

Q-Q, P-P 플롯의 변동 통계량에 대한 ROC 분석

이 제 영¹⁾ 이 성 원²⁾

요 약

정규분포에 관한 검정에 있어서 P-P 플롯과 Q-Q 플롯의 가시적인 변동을 이용한 통계량을 제시하고 이 통계량들과 Shapiro-Wilk의 W 통계량과의 비교를 정확도(accuracy)의 측면을 고려하여 실시하였다. 또한, 의학이나 임상에서 척도의 우수성을 검정하기 위해 많이 사용하는 Receiver Operating Characteristic (ROC) 분석 기법을 이용하여 제시된 통계량들에 관한 Power와 Accuracy는 물론 Best Cut-Off 측면에서의 효율성을 검정하였다.

1. 서 론

정규성 검정에 관한 연구는 통계학 연구에 있어서 고전적인 부분이면서 또한 최근까지 연구되는 중요한 주제이다. Shapiro와 Wilk (1965)가 W 검정을 제안하였고, 이어 Shapiro와 Francia (1972)가 W' 검정을 제안하였다. 계속해서 많은 학자들이 여러 가지 통계량들을 제시하여 연구를 하였으며, 최근에는 Lin과 Mudholker (1980)가 표본평균과 표본분산의 독립성을 기초로 Z 통계량을 제안하여 정규성 검정을 시도하였고, Looney (1995)는 다중정규성(multivariate normality)에 대한 연구를 하였고 Kim (1997)은 이변량 정규분포에 대한 연구를 하였다. 한편, 이와 같은 통계량을 이용한 정규성 검정 방법 외에도, 그래프에 의하여 시각적으로 표본의 정규성을 검정하는 방법이 통계학뿐만 아니라 의학부문에서도 계속 연구되어 왔다. 대표적인 그래프로는 정규화를 그레프를 위시하여 P-P, Q-Q 플롯을 들 수 있다. 이들 연구는 컴퓨터의 발달로 최근 들어 더욱 활발해 져서, Jackson *et al.* (1989), Endrenyi와 Patel (1991), Lee와 Rhee (1997), 그리고 Lee *et al.* (1998) 등에 의해 연구되었다.

한편 임상의학 부문에서는 진단법의 정확도 및 두 진단법의 우수성에 대한 검정 방법으로 널리 사용되는 통계적 기법인 Receiver Operating Characteristic (ROC) 분석에 대한 연구가 Metz (1978)에 의해 소개되었으며 Hanley와 McNeil (1982, 1983), Altman (1992), Zweig와 Campbell (1993), Schoonjans *et al.* (1995) 등에 의하여 계속 연구되었다.

본 논문에서는 먼저 ROC 분석 기법에 대하여 소개하고(2절), 정규성 검정에 대한 대표적인 그래프인 P-P 플롯과 Q-Q 플롯을 이용한 표본표준편차에 의한 새로운 검정 통계량을 제시하고(3

1) (712-749) 경상북도 경산시 대동 214-1 영남대학교 통계학과 조교수

2) (700-400) 대구시 중구 봉산동 255-5 국제조사연구소 선임연구원

절), 이들 통계량과 정규성 검정에 있어서 가장 널리 사용되는 Shapiro-Wilk의 W 통계량을 ROC 분석을 통하여 그들간의 효율성과 정확도를 검정하고자 한다(4절).

2. ROC (Receiver Operating Characteristic) 분석

의학 부문에서의 질병에 대한 진단(diagnostic decision)이나 통계학 부문에서의 가설에 대한 검정은 서로 의미가 일치하는 바가 있으므로 ROC 분석에 대하여 소개한다. 먼저, 몇 가지 기본적인 개념을 소개하기로 한다. 옳은 진단이나 검정을 내릴 비율을 정확도(accuracy)라고 정의하고, 특히 귀무가설이 참일 때 귀무가설을 채택할 비율을 특이도(specificity, Spec.)라고 하며 대립가설이 참일 때 대립가설을 채택할 비율을 민감도(sensitivity, Sens.)라고 한다. 이것은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \text{Specificity} &= \frac{\text{H}_0\text{를 옳게 채택할 경우의 수}}{\text{실제 H}_0\text{가 참인 경우의 수}}, \\ \text{Sensitivity} &= \frac{\text{H}_1\text{를 옳게 채택할 경우의 수}}{\text{실제 H}_1\text{가 참인 경우의 수}}, \\ \text{Accuracy} &= \frac{\text{옳은 진단을 한 경우의 수}}{\text{전체 경우의 수}} \\ &= \text{Sensitivity} \times (\text{전체에서 H}_1\text{이 참인 비율}) \\ &\quad + \text{Specificity} \times (\text{전체에서 H}_0\text{이 참인 비율}). \end{aligned}$$

