

독립인 두 모집단 설계에서의 표본수 비교

고해원¹ · 김동재²

¹²가톨릭대학교 의학통계학교실

접수 2010년 10월 1일, 수정 2010년 11월 18일, 게재확정 2010년 11월 23일

요약

임상시험을 시행하는 경우 위약과 신약과 비교하는 경우가 대다수이다. 기존에 독립인 두 모집단의 표본수를 계산하는 방법으로 모수적 방법에서는 t 검정을 이용하였고, 비모수적 방법에서는 Wilcoxon 순위합검정 (Wilcoxon, 1945)을 이용하였다. 본 논문에서는 Orban과 Wolfe (1982)가 제안한 선형위치통계량의 검정법과, Kim (1994)이 선형위치통계량에 기초하여 계산한 검정력의 결과를 이용하여 표본수 구하는 방법을 제안한다. 또한 앞서 제안한 방법의 표본수를 기존의 Wilcoxon 순위합검정을 이용하여 Wang 등 (2003)이 제안한 공식을 이용한 표본수, 그리고 모수적 방법을 이용한 t 검정의 표본수와 비교하였다.

주요용어: 비모수적 방법, 위치, 표본수.

1. 서론

신약을 개발하면 치료효과가 존재하는지 알아보기 위해 임상시험을 시행하여 효과유무를 검정한다. 이러한 임상시험에서는 독립인 두 모집단 설계가 일반적으로 사용되는데 그 예로 3상 시험 (phase III clinical trials)을 들 수 있다. 임상시험에서 만약 피험자수가 부족하면 실제 존재하는 중요한 효과를 알아낼 수 없고, 통계적 검정력이 떨어져서 의약품에 효과가 있음에도 불구하고 이를 찾아낼 수 없다. 반대로 피험자 수가 불필요하게 많게 된다면 시간과 돈이 많이 들고 안정성 및 윤리성이 입증되지 않은 의약품이라면 윤리적인 문제가 발생하게 된다. 따라서 피험자 수에 영향을 미칠 수 있는 유의수준, 검정력, 치료효과의 크기 등을 감안하여 피험자수를 최적화해야 하고, 적절한 피험자수를 정하는 것은 매우 중요한 문제라고 할 수 있다.

일반적으로 표본수와 검정력은 연관성이 높다. 표본수가 적은 경우 검정력은 낮아지고, 검정력을 높이기 위해서는 많은 표본수가 필요로 한다. 기존에는 비모수적 방법인 Wilcoxon 순위합과 모수적 방법의 t 검정을 이용하여 표본수를 계산하였다.

Kim과 Kim (2007)은 비열등성 시험에서 고정된 유의수준과 검정력하에서 여러 상황을 가정하여 비모수적 방법인 Wilcoxon의 순위합검정을 이용한 Wang 등 (2003)의 표본수 공식을 이용하여 표본수를 구하고, 모수적 방법의 t 검정을 이용하여 계산한 표본수와 비교하였다. 그 결과 Wilcoxon 순위합검정을 이용하였을 때가 t 검정을 이용하였을 때보다 표본수가 컸다.

또한 두 모집단 비교에서 Orban과 Wolfe (1982)은 비모수적 방법인 선형위치통계량을 이용한 검정법을 제안하였다. 그리고 Kim (1994)이 선형위치통계량을 이용한 검정법의 검정력을 계산하였다. 기

¹ (137-701) 서울시 서초구 반포동 505번지, 가톨릭대학교 의학통계학교실, 대학원생.

² 교신저자: (137-701) 서울시 서초구 반포동 505번지, 가톨릭대학교 의학통계학교실, 교수.

E-mail: djkim@catholic.ac.kr

존의 비모수적인 방법의 Wilcoxon 순위합검정을 이용한 경우 모수적 방법인 t 검정을 이용한 경우 보다 표본수가 커지므로 시간과 비용이 많이 드는 단점이 있다. 본 논문에서는 이러한 단점을 보완하기 위하여, 선형위치통계량을 이용한 검정법에서의 표본수를 제시한다. 또한 유의수준과 검정력을 고정하여 본 논문에서 제안하는 방법과 기존의 방법들을 이용하여 구한 표본수와 비교하여 본다.

