

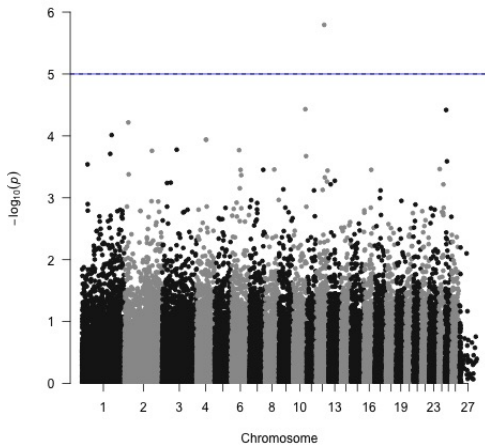
Multiple Comparisons

Department of Statistics, University of South Carolina

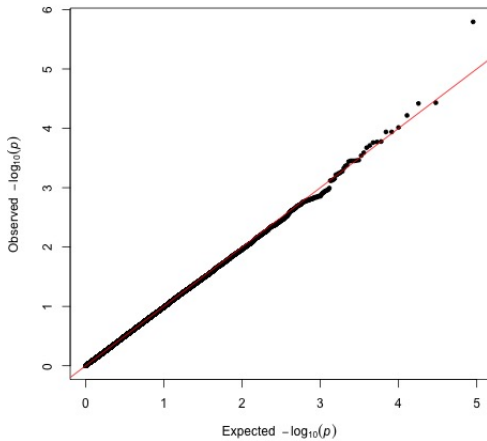
STAT718/BIOL703: Genomic Data Science

- Family-Wise Error Rate
- Bonferroni Correction
- False Discovery Rate
- q Value
- Benjamini Hochberg Adjustment

Manhattan plot

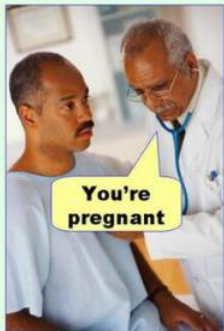


QQ plot



Performance of Statistical Tests

Type I error
(false positive)



Type II error
(false negative)



- Control false positive (Type I error, α) at 0.05
- Power (true positive rate): the ability to detect the signal when there is something going on

Approaches to Control Type I Errors

- Family-wise error rate (FWER): the probability of making at least one Type I errors among all the tests.
- False discovery rate (FDR): is the expected proportion of Type I errors among the rejected hypotheses.
- and many others ...

Approaches to Control Type I Errors

- Family-wise error rate (FWER): the probability of making at least one Type I errors among all the tests.
- False discovery rate (FDR): is the expected proportion of Type I errors among the rejected hypotheses.
- and many others ...

Which procedure is more stringent?

$$\alpha^* = \frac{\alpha}{m}.$$

Where m is the number hypothesis tests considered.

```
> 0.05/nrow(toptable2)
[1] 1.113487e-06
```


$$\alpha^* = \frac{\alpha}{m}.$$

Where m is the number hypothesis tests considered.

```
> 0.05/nrow(toptable2)
[1] 1.113487e-06
```

High probability of Type 2 errors.

- To control FDR at level δ :
 - ① Sort the p values in increasing order to get $p_{(1)}, p_{(2)}, \dots, p_{(m)}$
 - ② Define $k = \max\{j : p_{(j)} \leq \frac{j}{m} \times \delta\}$
 - ③ Reject $H_0^{(1)}, H_0^{(2)}, \dots, H_0^{(k)}$

$$p_{(j)} \leq \delta \times \frac{j}{m}$$

BH Example

Control FDR at $\delta = 0.05$

Rank (j)	p value	$(j/m) \times \delta$	Reject Ho?
1	0.001	0.005	1
2	0.009	0.010	1
3	0.165	0.015	0
4	0.205	0.020	0
5	0.396	0.025	0
6	0.450	0.030	0
7	0.641	0.035	0
8	0.781	0.040	0
9	0.900	0.045	0
10	0.993	0.050	0

BH Example

Control FDR at $\delta = 0.05$

Rank (j)	p value	$(j/m) \times \delta$	Reject Ho?
1	0.001	0.005	1
2	0.009	0.010	1
3	0.165	0.015	0
4	0.205	0.020	0
5	0.396	0.025	0
6	0.450	0.030	0
7	0.641	0.035	0

$$p \leq \delta \times \frac{j}{m}$$

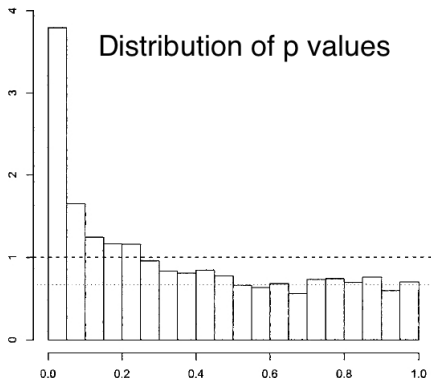
$$p \times \frac{m}{j} < \delta$$

$$p^* < \delta$$

In R,

```
> pvalue<-c(0.0008, 0.009, 0.165, 0.205, 0.396, 0.450,  
+ 0.641, 0.781, 0.900, 0.993)  
> adjp<-p.adjust(pvalue, method="BH")  
> adjp  
[1] 0.0080000 0.0450000 0.5125000 0.5125000 0.7500000  
[6] 0.7500000 0.9157143 0.9762500 0.9930000 0.9930000
```