이 Spec.과 Sens.은 통계적으로 $(1 - \text{Type I Error}) \times 100\%$ 와 $(1 - \text{Type II Error}) \times 100\%$ 의 의미와 유사하다. 특히, Sens.는 바로 Power를 의미하고 있다. 우리는 여기에서 진단의 정확도를 측정하고 가장 바람직한 진단의 기준을 파악하기 위하여 ROC 분석을 하게된다. 결국 진단의 기준(threshold)이나 검정 통계량의 기각역을 변화시켜가면서 좌표평면 위에 (1-Spec., Sens.) 좌표를 표시하는 그래프를 ROC 곡선이라고 부른다. 이 곡선과 x 축 사이의 면적을 AUC(Area Under the ROC Curve)라고 하는데, 이것은 정확도에 대한 척도로 사용된다. 즉, θ 를 아래와 같은 확률로 정의한다면, 이 θ 는 이 주어진 통계량이 $H_0: \text{normal}$ 하에서와 $H_1: \text{abnormal}$ 하에서 어느 정도 옳게 구별될 수 있는가에 대한 척도이다.

$$\begin{aligned} \theta &= \Pr[H_0\text{에서의 통계량 } < H_1\text{에서의 통계량 }], \\ \text{또는 } \theta &= \Pr[H_0\text{에서의 통계량 } > H_1\text{에서의 통계량 }]. \end{aligned}$$

이 AUC에 대하여 Hanley와 McNeil(1982)에 의해 다음과 같은 성질이 알려졌다.

$$\text{i) } AUC = \widehat{\theta},$$

$$\text{ii) } AUC = \frac{1}{n_A \cdot n_N} \sum_{i=1}^{n_A} \sum_{j=1}^{n_N} S(x_i^A, x_j^N),$$

이때, n_A 과 n_N 는 각각 H_1 과 H_0 하의 표본의 수를 말하고 x^A 과 x^N 는 각각 H_1 과 H_0 하의 통계량의 값을 말하며, S 는 다음과 같이 정의되는 함수를 의미한다.

$$S(x_i^A, x_j^N) = \begin{cases} 1 & \text{if } x_i^A > x_j^N, \\ \frac{1}{2} & \text{if } x_i^A = x_j^N, \\ 0 & \text{if } x_i^A < x_j^N. \end{cases}$$

결국, AUC는 주어진 통계량이 H_0 하에서와 H_1 하에서 어느 정도 옳게 구별이 될 수 있는가를 알아보는 확률 θ 를 재는 척도를 의미한다.

이때 우리는 ROC 분석을 통하여 일정한 유의수준하에서의 Power(Sens.) 등을 구할 수도 있으며, 또한 진단의 정확도를 가장 높일 수 있는 Best Cut-Off 값도 구할 수 있다. 이 Best Cut-Off 값은 정확도의 측면에서 발생하는 모든 오류(Type I Error + Type II Error)를 가장 최소로 하는 기각역으로 기존의 일정한 제1종 오류의 수준에서 제2종 오류의 수준을 줄여나가는 방법보다 많은 정보를 제공하여 주므로 Altman (1992)에 의하여 중요성이 강조되었다. 예를 들어 어떤 검정에 있어서 오류가 발생할 때 지불하는 비용이 제1종 오류와 제2종 오류에 관계없이 동일하다면 일정한 유의수준 하에서 정한 기각역보다 Best Cut-Off 하에서 정한 기각역에서 적은 비용이 발생하므로 Best Cut-Off를 이용하는 것이 더 바람직하다고 할 수 있다(Metz, 1978). 그리고 나아가서 두 진단법이나 검정법에 대한 비교에서도 AUC를 이용하여 정확도의 측면에서 우수성을 밝힐 수 있다.

