Multiple Testing and Related Topics

Joseph Abraham

Lecture VIII RBP5793

Outline of Lecture

- Screening Paradox Multiple Testing
- True and False Positives
- Bonferroni, Benjamimi-Hochberg, q-value

Reference: Larry Wasserman:
"All of Statistics" Springer 2003

Screening Paradox I

Imagine we have a town with 10000 inhabitants

1 % of them have a disease. We need to find these inhabitants.

We test the entire population (10000 tests performed). We have a test which is 98% sensitive. Only 2% of the affected individuals will test negative. The test has a specificity of 99%. Only 1% of the tests in healthy individuals will be positive.

We test all the individuals, and identify those who test positive.

Should we treat all those who test positive ?

Screening Paradox II

Number of diseased individuals is 100.

Among these individuals we expect positive results and negative results ?

Screening Paradox II

Number of diseased individuals is 100.

Among these individuals we expect positive results and negative results ? (98 and 2)

Till now 98 true positives and 2 false negatives.

Screening Paradox II

Number of diseased individuals is 100.

Among these individuals we expect positive results and negative results ? (98 and 2)

Till now 98 true positives and 2 false negatives.

Fomr the remaining 9900 individuals we expect negative results and positive results.

Screening Paradox II

Number of diseased individuals is 100.

Among these individuals we expect positive results and negative results ? (98 and 2)

Till now 98 true positives and 2 false negatives.

Fomr the remaining 9900 individuals we expect negative results and positive results. (9801 and 99). 9801 True negatives and 99 false positives.

Screening Paradox III

From previous results we have (98 + 99 = 197) positive tests.

But we had only 100 affected individuals ! About 50% of the individuals who test positive are not affected !

What went wrong ?

Screening Paradox III

From previous results we have (98 + 99 = 197) positive tests.

But we had only 100 affected individuals ! About 50% of the individuals who test positive are not affected !

What went wrong ? Imagine we have 100000 individuals and 0,5% of them are affected. How many true positives, false positives, false negatives and true negatives ?

Screening Paradox III

From previous results we have (98 + 99 = 197) positive tests.

But we had only 100 affected individuals ! About 50% of the individuals who test positive are not affected !

What went wrong ? Imagine we have 100000 individuals and 0,5% of them are affected. How many true positives, false positives, false negatives and true negatives ?

(490,

Screening Paradox III

From previous results we have (98 + 99 = 197) positive tests.

But we had only 100 affected individuals ! About 50% of the individuals who test positive are not affected !

What went wrong ? Imagine we have 100000 individuals and 0,5% of them are affected. How many true positives, false positives, false negatives and true negatives ?

(490, 995,

Screening Paradox III

From previous results we have (98 + 99 = 197) positive tests.

But we had only 100 affected individuals ! About 50% of the individuals who test positive are not affected !

What went wrong ? Imagine we have 100000 individuals and 0,5% of them are affected. How many true positives, false positives, false negatives and true negatives ?

(490, 995,10,&

Screening Paradox III

From previous results we have (98 + 99 = 197) positive tests.

But we had only 100 affected individuals ! About 50% of the individuals who test positive are not affected !

What went wrong ? Imagine we have 100000 individuals and 0,5% of them are affected. How many true positives, false positives, false negatives and true negatives ?

(490, 995,10,& 98505). Total Number of positive tests is 1484.

Proportion of True Positives is 33%.

Screening Paradox IV

With a very sensitive test we can ensure almost all who have the disease will test positive. However what we also need is that whenever someone tests positive more likely than not they have the disease. The second part depends not only on the test but also on the percentage of affecteds in the study population !

Screening Paradox IV

With a very sensitive test we can ensure almost all who have the disease will test positive. However what we also need is that whenever someone tests positive more likely than not they have the disease. The second part depends not only on the test but also on the percentage of affecteds in the study population ! More rigorous definition of diseased does not help, many mild cases will not be diagnosed.

Screening Paradox V

In the case of differential gene expression only a small fraction of genes are differentially expressed

False positives can arise easily. Cannot be too restrictive, (very small p-value) will discard many interesting genes. Need a procedure to control the proportion of false positives among all genes which appear to be differentially expressed.

