응용통계연구 제 12권 2호, 1999년, pp. 415–432

불완전한 짝자료에 대한 검정법*

이승묵[1) 박진경[2) 박태성[3)

요 약

짝자료(paired data)에서 결측값이 발생했을 때에 이 자료를 처리할 수 있는 여러 통계 검정 방법들을 고찰해보았다. 결측값들을 어떻게 처리하는 지에 따라 다섯 가지 방법으로 분류해 보았고, 이 방법들이 짝 t-검정에 미치는 효과를 모의실험을 통해 비교해 보았다. 결측값들에 대한 세 종류의 메커니즘을 고려하여 검정크기와 검정력을 구하였다.

1. 서론

결측값(missing observation)을 포함하는 불완전한 자료(incomplete data)는 사람을 대상으로 하는 임상시험(clinical trials) 자료에서 빈번하게 발생한다. 최근 들어 임상자료에 대한 분석기법을 주로 다루는 의학통계학 및 보건통계학 분야에서는 불완전한 자료를 분석하기 위한 많은 연구가 진행중이다. 그러나 현재까지도 이러한 자료를 잘 처리할 수 있는 보편적인 통계기법이 잘 개발되어 있지 않은 상태이다.

Rubin(1976)은 자료가 결측될 확률(혹은 미관측될 확률)이 관측된 자료값과 결측값에 영향을 받는지의 여부에 따라 결측자료를 세 종류로 구분하였다. 첫 번째는 결측될 확률이 관측된 자료값과 결측값에 모두 영향을 받지 않는 경우로 missing completely at random (MCAR) 이라고 부르고, 두 번째는 결측될 확률이 관측된 자료값에는 영향을 받지만 결측값에는 영향을 받지 않는 경우로서 missing at random (MAR)이라고 부르고, 마지막으로는 결측될 확률이 결측값에 영향을 받는 경우에는 nonignorable missing (NIM) 이라고 부른다.

결측값을 포함하는 불완전한 자료에 대한 연구는 Rubin (1976)이 결측자료(missing data)에 대한 정확한 정의를 내림으로써 시작되었다. Little and Rubin (1987)은 불완전한 자료를 분석하기 위한 통계적 분석 방법을 총괄적으로 정리하였다.

임상시험에서는 drop-out에 의해 불완전한 자료가 종종 발생한다. 그 동안 이런 자료들을 처리하기 위하여 몇몇 방법들이 제안되었다. Heyting등(1992)이 제안한 방법은 drop-out 경우를 가장 나쁜 결과로 간주하여 관측된 자료 중에서 가장 큰 순위 다음 순위를 불완전한 자료의 순위로 대치하는 방법이다.

---

* 본 연구는 1997년도 학술진흥재단 연구비 지원에 의한 결과임 (과제번호:1997-001-D00084)

1) (449-791) 경기도 용인시 모현면, 한국외국어대학교 통계학과, 대학원
2) (449-791) 경기도 용인시 모현면, 한국외국어대학교 통계학과, 대학원
3) (151-742) 서울시 관악구 신림동 산 56-1, 서울대학교 자연과학대학 통계학과, 부교수

Brown (1992)은 두 집단(비교집단과 처리집단)을 비교하는 임상실험에서 drop-out이 있는 경우에 처리효과를 검정하는 방법에 대해서 제안하였다. Brown은 중앙값을 사용하여 drop-out된 자료, 즉 불완전하게 관측된 자료를 대치하는 방법을 사용하였다.

표 1.1: 와우관이식(cochlea implant)수술 결과

원 자료			정렬된 자료		
환자	9개월	18개월	환자	9개월	18개월
1	0.00	0.90	1	0.00	0.90
2	0.00	0.00	2	0.00	0.00
3	0.00	.	3	20.79	27.42
4	24.42	.	4	12.67	28.80
5	20.79	27.42	5	45.50	43.32
6	12.67	28.80	6	96.08	97.47
7	28.34	.	7	41.01	51.15
8	45.50	43.32	8	81.30	71.20
9	96.08	97.47	9	8.76	16.59
10	41.01	51.15	10	0.00	0.00
11	81.30	71.20	11	14.98	9.68
12	8.76	16.59	12	44.47	62.90
13	0.00	0.00	13	41.40	64.00
14	14.98	9.68	14	43.55	48.16
15	44.47	62.90	15	25.00	30.88
16	29.72	.	16	0.00	.
17	41.40	64.00	17	24.42	.
18	43.55	48.16	18	28.34	.
19	0.00	.	19	29.72	.
20	60.00	.	20	0.00	.
21	25.00	30.88	21	60.00	.

표 1.1에 있는 자료는 임상시험에서 얻은 불완전한 자료의 실제 사례이다. 이 자료는 양쪽 귀가 들리지 않는 환자들을 대상으로 와우관이식(cochlea implant) 수술을 실시한 후에 결과를 정리한 것이다. 각 환자들이 수술을 실시한 후 9개월후에 치른 청력시험 결과와 18개월후에 치른 청력시험 결과가 나타나 있다. 이 자료에서 '.'를 결측값을 나타낸다. 이 자료는 같은 환자로부터 두 번 반복관측하여 자료를 수집하였으므로 짝자료(paired data)로 볼 수 있다. 짝자료의 특징은 같은 짝 내의 자료들간에는 서로 상관관계가 있으나 짝이 다른 자료들간에는 서로 독립적이라는 점이다.

Heyting등(1992)의 연구결과와 Brown(1992)의 연구결과 및 Little and Rubin(1987)이 소개한 대부분의 방법들은 모두 자료가 독립적으로 얻어진 경우에 대해서만 적용될 수 있는 방법이다. 본 논문의 목적은 짝자료에서 결측값이 발생했을 때에 이를 처리할 수 있는 통계적 방법들을 고찰하고 그 특징들을 살펴보는 것이다. 또한 모의실험을 통해서 검정력을 비교해 보고자 한다. 본 논문에서는 Little and Rubin(1987)이 정리한 다음의 방법들을 구체적으로 살펴보고자 한다.