3. 정규성 검정을 위한 Q-Q와 P-P 플롯의 변동통계량

잘 알려진 바와 같이 Q-Q와 P-P 플롯에서의 정규성 검정의 척도는 정규분포하에서 기대되는 직선에서의 벗어남의 정도이다. 이것은 개인이 시각적으로 파악해야 하므로 개인의 주관적인 판단에 크게 의존할 수밖에 없다. 하지만 이 벗어남의 정도를 평균에서 벗어남의 정도를 표준편차로 표현하듯이 수치적으로 표현하여 정규성 검정의 척도로 삼으면 개인의 주관적인 요소를 배제시킬 수 있다.

확률표본 X_1, \dots, X_n 의 순서통계량을 $X_{1:n}, \dots, X_{n:n}$ 이라 하고, 이 확률변수 X 가 표준정규분포를 따를 때 이 확률변수의 누적분포함수(CDF)를 $\Phi(x)$ 라고 하자. 그러면 Q-Q 플롯은 좌표평면 위에 $(\Phi^{-1}\left(\frac{i-c}{n-2c+1}\right), x_{i:n})$ 좌표로 구성된다. 이때, y 좌표의 기대값을 구하면 다음과 같다.

$$E[X_{i:n}] \cong \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right).$$

c 값에 대해서는 여러 가지 연구를 통하여 3/8일 때가 가장 바람직하다고 알려져 있다. 그러므로 표본이 정규분포를 따를 때, 그 표본의 관측값을 표준화(standardized)시켜 생각하면, 이 Q-Q 플롯에서의 기대되는 직선의 형태는 $y = x$ 이고, 우리는 이 직선으로부터의 흘어진 정도를 정규성 검정의 척도로 사용한다. 이 주관적인 흘어진 정도를 계산하기 위하여, 통계학에서 흔히 사용되는 변동에 대한 유용한 통계량인 표본의 표준편차를 이용하였다. 즉, 기대되는 직선으로부터 편차를 구하여 그 편차의 표본표준편차(sample standard deviation), S_{QQ} 를 아래와 같이 구할

수 있다.

표 1. Shapiro-Wilk W의 ROC 분석 결과

Sample Size	ROC Area	Shapiro-Wilk W						
		Best Cut-off				5% Sig.		
		Criteria	Sens. (Power)	Spec.	Accuracy	Criteria	Sens. (Power)	Accuracy
5	0.556	0.9023	49.0	61.4	55.20	0.7906	6.9	50.95
6	0.574	0.9039	52.4	61.4	56.90	0.8037	7.4	51.20
7	0.598	0.9183	63.4	53.5	58.45	0.8052	5.8	50.40
8	0.606	0.9186	57.0	59.6	58.30	0.8369	10.0	52.50
9	0.632	0.9307	65.4	55.9	60.65	0.8462	8.1	51.55
10	0.635	0.9228	55.0	66.5	60.75	0.8537	9.7	52.35
11	0.642	0.9323	63.0	59.3	61.15	0.8615	9.6	52.30
12	0.644	0.9441	73.2	48.7	60.95	0.8671	9.9	52.45
13	0.671	0.9280	53.8	72.2	63.00	0.8738	10.9	52.95
14	0.685	0.9428	69.3	58.5	63.90	0.8834	12.1	53.55
15	0.711	0.9360	62.1	69.7	65.90	0.8880	15.5	55.25
16	0.713	0.9404	65.6	67.4	66.50	0.8967	17.8	56.40
17	0.741	0.9417	67.3	70.4	68.85	0.8982	19.5	57.25
18	0.757	0.9495	78.3	62.9	70.60	0.9018	21.1	58.05
19	0.763	0.9452	72.2	70.4	71.30	0.9032	21.4	58.20
20	0.785	0.9540	84.6	60.3	72.44	0.9074	21.5	58.25

$$\begin{aligned}
 S_{QQ} &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left\{ \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right) - x_{(i)} - \overline{(\Phi^{-1}(\cdot)) - x} \right\}^2} \\
 &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left\{ \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right) - x_{(i)} - \frac{1}{n} \sum_{i=1}^n (\Phi^{-1}\left(\frac{i-c}{n-2c+1}\right) - \bar{x}) \right\}^2} \\
 &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left\{ \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right) - x_{(i)} \right\}^2}
 \end{aligned}$$

이렇게 구한 S_{QQ} 통계량을 DeWet과 Venter (1972)에 의해 성질이 밝혀진 표준화된 관측값에

대한 정규성의 검정 통계량 $L_n = \sum_{i=1}^n \left\{ \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right) - x_{i:n} \right\}^2$ 을 이용하여 표시하면, 아래와 같다.

