

# 임상 자료 분석을 위한 NORMAL TEST VARIABLE(NTV)의 고찰 \*

이제영<sup>†</sup> 우정수<sup>‡</sup> 최달우<sup>§</sup>

## 요약

임상자료의 정규성 검정에 있어서 NTV(Normal Test Variable)에 관한 몇가지 문제 점을 여러 가지 특정분포 즉, 정규분포, 두 정규분포의 혼합분포, 균일분포, 지수분포 그리고 오른쪽으로 치우친 베타분포에 관한 몬테칼로 모의실험 결과를 통하여 고찰하고 그 해결방안을 모색한다.

## 1. 서론

정규성 검정에 관한 연구는 많은 통계학자들의 관심의 대상으로 계속 연구되어 왔다. 이와 같은 분포의 정규성을 검정하는 대표적인 방법으로는 Kolmogorov-Smirnov 검정, Shapiro와 Wilk (1965)의 W 검정, Dyer (1974)의 Anderson과 Darling ( $An^2$ ) 검정 등이 있다. 이와 같이 대수적인 방법에 의한 통계량으로 분포의 정규성을 검정하는 방법 외에도, 한편에서는 그래프에 의하여 시각적으로 분포의 정규성을 검정하는 방법의 연구가 계속되어 왔다. 대표적인 방법으로는 P-P 플롯, Q-Q 플롯 그리고 정규확률 그래프 등이 있고, 이러한 연구는 Wilk와 Gnanadesikan (1968), LaBrecque (1977), Gnanadesikan (1980), Mage (1982), Neter *et al.* (1990), Johnson *et al.* (1992), Holmgren (1995) 그리고 Fang과 Case (1996) 등에 의해 발전되어 왔다.

한편, 의학실험에서는 일본인과 백인(Caucasian)의 debrisoquin과 mephenytoin hydroxylation이 Bimodal분포(1985, Nakamura *et al.*)로 밝혀지고 흑인(Ghanaians)과 독일인의 Spartein Oxidation의 분포(1985, M, Eichelbaum and N. M. Woolhouse)와 Cuna Amerindians의 sparteine metabolism의 분포(1988, Aris *et al.*) 등이 Bimodal분포를 나타낸다는 임상학적 연구가 밝혀지면서 정규분포와 Bimodal분포에 대하여 많은 관심을 갖게 되었다. 최근에 Jackson *et al.* (1989a, b)은 히스토그램과 Probit 플롯을 이용한 Bimodality의 검정에 대하여 연구를 했고, Endrenyi과 Patel (1991a, b)은 정규성과 Bimodality 검정을 위한 NTV(Normal Test Variable) 플롯을 개발하였다. 그리고 Lee와 Rhee (1997)는 정규성과 Bimodality의 검정에 사용되는 플롯들의 형태의 특징에 대한 연구를 하였다.

\* 본 연구는 한국과학재단 핵심전문연구비의 부분적 지원(971-0105-032-2)으로 수행되었음.