(1) 완전한 자료만을 이용하는 방법(complete data method, C)

(2) 평균대치법(mean imputation, MI)

(3) 회귀모형 대치법(regression imputation, RI)

(4) 확률적 회귀모형 대치법(stochastic regression imputation, SI)

(5) 최대우도법(maximum likelihood method, ML)

2절에서는 이 방법들에 대하여 자세히 설명하였고, 각 방법들에 대해서 표 1.1의 자료를 분석하였으며, 3절에서는 모의실험을 통하여 이 방법들에 대한 검정크기(size)와 검정력(power)을 비교해 보았다. 특히 불완전한 자료가 생성될 수 있는 세 종류의 메커니즘을 고려하여 모의실험을 실시하였다. 즉, i번째 개체로부터 얻은 관측값을 (y_{i1}, y_{i2})라고 할 때, y_{i2}가 결측될 확률이

(i) 모든 관측값에 대해 동일한 경우(MCAR)

(ii) y_{i1}의 함수에 비례하는 경우 (MAR)

(iii) y_{i2}의 함수에 비례하는 경우 (NIM)

로 구분하였다. 3절의 모의실험 결과들은 통계분석에서 결측값이 생성되는 메커니즘을 제대로 고려하지 않은 상태에서 검정을 하게 되면 잘못된 결론을 내릴 수 있다는 사실을 보여 주고 있다. 4절에서는 결론을 정리하였다.

2. 검정법

i번째 개체로부터 얻은 관측값을 (y_{i1}, y_{i2})라고 표시하자. 전체 자료의 수를 n, 이 중에서 완전한 자료의 수를 m이라고 하고 편의상 처음 m개의 자료가 완전한 자료라고 가정하자. 그러면 자료의 형태는 다음과 같다.

$$\text{완전한자료} : (y_{i1}, y_{i2}), i = 1, \cdots, m,$$

$$\text{불완전한자료} : (y_{i1}, \cdot), i = m+1, \cdots, n.$$

처음 관찰값의 평균을 μ_1으로, 두 번째 관찰값의 평균을 μ_2라고 표시하자. 그러면 두 처리를 비교하기 위한 가설은

$$H_0: \mu_1 = \mu_2, H_A: \mu_1 \neq \mu_2$$

가 된다. 만약 자료가 모두 관찰되었다면 이 가설을 검정하기 위한 짝 t-검정통계량(paired t-test statistics)은

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

이고, 여기서 $d_i = y_{i1} - y_{i2}$, $\bar{d} = \sum_i d_i/n$, $s_d^2 = \sum_i (d_i - \bar{d})^2/(n-1)$가 된다. 이제 앞에서 소개한 다섯 가지 불완전 짝자료에 대한 분석방법에 대해 살펴보도록 하자.

2.1. 완전한 자료만을 이용하는 방법(Complete data method, C)

이 방법은 결측자료를 제거한 후에 완전한 자료 m개만을 사용하여 검정하는 방법으로 이 때의 검정통계량은

$$t_C = \frac{\bar{d}_m}{s_{dm}/\sqrt{m}}$$

이고, 여기서 $\bar{d}_m = \sum_{i=1}^m d_i/m, s_{d_m}^2 = \sum_{i=1}^m (d_i - \bar{d}_m)^2/(m-1)$이 된다.

표 1.1의 완전하게 관찰된 15개의 자료로부터 d_i값을 구한 후에 $\bar{d}_m$을 계산하여 검정통계량을 구하면 $t_c = 2.2244$가 되고, 이 때의 $p-$값은 0.0431이 된다. 따라서 유의수준 0.05에서 두 처리간에 차이가 있다고 할 수 있다.

2.2. 평균대치법(Mean Imputation, MI)

평균대치법은 결측값 $y_{i2}(i = m+1, \cdots, n)$를 관측값 $y_{i2}(i = 1, \cdots, m)$의 평균값인 $\bar{y}_{2M} = \sum_{i=1}^m y_{i2}/m$ 으로 대치하는 방법이다. 즉, 이 대치값을 실제 관측값처럼 간주한 후에 검정하는 방법이다. 그러면 t-통계량의 분자 $\bar{d}_M$은

$$\bar{d}_M = [\sum_{i=1}^m (y_{i1} - y_{i2}) + \sum_{i=m+1}^n (y_{i1} - \bar{y}_{2M})]/n = \bar{y}_1 - \bar{y}_{2M}$$

가 된다. 따라서 t-통계량은

$$t_M = \frac{\bar{d}_M}{s_{dM}/\sqrt{n}}$$

이고, 여기서 $s_{d_M}^2 = \sum_{i=1}^n (d_{iM} - \bar{d}_M)^2/(n-1)$이며,

$$d_{iM} = \begin{cases} y_{i1} - y_{i2} & , i = 1, \cdots, m \\ y_{i1} - \bar{y}_{2M} & , i = m+1, \cdots, n \end{cases}$$

이 된다. 표 1.1에서 관측된 15개의 y_{i2}의 값으로부터 평균값을 계산하면 $\bar{y}_{2M} = \sum_{i=1}^{m} y_{i2}/m = 36.8313$가 되며, 이 값을 관측이 안된 나머지 6개의 y_{i2}에 대치하면 $\bar{d}_M = 7.4032$가 되고, 통계량은 $t_M = 2.4333$이 된다. 이때의 $p-$값은 0.0245로 유의수준 0.05에서 두 처리간에 차이가 있다고 할 수 있다.

2.3. 회귀모형 대치법(Regression Imputation, RI)

이 방법은 회귀모형을 이용하여 결측값 $y_{i2}(i = m+1, \cdots, n)$를 예측값으로 대치하는 방법이다. 먼저 m개의 완전한 자료로부터 y_{i2}를 종속변수로, y_{i1}을 독립변수로 간주하여 회귀모형을 적합한 다음에 $n-m$개의 관측된 $y_{i1}(i = m+1, \cdots, n)$으로부터 결측값 $y_{i2}(i = m+1, \cdots, n)$를 예측한다. 이 예측값을 실제 관측값처럼 간주하여 검정하면 된다. 먼저 m개의 완전한 자료로부터 구한 회귀모형의 모수추정값을 (b_0, b_1)이라고 하면

$$\hat{y}_{i2} = b_0 + b_1 y_{i1} \ (i = m+1, \cdots, n)$$

가 된다. 이 예측값을 이용하여 구한 검정통계량을 t_R로 나타낼 수 있다. 이런 방법으로 앞의 표 1.1의 15개의 완전한 자료로부터 y_{i2}를 종속변수로, y_{i1}을 독립변수로 간주하여 회귀모형을 적합하게 되면, 회귀식

$$\hat{y}_{i2} = 6.1565 + 0.9676 y_{i1}$$

을 얻게 된다. 이 결과를 이용하여 관측이 되지 않은 6개의 자료에 대한 예측값을 구한 후, 검정통계량을 구하면 $t_R = 3.1867$가 되고, $p-$값은 0.0046이 된다. 따라서 유의수준 0.05에서 두 처리간에 유의한 차이가 있다고 할 수 있다.

