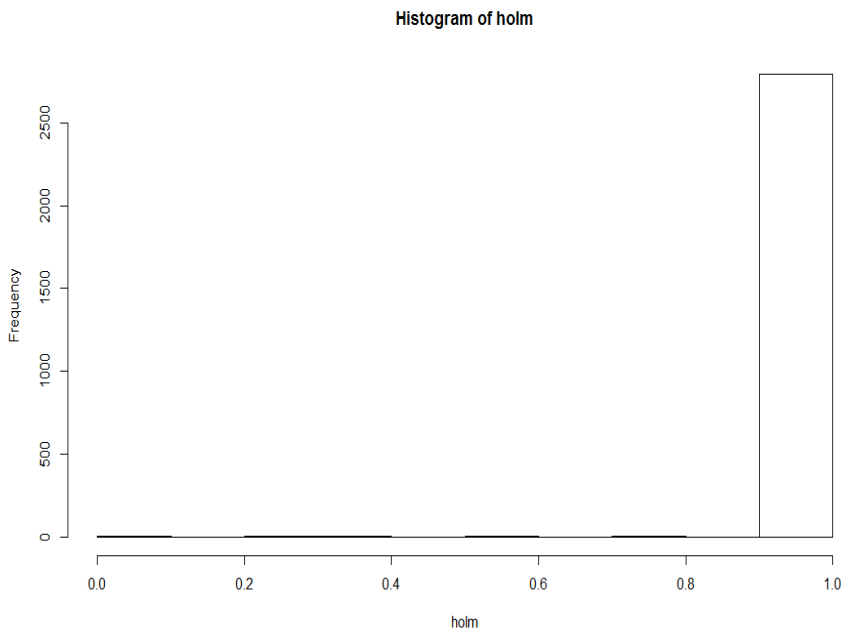
The BH procedure is valid when the tests are independent.

BH VS Holm

BH: 874 significant GO terms

Holm:1 significant GO terms





Benjamini Yekutieli's Correction

Procedures:

Divide α by $\sum_{i=1}^{m} \frac{1}{i}$ and use the BH procedures.

Advantages:

Controls FDR even if tests are dependent;

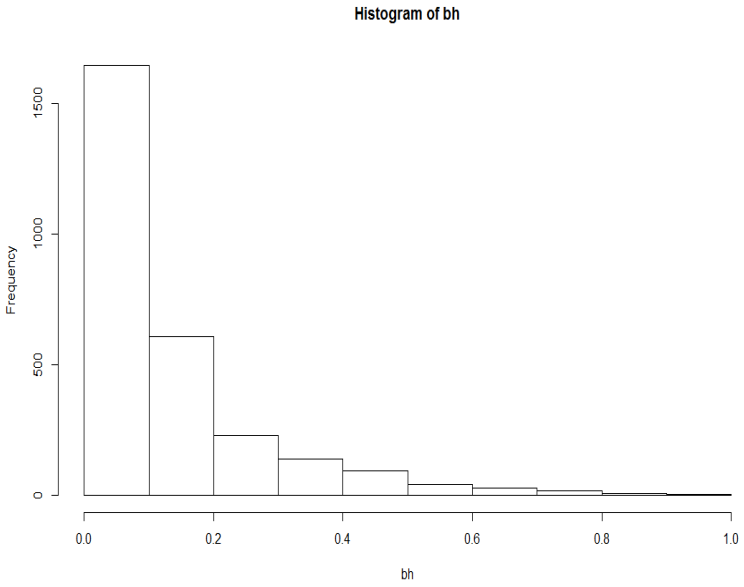
More conservative than BH.

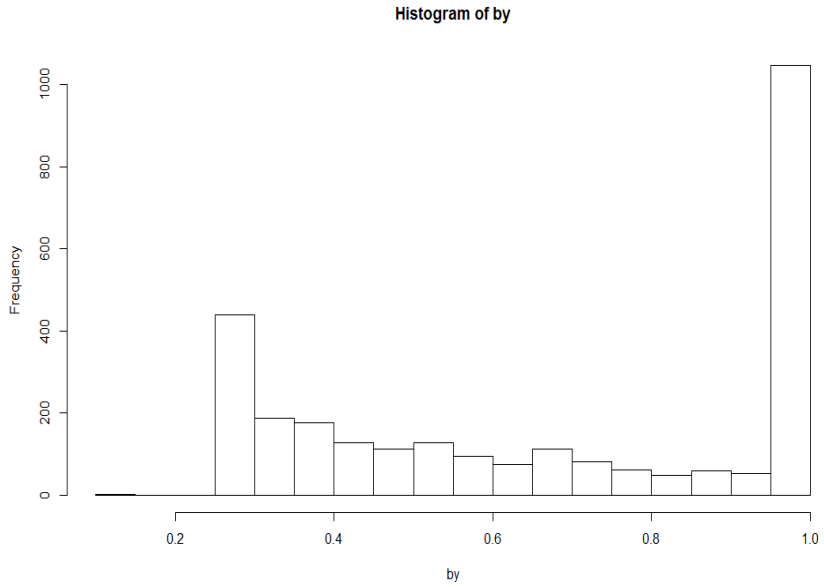
Disadvantages:

Less powerful than BH.

BH VS BH

BH: 874 significant GO terms BY:0 significant GO terms





Testing in GO Graph

- In global test, the null hypotheses are assumed to be a reflection of the relationships in GO graph.
- Two logical relationship assumptions:

1. If parent node isn't significant, the child node is not significant either;
2. Only if we rejected all the child nodes, can we reject the parent node.

Focus Level Method

- Make use of GO graph structures;
- A combination of Holm and the closed testing procedure;
- A sequence of procedures that depends on a chosen level to start.

Bottom-Up Procedure

Procedures:

- First looks at all the hypotheses corresponding to the end nodes of the GO graph
- Use the Holm's method to adjust the p-values.
- The parent node would be significant if at least one of its child nodes is significant.

Advantages

Strongly controls the FWER; Saves computation time;

Can easily find a single highly significant end node even when most of the other nodes are not significant.

Disadvantages

Multiple testing issues can still be severe;

It may fail to find out a significant parent node.

Top-Down Procedure

Procedures:

- Starts with the top node
- The test stops if it is not significant, otherwise keep on testing its offspring.

Advantages:

- All tests are done at α level;
- Good at finding the significant high level nodes where many offspring sets have small effects;
- Could be very efficient if there are not many significant effects.

Disadvantages:

- Can’t find a highly significant but isolated end node;
- The computation could be time consuming.

A More Balanced Procedure

Procedures:

- Reject all hypotheses in the focus level raw p-value ;
- For the hypotheses rejected in step 1, reject all their ancestors; (Upward)
- Add all the child nodes if their parents nodes have been rejected;(Downward)
- Recalculate Holm's factor h and repeat until there are no significant sets.

Advantages:

- This procedure controls family-wise error rate and more powerful than Holm;
- It is powerful detecting intermediate effects near the focus level;
- More flexibility.

Disadvantages:

- The significance of nodes far from the focus level are influenced a lot by the nodes at the focus level

Computational Issues

- Computationally expensive due to the enormous size of the expanded graph.

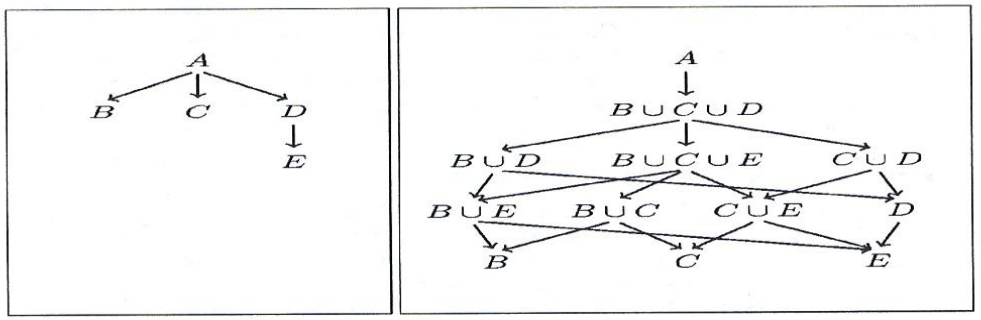


Fig. 1. Illustration of the expansion ('closing') of a graph for use in a Closed Testing procedure. Left: original graph. Right: expanded graph.