2. 두 모집단 검정에 대한 표본수 계산방법

확률표본 $X_i, i = 1, \dots, m$ 과 $Y_j, j = 1, \dots, n$ 을 각각 대조군의 누적분포함수 $F(x)$ 와 치료군의 누적분포함수 $G(x) = F(x - \theta)$ 에서 추출된 확률변수라고 하자. 이때 θ 는 두 모집단의 치료효과 차이를 나타내는 위치 모수이고, 검정에 대한 가설은 다음과 같이 나타낼 수 있다.

$$H_0 : \theta = 0 \text{ 대 } H_1 : \theta \neq 0$$

2.1. Wilcoxon 순위합검정에서의 표본수

Wilcoxon 순위합검정의 검정통계량 W 는 대조군과 치료군의 통합 순위 중 치료군에서의 순위합

$$\begin{aligned} W &= \sum_{j=1}^n R_j \\ &= \sum_{j=1}^n \left(\sum_{i=1}^n I\{Y_j \geq Y_i\} + \sum_{i=1}^m I\{Y_j \geq X_i\} \right) \\ &= n(n+1)/2 + \sum_{i=1}^m \sum_{j=1}^n I\{Y_j \geq X_i\} \end{aligned}$$

으로 정의할 수 있다. 여기서 R_j 는 치료군의 통합순위 ($i = 1, \dots, m, j = 1, \dots, n, i \neq j$)이다. 그리고 귀무가설 하에서 순위합 W 의 평균과 분산은

$$\begin{aligned} E(W) &= n(n+m+1)/2 \\ \text{var}(W) &= mn(m+n+1)/12 \end{aligned}$$

이다.

이를 이용하여 Wilcoxon 순위합의 검정통계량은 근사적으로 표준정규분포를 가지는 다음과 같은 W^* 로 나타낼 수 있다.

$$W^* = \frac{W - E(W)}{\sqrt{\text{var}(W)}} = \frac{W - n(n+m+1)/2}{\sqrt{mn(m+n+1)/12}}$$

그리고 $|W^*| \geq z_\alpha$ 을 만족하면 유의수준 α 하에서 귀무가설을 기각한다. 일반적으로 치료군의 순위합

W의 평균과 분산은

$$\begin{aligned} \mu_W &= n(n+1)/2 + mnp_1, \\ \sigma_W^2 &= \text{var}(W) = \text{var}\left(n(n+1)/2 + \sum_{j=1}^n \sum_{i=1}^m I(Y_j \geq X_i)\right) \\ &= \text{var}\left(\sum_{j=1}^n \sum_{i=1}^m I(Y_j \geq X_i)\right) \\ &= mn \text{var}(I(Y_1 \geq X_1)) + mn(m-1) \text{cov}(I(Y_1 \geq X_1), I(Y_1 \geq X_2)) \\ &\quad + mn(n-1) \text{cov}(I(Y_1 \geq X_1), I(Y_2 \geq X_1)) \\ &= mnp_1(1-p_1) + mn(m-1)(p_2 - p_1^2) + mn(n-1)(p_3 - p_1^2) \end{aligned}$$

이고, 여기서 p_1, p_2, p_3 은

$$\begin{aligned} p_1 &= P(Y_1 \geq X_1), \\ p_2 &= P(Y_1 \geq X_1, Y_1 \geq X_2) \\ p_3 &= P(Y_1 \geq X_1, Y_2 \geq X_1) \end{aligned}$$

이다. 따라서 대립가설이 참일 경우의 Wilcoxon 순위합검정의 검정력은

$$\begin{aligned} \text{Power} &= P(|W^*| > z_{\alpha/2} | H_1) \\ &\approx P(W^* > z_{\alpha/2} | H_1) \\ &= P\left(\frac{W - n(n+1)/2 - mnp_1}{\sigma_W} > \frac{z_{\alpha/2} \sqrt{mn(m+n+1)/12 + mn(1/2 - p_1)}}{\sigma_W}\right) \end{aligned}$$

이고, 두 군의 표본수가 동일할 때 검정력이 가장 크므로 $m = n$ 으로 놓으면 검정력과 각 군당 표본수는 각각

$$\begin{aligned} \text{Power} &= 1 - \phi\left(\frac{z_{\alpha/2} \sqrt{1/6} - \sqrt{n}(1/2 - p_1)}{\sqrt{p_2 + p_3 - 2p_1^2}}\right) \\ n &= \left(\frac{z_{\alpha/2} \sqrt{1/6} + z_{\beta} \sqrt{p_2 + p_3 - 2p_1^2}}{1/2 - p_1}\right)^2 \end{aligned} \tag{2.1}$$

이다. 여기서 $\phi(\cdot)$ 는 표준정규분포의 분포함수이다.