True and False Positives I

H_0: No differential expression

Type I Error (False Positive: Wrongly declare H_0 False)

Type II Error (False Negative: Declare H_0 True when H_0 False)

Basic Table of Possible Outcomes

	Not Reject H_0	Reject H_0	
H_0 True	U	V	m_0
H_0 False	T	S	m_1
Total	m-R	R	m

V is the number of False Positives (Type I Error)

T is the number of False Negatives (Type II Error)

Only R and (m-R) are known quantities !

True and False Positives II

First Guess; Try and reduce V. Almost all tests declared significant are actually significant. Define π the family wise error rate (FWER) to be our error threshold. If $\pi = 0,01$ then only 1% of all tests declared positive are false positives. To implement this we can use $\alpha = \frac{\pi}{m}$. (p value $\leq \alpha$ accepted)

This is called the Bonferroni correction

If we set $\pi = 0,01$ and m = 10000, what is the problem ?

False Discovery Rate I

Better approach: select some genes among those declared differentially expressed such that $\frac{V}{R}$ can be chosen to less than some user defined value. Various different techniques to implement this idea, False Discovery Rate (FDR)

Original paper (42346 citations !)
Benjamini, Y. and Hochberg, Y. (1995).
Controlling the false discovery rate:
a practical and powerful approach to multiple testing.
Journal of the Royal Statistical Society B 57, 289-300 (1995)

False Discovery Rate II

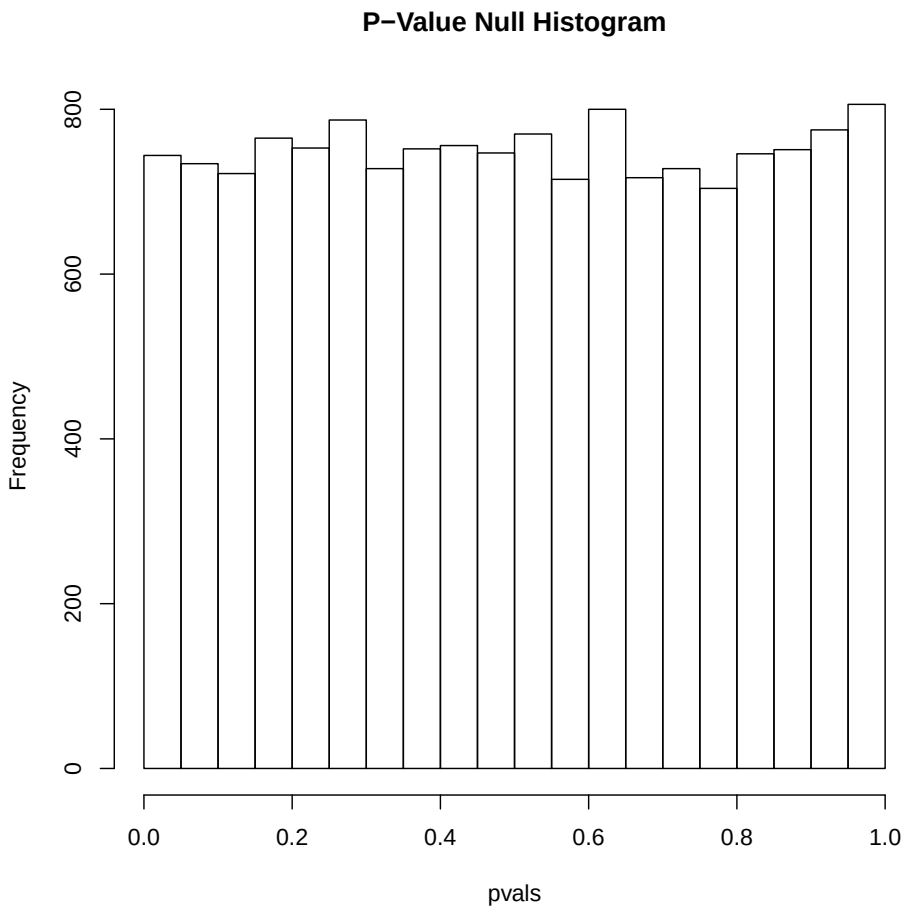
If we have n p-values, arrange in increasing order then given FDR δ we keep all $p(k) \leq \delta(\frac{k}{n})$ $k = 1, 2, \ldots n$