표 2. S_{QQ} 의 ROC 분석 결과

Sample Size	S_{QQ}							
	ROC Area	Best Cut-off				5% Sig.		
		Criteria	Sens. (Power)	Spec.	Accuracy	Criteria	Sens. (Power)	Accuracy
5	0.536	0.2876	54.5	53.0	53.75	0.4655	5.8	50.40
6	0.541	0.2414	74.5	32.6	53.55	0.4465	5.4	50.20
7	0.556	0.2693	59.0	52.4	55.70	0.4439	4.0	49.50
8	0.562	0.2615	57.7	54.1	55.90	0.4096	6.0	50.50
9	0.584	0.2613	55.6	59.2	57.40	0.3925	6.2	50.60
10	0.576	0.2517	57.3	55.2	56.25	0.3862	6.8	50.90
11	0.582	0.2257	69.3	43.8	56.55	0.3727	6.2	50.60
12	0.581	0.2511	49.8	63.0	56.40	0.3704	4.4	49.70
13	0.605	0.2372	56.3	61.4	58.85	0.3555	6.2	50.60
14	0.609	0.2501	47.1	69.9	58.50	0.3406	7.4	51.20
15	0.630	0.2247	62.6	59.4	61.00	0.3341	7.1	51.05
16	0.625	0.2240	59.6	61.9	60.75	0.3262	8.3	51.65
17	0.649	0.2095	68.6	56.7	62.65	0.3172	8.2	51.60
18	0.662	0.2028	71.6	54.7	63.15	0.3151	7.8	51.40
19	0.664	0.2008	72.2	54.7	63.45	0.3092	8.0	51.50
20	0.680	0.1888	79.1	48.9	64.00	0.3061	8.0	51.50

$$\begin{aligned} S_{QQ} &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left\{ \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right) - x_{i:n} \right\}^2} \\ &= \sqrt{\frac{1}{n-1} L_n} \end{aligned}$$

o) S_{QQ} 통계량은 Q-Q 플롯에서 좌표들의 $y=x$ 직선에 대한 잔차의 표본표준편차라고 할

수 있다.

같은 방법으로 P-P 플롯에서도 기대되는 직선의 잔차들의 표본표준편차, S_{PP} 를 다음과 같이 구할 수 있다.

표 3. S_{PP} 의 ROC 분석 결과

Sample Size	S_{PP}							
	ROC Area	Best Cut-off				5% Sig.		
		Criteria	Sens. (Power)	Spec.	Accuracy	Criteria	Sens. (Power)	Accuracy
5	0.548	0.1228	49.3	59.8	54.55	0.1708	7.2	51.10
6	0.572	0.1047	62.9	48.7	55.80	0.1553	7.0	51.00
7	0.605	0.1070	46.9	70.7	58.80	0.1399	9.1	52.05
8	0.617	0.0929	56.5	61.4	58.95	0.1302	9.7	52.35
9	0.641	0.0817	67.9	53.9	60.90	0.1197	11.9	53.45
10	0.647	0.0831	57.0	64.4	60.70	0.1134	12.6	53.80
11	0.665	0.0782	61.2	65.1	63.15	0.1080	13.4	54.20
12	0.672	0.0793	52.8	73.0	62.90	0.1056	12.7	53.85
13	0.685	0.0684	66.7	61.4	64.05	0.1006	14.5	54.75
14	0.701	0.0667	68.1	62.3	65.10	0.0982	12.4	53.70
15	0.714	0.0668	65.0	67.8	66.40	0.0915	16.9	55.95
16	0.717	0.0661	62.5	71.5	67.00	0.0909	15.9	55.45
17	0.741	0.0598	72.6	63.6	68.10	0.0859	19.3	57.15
18	0.745	0.0586	73.5	65.3	69.40	0.0846	18.1	56.55
19	0.751	0.0576	73.3	67.1	70.10	0.0820	19.7	57.35
20	0.764	0.0540	79.6	61.6	70.60	0.0806	18.4	56.70

$$S_{PP} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left\{ \left(\frac{i}{n} - \frac{n+1}{n} \times \Phi(x_{i:n}) \right) - \frac{1}{n} \sum_{i=1}^n \left(\frac{i}{n} - \frac{n+1}{n} \times \Phi(x_{i:n}) \right) \right\}^2}.$$

이렇게 구한 S_{PP} 통계량도 역시 P-P 플롯에서 좌표들이 기대되는 직선으로부터 많이 흩어져 있을수록 값이 커지는 표본표준편차가 된다.