따라서 Wilcoxon 순위합검정의 검정력은 모집단의 분포에 따라 p_1, p_2, p_3 를 다음과 같이 계산하여 구할 수 있다. 대조군과 치료군의 분포가 정규분포 $N(\mu_X = \theta_0, \sigma_X^2), N(\mu_Y = \theta_1, \sigma_Y^2)$ 을 따르는 경우에는 Simonoff 등 (1986)에 의해 제시된 방법으로

$$\begin{aligned} p_1 &= \phi((\mu_Y - \mu_X) / \sqrt{\sigma_X^2 + \sigma_Y^2}) \\ p_2 &= p_1^2 / (p_1^2 + p_1(1-p_1) + (1-p_1)^2) = p_1^2 / (p_1^2 - p_1 + 1) \\ p_3 &= p_1^2 / (p_1^2 - p_1 + 1) \end{aligned}$$

이다. 또 대조군과 치료군의 분포가 코쉬분포 $Cauchy(\theta_0, \gamma)$, $Cauchy(\theta_1, \gamma)$ 을 따르는 경우 p_1, p_2, p_3 은

$$p_1 = \int_{-\infty}^{\infty} \int_{-\infty}^y \frac{1}{\pi} \left[\frac{\gamma}{(x - \theta_0)^2 + \gamma^2} \right] \frac{1}{\pi} \left[\frac{\gamma}{(y - \theta_1)^2 + \gamma^2} \right] dx dy$$

$$p_2 = p_1^2 / [p_1^2 + p_1(1 - p_1) + (1 - p_1)^2] = p_1^2 / (p_1^2 - p_1 + 1)$$

$$p_3 = p_1^2 / (p_1^2 - p_1 + 1)$$

이고, 대조군과 치료군의 분포가 이중지수분포 $Laplace(\theta_0, b)$, $Laplace(\theta_1, b)$ 을 따르는 경우의 p_1, p_2, p_3 는

$$p_1 = 1 + (\sigma_X / 4(\sigma_X + \sigma_Y) - 1/2) \exp((\mu_X - \mu_Y) / \sigma_Y)$$

$$+ ((\mu_X - \mu_Y) / 4\sigma_Y) \exp((\mu_X \sigma_Y - \mu_Y \sigma_X) / \sigma_X \sigma_Y)$$

$$- \sigma_X / [4(\sigma_X + \sigma_Y)] \exp((\mu_X - \mu_Y) / \sigma_Y), \quad \sigma_X^2 = \sigma_Y^2$$

$$p_2 = p_1^2 / [p_1^2 + p_1(1 - p_1) + (1 - p_1)^2] = p_1^2 / (p_1^2 - p_1 + 1)$$

$$p_3 = p_1^2 / (p_1^2 - p_1 + 1)$$

이다. 따라서 모집단의 분포에 따라 계산된 p_1, p_2, p_3 를 이용하여 검정력과 표본수를 식 (2.1)에 대입하여 구할 수 있다.

2.2. 선형위치통계량을 이용한 검정법에서의 표본수

선형위치통계량을 이용한 검정법에서 위치 (placement) 확률변수 U_1, U_2, \dots, U_n 은 Orban과 Wolfe (1982)에서와 같이

$$mU_i = [\text{number of } X' \text{'s} \leq Y_i], i = 1, 2, \dots, n$$

로 정의된다. 여기서 m 은 대조군의 표본수이다. 선형위치통계량은

$$S_{n,m}^P = \sum_{i=1}^n \varphi_m(U_i)$$

이다. 여기서 $\varphi_m(x)$ 는 $[0,1]$ 에서 정의된 실수 값인 점수함수이다. 그리고 $S_{n,m}^P \geq s_\alpha$ 를 만족하면 유의 수준 α 하에서 귀무가설을 기각한다. 여기서 s_α 는 $S_{n,m}^P$ 의 귀무가설하에서 분포에 대한 $(100 \times \alpha)\%$ 백분위수이다. 본 논문에서는 균일 점수함수와 이중지수 점수함수를 이용한 선형위치통계량을 고려하며, 이때 점수함수 형태는 다음과 같다.