Example (unadjusted p-values on left, adjusted on right)

0.0001	0.005 (Reject H_0)
0.004	0.010 (Reject H_0)
0.007	0.015 (Reject H_0)
0.009	0.020 (Reject H_0)
0.012	0.025 (Reject H_0)
0.336	0.030 (Fail to Reject H_0)
0.393	0.035 (Fail to Reject H_0)
0.539	0.040 (Fail to Reject H_0)
0.581	0.045 (Fail to Reject H_0)
0.986	0.050 (Fail to Reject H_0)

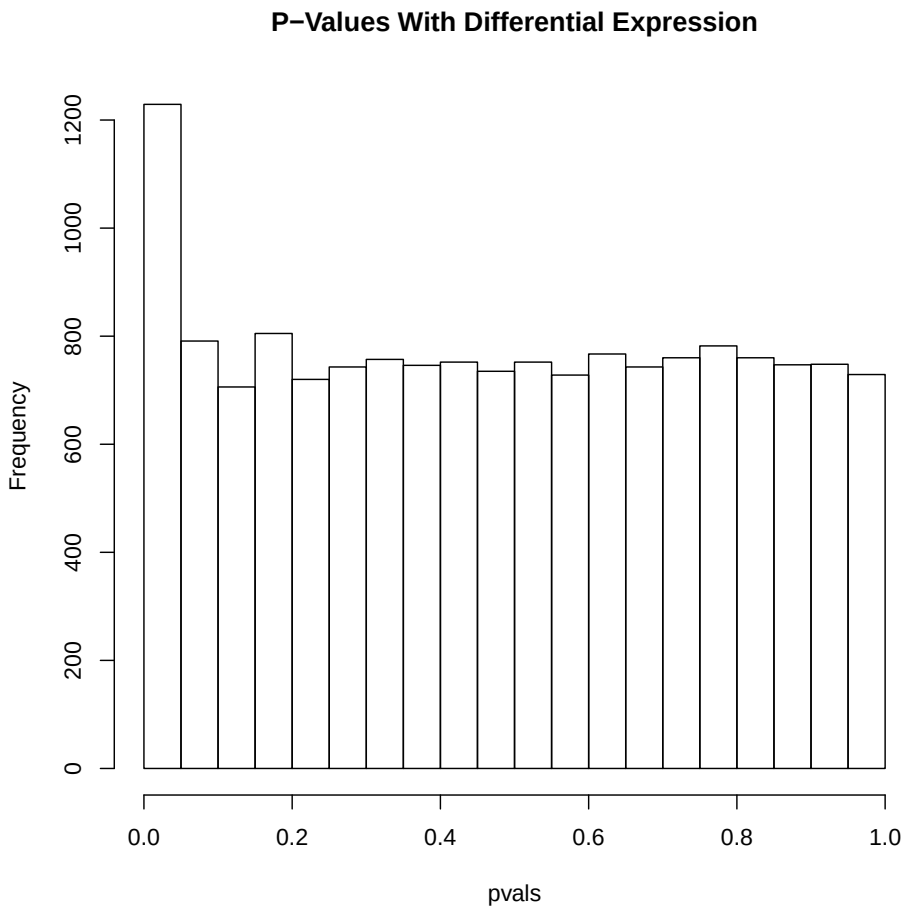
False Discovery Rate III

With no differential expression , uniform p-value distribution



False Discovery Rate III

With differential expression , non-uniform p-value distribution



False Discovery Rate IV

Can try and separate the p-values into 2 components

a uniform component and a component with

peak at small values. Combine a uniform and β

distribution and fit the distribution to the mixture.

Can separate those small p-values from H_0 and

those from differentialy expressed genes.

Allison, D. B., G. L. Gadbury, M. Heo, J. R. Fernandez, C.-K. Lee, T. A. Prolla, & R. Weindruch A mixture model approach for the analysis of microarray gene expression data. Computational Statistics and Data Analysis 39: 1-20. (2002)

q-values

Q: Suppose I have the p-values for all genes.

I retain a gene and all genes with p-values

smaller than the choosen gene. Among all these

what is the FDR among all these genes ?

Determined by the q-value of the starting gene.

Storey, J. D., and R. Tibshirani
Statistical significance for genomewide studies.
PNAS 100: 9440-9445. (2003)