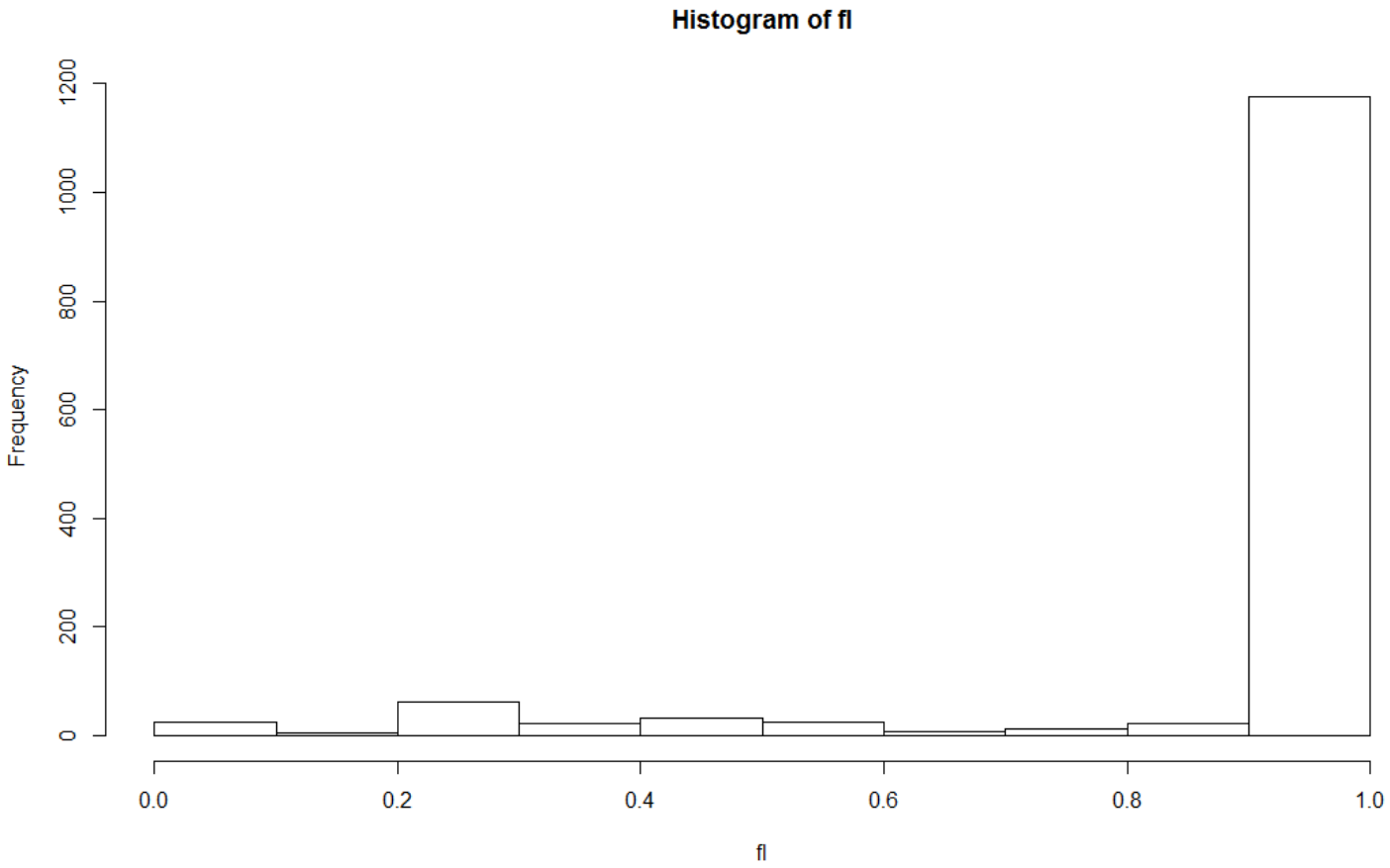
- To reduce the size of the expanded graph, a small number of atom sets, whose unions construct all offspring sets, are built in each subgraph

