

Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

binGroup: A Package for Group Testing

Christopher R. Bilder¹

Boan Zhang¹

Frank Schaarschmidt²

Joshua M. Tebbs³

¹University of Nebraska-Lincoln

²Leibniz Universität Hannover, Germany

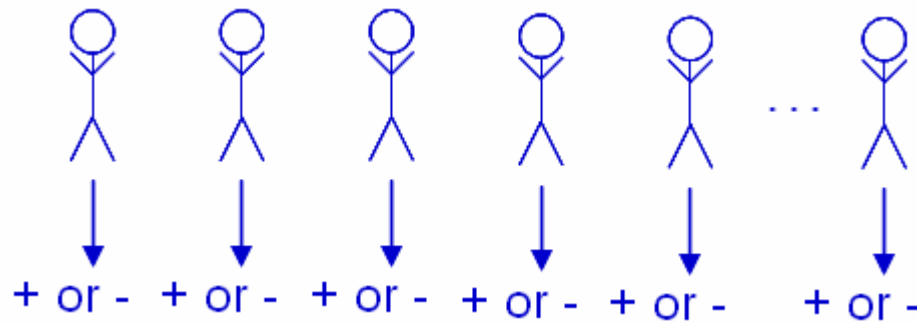
³University of South Carolina

What is group testing?

Outline

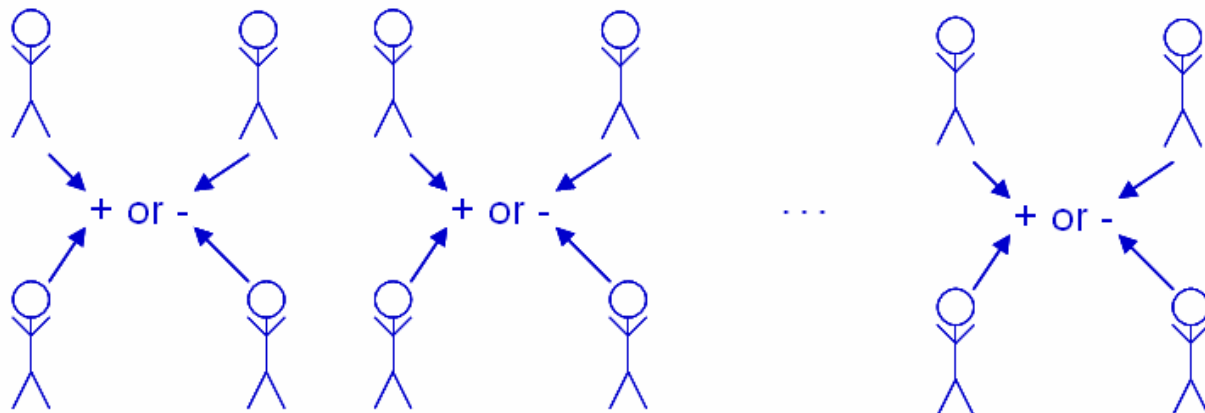
- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- Testing a person for a disease
- Individual testing



– Problem: Cost and Time

- Group testing



What is group testing?

Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- Group testing (continued)
 - Save time and resources
 - Applications in screening blood donations, drug discovery experiments, veterinary and public health studies
 - Estimation
 - Probability of disease
 - Identification
 - Which individuals are positive

Homogeneous Population

Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

• Notation

– Individual responses

- Y_{ik} are independent Bernoulli(p) random variables for item i in group k ($i = 1, \dots, I_k$, $k = 1, \dots, K$)
- Need to estimate $p = P(Y_{ik} = 1)$
- p is the “prevalence in a population”
- Y_{ik} are unobserved

– Group responses

- Z_k are independent Bernoulli(θ_k) random variables
- $\theta_k = P(Z_k = 1)$
- $\theta_k = 1 - (1 - p)^{I_k}$

Homogeneous Population

Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- Estimate p

- Likelihood function

$$L(p) = \prod_{k=1}^K \theta_k^{z_k} (1 - \theta_k)^{1 - z_k} = \prod_{k=1}^K \left[1 - (1 - p)^{I_k} \right]^{z_k} (1 - p)^{I_k (1 - z_k)}$$

- Common group size I

- MLE for p : $\hat{p} = 1 - \left(1 - \sum_{k=1}^K z_k / K \right)^{1/I}$

- Unequal group sizes

- Iterative numerical methods needed

Homogeneous Population

Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- `bgtCI ()`
 - Calculates point estimate and confidence interval for p
 - Common group size
 - Different types of confidence intervals
- Example (Ornaghi et al., 1999)
 - The purpose is to estimate the probability that female planthopper transfer the MRC virus to maize crops

Homogeneous Population

Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- Example (continued)

- 24 plants with 7 planthoppers on each
 - 3 plants test positive for the virus
- ```
> bgtCI(n=24, y=3, s=7, conf.level=0.95,
+ alternative="two.sided", method="Score")
```