$$\varphi(x) = x \quad \text{균일 점수함수}$$

$$\varphi(x) = -\ln(1 - x) \quad \text{이중지수 점수함수}$$

단, 균일 점수함수를 사용하는 경우에는 선형위치통계량을 이용한 검정은 Wilcoxon 순위합검정과 같음이 알려져 있다 (Orban과 Wolfe, 1982). 선형위치통계량의 평균과 분산은

$$\bar{\varphi} = E(\varphi[F(Y_i)]) = \int \varphi[F(y)] dG(y) \quad (2.2)$$

$$V_\varphi = Var(\varphi[F(Y_i)]) = \int [\varphi[F(y)] - \bar{\varphi}]^2 dG(y) \quad (2.3)$$

이고, Kim (1994)에 의해서 계산된 선형위치통계량을 이용한 검정의 검정력은

$$\begin{aligned}\beta(\theta) &= P(S_{n,m}^P > S_{\alpha/2} : H_1 \text{ is true}) \\ &\approx P(Z \geq [n\bar{\varphi}_0 + z_{\alpha/2}\sqrt{nV_{\varphi_0}} - n\bar{\varphi}_\theta] / \sqrt{nV_{\varphi_0}}) \\ &= 1 - \phi(z_{\alpha/2} + \sqrt{n}(\bar{\varphi}_0 - \bar{\varphi}_\theta) / \sqrt{V_{\varphi_0}})\end{aligned}$$

이고, $\bar{\varphi}_0$ 와 V_{φ_0} 는 각각 귀무가설하에서의 평균과 분산이다. 여기서 식(2.1)과 (2.2)을 귀무가설하에서 대입하면

$$\begin{aligned}&\sqrt{n}(\bar{\varphi}_0 - \bar{\varphi}_\theta) / \sqrt{V_{\varphi_0}} \\ &= \sqrt{n} \left(\int \varphi(F(y))dF(y) - \int \varphi(F(y))dF(y - \theta) \right) / \sqrt{\int \varphi(F(y) - \bar{\varphi}_0)^2 dF(y)} \\ &= \sqrt{n} \left(\int \varphi(F(y))dF(y) - \int \varphi(F(y + \theta))dF(y) \right) / \sqrt{\int \varphi(F(y) - \bar{\varphi}_0)^2 dF(y)} \\ &\approx -\theta\sqrt{n} \left(\int (\varphi F)'(y)dF(y) \right) / \sqrt{\int \varphi(F(y) - \bar{\varphi}_0)^2 dF(y)}\end{aligned}$$

이고, 이를 이용하여 검정력과 각 군의 표본수는

$$\begin{aligned}\beta(\theta) &\approx 1 - \phi \left(z_{\alpha/2} - \frac{\theta\sqrt{n}(\int (\varphi F)'(y)dF(y))}{\sqrt{\int \varphi(F(y) - \bar{\varphi}_0)^2 dF(y)}} \right) \\ n &= [(z_{\alpha/2} + z_\beta) \sqrt{\int \varphi(F(y) - \bar{\varphi})^2 dF(y) / \theta (\int (\varphi F)'(y)dF(y))}]^2\end{aligned}\quad (2.4)$$

이다.

3. 표본수 계산 결과 및 결론

본 논문에서는 기존의 방법인 Wilcoxon 순위합검정을 이용한 표본수 계산방법과 t 검정법을 이용한 표본수 계산방법을, 제안한 선형위치통계량을 이용한 검정법에서의 표본수 계산방법과 비교한다.

