

그룹 검사를 이용한 효율적인 혈액 검사 절차에 대한 연구

양 원 석¹⁾, 채 경 철²⁾

요 약

혈액 검사를 통해 어떤 질병의 감염 여부를 확인하려 한다. g 명의 혈액을 검사한다고 하자. g 명을 개별적으로 검사하면 g 번의 검사가 필요하다. 그러나 g 명의 혈액을 섞어서 검사하는 방법도 가능하다. 이처럼 여러 개를 한꺼번에 검사하는 방법을 그룹 검사라고 한다. 본 논문에서는, 질병률이 알려진 경우에 대한 그룹 검사 방법 중에서 계층적 도프만 절차를 이용한다. 그리고 검사의 종속성 때문에 발생하는 단축 효과를 고려한다. 이를 통해 개별 검사보다 매우 효율적인 혈액 검사 절차를 제안하겠다. 아울러 실제 상황에 쉽고 간단하게 적용할 수 있는 실용적인 방법을 제안한다.

1. 서론

어떤 질병 A를 가지고 있는 모집단이 있다. 모집단의 질병률이 p 이고, 질병 A는 혈액 검사를 통해 발견할 수 있다. 질병률은 알려져 있고, 전체 검사 대상자의 수는 충분히 크다고 가정한다. 우리는 다음의 두 가지 방법으로 혈액 검사를 할 수 있다.

첫번째 방법 : 전체 검사 대상자를 1명씩 개별 검사한다.

두번째 방법 : 전체 검사 대상자를 g 명씩 묶어 각각을 그룹으로 검사한다. 먼저, g 명의 혈액을 섞어서 그 섞은 혈액을 검사한다. 검사 결과가 음성이면, g 명이 모두 건강함을 의미한다. 이 경우에는 1번의 검사로 g 명에 대한 검사를 종료할 수 있다. 반대로 검사 결과가 양성이면, g 명을 다시 개별 검사한다. 이 경우에는 검사 횟수는 $(1+g)$ 번이다.

각 방법에 대해 1명을 검사하는데 소요되는 평균 검사 횟수인 일인당 기대 검사 횟수를 구해보면, 첫번째 방법은 1이고 두번째 방법은 $1 + \frac{1}{g} - (1-p)^g$ 이다. 예를 들어, $p=0.01$, $g=10$

인 경우, 두번째 방법으로 검사하면 일인당 기대 검사 횟수가 0.1956번으로 개별 검사를 하는 첫번째 방법보다 일인당 기대 검사 횟수가 훨씬 작아진다.

두번째로 제시한 방법을 도프만 절차(dorfman procedure)라 부른다. Dorfman[1943]은 미군의 매독균 검사를 위한 혈액 검사에 이 방법을 적용하였다. 그리고 이와 같이 여러 개의 혈액을 섞어서 먼저 검사해 보는 방법을 그룹 검사(group testing)라 부른다. 일반적으로 질병률이 낮

1) (305-701) 대전시 유성구 구성동 373-1, 한국과학기술원 산업경영학과 박사과정.
2) (305-701) 대전시 유성구 구성동 373-1, 한국과학기술원 산업경영학과 부교수.

은 경우 그룹 검사를 이용하면 기대 검사 횟수가 작아진다. 검사 비용이 크다면 그룹 검사는 경제적으로 바람직하다. 최근 그룹 검사는 AIDS의 항체 검사에 주로 응용된다.

그룹 검사는 혈액 검사 이외의 다른 분야에도 응용될 수 있다. 불량품의 무게가 기준이 되는 무게에 비해 모두 함량 미달이거나 혹은 모두 함량 초과인 경우에 그룹 검사를 적용할 수 있다. (건진지의 불량품 검사 때는 불량품의 전압이 모두 기준 전압보다 높거나 낮은 경우) 일반적으로 여러 개의 부품을 한데 묶거나 혹은 섞어서 검사하여, 모든 부품에 결함이 없거나 혹은 그 중에 적어도 한 개의 부품에 결함이 있는 것을 발견할 수 있는 경우에 그룹 검사를 적용할 수 있다.

모집단으로부터 여러 개의 혈액을 추출한 경우, 엄격히 정의하면 그 전체가 하나의 표본(sample)이다. 그러나 편의상 통상 병원 등에서 사용하는 용어를 사용하여 혈액 각각을 검사용 샘플이라 부르도록 한다. 또한 혈액 검사인 경우에는 검사 결과가 양성 또는 음성으로, 불량품 검사인 경우에는 결함이 있다(defective)거나 결함이 없다(nondefective)로 불리게 되는데, 본 논문에서는 이를 모두 양성 또는 음성으로 부르도록 하겠다. 또한 g 개의 샘플을 섞은 혈액 샘플의 검사를 g 개 샘플의 그룹검사라 부르도록 하겠다. 아울러 그룹검사를 하게 되는 g 개의 혈액 샘플을 크기가 g 인 그룹이라 부르도록 하겠다.