여기에서 정의된 S_{QQ} 와 S_{PP} 통계량에 대한 효율성을 ROC 분석을 통하여 Power 측면과

Accuracy 측면에서 검정하고, 특히 AUC를 이용하여 이들 통계량들과 Shapiro-Wilk의 W 통계량 사이의 우수성을 검정하고자 한다.

표 4. ROC 분석의 AUC를 통한 통계량의 검정 결과

Sample Size	W vs S_{QQ}		W vs S_{PP}		S_{PP} vs S_{QQ}	
	Difference ¹⁾	P-value	Difference ²⁾	P-value	Difference ³⁾	P-value
5	0.019	< 0.001	0.008	0.008	0.011	0.010
6	0.033	< 0.001	0.002	0.659	0.032	< 0.001
7	0.043	< 0.001	-0.006	0.147	0.049	< 0.001
8	0.045	< 0.001	-0.011	0.021	0.055	< 0.001
9	0.048	< 0.001	-0.009	0.072	0.056	< 0.001
10	0.059	< 0.001	-0.011	0.026	0.070	< 0.001
11	0.061	< 0.001	-0.022	< 0.001	0.083	< 0.001
12	0.063	< 0.001	-0.028	< 0.001	0.090	< 0.001
13	0.066	< 0.001	-0.014	0.003	0.080	< 0.001
14	0.076	< 0.001	-0.016	0.002	0.092	< 0.001
15	0.081	< 0.001	-0.003	0.545	0.084	< 0.001
16	0.089	< 0.001	-0.004	0.357	0.093	< 0.001
17	0.091	< 0.001	-0.001	0.845	0.092	< 0.001
18	0.095	< 0.001	0.012	0.014	0.083	< 0.001
19	0.099	< 0.001	0.012	0.012	0.087	< 0.001
20	0.104	< 0.001	0.020	< 0.001	0.084	< 0.001

1) Difference = W 의 AUC - S_{QQ} 의 AUC

2) Difference = W 의 AUC - S_{PP} 의 AUC

3) Difference = S_{PP} 의 AUC - S_{QQ} 의 AUC

4. W , S_{QQ} , S_{PP} 통계량의 ROC 분석

이미 정규성 검정의 척도로 잘 알려진 Shapiro-Wilk의 W 통계량과 앞에서 제시한 두 통계량을 여러 가지 표본의 크기($n=5, 6, \dots, 500$)에 대하여 정규분포와 균일분포의 표본을 모의실험을 통하여 각각 2000표본을 생성하여 ROC 분석 기법으로 Power 측면과 Best Cut-Off 측면에서 효율성을 살펴보고, 또한 AUC 측면에서 정확도에 대한 우수성을 검정하도록 한다. 이때 정규분포에 대한 비교를 균일분포에 대하여 한정시킨 것은 ROC분석이 두 개의 집단에 대한 분석이고, 또

한 지수분포나 코쉬분포 등과 같이 손쉽게 구할 수 있는 잘 알려진 특정분포들 중에서 소표본에서 구별이 어려운 분포이면서 임상자료에서 자주 비교되기 때문이다(Shapiro and Wilk, 1965; Fang and Case, 1996).

먼저, 소표본($n=5, 6, \dots, 20$)인 경우에 대하여 살펴보기로 한다. 표 1에서는 Shapiro-Wilk의 W 통계량에 대한 ROC 분석을 통하여 요약한 결과를 나타낸 것이다. 5%의 유의수준 측면에서의 Power는 약간 불안정하지만($n=7, 8, 10$) 표본의 크기가 증가함에 따라 커지는 것은 이미 알려진 바와 같게 나타났으며, 이때의 정확도(accuracy)도 같이 증가한다. 그리고 Best Cut-Off 측면에서의 정확도도 몇 가지 경우($n=8, 12$)만 제외하면 역시 표본의 크기가 증가함에 따라 커지는 것을 알 수 있다. 그러나 이 Best Cut-Off 지점에서의 정확도는 5% 유의수준 측면에서의 정확도보다 모두 높은 값을 가진다. 이것은 가장 정확도가 높아지는 기각역을 Best Cut-Off 지점으로 정하기 때문이다. 일반적으로 일정한 유의수준에서 Power가 높은 검정법을 구하는 것이 통계적 가설검정에 있어서의 주된 명제이다. 이렇게 구한 검정법의 기각역으로 구한 정확도는 최대가 되기 힘들다. 가령, 오류의 손실 또는 비용이 제1종이나 제2종 오류에서 동일하다면 가장 높은 정확도를 얻을 수 있는 기각역을 구하는 것이 바람직한 일이 된다. 이러한 관점에서 ROC 분석을 통하여 정확도가 가장 높은 Best Cut-Off 기각역을 구할 수 있다.