먼저 유의수준과 검정력은 $\alpha = 0.05$, $1 - \beta = 0.8$ 로 고정하였다. 그리고 대조군과 치료군간의 치료효과 차이 θ 는 0.3, 0.5, 0.7, 1, 1.3, 1.5, 1.7, 2로 변화시켜 계산하였다. 모수적 방법에서는 독립표본에서의 t 검정을 이용하여 정확한 (exact) 표본수와 근사적인 표본수를 계산하였다. 근사적인 방법의 표본수는 $n = 2(z_{\alpha/2} + z_\beta)^2 \sigma^2 / \epsilon^2$, $\sigma_X^2 = \sigma_Y^2$ 을 이용하였다. 그리고 Wilcoxon 순위합검정에서의 표본수는 식 (2.1), 제안한 방법인 표본수는 식 (2.4)을 이용하여 계산하며, 그 결과를 표 3.1과 표 3.2로 정리하였다.

코쉬분포의 경우에는 평균과 분산값을 가지지 않는다. 그래서 코쉬분포의 경우에는 크기모수를 γ 로 표기하였다. 각 표에서 σ^2 가 커질수록 또한 θ 가 작을수록 표본수가 커지는 경향을 보인다. 그리고 코쉬

표 3.1 $\alpha = 0.05, 1 - \beta = 0.80$ 인 경우 표본 수 계산결과 ($\theta: 0.3, 0.5, 0.7, 1$)

분포	θ	$\sigma_X^2 = \sigma_Y^2$ (γ : 코쉬)	비모수적 방법		Wilcoxon	모수적 방법	
			위치 (placement)			T검정	
			균일	지수		정확값	근사값
정규	0.3	1	92	182	202	176	175
		2	184	364	392	350	349
		3	275	545	581	525	523
이중지수	0.3	1	59	91	258	176	175
		2	117	182	499	350	349
		3	175	273	738	525	523
코쉬	0.3	1	288	581	620	-	-
		2	1151	1151	2474	-	-
정규	0.5	1	33	66	78	64	63
		2	66	131	148	127	126
		3	99	197	218	190	189
이중지수	0.5	1	21	33	101	64	63
		2	42	66	190	127	126
		3	63	99	622	190	189
코쉬	0.5	1	104	210	235	-	-
		2	415	415	896	-	-
정규	0.7	1	17	34	42	34	32
		2	34	67	79	66	64
		3	51	101	115	98	96
이중지수	0.7	1	11	17	56	34	32
		2	22	34	102	66	64
		3	33	51	148	98	96
코쉬	0.7	1	53	107	127	-	-
		2	212	212	465	-	-
정규	1	1	9	17	23	17	16
		2	17	33	42	33	32
		3	25	50	60	49	48
이중지수	1	1	6	9	31	17	16
		2	11	17	55	33	32
		3	16	25	78	49	48
코쉬	1	1	26	53	69	-	-
		2	104	104	237	-	-

분포도 다른 분포가 분산이 커질수록 표본수가 커지는 것처럼 크기모수인 γ 가 커질수록 표본수가 커진다.

세 가지 방법을 비교하면, Wilcoxon 순위합검정을 이용한 경우 표본수가 제일 크다. 그 다음으로 위치 (placement) 방법에서 지수분포를 가정한 경우, 세 번째로는 t 검정을 이용한 경우이고, 위치 (placement)의 균일분포를 가정한 경우의 방법인 경우 표본수가 제일 작다. 모수적 방법의 경우 t 검정 근사값과 정확한 (Exact)값은 차이가 없다. t 검정을 이용하여 표본수를 구하는 경우에서 코쉬분포의 경우 평균과 분산을 알 수 없기 때문에 계산할 수가 없다.

모집단의 분포를 고려하여 표본수를 비교해보면 모집단의 분포가 코쉬분포, 정규분포 그리고 이중지수분포 순으로 표본수가 크다. 그러나 여기서 모수적 방법의 t 검정법을 이용한 경우에는 모집단의 분포에 상관없이 같은 표본수 공식을 사용하므로 항상 같은 표본수를 갖는다.

비모수적 방법에서 균일 점수함수를 사용한 위치 (placement)방법은 Wilcoxon 순위합검정과 같은 검정 방법임에도 불구하고 표본수는 차이가 많이 난다. 이는 검정력을 근사적으로 계산한 결과이고 추후 소표본에서의 정확도는 연구해야 할 문제라고 생각한다.