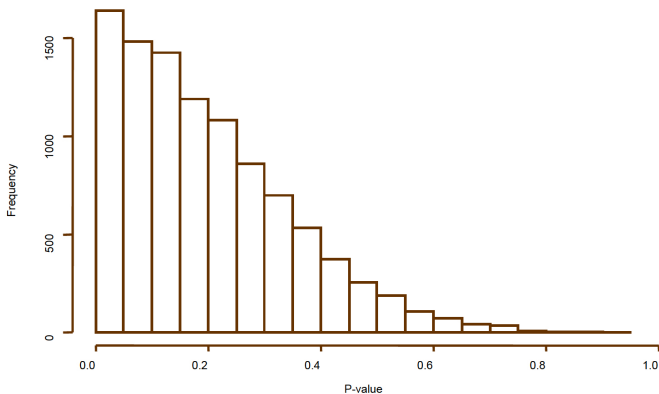
Estimating FDR



A density histogram of the 3,170 p values from the Hedenfalk et al. (14) data. The dashed line is the density histogram we would expect if all genes were null (not differentially expressed). The dotted line is at the height of our estimate of the proportion of null p values.

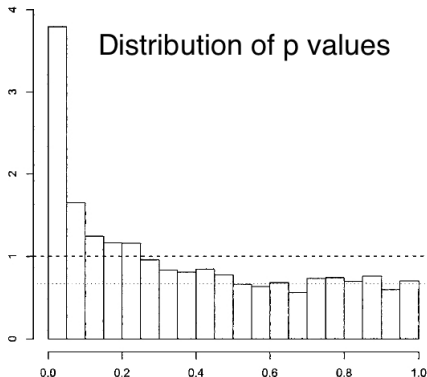
The distribution of p values under alternative hypothesis

Under the alternative hypothesis, p values are are skewed toward 0.



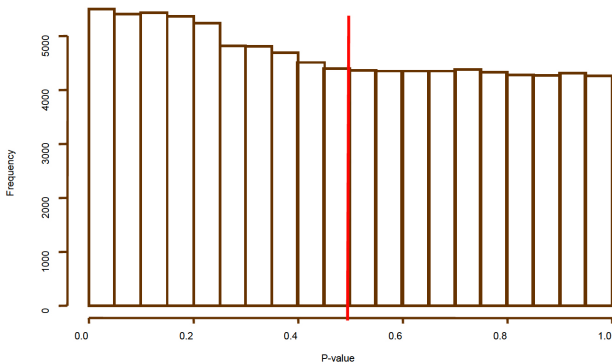
Estimating FDR

Combined distribution is a mixture of p values from the null and alternative hypotheses.



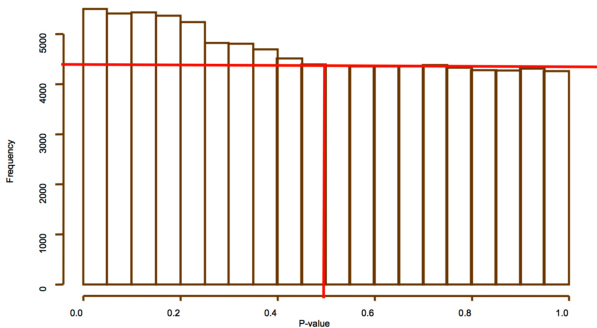
Estimating FDR

For p values greater than say 0.5, we can assume they mostly represent observations from the null hypothesis.



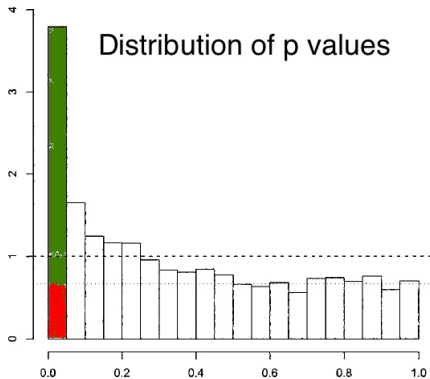
Estimating FDR

The proportion of truly null tests



Estimating FDR

$$\text{FDR} = \frac{\text{red}}{\text{green} + \text{red}}$$



estimating q value in R

q value is the minimum FDR that can be attained when calling that feature significant.

```
> library(qvalue)
> qobj<-qvalue(toptable2[,5])
> str(qobj)
List of 8
 $ call      : language qvalue(p = toptable2[, 5])
 $ pi0       : num 1
 $ qvalues   : num [1:44902] 0.0722 0.571 0.571 0.6809 0.738 ...
 $ pvalues   : num [1:44902] 1.61e-06 3.70e-05 3.81e-05 6.07e-05 9.68e-05 ...
 $ lfdr      : num [1:44902] 0.139 0.737 0.737 0.76 0.795 ...
 $ pi0.lambda: num [1:19] 1 1 1 1 1.01 ...
 $ lambda    : num [1:19] 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 ...
 $ pi0.smooth: num [1:19] 1 1 1 1 1.01 ...
 - attr(*, "class")= chr "qvalue"
> gqvalue<-qobj$qvalues
> gqvalue[1:10]
[1] 0.07223014 0.57099786 0.57099786 0.68086267 0.73799419
[6] 0.73799419 0.73799419 0.78130669 0.78130669 0.78130669
```