2.4. 확률적 회귀모형 대치법(Stochastic regression Imputation, SI)

이 방법은 회귀모형 대치법을 이용하여 구한 y_{i2}의 예측값에다 추가로 확률변수값을 더하여 임의성(randomness)을 높이는 방법이다. 즉, y_{i2}의 예측값이

$$\hat{y}_{i2} = b_0 + b_1 y_{i1} + e_i \ , \ e_i \sim N(0, MSE) \ (i = m+1, \cdots, n)$$

이 된다. 여기서 MSE는 회귀모형의 MSE를 의미한다. 이 예측값을 이용하여 구한 검정통계량을 t_S라 하자. 이와 같이 표 1.1에 대한 y_{i2}의 예측값은

$$\hat{y}_{i2} = 6.1565 + 0.9676 y_{i1} + e_i, \ e_i \sim N(0, MSE)$$

이 되며, MSE는 85.0059이다. 실제로 정규분포를 따르는 확률변수를 생성시킨 후에 검정통계량을 구하면 $t_S = 2.9012$이 되고, $p-$값은 0.0088이 된다. 따라서, 유의수준 0.05에서 두 처리간에 유의한 차이가 있다고 할 수 있다.

2.5. 최대우도법(Maximum Likelihood method, ML)

이 방법은 관측된 자료가 이변량정규분포(bivariate normal distribution)를 따른다고 간주하여 우도함수를 유도한 후에 이 우도함수를 최대로 만드는 최대우도추정법을 이용하여 검정하는 방법이다. 여기에서는 자유도를 수정한 최대우도추정법(restricted maximum likelihood method, REML)을 사용하였다. 즉, i번째 관측값 (y_{i1}, y_{i2})이

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma \right]$$

로서, 평균벡터(μ_1, μ_2)와 공분산행렬 Σ를 갖는 이변량정규분포로부터 관측된 값으로 간주한 후에 (μ_1, μ_2)의 MLE를 구하여 가설을 검정하는 방법이다.

Anderson(1984)은 y_{i1}과 y_{i2}의 결합분포를 y_{i1}의 주변분포와 $y_{i2}|y_{i1}$의 조건부 분포로 분해하여 모수들에 대한 MLE를 구하였다. y_{i2}의 평균에 대한 ML추정값 $\hat{\mu}_2$은

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{m} y_{i2} + \sum_{i=m+1}^{n} \hat{y}_{i2}$$

이고, 이 때,

$$\hat{y}_{i2} = \bar{y}_2 + b_1(y_{i1} - \bar{y}_1)$$

$$\bar{y}_2 = m^{-1}\sum_{i=1}^{m} y_{i2}\ ,\ \bar{y}_1 = n^{-}1\sum_{i=1}^{n} y_{i1}$$

$$b_1 = s_{12}/s_{11}$$

$$s_{jk} = (m-1)^{-}1\sum_{i=1}^{m}(y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k)$$

이다. MLE를 이용하여 구한 검정통계량을 t_L로 나타내자. 이 때 t_L은 다음과 같다.

$$t_L = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\hat{Var}(\hat{\mu}_1 - \hat{\mu}_2)}}$$

이와 같은 방법으로 앞의 표 1.1에 대해서 나타내면 평균벡터 $(\mu_1, \mu_2)'$의 MLE는 $(29.4281, 34.6323)$이고, 공분산행렬 Σ의 MLE는

$$\begin{pmatrix} 719.8219 & 696.5333 \\ 696.5333 & 752.9303 \end{pmatrix}$$

이다. 이를 이용하여 검정통계량 t_L을 구하면 $t_L = 2.26$이고 $p-$값은 0.0351로 서 두 처리 간에 유의한 차가 있다고 할 수 있다.

표 2.1는 표 1.1의 와우관이식 수술 결과 자료에 대해서 위에 설명된 다섯 가지 방법들의 통계량 값과 $p-$값을 정리한 것이다. 완전한 자료만을 이용하는 방법(C)을 포함하여 나머지 4가지 방법 (MI, RI, SI, ML)에 대해서 두 처리간의 차이가 있는 것으로 나타났다. 표 2.1에서 나타나듯이 불완전한 자료값에 회귀모형 대치법과 확률적 회귀모형 대치법들의 $p-$값이 완전한 자료만을 이용하는 방법이나 최대우도법의 $p-$값보다 더 작아짐을 알 수 있는데, 이는 불완전한 자료값에 대치된 값을 실제값으로 간주하여 처리하기 때문에 분산이 과소추정되었기 때문이다. 이러한 문제에 대해 보다 체계적인 비교연구를 하기 위해 3절에서는 여러 조건에서 발생하는 결측치들을 생성하여 모의실험을 실시하였다.

표 2.1: 불완전자료 검정법에 대한 결과

검정방법	통계량	자유도	$p-$값
완전한 자료만을 이용하는 방법 (C)	2.2244	14	0.0431
평균대치법 (MI)	2.4333	20*	0.0245
회귀모형 대치법 (RI)	3.1867	20*	0.0046
확률적 회귀모형 대치법 (SI)	2.9012	20*	0.0088
최대우도법 (ML)	2.2600	20*	0.0351

* - 추정된 자료들을 실제 자료인 것처럼 간주하여 계산한 것

3. 모의실험

3절에서는 이변량정규분포를 따르는 모의자료에서 i번째 관측값 (y_{i1}, y_{i2})을 n (=15, 20, 30, 50, 100)개 생성하여, 세 가지 형태의 메커니즘에 의한 결측자료를 만들어 모의실험을 5000번 실시하였다. 각 짝자료에 대한 상관계수 ρ값은 0.3, 0.5, 0.7로 정하였고, 전체자료 중에서 결측자료의 비율을 p로 나타내고, p의 값은 0.0, 0.1, 0.2, 0.3으로 가정하여 유의수준 0.05에서 가설 $H_0 : \mu_1 = \mu_2$에 대한 검정크기(size)와 검정력(power)을 비교하였다. 여기서, 검정크기는 귀무가설하에서 귀무가설이 기각되는 경험적 확률로 정의된다. 유의수준을 0.05로 정했을 때에 검정크기가 보존되는 경우는 5000개의 모의실험 중에서 가설을 기각하는 비율에 대한 95%신뢰구간을 구했을 때에 0.05가 포함되는 경우로 간주하였다. 즉, 검정크기가 0.0563보다 작을 경우는 검정크기가 보존되는 것으로 간주하였다.