<sup>†</sup> (712-749) 경상북도 경산시 대동 214-1, 영남대학교 통계학과 조교수

<sup>‡</sup> (712-749) 경상북도 경산시 대동 214-1, 영남대학교 통계학과 교수

<sup>§</sup> (702-701) 대구시 북구 산격동 1370, 경북대학교 통계학과 강사

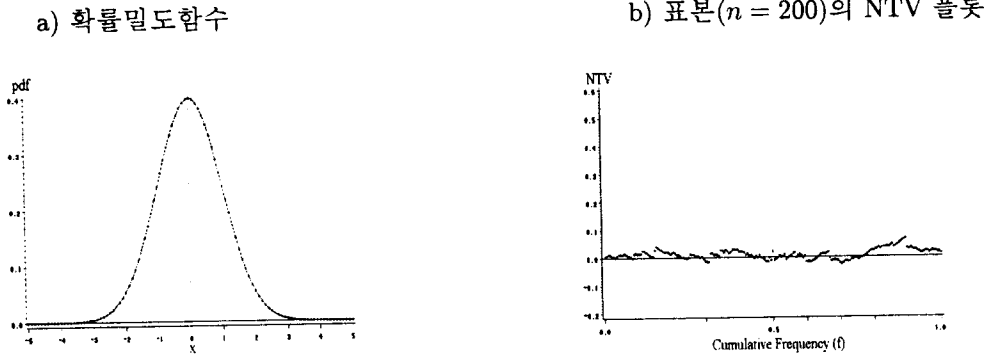


그림 2.1: 정규분포( $N(0, 1)$ )의 확률밀도함수와 확률표본에 대한 NTV 플롯

본 논문에서는 임상자료의 정규성 검정에 대해 최근에 발표된 NTV 플롯에서 비롯된 몇 가지 문제점과 그에 대한 대안에 대하여 논의해 보고자 한다. 먼저 NTV 플롯에 대한 통계학적인 소개를 한다(2절). 이때  $NTV_i$ 의 값이 대략 0 주위의 값, 즉  $NTV=0$  주변의 값을 갖는다면 표본이 정규분포를 따른다고 기대되고  $NTV_i$ 의 최저값과 최고값이 정규성 검정의 측도가 된다고 알려져 있으며, 특히  $NTV_i$  값이  $-0.03$ 보다 작은 값을 가질 때에는 주어진 분포가 Bimodality를 나타낸다고 알려져 있다. 여기에서 본 논문은 NTV 플롯과  $NTV_i$ 가 나타내는 수리적인 값인  $-0.03$ 의 기준점을 중심으로 특징적인 문제점을 모의실험 결과를 바탕으로 여러 특정 분포별로 제시하고 그 해결 방안을 살펴본다. 즉, 여러 확률분포로부터 SAS 프로그램을 통하여 생성된 표본들을 사용해서 Shapiro-Wilk 검정과 비교하여 NTV 플롯의 특징과 문제점을 서로 고찰해 보고(3절), 더 나아가 임상학적 예제로서 설명하며(4절), 마지막으로 이들 문제점에 대한 대안을 제시한다(5절).

## 2. NTV 플롯의 이론적 배경

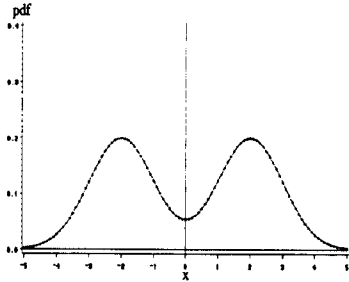
NTV 플롯은 Endrenyi와 Patel (1991a)에 의해 제시된 것으로 정규성 검정뿐만 아니라, Bimodality 검정에도 유용하다고 알려져 의학통계에서 많이 사용되고 있다. NTV 플롯을 구현하는 절차를 통계학적으로 알아보면 다음과 같다. 먼저, 표본  $x_1, \dots, x_n$ 의 중위수  $m$ 과 표준편차  $s$ 를 계산하고, 표본의 순서통계량  $x_{1:n}, \dots, x_{n:n}$ 을 구하여

$$y_{i:n} = \frac{x_{i:n} - m}{s}$$

$$f = \frac{i}{n}$$

$$z_i = f - (1 - f)e^{1.6y_{i:n}}$$

a) 확률밀도함수



b) 표본(n = 200)의 NTV 플롯

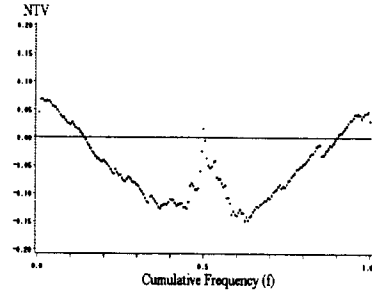


그림 2.2: Bimodal분포( $\frac{1}{2}N(-2, 1) + \frac{1}{2}N(2, 1)$ )의 확률밀도함수와 확률표본에 대한 NTV 플롯