Choose a Focus Level

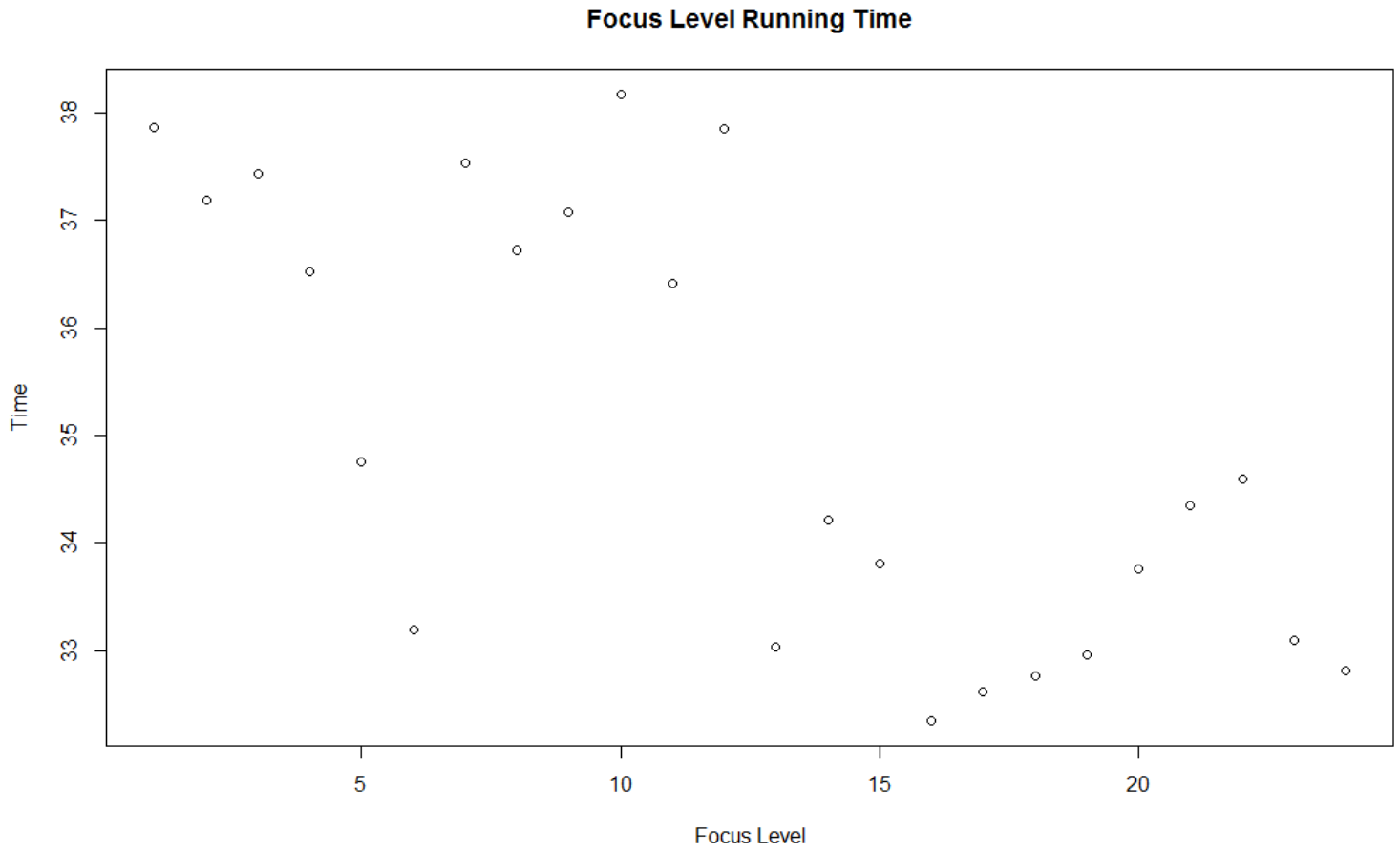
- The major interest of the research and the computation cost should be taken into account.
- The default focus level in gtGO function in R is 10, since it has a good combination of power and reasonable computation time.
- For the chosen level, we get a collection of GO terms with no descendent relationships with each other. All other GO terms are either ancestors or offspring of the focus level nodes.

P-values of Focus Level Method

No significant GO Items



Computing Time



```
time <- rep(0, 24)
sig.fl <- rep(0, 24)

for (i in 1: 24){
  print(i)
  gt.GO.fl <- gtGO(trt, Eset, multtest="focuslevel",
     ontology="BP",minsize=20,maxsize=200, focuslevel = i)
  fl<-gt.GO.fl@extra[, 1]
  sig.fl[i] <- sum(fl < .05)
   print(summary(fl))
  hist(fl, main = paste('fl', i, sep = ''))
  timemore[i] <- system.time(gtGO(trt, Eset, multtest="focuslevel",
     ontology="BP",minsize=20,maxsize=200, focuslevel = i))[[1]]
}

plot(time, main = 'Focus Level Running Time', xlab = 'Focus Level', ylab =
     'Time')
```

Summary

- Selecting a p-value correction method is subjective but important.

- It depends on the goal of the test, what type of error rate you want to control and whether the tests are independent or not.

Reference

- Goeman and Mansmann(2008), Multiple testing on the directed acyclic graph of gene ontology, bioinformatics, Vol. 24 no. 4 2008, pages 537 – 544
- Benjamini and Hochberg (1995), Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society, Series B 57, No. 1, pp. 289-300
- Holm (1978), A Simple Sequentially Rejective Multiple Test Procedure, Scand J Statist 6: 65-70
- Data website: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9445

Thank You!