표 3.2 $\alpha = 0.05, 1 - \beta = 0.80$ 인 경우 표본 수 계산결과 ($\theta: 1.3, 1.5, 1.7, 2$)

분포	θ	$\sigma_X^2 = \sigma_Y^2$ (γ : 코쉬)	비모수적 방법		Wilcoxon	모수적 방법	
			위치 (placement)			T검정	
			균일	지수		정확값	근사값
정규	1.3	1	5	10	15	11	10
		2	10	20	26	20	19
		3	15	30	37	29	28
이중지수	1.3	1	4	5	20	11	10
		2	7	10	35	20	19
		3	10	15	49	29	28
코쉬	1.3	1	16	31	45	-	-
		2	62	62	146	-	-
정규	1.5	1	4	8	12	9	7
		2	8	15	21	15	14
		3	11	22	29	22	21
이중지수	1.5	1	3	4	16	9	7
		2	5	8	28	15	14
		3	7	11	39	22	21
코쉬	1.5	1	12	24	37	-	-
		2	47	47	113	-	-
정규	1.7	1	3	6	9	7	6
		2	6	12	17	12	11
		3	9	17	23	18	17
이중지수	1.7	1	2	3	14	7	6
		2	4	6	23	12	11
		3	6	9	32	18	17
코쉬	1.7	1	9	19	31	-	-
		2	36	36	91	-	-
정규	2	1	3	5	7	6	4
		2	5	9	13	9	8
		3	7	13	18	13	12
이중지수	2	1	2	3	11	6	4
		2	3	5	18	9	8
		3	4	7	24	13	12
코쉬	2	1	7	14	25	-	-
		2	26	26	69	-	-

적절한 표본수를 구하는 것은 매우 중요하기 때문에 모수적 방법과 비모수적 방법을 이용하여 같은 조건하에서 표본수를 구하여 비교하여 보았다. 유의수준과 검정력이 같은 경우, 기존의 방법과 본 논문에서 제시하는 방법을 비교하여 본 결과 본 논문에서 제안하는 위치 (placement) 방법에서 균일 점수함수를 사용하는 경우의 표본수가 가장 작다. 지금까지 Wilcoxon 순위합검정, 선형위치통계량, t검정 등을 이용하여 표본수를 비교하여 보았다. 추후에는 지금까지의 방법 이외의 기존의 다른 통계적 방법을 이용하여 표본수를 제시하고 비교하는 것이 좋겠다.

참고문헌

Kim, D. (1994). A comparison of distribution-free two-sample procedures based on placements or ranks. *Journal of the Korean Statistical Society*, **23**, 135-149.

Kim, D-W. and Kim, D. (2007). Sample size comparison for non-inferiority trials. *Journal of Korean Data & Information Science Society*, **18**, 411-418.

Orban, J. and Wolfe, D. A. (1982). A class of distribution-free two-sample tests based on placements. *Journals of the American Statistical Association*, **77**, 666-671.

- Simonoff, J. S., Hochberg, Y. and Reiser, B. (1986). Alternative estimation procedures for $\Pr(X < Y)$ in categorized data. *Biometrics*, **42**, 895-907.
- Wang, H., Chen, B. and Chew, S. C. (2003). *Sample size calculation in clinical research*, Marcel Dekker, INC., New York.
- Wilcoxon, F. (1945). Individual comparison by ranking methods. *Biometrics*, **1**, 80-83.

Sample size comparison for two independent populations

Haewon Ko¹ · Dongjae Kim²

^{1,2}Department of Biostatistics, The Catholic University of Korea

Received 1 October 2010, revised 18 November 2010, accepted 23 November 2010

Abstract

For clinical trials, it is common to compare the placebo and new drug. The method of calculating a sample size for two independent populations are the t-test that is used for parametric methods, and the Wilcoxon rank-sum test that is used in the non-parametric methods. In this paper, we propose a method that is using Kim's (1994) statistic power based on the linear placement statistic, which was proposed by Orban and Wolfe (1982). We also compare the sample size for the proposed method with that for using Wang *et al.* (2003)'s sample size formula which is based on Wilcoxon rank-sum test, and with that of t-test for parametric methods.

Keywords: Non-parametric method, placement, sample size.

¹ Researcher, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701, Korea.

² Corresponding author: Professor, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701, Korea. E-mail: djkim@catholic.ac.kr