를 계산한 후, NTV를 다음과 같이 정의한다:

$$NTV_i = \begin{cases} -z_i, & f \leq 0.5 \\ \frac{z_i}{e^{1.6z_i/n}}, & f > 0.5. \end{cases}$$

이때, NTV 플롯은  $(x, y)$  평면상에,

$$(f, NTV_i)$$

를 표시하는 그래프를 말한다. 이것이 정의된 합리적인 임상학적 배경은 특히 Michaelis-Menten 방정식에 의존하며 구체적인 해석은 Endrenyi와 Patel (1991a, b)을 참조하기 바란다.

실제로 표준정규분포( $N(0, 1)$ )로부터 생성된 크기가 200( $n = 200$ )인 확률표본으로 그려본 NTV 플롯의 그래프는 대강  $x$ 축 주위로 퍼져 있으며 이때 모집단은 정규분포를 따른다고 결론을 내린다(그림 2.1). 또 한편 모분산이 같고 모평균이 다른 두 정규분포에서 같은 비율로 생성된 크기 200인 확률표본에서 나타난 NTV 플롯은 그림 2.2와 같으며 누적빈도=0.5를 중심으로 좌우에 아래쪽으로 크게 하나씩의 정상점(peak)을 가진 W자 형태를 형성하였다. 여기에서 Endrenyi와 Patel (1991a)은 이와같이 Bimodal분포에 대한 NTV가 Q-Q 플롯과 같은 정규 그래프에서 나타나는 일반적인 단락점(break-point)보다 훨씬 민감함을 보였다. 이것과 관련된 여러 특징적 분포에 대한 연구가 Lee와 Rhee (1997)에 의해 연구되었다. 하지만 본 논문은 Endrenyi와 Patel (1991a, b)이 정의한 NTV 플롯에 대한 여러가지 그래프적인 특징과 민감도에는 동의하면서도 한편으로  $NTV_i$ 의 최소값이 그들이 제시한 경계값인 -0.03보다 작을 때 분포가 Bimodality를 나타낸다는 점에 의아심을 갖고 이들에 대한 검정을 모의실험 결과를 통하여 점검해 보고 그 문제점을 밝힌다(3절).

표 3.1: 정규분포에서 생성된 확률표본들에 대한 Shapiro-Wilk(S-W) 정규성 검정 결과

Generated Distribution	Sample Size	Level of Significance	
		0.1	0.05
Normal(0,1)	200	911	955

표 3.2: 정규분포에서 생성된 확률표본들에 대한 NTV 최소값과 S-W 검정의 비교

Generated Distribution	Sample Size	Whole Samples		Normal Samples Accepted S-W Test at Level of Significance			
				0.1		0.05	
		minimum NTV < 0	< -0.03	minimum NTV < 0	< -0.03	minimum NTV < 0	< -0.03
Normal(0, 1)	200	1000	822	911	738	955	780

### 3. 모의실험 결과에 의한 NTV 플롯의 문제점

우리는 모의실험 결과(simulation results)를 통하여 위에서 제시한 Endrenyi와 Patel (1991a)의 결론에 내재된 몇가지 문제점을 제시한다.

#### 3.1. 정규성 검정에 대한 모의실험 결과

먼저 정규성 검정에 관해 알아본다. SAS 프로그램을 통하여 생성된 크기가 200인 표본에 대한  $NTV_i$ 를 계산하고 그것들 중에서 최소값을 구하여 0과 -0.03과 비교하여 결과를 기록한다. 이와같은 몬테칼로 모의실험을 1000번씩 반복하였다. 만약 Endrenyi와 Patel이 제시한 정규성의 검정 기준인 0, 혹은 Bimodality 검정의 기준인 -0.03이 특별한 의미를 가지고 있다면 모의실험 결과와 어떤 연관성이 있다고 기대할 수 있다. 먼저 샘플링된 표본에 대하여 Shapiro-Wilk(S-W) 정규성 검정을 통하여 이 표본들이 정규분포에서 샘플링된 것임을 보였다.