귀무가설하에서는 각 그룹의 평균을 $\mu_1 = \mu_2 = 0$으로 설정하였고 대립가설하에서는 $\mu_1 = \mu_2 + \delta(\delta > 0)$로 설정하여, δ를 두 평균간의 차이로 표시하였다. 두 그룹의 분산은 각각 1로 설정하였고, 자료 생성에 대한 모수설정은 다음과 같다.

$$\mu_1 = \mu_2 + \delta(\delta = 0, 0.3, 0.5, 1.0)$$

$$Var(y_{i1}) = 1, Var(y_{i2}) = 1$$

$$Cov(y_{i1}, y_{i2}) = Corr(y_{i1}, y_{i2}) = \rho$$

결측값을 생성하는 메커니즘은 앞에서 살펴본 것과 같이 세 가지로 구분할 수 있는데, 첫 번째는 임의의 y_{i2}가 결측될 확률이 모든 관측값에 대해 동일한 경우(MCAR)이고, 두 번째는 결측될 확률이 y_{i1}의 함수에 비례하는 경우 (MAR), 세 번째는 y_{i2}에 비례하는 경우 (NIM)이다. 즉, p_i가 y_{i2}의 결측될 확률일 때, p_i가 모든 i 에 대해 $p_i = p$로 동일하면 MCAR이고, $p_i \propto f_1(y_{i1})$이면 MAR이고, $p_i \propto f_2(y_{i2})$이면 NIM이 된다. 여기서 f_1과 f_2는 특정의 함수를 나타낸다.

모의실험에서는 Brown(1992)의 방법에 따라 편의상 MAR의 경우에는 p_i가 y_{i1}의 순위 (rank)에 비례하도록 가정했으며, NIM의 경우에는 p_i가 y_{i2}의 순위 (rank)에 비례하도록 가정했다. 표 3.1부터 표 3.6까지는 모의실험 결과를 나타낸다. 모의실험 결과 n이 20인 경우는 n이 15인 경우와 유사하게 나오고 n이 50과 100인 경우는 n이 30인 경우와 유사하게 나오기 때문에 본 논문에서는 n이 15인 경우와 n이 30인 경우만을 정리하였다. 각 표에서 $\delta = 0$인 경우는 귀무가설이 성립되는 경우로, 검정크기를 나타내며, $\delta > 0$인 경우는 검정력을 의미한다. 그리고, 표 3.1부터 표 3.6까지의 결과를 표 3.7에 요약하여 정리하였다.

표 3.1과 표 3.2는 임의의 y_{i2}가 결측이 될 확률이 모든 관측값에 대해 동일한 경우인 MCAR에 대한 결과이다. 이 경우에는 ρ에 상관없이 $p = 0.1$인 경우에는 완전한 자료만을 이용하는 방법(C), 평균대치법(MI), 최대우도법(ML)의 검정크기가 보존되었고, 그 중 대부분의 경우에서 ML의 검정력이 가장 큰 경향을 보였고, 그 다음으로는 C, MI의 순이였다. $p = 0.2$인 경우에서는 C, MI와 ML의 검정크기가 보존되었고, 이 경우 검정력은 ML이 가장 높게 나타났다. $p = 0.3$인 경우에서는 $n = 15$일때는 대략적으로 C와 MI의 검정크기가 보존되었고, $n = 30$인 경우에는 오히려 C와 ML, 혹은 MI의 검정크기도 보존되는 경향을 보였다. 이 경우들의 검정력을 비교하면, $n = 15$인 경우에서는 C와 MI 중에 어느 한 방법이 좋다고 말할 수 없었고, $n = 30$인 경우에서는 ML의 검정력이 C나 MI보다는 큰 경향을 보였다.

표 3.3과 표 3.4은 임의의 y_{i2}가 결측될 확률이 y_{i1}의 순위에 비례하는 경우 (MAR)로, ρ가 0.3과 0.5이고 $p = 0.1$인 경우일 때, C, MI과 ML의 검정크기가 보존되었다. 이 경우에서는 ML의 검정력이 가장 크게 나타났다. ρ가 0.3일때, $p = 0.2$인 경우에 검정크기는 $n = 15$인 경우에는 C, MI, ML, $n = 30$인 경우에는 MI, ML이 보존되었고, 검정력은 MI가 가장 크게 나타나는 경향을 보였다. 그리고 $p = 0.3$인 경우에는 검정크기는 ML만이 보존되었다. 또한, ρ가 0.5이고 $p = 0.2$와 ρ 가 0.7이고 $p = 0.1$, $p = 0.2$인 경우에는 검정크기가 보존되는 방법이 MI와 ML이었고, 검정력은 ML이 크게 나타났다. ρ가 0.5와 0.7이고 $p = 0.3$인 경우에는 MI과 ML의 검정크기가 보존되었고, 이 경우에서의 검정력을 비교하면 ML보다는 MI가 큰 경향을 보였다.

표 3.5과 표 3.6은 임의의 y_{i2}가 결측이 될 확률이 y_{i2}의 순위에 비례하는 경우 (NIM)로, 이 경우에서는 ρ가 0.3인 경우에는 $p = 0.1$인 경우를 제외한 대부분의 경우에서 검정크기가 보존되지 않는다. $n = 15$이고 ρ가 0.3, 0.5인 경우에는 대부분 ML의 검정력이 크게 나타났지만, ρ 가 0.7인 경우에는 n과 p에 따라 검정크기과 검정력이 다르게 나타났다.

