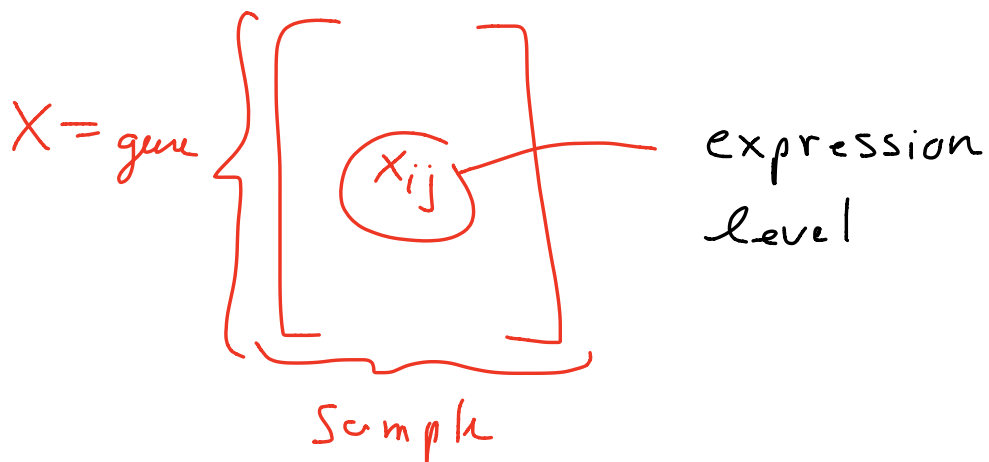


# Large Scale Hypothesis Testing

## 2.1 Microarray Example

Goal: Identify "interesting genes"

Data:



Partition the samples into controls versus infected spaces.

$\bar{X}_i(k)$  = ave. expression for  
 $k=1,2$  controls / infected

The two sample  $t$ -stat for testing  
gene  $i$

$$t_i = \frac{\bar{x}_i(2) - \bar{x}_i(1)}{s_i}$$

$$s_i^2 = \frac{\sum_{\text{controls}} (x_{ij} - \bar{x}(1))^2 + \sum_{\text{infected}} (x_{ij} - \bar{x}(2))^2}{100} \left( \frac{1}{s_0} + \frac{1}{s_2} \right)$$

Here we would test based on the  $T$ -distribution. But generally we want to consider everything in normal theory

$$z_i = \Phi^{-1}(F_{100}(t_i))$$

In general we have  $N \gg$  tests so we want to compare these  $\{z_i, 1 \leq i \leq n\}$  to a general normal without making multiple testing mistakes.

The conservative approach: Bonferroni

Reject at  $\alpha/N$  level.

More Modern approach: Bayesian Approach

## 2.2 Bayesian Approach

The Two group Model

- $N$  cases either null or non-null with prob  $\pi_0, \pi_1$
- $z$  values having density either  $f_0(z), f_1(z), F_0(z), F_1(z)$
- The mixture model describing the model is

$$F(z) = \pi_0 F_0(z) + \pi_1 F_1(z)$$

That is  $z \sim F$

- Suppose we observe  $z \in \mathbb{Z}$  and wonder if it corresponds to the

null or non-null

$$\begin{aligned}\phi(z) &= P(\text{null} | z \in Z) \\ &= \pi_0 \frac{F_0(z)}{F(z)}\end{aligned}$$

Bayes False Discovery rate  
for  $z$

### 2.3 Empirical Bayesian Methods