표 3.1은 각각의 유의수준에서 정규성을 기각하지 못한 표본의 수를 표시한 것이다. 유의수준 0.1과 0.05에서 가정된 분포가 정규분포이면 정규성을 나타내는 표본의 수가 각각 900과 950 정도가 기대되며, 표 3.1의 수치는 이것을 조금 넘어서는 결과를 나타내고 있어서 가정된 분포가 정규성을 나타낸다고 볼 수 있다. 표 3.2는 각 표본의 최소 NTV 값과 Shapiro-Wilk 검정 결과와의 관계이다. 이것을 살펴보면 표본 모두 NTV 값이 0이하의

표 3.3: Bimodal분포에서 생성된 확률표본들에 대한 NTV 최소값과 S-W 검정의 비교

Generated Distribution	Sample Size	Whole Samples		Normal Samples Accepted S-W Test at Level of Significance			
				0.1		0.05	
		minimum NTV < 0	< -0.03	minimum NTV < 0	< -0.03	minimum NTV < 0	< -0.03
Mixture Normal*	200	1000	988	493	482	630	618
Mixture Normal**	200	1000	824	897	724	943	768

\*  $\frac{1}{2}N(-1, 1) + \frac{1}{2}N(1, 1)$

\*\*  $\frac{1}{2}N(-0.5, 1) + \frac{1}{2}N(0.5, 1)$

값을 가짐을 알 수 있고 아울러 -0.03 이하의 값도 80 % 이상(1000 경우중 822)의 표본에서 나타났다. 이것은 표 3.1의 Shapiro-Wilk 검정 결과와는 다름을 알 수 있다. 예를 들어 Shapiro-Wilk 검정 결과, 유의수준 0.1에서 정규성을 받아들인 표본 911개에 대하여 911 표본 모두 0 이하의 NTV 값이 나타나 정규성을 의심하는 결론을 나타내고, 그중에서 738개가 -0.03이하의 값을 가져 Shapiro-Wilk 검정과는 아주 다른 Bimodality를 나타낸다는 결론을 내리게 된다.

### 3.2. BIMODALITY 검정에 대한 모의실험 결과

다음으로 Bimodality 검정에 대한 모의실험을 위하여 몇가지 혼합정규분포(Mixture Normal Distribution)를 SAS 프로그램에서 구현하여 각 분포에 대하여 크기가 200인 확률표본을 1000개 생성하였다. 앞의 모의실험과 같이 각 표본에 대하여 NTV<sub>i</sub>를 계산하고, 그것들 중에서 최소값을 구하여 그 값을 0과 -0.03과 비교하여 결과를 기록한다. 앞에서와 같이 Endrenyi와 Patel (1991a)이 제시한 정규성의 검정 기준인 0 혹은 Bimodality 검정의 기준인 -0.03과 모의실험 결과를 비교하여 표 3.3에 나타내었다. 표 3.3에서는 잘 나타나 있지 않지만 가장 오른쪽에 있는 두개의 열(column)들은 Shapiro-Wilk 검정에서 각각의 유의수준에서 정규분포를 기각하지 못한 표본에 대하여 NTV의 최소값을 비교한 부분이다. 이것을 보면 Bimodal분포에서 두개의 최빈값(mode)이 가까워지면 Shapiro-Wilk 검정은 정규성을 잘 기각하지 못하는 경향을 보이고 있다. 이때 NTV 플롯에서는 상당한 수의 표본들은 -0.03 이하의 NTV 값을 포함하고 있으므로 Bimodality를 나타낸다고 볼 수 있다. 이것은 가정된 분포가 혼합정규분포이므로 NTV 플롯에서의 결과가 훨씬 민감함을 알 수 있었다.