그룹 검사는 크게 질병률이 알려진 경우와 알려지지 않은 경우의 두 분야로 나뉜다. 그리고 검사의 오류가 있는가 없는가에 따라 여러 분야로 나뉘게 된다. 본 논문은 질병률이 알려지고 검사의 오류가 없는 분야를 다룬다.

그룹 검사에 의해 기대 검사 횟수가 크게 줄어들기 때문에 많은 사람들이 이 문제에 관심을 가지게 되었다. Sterrett[1957]는 도프만 절차를 변형한 방법을 제안하였다. 그룹 검사가 양성인 경우, 하나의 양성 샘플이 발견될 때까지 개별 검사하고, 검사하지 않은 나머지 샘플들을 섞어서 다시 그룹으로 검사하는 방법을 제안하였다. Sobel & Groll[1959]은 검사 대상 샘플의 수가 작은 경우, 기대 검사 횟수에 대한 축차 방정식을 이용하여 최적 검사 절차에 대해 광범위한 연구를 하였다.

이제, g 개 샘플의 그룹검사가 양성인 경우를 생각해 보자. 도프만 절차를 따르게 되면 g 개를 개별 검사한다. 그러나 g 개의 샘플을 각각 크기가 g_1, g_2, \dots, g_k 인 k 개의 소그룹(subgroup)으로 쪼갠 후 ($g = \sum_{i=1}^k g_i$), 그 소그룹들에 대해 다시 도프만 절차를 적용할 수 있다. 이와 같이 여러 단계에 걸쳐 도프만 절차를 적용하는 방법을 계층적 도프만 절차(hierarchical Dorfman procedure)라 부른다. Finucan[1964]은 각 소그룹중 그룹검사가 양성인 소그룹의 모든 혈액을 다시 모아 그것을 다시 소그룹으로 나누어 검사하고, 마지막 단계(stage)에서만 개별 검사하는 절차를 제안하였다. 그리고 근사적인 방법을 이용하여 기대 검사 횟수를 구하였다. Mundel[1984]은 계층적 도프만 절차의 기대 검사 횟수의 상한치를 제공하는 방법을 제안하였다. 그리고 Kotz & Johnson[1982]은 검사에 오류가 있는 경우 계층적 도프만 절차에 대한 연구를 하였다.

다시 도프만 절차로 돌아가, g 개 샘플의 그룹검사가 양성인 경우를 생각해 보자. (이 결과는 g 개의 샘플중 적어도 1개의 샘플이 양성임을 의미한다.) 도프만 절차를 따르면 g 개를 다시 개별 검사해야 하므로, 총 검사 횟수는 $(g+1)$ 번이 된다. 그런데 g 개를 다시 개별 검사하는 과

정에서, 첫 샘플부터 $(g-1)$ 번째 샘플의 검사 결과가 모두 음성이었다고 하자. 이런 경우엔, g 번째 샘플은 검사하지 않아도 양성임을 알 수 있다. 그러므로 검사 횟수를 1번 절약할 수 있다. 이와 같이 검사의 종속성 때문에 검사 횟수가 줄어드는 것을 단축 효과(cessation effect)라 부르도록 하자. 이 단축 효과에 대해서는 Sobel & Groll[1959]이 처음으로 언급하였다. 그런데 이러한 효과는 개별 검사뿐만이 아니라 소그룹으로 쪼갠 후 각 소그룹을 그룹검사하는 경우에도 발생한다. g 개의 샘플을 각각 크기가 g_1, g_2, \dots, g_k 인 k 개의 소그룹으로 쪼갠 후, 각 그룹의 혈액을 섞어서 검사한다. 만일 $(k-1)$ 번째까지 소그룹의 그룹검사가 모두 음성이라면, 마지막 k 번째 소그룹에는 양성 샘플이 1개 이상 들어 있게 된다. 그러므로 k 번째 소그룹은 혈액을 섞어서 검사해 볼 필요가 없다. 도프만 절차에서는 단축 효과에 의해 검사 횟수가 줄어드는 양이 전체 검사 횟수에 크게 영향을 끼치지 않기 때문에 이 효과를 무시할 수 있다. 그러나 소그룹으로 나누어 여러 단계의 검사를 하면, 단축 효과에 의해 검사 횟수가 줄어드는 정도가 더욱 커지게 되므로 단축 효과를 무시할 수 없다.

기존의 계층적 도프만 절차에 대한 연구에서는 이 단축 효과가 무시되었다. 본 논문에서는 검사 대상 샘플의 수가 큰 경우, 단축 효과를 고려한 계층적 도프만 절차에 대해서 살펴보고자 하겠다. 2절에서는 소그룹의 크기가 같다는 제약을 첨가한 경우 소그룹핑 방법에 대해 살펴본다. 그리고 샘플당 기대 검사 횟수의 축차 방정식을 유도한다. 그리고 3절에서는 동적 계획법을 이용하여 최적 그룹의 크기를 구한다. 그리고 실제 상황에 간단하고 쉽게 적용할 수 있는 실용적인 방법을 제안한다.