각 메커니즘에 대해서 검정크기가 보존되는 경우와 검정력이 높은 검정법을 요약하면

< 표9 >와 같다. 표 3.7에서 보듯이 결측치를 분석하는 방법들은 결측생성에 대한 각 메커니즘에 따라서 검정크기가 보존되는 경우와 검정력이 높은 검정법가 다르게 나타남을 알 수 있다.

4. 결 론

본 논문에서는 결측값을 생성하는 세 가지 메커니즘인 MCAR, MAR, NIM에 의해 결측값이 발생했을 때에 이러한 자료를 처리하는 분석방법들인 완전한 자료만을 이용하는 방법(C)과 평균대치법(MI), 회귀모형 대치법(RI), 그리고 확률적 회귀모형대치법(SI)과 최대우도법(ML)을 비교해 보았다.

소표본 짝자료에 대해 두 그룹의 평균을 비교하는 짝 t-검정의 검정크기와 검정력은 결측생성에 대한 메커니즘에 따라 추천할 수 있는 방법이 다르게 나타났다. MCAR의 경우, ρ에 상관없이 p가 작은 경우에는 C, MI과 ML의 검정크기가 보존되었고, 이 경우에서는 ML의 검정력이 가장 크게 나타났고, p가 큰 경우에도 n이 작을 때와 다르게 n이 커질수록 ML의 검정크기가 보존되며 검정력도 크게 나타났다. MAR인 경우에 검정크기가 보존되는 검정법 중에서 검정력이 크게 나타나는 검정법은 각각의 경우에 따라서 대체적으로 ML, MI이었다. NIM의 경우에는 ρ가 작은 경우에는 대부분의 검정크기가 보존되지 않았고, ρ가 큰 경우에는 n과 p에 따라 검정크기와 검정력이 다르게 나타났다. 일반적으로 MCAR은 n이 커짐에 따라 ML을 추천할 수 있고, MAR은 MI과 ML을 추천할 수 있다. 그러나, NIM의 경우에는 n이 커져도 특정한 방법을 추천하기에 어려움이 있다.

그러나, 실제 결측값이 어떤 메커니즘을 따르는지 알 수 없으므로 C, MI, ML를 비교하여 사용하는 것을 추천할 수 있다. 이렇게 각 메커니즘에 따라서 추천할 수 있는 방법이 다르게 나타나는 이유는 결측치를 분석하는 다섯 가지의 방법들이 결측값 생성에 대한 메커니즘(missing data mechanism)을 고려하지 않았으며, y_{i2}의 결측확률 p와 두 표본들간의 상관계수 ρ에 대해서 민감하게 나타났기 때문이다. 이와 같은 문제를 해결하기 위해서는 결측메커니즘을 고려하는 방법의 개발이 요구된다. 또한, 결측값에 하나의 값 이상을 대치하는 다중대치법(multiple imputation)(Rubin, 1987)의 비교도 앞으로의 연구과제로 남겨둔다.

표 3.1: 모의실험결과 (MCAR인 경우, $n = 15$)

ρ	δ	p	C	MI	RI	SI	ML
0.3	0.0	0.1	0.0466	0.0412	0.0638	0.0594	0.0460
		0.2	0.0514	0.0544	0.0930	0.0850	0.0544
		0.3	0.0486	0.0742	0.1338	0.1194	0.0612
	0.3	0.1	0.1322	0.1178	0.1644	0.1620	0.1342
		0.2	0.1260	0.1404	0.2056	0.1850	0.1402
		0.3	0.1134	0.1520	0.2356	0.2050	0.1300
	0.5	0.1	0.2986	0.2942	0.3618	0.3480	0.3202
		0.2	0.2662	0.2880	0.3862	0.3566	0.2958
		0.3	0.2234	0.2936	0.4078	0.3612	0.2708
	1.0	0.1	0.8098	0.8106	0.8618	0.8434	0.8320
		0.2	0.7386	0.7816	0.8582	0.8230	0.7934
		0.3	0.6688	0.7918	0.8604	0.8110	0.7606
0.5	0.0	0.1	0.0442	0.0330	0.0606	0.0574	0.0450
		0.2	0.0492	0.0368	0.0882	0.0832	0.0526
		0.3	0.0546	0.0460	0.1256	0.1104	0.0598
	0.3	0.1	0.1600	0.1290	0.2008	0.1944	0.1706
		0.2	0.1508	0.1182	0.2280	0.2140	0.1656
		0.3	0.1318	0.1354	0.2636	0.2390	0.1552
	0.5	0.1	0.3674	0.3054	0.4306	0.4120	0.3900
		0.2	0.3162	0.2780	0.4528	0.4152	0.3526
		0.3	0.2720	0.2798	0.4650	0.4242	0.3262
	1.0	0.1	0.8800	0.8506	0.9212	0.9102	0.9020
		0.2	0.8336	0.8228	0.9246	0.8974	0.8834
		0.3	0.7490	0.8086	0.9172	0.8800	0.8394
0.7	0.0	0.1	0.0420	0.0234	0.0870	0.0774	0.0448
		0.2	0.0528	0.0238	0.0782	0.0728	0.0496
		0.3	0.0540	0.0308	0.1002	0.0914	0.0566
	0.3	0.1	0.1726	0.1186	0.2076	0.2044	0.1854
		0.2	0.1576	0.1030	0.2284	0.2136	0.1806
		0.3	0.1442	0.0998	0.2440	0.2258	0.1686
	0.5	0.1	0.3824	0.2934	0.4394	0.4328	0.4082
		0.2	0.3566	0.2648	0.4642	0.4418	0.3988
		0.3	0.2964	0.2498	0.4726	0.4398	0.3658
	1.0	0.1	0.9112	0.8586	0.9486	0.9414	0.9382
		0.2	0.8746	0.8414	0.9470	0.9364	0.9272
		0.3	0.8110	0.8146	0.9416	0.9270	0.9026

표 3.2: 모의실험결과 (MCAR인 경우, $n = 30$)