표 3.4: 특정분포에서 생성된 확률표본들에 대한 NTV 최소값

Generated Distribution	Sample Size	Minimum NTV < 0	Minimum NTV < -0.03	Shape of Distribution
Uniform(0,1)	200	1000	1000	Short-tailed
Exponential(1)	200	1000	840	Skewed-left
Beta(2,1)	200	1000	1000	Skewed-right

### 3.3. 특정 형태의 분포에 관한 모의실험 결과

여기에서 다루는 특정 분포는 정규분포와 비교하여 분포의 형태가 다른 3 가지 분포를 말한다. 즉, 분포가 짧은 꼬리를 가지는 분포(Uniform(0, 1)), 왼쪽으로 치우친 분포(Exponential(1)), 그리고 오른쪽으로 치우친 분포(Beta(2, 1))이다. 앞의 모의실험과 같은 실험을 한 결과를 표 3.4에 나타내었다. 여기서는 정규분포가 아니므로 Shapiro-Wilk의 정규성 검정 결과는 생략하였다. 이 표를 보면 분명 모집단은 Bimodal 분포가 아닌데도 Bimodality를 나타내는 -0.03 이하의 NTV 값이 거의 모든 표본에서 관측됨을 알 수 있었다.

## 4. 임상자료의 실례

실제 임상실험의 자료인 표 4.1은 Hodgkin's 병에서 회복된 환자와 Non-Hodgkin's 병에서 회복된 환자 각각 20명에 대한 혈액 샘플에서  $mm^3$  당 T4 세포의 숫자를 Log 변환시킨 값을 나타낸 것이다. Alterman (1991)은 이 변환에 의한 표본이 정규분포를 따른다고 주장하였고, 실제로 Shapiro-Wilk 정규성 검정 결과, P-값이 각각 0.7950과 0.4354로 구해져서 표본이 정규분포를 따른다는 주장을 받아들인다. 그러나 그림 4.1에서 보는 바와같이 NTV 플롯의 형태는 Endrenyi와 Patel (1991a)이 밝힌 정규분포에 대한 그래프와는 다소 차이가 있으며, NTV의 최소값 또한 Bimodality를 나타낸다는 -0.03보다 훨씬 적은 값을 가진다는 사실을 알 수 있다. 이 예제를 통해 여기에서도 Endrenyi와 Patel (1991a)의 -0.03 기준값은 그 적용이 어렵다는 사실을 확인할 수 있다.

## 5. 결론

Endrenyi와 Patel (1991a)에 따르면, 정규분포에 대한 NTV 플롯에서의  $NTV_i$  값은 수평축( $NTV = 0$ )에 거의 접근되어야만 하고 이것을 통하여 그래프적인 측면에서 매우 민감함을 보여주었으며, 또한 실제로 Probit 플롯과의 비교를 통해서도 상세히 분석되었다

표 4.1: Hodgkin's 병과 Non-Hodgkin's 병에서 회복된 환자들에 대한 혈액 표본에서  $mm^3$  당 T4 세포의 수와 그것의 Log 변환값 (Alterman, 1992)

Subject	Raw Sample		Log Transformation	
	Hodgkin's Disease	Non-Hodgkin's Disease	Hodgkin's Disease	Non-Hodgkin's Disease
1	396	375	2.59770	2.57403
2	568	375	2.75435	2.57403
3	1212	752	3.08350	2.87622
4	171	208	2.23300	2.31806
5	554	151	2.74351	2.17898
6	1104	116	3.04297	2.06446
7	257	736	2.40993	2.86688
8	435	192	2.63849	2.28330
9	295	315	2.46982	2.49831
10	397	1252	2.59879	3.09760
11	288	675	2.45939	2.82930
12	1004	700	3.00173	2.84510
13	431	440	2.63448	2.64345
14	795	771	2.90037	2.88705
15	1621	688	3.20978	2.83759
16	1378	426	3.13925	2.62941
17	902	410	2.95521	2.61278
18	958	979	2.98137	2.99078
19	1283	377	3.10823	2.57634
20	2415	503	3.38292	2.70157