2. 본론

각 검사 대상 샘플이 양성일 확률이 p 이고 음성일 확률이 $q(=1-p)$ 라 하자. p 는 알려져 있고, 전체 검사 대상자의 수는 충분히 크다고 가정한다. 이제 전체 검사 대상 샘플을 각각 g 개씩 묶어, g 개의 샘플을 계층적 도프만 절차에 의해 검사한다. 처음으로 그룹검사를 적용하는 g 개의 샘플을 초기 그룹, 그리고 그 크기(g)를 초기 그룹의 크기라고 부르도록 하자.

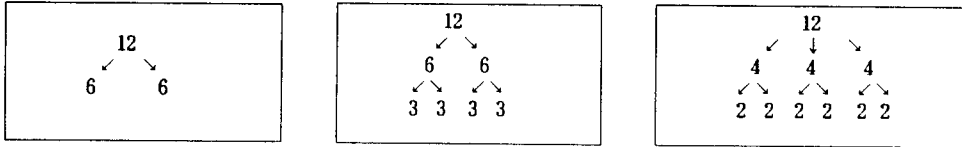
서론에서 언급한 바와 같이, 본 논문에서는 여러 단계에 걸쳐 도프만 절차를 적용하는 계층적 도프만 절차를 이용한다. 이제 어떻게 소그룹으로 쪼개서 검사할 것인가에 대한 소그룹핑(subgrouping) 방법에 대해서 살펴보고자 하자.

g 개 샘플의 그룹검사가 양성이면 g 개의 샘플을 개별 검사하거나, g 개의 샘플을 몇 개의 소그룹으로 쪼개서 검사할 수 있다. g 개의 샘플을 각각 크기가 g_1, \dots, g_k 인 k 개의 소그룹으로 쪼갠다면, 식 (1)을 만족하는 임의의 정수 $g_i (i=1, \dots, k)$ 를 모두 소그룹의 크기로 할 수 있다.

$$g = \sum_{i=1}^k g_i, \quad g_i > 2 \tag{1}$$

그런데 이와 같이 임의 크기의 소그룹으로 쪼개면, 초기 그룹의 크기가 커짐에 따라 가능한 소그룹핑 방법이 기하급수적으로 증가한다. 따라서 최적 그룹의 크기를 구할 수 없을 뿐만 아니라, 초기 그룹의 크기가 주어진 경우에 최적 소그룹핑 방법을 찾기도 불가능하다. 그러므로 본 논문에서는, 소그룹의 크기가 같다는 제약을 첨가하도록 하겠다.

<그림 1, 2, 3>은 소그룹의 크기가 같다는 제약을 첨가한 경우, $g=12$ 일 때, 가능한 소그룹핑 방법의 몇 가지 예이다.



<그림 1> 소그룹핑 방법 1 <그림 2> 소그룹핑 방법 2 <그림 3> 소그룹핑 방법 3

* 참고 : 소그룹핑 방법 1에서는 6개 샘플의 그룹검사가 양성인 경우 6개의 샘플을 개별 검사한다. 반면 소그룹핑 방법 2에서는 6개의 샘플을 크기가 3인 두개의 소그룹으로 쪼개 검사한다. 그리고 3개 샘플의 그룹검사가 양성이면 3개 샘플을 개별 검사한다.

이제 소그룹핑의 각 단계를, 마지막 단계를 '단계 1'로 하여, 거꾸로 한 단계씩 증가시켜 맨 처음 단계를 '단계 n '으로 한다. 아울러 각 단계 i 에서 소그룹의 크기를 x_i 이라 하자(<그림 4> 참조).

| 소그룹핑 방법 | 단계 | 소그룹의 크기 |
|---------|----|------------|
| | 3 | $x_3 = 12$ |
| | 2 | $x_2 = 6$ |
| | 1 | $x_1 = 2$ |

<그림 4> $g=12$, $n=3$ 인 경우, 가능한 소그룹핑 방법

x_n 은 초기 그룹의 크기(g)를 나타낸다. 그리고 x_1 은 그룹검사가 양성인 경우 개별 검사하는 그룹의 크기를 나타낸다. 또한 x_m/x_{m-1} 은 $(m-1)$ 단계에서 소그룹의 갯수를 의미한다. ($m=2, \dots, n$) 전체 검사 절차를 간략히 요약하면 다음과 같다.