ρ	δ	p	C	MI	RI	SI	ML
0.3	0.0	0.1	0.0520	0.0424	0.0686	0.0664	0.0520
		0.2	0.0474	0.0422	0.0844	0.0828	0.0480
		0.3	0.0490	0.0610	0.1192	0.1034	0.0490
	0.3	0.1	0.2404	0.2190	0.2972	0.2844	0.2590
		0.2	0.2204	0.2282	0.3262	0.3018	0.2466
		0.3	0.1934	0.2370	0.3438	0.2986	0.2162
	0.5	0.1	0.5600	0.5234	0.6256	0.6034	0.5806
		0.2	0.4930	0.5118	0.6278	0.5864	0.5348
		0.3	0.4240	0.5086	0.6360	0.5706	0.4868
	1.0	0.1	0.9848	0.9828	0.9932	0.9900	0.9884
		0.2	0.9734	0.9818	0.9912	0.9828	0.9810
		0.3	0.9466	0.9746	0.9888	0.9756	0.9696
0.5	0.0	0.1	0.0504	0.0262	0.0670	0.0642	0.0510
		0.2	0.0500	0.0266	0.0818	0.0806	0.0480
		0.3	0.0486	0.0392	0.1158	0.0966	0.0524
	0.3	0.1	0.2974	0.2084	0.3530	0.3358	0.3128
		0.2	0.2586	0.1924	0.3688	0.3364	0.2858
		0.3	0.2240	0.2104	0.3938	0.3502	0.2686
	0.5	0.1	0.6502	0.5416	0.7164	0.6910	0.6766
		0.2	0.5940	0.5222	0.7218	0.6788	0.6398
		0.3	0.5368	0.5302	0.7326	0.6836	0.6128
	1.0	0.1	0.9970	0.9926	0.9990	0.9986	0.9984
		0.2	0.9878	0.9874	0.9968	0.9952	0.9940
		0.3	0.9792	0.9880	0.9986	0.9940	0.9926
0.7	0.0	0.1	0.0484	0.0174	0.0620	0.0590	0.0504
		0.2	0.0472	0.0152	0.0702	0.0608	0.0462
		0.3	0.0490	0.0196	0.0836	0.0790	0.0456
	0.3	0.1	0.3216	0.1966	0.3768	0.3622	0.3498
		0.2	0.2884	0.1756	0.3946	0.3656	0.3300
		0.3	0.2640	0.1816	0.3982	0.3676	0.3082
	0.5	0.1	0.7084	0.5492	0.7616	0.7506	0.7410
		0.2	0.6516	0.5074	0.7656	0.7372	0.7174
		0.3	0.5756	0.4996	0.7624	0.7218	0.6832
	1.0	0.1	0.9990	0.9934	0.9994	0.9992	0.9992
		0.2	0.9962	0.9944	0.9994	0.9994	0.9992
		0.3	0.9902	0.9902	0.9992	0.9978	0.9974

표 3.3: 모의실험결과 (MAR인 경우, $n = 15$)

ρ	δ	p	C	MI	RI	SI	ML
0.3	0.0	0.1	0.0516	0.0460	0.0718	0.0702	0.0552
		0.2	0.0550	0.0548	0.0922	0.0846	0.0498
		0.3	0.0798	0.0642	0.1408	0.1226	0.0556
	0.3	0.1	0.1112	0.1382	0.1772	0.1652	0.1402
		0.2	0.0880	0.1588	0.2024	0.1840	0.1338
		0.3	0.0616	0.1830	0.2314	0.1968	0.1142
	0.5	0.1	0.2494	0.3046	0.3532	0.3338	0.3048
		0.2	0.1902	0.3230	0.3836	0.3506	0.2842
		0.3	0.1354	0.3620	0.4084	0.3622	0.2518
	1.0	0.1	0.7608	0.8106	0.8518	0.8356	0.8224
		0.2	0.6624	0.8114	0.8696	0.8382	0.7908
		0.3	0.5286	0.8368	0.8706	0.8240	0.7380
0.5	0.0	0.1	0.0500	0.0362	0.0670	0.0660	0.0498
		0.2	0.0658	0.0388	0.1002	0.0896	0.0558
		0.3	0.0700	0.0452	0.1224	0.1058	0.0450
	0.3	0.1	0.1332	0.1444	0.1974	0.1914	0.1666
		0.2	0.0948	0.1378	0.2138	0.1944	0.1436
		0.3	0.0728	0.1790	0.2470	0.2240	0.1334
	0.5	0.1	0.3000	0.3144	0.4056	0.3924	0.3672
		0.2	0.2306	0.3240	0.4308	0.3964	0.3310
		0.3	0.1638	0.3658	0.4690	0.4168	0.3016
	1.0	0.1	0.8598	0.8722	0.9316	0.9224	0.9132
		0.2	0.7472	0.8570	0.9210	0.8990	0.8702
		0.3	0.6216	0.8686	0.9262	0.8888	0.8220
0.7	0.0	0.1	0.0816	0.0290	0.0946	0.0876	0.0398
		0.2	0.0622	0.0256	0.0794	0.0728	0.0476
		0.3	0.0860	0.0278	0.1048	0.0960	0.0452
	0.3	0.1	0.1426	0.1360	0.2162	0.2090	0.1862
		0.2	0.0972	0.1214	0.2140	0.2038	0.1576
		0.3	0.0718	0.1582	0.2492	0.2270	0.1460
	0.5	0.1	0.3294	0.3226	0.4492	0.4314	0.4102
		0.2	0.2480	0.3244	0.4700	0.4478	0.3938
		0.3	0.1634	0.3554	0.4772	0.4400	0.3368
	1.0	0.1	0.8824	0.8766	0.9484	0.9432	0.9380
		0.2	0.7988	0.8778	0.9550	0.9434	0.9240
		0.3	0.6706	0.8844	0.9494	0.9326	0.8828

표 3.4: 모의실험결과 (MAR인 경우, $n = 30$)