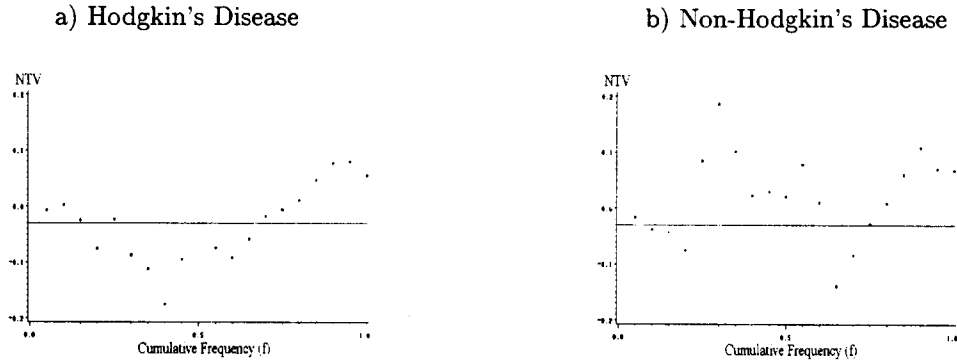


그림 4.1: Hodgkin's 병과 Non-Hodgkin's 병에서 회복된 환자 각각 20명에 대한 혈액 표본에서  $mm^3$  당 T4 세포의 숫자를 Log 변환시킨 표본에 대한 NTV 플롯

(Endrenyi and Patel, 1991a 그리고 Lee and Rhee, 1997). 그러나 그들이 Bimodality를 나타내는 경계값으로 주장한  $-0.03$ 이 앞의 표 3.2와 표 3.4에서와 같이 Bimodality만을 나타낸다고 볼 수 없게 되었다. 수치적인 해석을 위하여 몇가지 크기의 표본에 대하여 모의실험을 실시하였다. 가령, 표본의 크기가 100과 200인 경우 정규분포로부터 생성한 표본으로 일만(10000)번의 모의실험을 한 결과, 표본의 크기가 100인 경우에는 약 5% 정도가  $-0.11$  이하의  $NTV_i$  값을 포함하고 있었고, 표본의 크기가 200인 경우에는 약 5% 정도가  $-0.08$  이하의  $NTV_i$  값을 포함하고 있었다. 이것은 Endrenyi와 Patel (1991a)이 주장한  $-0.03$ 이 Bimodality 뿐만아니라 정규성을 검정하는데도 좋은 경계값이 되지 못함을 수치적으로도 보여주고 있다. 여기에서 특히 정규성을 검정하는 기준값이 표본의 크기에 따라서 민감하게 변화하고 있음을 알게되었다. 그래서 본 연구에서는 다음과 같은 결론을 얻을 수 있었다. NTV의 최대값과 최소값은 표본의 크기에 매우 의존적인 성질을 나타내고 있어서 정규성이나 Bimodality를 검정할 경우에는 Endrenyi와 Patel (1991a) 그리고 Lee와 Rhee (1997)가 보여준 NTV 플롯의 그래프적인 측면에서만 판단을 해야한다.

## 참고문헌

- [1] Alterman, D.G. (1992). *Practical statistics for medical research*. Chapman and Hall. London.
- [2] Arias, T. D., Jorge, L. F., Lee, D. Barrantes, R. and Inaba, T. (1988). The oxidative metabolism of sparteine in the Cuna Amerindians of Panama: Absence of evidence for deficient metabolizers. *Clin. Pharmacol. Ther.* Vol. 43. 456-465.
- [3] Dyer, A.R. (1974). Comparison of tests for normality with a cautionary note. *Biometrika*. Vol. 61. 185-189.