< 검사 절차 >

| | |
|-----------|-----|
| 초기 그룹의 크기 | g |
| 소그룹핑 단계 | n |

| | |
|------|------------------|
| 절차 1 | $m = n, x_m = g$ |
|------|------------------|

↓

| | |
|------|----------------------|
| 절차 2 | x_m 개의 샘플을 그룹검사한다. |
|------|----------------------|

↓

| | | |
|-----------|-------------|-----------|
| 그룹검사 결과 ? | --- 음성 ---> | 검사를 종료한다. |
|-----------|-------------|-----------|

↓ 양성

| | | |
|-----------|--------------|-----------------------|
| $m = 1$? | --- Yes ---> | x_1 개의 샘플을 개별 검사한다. |
|-----------|--------------|-----------------------|

↓ No

| | |
|------|--|
| 절차 3 | <ol style="list-style-type: none"> x_m개를 크기가 x_{m-1}인 소그룹으로 쪼갬다. (x_{m-1}은 1과 x_m을 제외한 x_m의 약수이다.) $m \leftarrow m-1$ 각 소그룹들 절차 2 이후의 과정 통해 검사한다. |
|------|--|

2.1 샘플당 기대 검사 횟수에 대한 축차 방정식

임의 중간 단계에서 그룹의 크기를 y 라 하자. 그리고 크기가 y 인 소그룹을 계층적 도프만 절차에 의해 검사한 경우, 검사 횟수의 기대치를 $HT(y)$ 라 하자.

<정리 1> 크기가 y 인 소그룹을 계층적 도프만 절차에 의해 검사한 경우, 검사 횟수의 기대치 $HT(y)$ 를 다음과 같은 축차 방정식으로 나타낼 수 있다.

i) y 개 샘플의 그룹검사가 양성인 경우, y 개의 샘플을 개별 검사할 때

$$HT(y) = 1 + y(1 - q^y) - tq^{y-1} \quad (2.1)$$

ii) y 개 샘플의 그룹검사가 양성인 경우, y 개의 샘플을 각각의 크기가 y_1, \dots, y_k 인 k 개의 소그룹으로 쪼개서 검사할 때

$$HT(y) = 1 - kq^y - q^{y-y_1}(1 - q^{y_1}) + \sum_{i=1}^k HT(y_i) \quad (2.2)$$

$$y = \sum_{i=1}^k y_i, \quad y_i \geq 2, \quad y_i \text{는 정수.}$$

정리 1에 대한 증명은 부록 1에서 다룬다. 각 단계에서 소그룹의 크기가 $(x_n, x_{n-1}, \dots, x_1)$ 인 경우, '단계 m '에서 x_m 개 샘플의 기대 검사 횟수를 $T_m(x_m, \dots, x_1)$, 샘플당 기대 검사 횟수를 $I_m(x_m, \dots, x_1)$ 이라 하자 ($m=1, \dots, n$).

이제 '단계 m '에서 x_m 개의 샘플을 검사한다고 하자. 먼저, '단계 1'에서는 x_1 개의 샘플을 개별 검사한다. 식 (2.1)에서 $y=x_1$ 인 경우이므로

$$T_1(x_1) = 1 + x_1(1 - q^{x_1}) - pq^{x_1-1} \quad (3)$$

이다. $m \geq 2$ 인 '단계 m '에서는 x_m 개의 샘플을 검사하므로, 식 (2.2)에서 $y=x_m$, $y_i=x_{m-1}$ ($i=1, \dots, k$)인 경우이다. 그러므로 식 (2.2)에서 좌변의 $HT(y)$ 는 $T_m(x_m, \dots, x_1)$ 로, 우변의 $\sum_{i=1}^k HT(y_i)$ 는 $\sum_{i=1}^k T_{m-1}(x_{m-1}, \dots, x_1)$ 로 나타낼 수 있다. 또한 x_m 개 샘플의 그룹검사가 양성이면, x_m 개의 샘플을 크기가 x_{m-1} 인 소그룹으로 쪼개 검사하므로, 식 (2.2)의 우변에서 소그룹의 갯수 k 를 x_m/x_{m-1} 으로 나타낼 수 있다. 그러므로 식 (2.2)를 $T_m(x_m, \dots, x_1)$ 의 식으로 고치면 다음을 얻는다.

$$T_m(x_m, \dots) = 1 - kq^{x_m} - q^{(k-1)x_{m-1}}(1 - q^{x_{m-1}}) + kT_{m-1}(x_{m-1}, \dots), \quad m=2, \dots, n \quad (4)$$

$k = x_m/x_{m-1}$

또한, x_m 개 샘플의 샘플당 기대 검사 횟수는

$$I_m(x_m, \dots, x_1) = \frac{T_m(x_m, \dots, x_1)}{x_m}, \quad m=1, \dots, n \quad (5)$$

이므로, 식 (3), (4)의 양변을 x_m 으로 나누고 식 (5)를 이용하면 샘플당 기대 검사 횟수에 대한 축차 방정식을 얻을 수 있다.

$$I_n(x_n, \dots, x_1) = \frac{1}{x_n} + \left\{ \frac{1}{x_n} - \frac{1}{x_{n-1}} - \frac{1}{x_n q^{x_{n-1}}} \right\} q^{x_n} + I_{n-1}(x_{n-1}, \dots, x_1) \quad (6)$$

$$I_1(x_1) = \frac{1}{x_1} + 1 - q^{x_1} - \frac{pq^{x_1-1}}{x_1}$$

여기에서, x_{n-1} 은 1과 x_n 을 제외한 x_n 의 약수라는 제약 조건이 따른다 ($n=2, \dots$).