The 95 percent Score confidence interval is:

```
[0.006325 0.05164]
```

```
point estimator = 0.0189
```

# Homogeneous Population

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- `estDesign()`
  - Find the optimal group size
  - Example
 

```
> estDesign(n = 24, smax = 100, p.tr = 0.0189)

group size s with minimal mse(p) = 43
$varp [1] 3.239869e-05
$mse [1] 3.2808e-05
$bias [1] 0.0006397784
$exp [1] 0.01953978
```
- Other functions include:
  - `bgtvs()`, `bgtTest()`, `bgtPower()`, `nDesign()`, `sDesign()`



# Heterogeneous Population

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

## • Notation

### – Individual responses

- $Y_{ik}$  are independent Bernoulli( $p_{ik}$ ) random variables for item  $i$  in group  $k$  ( $i = 1, \dots, I_k$ ,  $k = 1, \dots, K$ )
- Need to estimate  $p_{ik} = P(Y_{ik} = 1)$
- $Y_{ik}$  are unobserved

### – Group responses

- $Z_k$  are independent Bernoulli( $\theta_k$ ) random variables
- $\theta_k = P(Z_k = 1)$  for group  $k$
- $\theta_k = 1 - \prod_{i=1}^{I_k} (1 - p_{ik})$

# Heterogeneous Population

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- Notation (continued)
  - Covariates
    - $x_{ik1}, x_{ik2}, \dots, x_{ikp}$  for the  $i^{\text{th}}$  item in the  $k^{\text{th}}$  group
    - Incorporate factors which influence disease status
- Model:  $\text{logit}(p_{ik}) = \beta_0 + \beta_1 x_{ik1} + \dots + \beta_p x_{ikp}$
- Estimation of  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 
  - Note that  $Y_{ik}$  are not directly observed
  - Vansteelandt et al. (*Biometrics*, 2000)
    - Likelihood function is written in terms of the  $Z_k$

$$\begin{aligned}
 L &= \prod_{k=1}^K \theta_k^{z_k} (1 - \theta_k)^{1 - z_k} \\
 &= \prod_{k=1}^K \left[ 1 - \prod_{i=1}^{I_k} (1 - p_{ik}) \right]^{z_k} \left[ \prod_{i=1}^{I_k} (1 - p_{ik}) \right]^{1 - z_k}
 \end{aligned}$$

# Heterogeneous Population

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- Estimation of  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  (continued):

– Xie (*Statistics in Medicine*, 2001)

- Likelihood function is written in terms of the unobserved  $Y_{ik}$

$$L = \prod_{k=1}^K \prod_{i=1}^{I_k} p_{ik}^{y_{ik}} (1 - p_{ik})^{1 - y_{ik}}$$

- EM algorithm used

$$E(Y_{ik} \mid Z_k = 1) = \frac{p_{ik}}{1 - \prod_{i=1}^{I_k} (1 - p_{ik})}$$

$$E(Y_{ik} \mid Z_k = 0) = 0$$

# Heterogeneous Population

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- HIV surveillance (Verstraeten et al., *Tropical Medicine and International Health*, 1998)

| AGE | EDUC . | groupres | gnum |
|-----|--------|----------|------|
| 21  | 4      | 0        | 1    |
| 16  | 2      | 0        | 1    |
| 17  | 1      | 0        | 1    |
| 17  | 2      | 0        | 1    |
| 18  | 1      | 0        | 1    |
|     |        |          | ⋮    |
| 25  | 2      | 1        | 85   |
| 29  | 3      | 1        | 85   |
| 17  | 2      | 1        | 85   |
| 18  | 2      | 1        | 85   |
| 18  | 2      | 1        | 85   |

# Heterogeneous Population

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- Model HIV with two covariates: AGE and EDUC.

```
> fit1 <- gtreg(formula = groupres ~ AGE + EDUC.,
 data = hivsurv, groupn = gnum, sens = 1, spec = 1
 method = "Vansteelandt")
```

- Result is a list that contains a lot of information:

```
> class(fit1)
[1] "gt"
> names(fit1)
[1] "coefficients" "hessian" "fitted.group.values"
[4] "deviance" "df.residual" "null.deviance"
[7] "df.null" "aic" "counts"
[10] "residuals" "z" "call"
[13] "formula" "method" "link"
[16] "terms"
```

- Summarize the results:

```
> summary(fit1)
```

# Heterogeneous Population

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

(Continued)

Call:

```
gtreg(formula = groupres ~ AGE + EDUC., data = hivsurv, groupn
 = gnum, linkf = "logit", method = "Vansteelandt")
```

Deviance Residuals:

|  | Min     | 1Q      | Median  | 3Q     | Max    |
|--|---------|---------|---------|--------|--------|
|  | -1.1868 | -0.9376 | -0.8197 | 1.3223 | 1.6826 |

