

Estimation of the overall prevalence

Despite observing group information only, the overall prevalence, p , can be estimated. For a homogenous population (the classical assumption in group testing), this is also the individual probability of having the trait. Suppose a group size of I is used. The probability a group is positive is $\theta = 1 - P(\text{all individuals are negative}) = 1 - (1 - p)^I$. Solving for p results in $p = 1 - (1 - \theta)^{1/I}$. The maximum likelihood estimator (MLE) of θ is simply the proportion of initial groups that test positive, say $\hat{\theta}$. By the invariance property of MLEs, the MLE of p is $p = 1 - (1 - \hat{\theta})^{1/I}$. When there are unequal group sizes, one needs to examine the likelihood function

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

where K is the number of groups, I_k is the group size for group k , and $z_k = 0$ or 1 is the Bernoulli(θ_k) group response for $k = 1, \dots, K$. This likelihood function can be maximized numerically to obtain the MLE for p . Standard MLE results lead to finding the large sample variance estimate:

$$\widehat{Var}(\hat{p}) = \left\{ \sum_{k=1}^K \frac{z_k I_k (1 - \hat{p})^{I_k - 2}}{1 - (1 - \hat{p})^{I_k}} \left[I_k - 1 + \frac{I_k (1 - \hat{p})^{I_k}}{1 - (1 - \hat{p})^{I_k}} \right] + \frac{(1 - z_k) I_k}{(1 - \hat{p})^2} \right\}^{-1}.$$

An approximate $(1 - \alpha)100\%$ confidence interval for p is $\hat{p} \pm Z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{p})}$ where $Z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile from a standard normal distribution. If all groups are of equal size I , the variance formula above simplifies to

$$\widehat{Var}(\hat{p}) = \frac{1 - (1 - \hat{p})^I}{KI^2(1 - \hat{p})^{I-2}}.$$

Additional references:

- Dorfman, R. (1943). "The detection of defective members of large populations." *Annals of Mathematical Statistics* 14:436-440.
- Kennedy, J., Mortimer, R., and Powers, B. (2006). "Reverse transcription-polymerase chain reaction on pooled samples to detect bovine viral diarrhoea virus by using fresh ear-notch-sample supernatants." *Journal of Veterinary Diagnostic Investigation* 18:89-93.
- Litvak, E., Tu, X., and Pagano, M. (1994). "Screening for the presence of a disease by pooling sera samples." *Journal of the American Statistical Association* 89:424-434.
- Phatarfod, R. and Sudbury, A. (1994). "The use of a square array scheme in blood testing." *Statistics in Medicine* 13:2337-2343.
- Sterrett, A. (1957). "On the detection of defective members of large populations." *Annals of Mathematical Statistics* 28:1033-1036.