2.2 최적 그룹의 크기를 구하는 방법

샘플당 기대 검사 횟수가 최소화되는 최적 단계 n 을 n^* 라 하자. 그리고 최적 그룹의 크기를 $(x_1^*, x_2^*, \dots, x_{n^*}^*)$ 라 하자. 아울러 x_n 이 주어졌을 때, 샘플당 기대 검사 횟수의 최소값을

$I_n^*(x_n)$ 라 하자. 식 (6)에서 $n \geq 2$ 인 경우에는

$$\begin{aligned} I_n^*(x_n) &= \text{Min} \left[\frac{1}{x_n} + \left\{ \frac{1}{x_n} - \frac{1}{x_{n-1}} - \frac{1}{x_n q^{x_{n-1}}} \right\} q^{x_n} + I_{n-1}(x_{n-1}, \dots, x_1) \right] \\ &= \text{Min} \left[\frac{1}{x_n} + \left\{ \frac{1}{x_n} - \frac{1}{x_{n-1}} - \frac{1}{x_n q^{x_{n-1}}} \right\} q^{x_n} + I_{n-1}^*(x_{n-1}) \right] \end{aligned} \quad (7)$$

를 얻는다. 여기에서, x_{n-1} 은 1과 x_n 을 제외한 x_n 의 약수라는 제약 조건이 따른다. 또한 $n=1$ 인 경우에는

$$I_1^*(x_1) = \frac{1}{x_1} + 1 - q^{x_1} - \frac{pq^{x_1-1}}{x_1} \quad (8)$$

이므로, 식 (7), (8)로 부터 $I_n^*(x_n)$ 에 관한 축차방정식을 얻을 수 있다.

$$\begin{cases} I_n^*(x_n) = \text{Min} \left[\frac{1}{x_n} + \left\{ \frac{1}{x_n} - \frac{1}{x_{n-1}} - \frac{1}{x_n q^{x_{n-1}}} \right\} q^{x_n} + I_{n-1}^*(x_{n-1}) \right] \\ I_1^*(x_1) = \frac{1}{x_1} + 1 - q^{x_1} - \frac{pq^{x_1-1}}{x_1} \end{cases} \quad (9)$$

여기에서, $x_n \geq 2$ 이고 x_n 은 정수이다 ($n=1, 2, \dots$). 그리고 x_{n-1} 은 1과 x_n 을 제외한 x_n 의 약수이다 ($n=2, \dots$).

모든 가능한 x_n 에 대해, $I_n^*(x_n)$ 의 최소값을 $I_n^*(x_n^*)$ 이라 하자. 그리고 그때 그룹의 크기를 x_n^* 이라 하면,

$$I_n^*(x_n^*) = \text{Min} \{ I_n^*(x_n) \}, \text{ for } x_n \ (n=1, 2, \dots) \quad (10)$$

이다. 그리고 최적 단계 n^* 는,

$$n^* = i \quad \text{s.t.} \quad \text{Min} \{ I_i^*(x_i^*) \}, \ (i=1, 2, \dots) \quad (11)$$

이다. 동적 계획법을 이용하여 식 (9), (10), (11)을 풀면 샘플당 기대 검사 횟수를 최소화하는 그룹의 크기와 단계를 구할 수 있다. (간단한 예제를 부록 2에서 다룬다.)

3. 풀이 결과

<표-1>은 초기 그룹의 크기가 10000이하인 범위에서, 질병률을 변화시켜가며 식 (9), (10), (11)을 풀 결과이다. <표-1>은 최적 그룹의 크기가 2^n ($n=1, 2, \dots$)의 형태임을 보여준다. 소 그룹의 크기가 같다는 제약하에서 계층적 도프만 절차를 이용하면 다음의 검사 절차가 최적이다.

< 검사 절차 >

질병률이 0.381보다 큰 경우 : 혈액 샘플을 개별 검사한다.

질병률이 0.381보다 작은 경우 :

1. <표-2>에서 초기 그룹의 크기를 결정한다.
2. 그룹검사가 양성이면, 그 그룹을 같은 크기의 소그룹 두개로 쪼개 검사한다.
3. 소그룹의 크기가 2이고 그룹검사가 양성이면, 혈액 샘플들을 개별 검사한다.