Coefficients:

|             | Estimate | Std. Error | z value | Pr(> z ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | -2.78115 | 1.45576    | -1.910  | 0.0561 . |
| AGE         | -0.04921 | 0.06224    | -0.791  | 0.4292   |
| EDUC.       | 0.67646  | 0.40087    | 1.687   | 0.0915 . |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

$$\hat{p} = \frac{\exp(-2.781 - 0.049age + 0.676educ)}{1 + \exp(-2.781 - 0.049age + 0.676educ)}$$

# Heterogeneous Population

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

## (Continued)

```
Null deviance: 112.1 on 428 degrees of freedom
Residual deviance: 109.3 on 425 degrees of freedom
AIC: 115.3
```

```
Number of iterations in optim(): 194
```

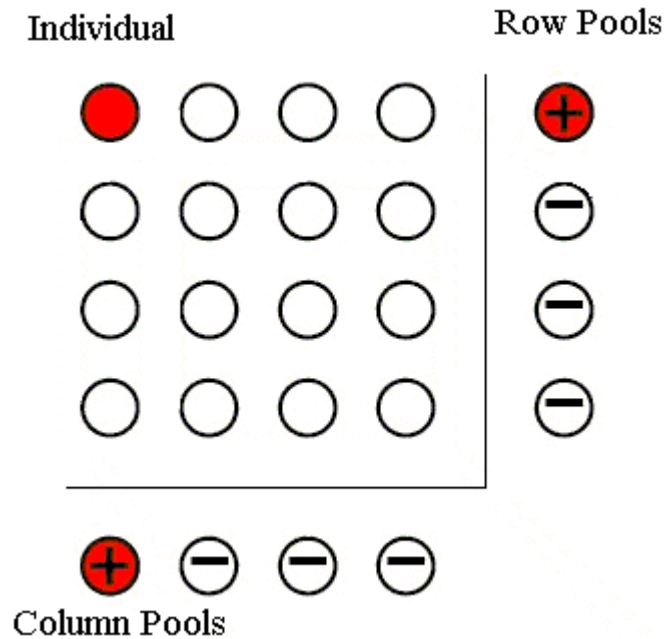
- `summary.gt()`, `predict.gt()`, `residuals.gt()`
  - Similar to those method functions for `glm` class
- `sim.g()`
  - Simulates data in group testing form

# Matrix Pooling

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- Regression in the matrix pooling group testing scheme
  - Put individual specimens in square arrays and test each row and column pool (Phatarfod and Sudbury, 1994)
  - A simple example of  $4 \times 4$  square array:





# Matrix Pooling

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- Regression in the matrix pooling group testing scheme (continued)
  - EM algorithm
  - $E(Y_{ij} \mid \text{Row and column responses})$  in the E-step cannot be explicitly expressed
  - Gibbs sampling in each E-step, suggested by Xie (*Statistics in Medicine*, 2001)

# Matrix Pooling

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

- `gtreg.mp()`
  - Fit the group testing regression model in the matrix pooling setting
  - Individual retests can be included
- A set of method functions available

## • Example

```
> head(sa1)
```

|   | x        | col.resp | row.resp | coln | rown | arrayn | retest |
|---|----------|----------|----------|------|------|--------|--------|
| 1 | 29.96059 | 0        | 0        | 1    | 1    | 1      | NA     |
| 2 | 61.28205 | 0        | 1        | 1    | 2    | 1      | NA     |
| 3 | 34.27341 | 0        | 1        | 1    | 3    | 1      | NA     |
| 4 | 46.19001 | 0        | 0        | 1    | 4    | 1      | NA     |
| 5 | 39.43801 | 0        | 1        | 1    | 5    | 1      | NA     |
| 6 | 45.88038 | 1        | 0        | 2    | 1    | 1      | NA     |

## Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

## • Example (continued)

```
> fit1 <- gtreg.mp(formula =
 cbind(col.resp, row.resp) ~ x,
 data = sal, coln = coln, rown = rown,
 arrayn = arrayn, sens = 0.95, spec = 0.95,
 n.gibbs = 2000, trace = TRUE)
```

```
beta is -6.4126 0.088847 diff is 0.091304
beta is -6.2670 0.086828 diff is 0.022727
beta is -6.2053 0.085777 diff is 0.012097
beta is -6.2486 0.086816 diff is 0.012102
beta is -6.2398 0.086598 diff is 0.0025023
```

Number of minutes runing: 1.43

$$\hat{p} = \frac{\exp(-6.2398 + 0.0866x)}{1 + \exp(-6.2398 + 0.0866x)}$$

Outline

- What is group testing?
- Homogeneous Population
- Heterogeneous Population
- Matrix Pooling

# **binGroup: A Package for Group Testing**

Christopher R. Bilder<sup>1</sup>

Boan Zhang<sup>1</sup>

Frank Schaarschmidt<sup>2</sup>

Joshua M. Tebbs<sup>3</sup>

<sup>1</sup>University of Nebraska-Lincoln

<sup>2</sup>Leibniz Universität Hannover, Germany

<sup>3</sup>University of South Carolina