ρ	δ	p	C	MI	RI	SI	ML
0.3	0.0	0.1	0.0530	0.0386	0.0652	0.0664	0.0512
		0.2	0.0654	0.0492	0.0846	0.0824	0.0508
		0.3	0.0982	0.0708	0.1308	0.1146	0.0504
	0.3	0.1	0.1856	0.2316	0.2878	0.2672	0.2490
		0.2	0.1192	0.2648	0.3082	0.2840	0.2242
		0.3	0.0794	0.3260	0.3592	0.3044	0.2068
	0.5	0.1	0.4572	0.5364	0.6028	0.5830	0.5556
		0.2	0.3380	0.5698	0.6290	0.5814	0.5242
		0.3	0.2342	0.6282	0.6544	0.5880	0.4990
	1.0	0.1	0.9786	0.9876	0.9940	0.9918	0.9906
		0.2	0.9326	0.9880	0.9928	0.9854	0.9816
		0.3	0.8494	0.9882	0.9928	0.9852	0.9736
0.5	0.0	0.1	0.0536	0.0268	0.0666	0.0616	0.0492
		0.2	0.0664	0.0316	0.0836	0.0826	0.0444
		0.3	0.1110	0.0452	0.1148	0.1082	0.0468
	0.3	0.1	0.2300	0.2344	0.3492	0.3342	0.3062
		0.2	0.1454	0.2706	0.3642	0.3312	0.2744
		0.3	0.0876	0.3440	0.3934	0.3450	0.2454
	0.5	0.1	0.5650	0.5740	0.7122	0.6902	0.6716
		0.2	0.4290	0.6094	0.7258	0.6848	0.6356
		0.3	0.2834	0.6710	0.7290	0.6744	0.5694
	1.0	0.1	0.9926	0.9942	0.9982	0.9972	0.9970
		0.2	0.9736	0.9944	0.9980	0.9958	0.9958
		0.3	0.9130	0.9954	0.9964	0.9922	0.9876
0.7	0.0	0.1	0.0582	0.0190	0.0596	0.0594	0.0490
		0.2	0.0810	0.0238	0.0704	0.0720	0.0458
		0.3	0.1312	0.0352	0.0946	0.0896	0.0440
	0.3	0.1	0.2430	0.2224	0.3682	0.3586	0.3368
		0.2	0.1478	0.2476	0.3854	0.3704	0.3232
		0.3	0.0908	0.3200	0.4054	0.3722	0.2896
	0.5	0.1	0.6174	0.5916	0.7626	0.7508	0.7374
		0.2	0.4490	0.6120	0.7598	0.7388	0.6994
		0.3	0.2874	0.6892	0.7732	0.7336	0.6602
	1.0	0.1	0.9968	0.9964	0.9992	0.9990	0.9992
		0.2	0.9856	0.9980	0.9994	0.9994	0.9994
		0.3	0.9322	0.9976	0.9992	0.9980	0.9970

표 3.5: 모의실험결과 (NIM인 경우, $n = 15$)

ρ	δ	p	C	MI	RI	SI	ML
0.3	0.0	0.1	0.0540	0.0480	0.0762	0.0712	0.0554
		0.2	0.0530	0.0726	0.1150	0.1006	0.0626
		0.3	0.0618	0.1174	0.1726	0.1586	0.0898
	0.3	0.1	0.1628	0.1668	0.2152	0.2068	0.1788
		0.2	0.1742	0.2234	0.3002	0.2676	0.2180
		0.3	0.1940	0.3330	0.4120	0.3606	0.2740
	0.5	0.1	0.3362	0.3486	0.4156	0.3954	0.3666
		0.2	0.3434	0.4184	0.5118	0.4802	0.4130
		0.3	0.3402	0.5184	0.6060	0.5506	0.4610
	1.0	0.1	0.8410	0.8574	0.8908	0.8774	0.8712
		0.2	0.8114	0.8748	0.9226	0.8944	0.8800
		0.3	0.7874	0.9142	0.9460	0.9172	0.8916
0.5	0.0	0.1	0.0478	0.0378	0.0686	0.0662	0.0528
		0.2	0.0496	0.0440	0.0958	0.0898	0.0556
		0.3	0.0558	0.0812	0.1550	0.1386	0.0792
	0.3	0.1	0.1718	0.1496	0.2304	0.2176	0.1956
		0.2	0.1768	0.2020	0.3160	0.2910	0.2344
		0.3	0.1718	0.2760	0.4118	0.3668	0.2768
	0.5	0.1	0.3896	0.3682	0.4836	0.4668	0.4338
		0.2	0.3632	0.3894	0.5474	0.5054	0.4554
		0.3	0.3518	0.4986	0.6488	0.5910	0.5020
	1.0	0.1	0.8954	0.8816	0.9384	0.9298	0.9268
		0.2	0.8762	0.8962	0.9570	0.9390	0.9278
		0.3	0.8368	0.9336	0.9728	0.9566	0.9410
0.7	0.0	0.1	0.0530	0.0464	0.1118	0.1018	0.0596
		0.2	0.0490	0.0332	0.0818	0.0780	0.0536
		0.3	0.0454	0.0500	0.1116	0.1030	0.0576
	0.3	0.1	0.1750	0.1464	0.2384	0.2314	0.2132
		0.2	0.1594	0.1678	0.3012	0.2836	0.2416
		0.3	0.1390	0.2178	0.3628	0.3286	0.2590
	0.5	0.1	0.3940	0.3514	0.4932	0.4788	0.4574
		0.2	0.3538	0.3670	0.5480	0.5252	0.4786
		0.3	0.3106	0.4380	0.6250	0.5894	0.5106
	1.0	0.1	0.9158	0.9014	0.9578	0.9530	0.9512
		0.2	0.8728	0.9020	0.9726	0.9642	0.9554
		0.3	0.8126	0.9170	0.9776	0.9652	0.9518

표 3.6: 모의실험결과 (NIM인 경우, $n = 30$)