<표-1> 최적 그룹의 크기

| 질병률 | 샘플당 기대 검사 횟수 | 최적 그룹의 크기 |
|------------------|----------------------|---|
| 0.0001 0.0002 | 0.001984 0.003668 | 8192 4096 2048 1024 512 256 128 64 32 16 8 4 2 4096 2048 1024 512 256 128 64 32 16 8 4 2 |
| 0.0003 - 0.0004 | 0.005234 - 0.006738 | 2048 1024 512 256 128 64 32 16 8 4 2 |
| 0.0005 - 0.0009 | 0.008188 - 0.013607 | 1024 512 256 128 64 32 16 8 4 2 |
| 0.0010 - 0.0018 | 0.014880 - 0.024524 | 512 256 128 64 32 16 8 4 2 |
| 0.0019 - 0.0037 | 0.025678 - 0.044707 | 256 128 64 32 16 8 4 2 |
| 0.0038 - 0.0074 | 0.045693 - 0.078447 | 128 64 32 16 8 4 2 |
| 0.0075 - 0.014 | 0.079314 - 0.129281 | 64 32 16 8 4 2 |
| 0.015 - 0.029 | 0.136554 - 0.223865 | 32 16 8 4 2 |
| 0.030 - 0.058 | 0.229497 - 0.365416 | 16 8 4 2 |
| 0.059 - 0.113 | 0.369514 - 0.561400 | 8 4 2 |
| 0.114 - 0.213 | 0.564198 - 0.796069 | 4 2 |
| 0.214 - 0.381 | 0.798102 - 0.998920 | 2 |
| 0.382 - 1.00 | | 개별 검사 |

* 참고 : 질병률이 p 일 때 초기 그룹의 크기를 $g^*(p)$ 라 하자. <표-1>을 살펴보면 $0.0005 \leq p \leq 0.381$ 범위에서 대략적으로 $p \cdot g^*(p) \geq 0.5$ 를 만족한다. 이 결과를 토대로 실제 상황에 초기 그룹의 크기를 쉽고 간단하게 구할 수 있는 실용적인 방법을 다음과 같이 제안한다.

< 실용적인 방법 >

$$g^*(p) \approx 2^{n^*}, \quad n^* = \min \left\{ n \mid 2^n \geq \left(\frac{0.5}{p} \right) \right\}$$

<표-2> 초기 그룹의 크기

| 질병률 | 초기 그룹의 크기 | 질병률 | 초기 그룹의 크기 |
|-----------------|-----------|----------------|-----------|
| 0.0001 | 2^{13} | 0.0075 ~ 0.014 | 2^6 |
| 0.0002 | 2^{12} | 0.015 ~ 0.029 | 2^5 |
| 0.0003 ~ 0.0004 | 2^{11} | 0.030 ~ 0.058 | 2^4 |
| 0.0005 ~ 0.0009 | 2^{10} | 0.059 ~ 0.113 | 2^3 |
| 0.0010 ~ 0.0018 | 2^9 | 0.114 ~ 0.213 | 2^2 |
| 0.0019 ~ 0.0037 | 2^8 | 0.214 ~ 0.381 | 2 |
| 0.0038 ~ 0.0074 | 2^7 | 0.382 ~ 1 | 1 |

4. 맺음말

본 논문에서는, 그룹 검사 방법 중에서 계층적 도프만 절차를 다루었다. 그리고 기존 연구에서 무시되었던 단축 효과를 고려하였다. 소그룹의 크기가 같다는 제약을 첨가하여 샘플당 기대 검사 횟수에 대한 축차 방정식을 유도하였다. 그리고 이를 이용하여 샘플당 기대 검사 횟수를 최소화하는 소그룹의 크기를 구하였다. 계층적 도프만 절차를 이용하여 소그룹의 크기가 같다는 제약 및 초기 그룹의 크기가 10000이하인 범위에서, 최적 그룹의 크기가 항상 2^n ($n=1,2,\dots$)의 형태임을 밝혀냈다. 이 결과를 토대로 실제 상황에 적용하기 쉬운 실용적인 방법을 제안하였다.

참 고 문 헌

- [1] Dorfman, R. (1943). The Detection of Defective Members of Large Populations, *Annals of Mathematical Statistics*, Vol. 14, 436-440.
- [2] Eugene Livax, Xin M. Tu, and Marcello Pagnao (1994). Screening for the Presence of a Disease by Pooling Sera Samples, *Journal of the American Statistical Association*, Vol. 89, 424-434.
- [3] Finucan, H. M. (1964). The Blood Testing Problem, *Applied Statistics*, Vol. 13, 43-50.
- [4] Kotz, S. and Johnson, L. N. (1982). Errors in Inspection and Grading: Distributional Aspects of Screening and Hierarchical Screening, *Communications in Statistics - Theory and Meth.*, Vol. 11, 1997-2016.
- [5] Mundel, A. B. (1984). Group Testing, *Journal of Quality Technology*, Vol. 16, 181-188.
- [6] Sobel, M. and Groll, P. A. (1959). Group Testing To Eliminate Efficiently All Defectives in a Binomial Sample, *Bell System Technical Journal*, Vol. 38, 1179-1252.
- [7] Sterrett, A. (1957). On The Detection of Defective Members of Large Populations, *Annals of Mathematical Statistics*, Vol. 28, 1033-1036.