- [4] Eichelbaum, M. and Woolhouse, N. M. (1985). Interethnic difference in sparteine oxidation among Ghanaians and Germans. *Eur. J. Clin. Pharmacol.* Vol. 28. 79-83.
- [5] Endrenyi, L. and Patel, M. (1991a). A new, sensitive graphical method for detecting deviations from the normal distribution of drug responses: the NTV plot. *Br. J. Clin. pharmacol.* Vol. 32. 159-166.
- [6] Endrenyi, L. and Patel, M. (1991b). Evaluation of two assumptions: single straight line, and single normal distribution. *Trends. Pharmacol. Sci.* Vol. 12. 293-296.
- [7] Fang, J. and Case, K. E. (1996). Using a correlation test for normality. *Journal of Quality Technology.* Vol. 28. 356-362.
- [8] Gnanadesikan, R. (1980). *Graphical methods for internal comparisons in ANOVA and MANOVA.* In P. R. Krishnaiah (Ed.). Handbook of statistics. Vol. 1. 133-177. North-Holand: Amsterdam.
- [9] Holmgren, E. B. (1995). The P-P plot as a method for comparing treatment effects. *J. of Am. Stat. Assoc.* Vol. 90. 360-365.
- [10] Jackson, P.R., Tucker, G.T. and Woods, H.F. (1989a). Testing for bimodality in frequency distributions of data suggesting polymorphisms of drug metabolism -histograms and probit plots. *Br. J. Clin. Pharmacol.* Vol. 28. 647-653.
- [11] Jackson, P.R., Tucker, G.T. and Woods, H.F. (1989b). Testing for bimodality in frequency distributions of data suggesting polymorphisms of drug metabolism -hypothesis testing. *Br. J. Clin. Pharmacol.* Vol. 28. 655-662.
- [12] Johnson, R. A. and Wichern, D. W. (1992). *Applied multivariate statistical analysis* (3rd ed.). Englewood cliffs, NJ; Printice-Hall.
- [13] LaBrecque, J. (1977). Goodness-of-fit tests based on nonlinearity in probability plots. *Technometrics.* Vol. 19. 293-306.
- [14] Lee, J. and Rhee S. (1997). 특정분포에 따른 확률 Plot들의 정규성과 Bimodality 비교. <한국통계학회논문집>. 제4권 1호. 243-254.
- [15] Mage, D. T. (1982). An objective graphical method for testing normal distributional assumptions using probability plots. *Am. Stat.* Vol. 36. 116-120.
- [16] Nakamura, K., Goto, F., Ray, W. A., McAllister, C. B., Jacqz, E., Wilkinson, G. R., and Branch, R. A. (1985). Interethnic differences in genetic polymorphism of debrisoquin and mephenytoin hydroxylation between Japanese and Caucasian populations. *Clin. Pharmacol. Ther.* Vol. 38. 402-408.
- [17] Neter, J., Wasserman, W. and Kunter, M. (1990). *Applied linear statistical Models* (3rd ed.). Richard D. Irwin, Inc., Homewood, Illinois.
- [18] Shapiro, S.S. and Wilk, M.B. (1965). An analysis-of-variance test for normality (complete sample). *Biometrika.* Vol. 52. 591-611.
- [19] Shapiro, S.S. and Francia, R.S. (1972). An approximation analysis of variance test for normality. *J. Am. Stat. Assoc.* Vol. 67. 215-216.

- [20] Wilk, M. B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*. Vol. 55. 1-17.

[ 1997년 4월 접수, 1997년 8월 최종수정 ]

## Using a Normal Test Variable(NTV) for Clinical Research

Jea-Young Lee †, Jungsoo Woo ‡, Dal Woo Choi §

### ABSTRACT

This article examines the use and some difficulties of Normal Test Variables (NTV) plot for clinical research. Monte Carlo Simulation results are presented based on Normal, Bimodal, Uniform, Exponential and skewed-right distributed Beta Distributions. Further, some solutions are presented and illustrated.

---

† Department of Statistics, Yeungnam University, Kyongsan, Kyongbuk, Korea.

‡ Department of Statistics, Yeungnam University, Kyongsan, Kyongbuk, Korea.

§ Department of Statistics, Kyungpook National University, Taegu, Korea.