ρ	δ	p	C	MI	RI	SI	ML
0.3	0.0	0.1	0.0480	0.0446	0.0690	0.0656	0.0534
		0.2	0.0630	0.0822	0.1288	0.1186	0.0782
		0.3	0.0900	0.1646	0.2338	0.2030	0.1364
	0.3	0.1	0.2914	0.2946	0.3786	0.3578	0.3258
		0.2	0.3382	0.4158	0.5188	0.4754	0.4220
		0.3	0.3874	0.5846	0.6716	0.6120	0.5322
	0.5	0.1	0.6130	0.6116	0.7060	0.6822	0.6640
		0.2	0.6364	0.7126	0.7962	0.7500	0.7158
		0.3	0.6638	0.8264	0.8820	0.8348	0.7982
	1.0	0.1	0.9910	0.9922	0.9956	0.9938	0.9944
		0.2	0.9898	0.9956	0.9982	0.9970	0.9962
		0.3	0.9904	0.9994	0.9992	0.9980	0.9986
0.5	0.0	0.1	0.0516	0.0282	0.0680	0.0718	0.0546
		0.2	0.0622	0.0536	0.1180	0.1086	0.0772
		0.3	0.0676	0.1162	0.2056	0.1786	0.1142
	0.3	0.1	0.3234	0.2762	0.4150	0.3974	0.3686
		0.2	0.3386	0.3730	0.5376	0.4976	0.4490
		0.3	0.3570	0.5282	0.6770	0.6156	0.5454
	0.5	0.1	0.6842	0.6320	0.7764	0.7576	0.7450
		0.2	0.6742	0.7194	0.8462	0.8132	0.7914
		0.3	0.6546	0.8302	0.9144	0.8730	0.8466
	1.0	0.1	0.9966	0.9952	0.9990	0.9992	0.9982
		0.2	0.9964	0.9986	0.9994	0.9992	0.9992
		0.3	0.9916	0.9986	0.9996	0.9994	0.9992
0.7	0.0	0.1	0.0482	0.0206	0.0582	0.0580	0.0512
		0.2	0.0512	0.0314	0.0892	0.0848	0.0594
		0.3	0.0516	0.0730	0.1372	0.1266	0.0842
	0.3	0.1	0.3294	0.2544	0.4286	0.4218	0.3994
		0.2	0.2968	0.3338	0.5286	0.4982	0.4586
		0.3	0.2602	0.4584	0.6162	0.5830	0.5144
	0.5	0.1	0.7108	0.6306	0.8088	0.7974	0.7872
		0.2	0.6532	0.6954	0.8584	0.8408	0.8210
		0.3	0.6118	0.8030	0.9088	0.8858	0.8602
	1.0	0.1	0.9994	0.9982	1	1	1
		0.2	0.9944	0.9984	0.9996	0.9994	0.9996
		0.3	0.9904	0.9984	1	1	0.9998

표 3.7: 모의 실험 결과

MCAR					
		$n = 15$		$n = 30$	
		size	power	size	power
$\rho = 0.3$	p=0.1	C, MI, ML	ML > C > MI*	C, MI, ML	ML > C > MI
	p=0.2	C, MI, ML	ML > MI>C***	C, MI, ML	ML > MI > C*
	p=0.3	C	C	C, ML	ML > C
$\rho = 0.5$	p=0.1	C, MI, ML	ML > C > MI	C, MI, ML	ML > C > MI*
	p=0.2	C, MI, ML	ML > C > MI	C, MI, ML	ML > C > MI
	p=0.3	C, MI	MI > C	C, ML	ML > C >MI**
$\rho = 0.7$	p=0.1	C, MI, ML	ML > C > MI	C, MI, ML	ML > C > MI
	p=0.2	C, MI, ML	ML > C > MI	C, MI, ML	ML > C > MI*
	p=0.3	C, MI	C > MI *	C, MI, ML	ML > C > MI*
MAR					
		$n = 15$		$n = 30$	
		size	power	size	power
$\rho = 0.3$	p=0.1	C, MI, ML	ML > MI > C	C, MI, ML	ML > MI > C
	p=0.2	C, MI, ML	MI > ML > C	MI, ML	MI > ML
	p=0.3	ML	ML	ML	ML
$\rho = 0.5$	p=0.1	C, MI, ML	ML > MI > C	C, MI, ML	ML > MI > C
	p=0.2	MI, ML	ML > MI	MI, ML	ML > MI
	p=0.3	MI, ML	MI > ML	MI, ML	MI > ML
$\rho = 0.7$	p=0.1	MI, ML	ML > MI	MI, ML	ML > MI*
	p=0.2	MI, ML	ML > MI	MI, ML	ML > MI
	p=0.3	MI, ML	MI > ML	MI, ML	MI > ML*
NIM					
		$n = 15$		$n = 30$	
		size	power	size	power
$\rho = 0.3$	p=0.1	C, MI, ML	ML > MI > C	C, MI, ML	ML > MI > C**
	p=0.2	C	C		
	p=0.3				
$\rho = 0.5$	p=0.1	C, MI, ML	ML > C > MI	C, MI, ML	ML > C > MI
	p=0.2	C, MI, ML	ML > MI > C	MI	MI
	p=0.3	C	C		
$\rho = 0.7$	p=0.1	C, MI	C > MI*	C, MI, ML	ML > C > MI*
	p=0.2	C, MI, ML	ML > MI > C	C, MI	MI > C
	p=0.3	C, MI	MI > C	C	C

* ; except $\delta = 1.0$

** ; except $\delta = 0.5$

*** ; except $\delta = 0.3$

감사의 글

본 논문을 심사해주신 심사위원님들께 감사를 드립니다.

참고문헌

[1] Anderson, T. W. (1984) . *An Introduction to Multivariate Statistical Methods* (2nd ed.), John Wiley & Sons, New York.

[2] Brown, M. B. (1992). A test for difference between two treatments in a continuous measure of outcome when there are dropouts. *Controlled Clinical Trials*, 13, 213-225.

[3] Heyting, A., Tolboom, J,T.B.M and Essers, J.G.A. (1992). Statistical handling of dropouts in longitudinal clinical trials, *Statistics in Medicine*, 11, 2043-2061.

[4] Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* John Wiley & Sons, New York.

[5] Rubin, D. B. (1976). Inference and missing data, *Biometrika*, 63, 581-92.

[6] Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Survey.* John Wiley & Sons, New York.

[1998년 6월 접수, 1999년 4월 최종수정]

Tests for Incomplete Paired Data

Seung-Mook Lee[1)] Jin-Kyung Park[2)] Taesung Park[3)]

ABSTRACT

We consider five test procedures for analyzing paired data in the presence of missing observations. They are classified according to how to handle missing data. Through simulation studies, we compare them based on empirical sizes and powers under the three missing data mechanisms.

1) Graduate Student, Department of Statistics, Hankuk University of Foreign Studies
2) Graduate Student, Department of Statistics, Hankuk University of Foreign Studies
3) Associate Professor, Department of Statistics, Seoul National University