부 록 1 : 정리 1의 증명

양성일 확률이 p 이고 음성일 확률이 $q(=1-p)$ 인 샘플이 y 개 있다. 계층적 도프만 절차에 의해 y 개의 샘플을 검사한 경우, 검사 횟수의 기대치를 $HT(y)$ 라 하자. 그리고 y 개 샘플중 적어도 한 개의 샘플이 양성이라는 조건하에서 y 개 샘플의 기대 검사 횟수를 $G(y)$ 라 하자. 먼저 y 개 샘플을 그룹검사한다. 검사 결과가 음성이면 1번의 검사로 y 개 샘플의 검사가 종료된다. 그러나 검사 결과가 양성이면 y 개의 샘플을 다시 검사해야 한다.

$$HT(y) = 1 + (1-q^y)G(y) \quad (1)$$

이제 y 개 샘플의 그룹검사가 양성일 때, y 개의 샘플을 개별 검사하는 경우와 여러 개의 소그룹으로 쪼개 검사하는 경우에 대해 $G(y)$ 를 구해보자.

i) y 개의 샘플을 개별 검사하는 경우

처음부터 $(y-1)$ 개까지의 샘플이 모두 음성이고 마지막 y 번째 샘플이 양성이면, 마지막 샘플은 검사해보지 않아도 양성임을 알 수 있다. 이런 경우에는 검사 횟수가 $(y-1)$ 번이다. (검사 횟수가 1번 줄어드는 단축 효과가 발생한다.) 그리고 나머지 경우는 모두 y 번이므로,

$$G(y) = \frac{1}{1-q^y} \{y(1-q^y) - pq^{y-1}\} = y - \frac{pq^{y-1}}{1-q^y} \quad (2)$$

ii) y 개의 샘플을 k 개의 소그룹으로 나누어 검사하는 경우 ($k \geq 2$)

각 소그룹의 크기를 y_i ($i=1, 2, \dots, k$)라 하자. 그리고 A_i 를 i 번째 소그룹에 양성 샘플이 들어있는지의 유무를 나타내는 변수라 하자.

$$y = \sum_{i=1}^k y_i, \quad y_i \geq 2$$

$$A_i = \begin{cases} 1 & : i\text{번째 소그룹에 적어도 한개의 양성 샘플이 있다} \\ 0 & : i\text{번째 소그룹이 모두 음성 샘플이다.} \end{cases}$$

a) $A_1 = \dots = A_{k-1} = 0, A_k = 1$ 인 경우

이 경우엔 처음부터 $(k-1)$ 번째까지 소그룹의 그룹검사가 모두 음성으로 판정된다. 그런데 우리는 y 개 샘플중 적어도 1개가 양성 샘플임을 알고 있다. 그러므로 마지막 k 번째 소그룹은 그룹검사하지 않아도 적어도 1개의 양성 샘플이 있음을 알 수 있다. (검사 횟

수가 1번 줄어드는 단축 효과가 발생한다.) 그러므로 기대 검사 횟수는 $(k-1)+G(y_k)$ 이다. 그리고 이 경우의 확률은 $\{q^{y-y_k}(1-q^{y_k})\}/(1-q^y)$ 이다.

b) 나머지 경우

기대 검사 횟수가 $k + \sum_{i=1}^k G(y_i)A_i$ 이고, 확률은 $\left\{ q^{y-\sum_{i=1}^k y_i A_i} \prod_{i=1}^k (1-q^{y_i A_i}) \right\} / (1-q^y)$ 이다.

그러므로 $G(y)$ 는,

$$G(y) = \frac{1}{1-q^y} \left\{ \left(\sum_{(\sum_{i=1}^k A_i \geq 1 \text{인 모든 } A_i)} [k + G(y_i)A_i] \left[q^{y-\sum_{i=1}^k y_i A_i} \prod_{i=1}^k (1-q^{y_i A_i}) \right] \right) - q^{y-y_k}(1-q^{y_k}) \right\}$$

정리하면,

$$G(y) = \frac{1}{1-q^y} \left\{ k(1-q^y) - q^{y-y_k}(1-q^{y_k}) + \sum_{i=1}^k G(y_i)(1-q^{y_i}) \right\} \quad (3)$$

식 (1), (2), (3)을 이용하면 <정리 1>의 식 (2.1) (2.2)를 얻을 수 있다.

부 록 2 : 예 제

다음은 질병률이 0.1이고, 각 단계에서 그룹의 크기를 24이하로 제한하며, 3단계 이하로 소그룹핑하는 경우, 동적 계획법을 이용하여 식 (9), (10), (11)을 풀 예제이다. ($p=0.1$, $n \leq 3$, $x_i \leq 24$ ($i=1, 2, 3$)).

i) $n=1$

| x_1 | 2 | 3 | 4 | 5 | 6 | 7 | ... |
|--------------|--------|----------|----------|----------|----------|----------|-----|
| $I_1^*(x_1)$ | 0.6450 | 0.577333 | 0.575674 | 0.596388 | 0.625384 | 0.656968 | ... |

$$I_1^*(x_1^*) = \text{Min}\{I_1^*(x_1)\} \quad , \quad \text{for } x_1 = 2, 3, 4, \dots$$

$$\therefore x_1^* = 4 \quad , \quad I_1^*(x_1^*) = 0.575674.$$

ii) $n=2$

| x_2 | x_1 | | | | | | $I_2^*(x_2)$ | x_1^* |
|-------|----------|----------|----------|----------|----------|-----|--------------|---------|
| | 2 | 3 | 4 | 5 | 6 | ... | | |
| 4 | 0.528475 | - | - | - | - | ... | 0.528475 | 2 |
| 6 | 0.525170 | 0.533926 | - | - | - | ... | 0.525170 | 2 |
| 8 | 0.542145 | - | 0.564854 | - | - | ... | 0.542145 | 2 |
| 9 | - | 0.543302 | - | - | - | ... | 0.543302 | 3 |
| 10 | 0.562482 | - | - | 0.602471 | - | ... | 0.562482 | 2 |
| 12 | 0.581598 | 0.557774 | 0.576064 | - | 0.640895 | ... | 0.557774 | 3 |
| 14 | 0.598212 | - | - | - | - | ... | 0.598212 | 2 |
| 15 | - | 0.570267 | - | 0.612357 | - | ... | 0.570267 | 3 |
| 16 | 0.612132 | - | 0.585779 | - | - | ... | 0.585779 | 4 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |

$$\begin{aligned} \therefore I_2^*(x_2^*) &= \text{Min}\{I_2^*(x_2)\}, \text{ for } x_2=4,6,8,\dots \\ &= \text{Min}\{0.528475, 0.525170, 0.542145, \dots\} \\ &= 0.525170 \text{ 그리고 } x_2^*=6, x_1^*=2. \end{aligned}$$

iii) $n=3$

| x_3 | x_2 | | | | | | $I_3^*(x_3)$ | x_2^* |
|-------|----------|----------|----------|----------|----------|-----|--------------|---------|
| | 4 | 6 | 8 | 9 | 10 | ... | | |
| 8 | 0.517654 | - | - | - | - | ... | 0.517654 | 4 |
| 12 | 0.528864 | 0.540680 | - | - | - | ... | 0.528864 | 4 |
| 16 | 0.538579 | - | 0.566159 | - | - | ... | 0.538579 | 4 |
| 18 | - | 0.548358 | - | 0.568996 | - | ... | 0.548358 | 6 |
| 20 | 0.544895 | - | - | - | 0.588967 | ... | 0.544895 | 4 |
| 24 | 0.548458 | 0.550612 | 0.569443 | - | - | ... | 0.548458 | 4 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |

$$\begin{aligned} \therefore I_3^*(x_3^*) &= \text{Min}\{I_3^*(x_3)\}, \text{ for } x_3=8,12,16,\dots \\ &= \text{Min}\{0.517654, 0.528864, 0.538864, \dots\} \\ &= 0.517654 \text{ 그리고 } x_3^*=8, x_2^*=4, x_1^*=2. \end{aligned}$$

$$\{I_1^*(x_1^*), I_2^*(x_2^*), I_3^*(x_3^*)\} = \{0.575674, 0.525170, 0.517654\}$$

$$\therefore n^*=3, x_3^*=8, x_2^*=4, x_1^*=2.$$

그러므로 질병률이 0.1이고, 각 단계에서 그룹의 크기를 24이하로 제한하며, 3단계 이하로 소그룹평하는 경우, 검사 대상 샘플을 (8,4,2)의 크기로 소그룹평하여 검사하면 기대 검사 횟수를 최소화할 수 있다.

An Efficient Blood Testing Procedure by Using Group Testing

W. S. Yang³⁾, K. C. Chae⁴⁾

Abstract

Suppose we have g items classified as defective or as nondefective. If we test each item separately, then g tests are required. However, it is possible to test a group of g items so as to detect whether at least one of them is defective, and to proceed to individual testing only if the presence of at least one defective item is indicated. The method is called Dorfman procedure. In this paper, we consider a variant of Dorfman procedure called a hierarchical Dorfman procedure. We also consider the curtailment effect resulted from the dependency of the tests. We propose an efficient blood testing procedure. We also present a heuristic rule of thumb which is simple and easily implementable.

3) Ph. D Candidate, Dept. of Industrial Management, KAIST, Taejon-shi, 305-701, Korea.

4) Associate Professor, Dept. of Industrial Management, KAIST, Taejon-shi, 305-701, Korea.