표 2는 S_{QQ} 통계량에 대한 ROC 분석 결과이고, 표 3은 S_{PP} 통계량에 대한 ROC 분석 결과이다. 여기에서도 앞의 W 통계량과 마찬가지로 5%의 유의수준 측면에서 Power와 정확도는 다소 불안정하지만 표본의 크기가 증가함에 따라 커지는 것으로 나타났으며, Best Cut-Off 측면의 정확도도 역시 표본의 크기가 증가함에 따라 유사하게 커지는 것을 알 수 있다.

세 통계량에 대한 비교를 하자면, 표본의 크기가 약 15정도이하에서는 S_{PP} 가 다른 두 통계량보다 Power나 정확도 면에서 더 좋은 것으로 나타났고, 크기가 20에 접근하면 W 가 다른 두 통계량보다 Power나 정확도 면에서 더 좋은 것으로 나타났다.

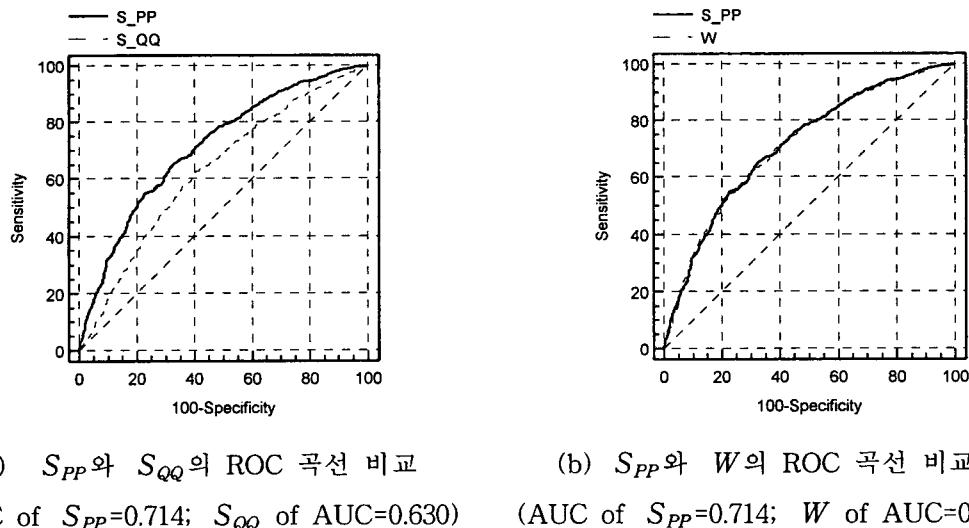


그림 1. 표본의 크기 15에서의 세 통계량의 ROC Curve 분석

그리고 AUC 측면에서 살펴보면, W 와 S_{PP} 는 표본의 크기가 증가할수록 AUC도 따라서 안정적으로 증가하지만 S_{QQ} 는 어느 정도 불안정하게 증가함을 알 수 있다. AUC를 이용한 검정의 결과인 표 4에 나타난 것과 같이 W 와 S_{PP} 가 S_{QQ} 보다는 20이하의 표본크기에서 통계적으로 더 우수한 통계량임을 나타내고 있으며, 표본의 크기가 8에서 14까지는 S_{PP} 가 W 보다 우수하고, 표본의 크기가 5인 경우와 18에서 20까지에서는 W 가 S_{PP} 보다 우수하였으며, 나머지 표본의 크기, $n=6, 7, 15, 16, 17$ 에서는 두 통계량이 비슷한 것으로 나타났다.

대표본에서의 분석을 알아보면, 표 5에 나타난 것과 같이 표본의 크기가 30 이상에서 W 가 다른 두 통계량보다 Power나 정확도, AUC 면에서 더 좋은 것으로 나타났고, 표본의 크기가 30에서 50까지는 S_{PP} 가 S_{QQ} 보다 우수하고, 70이상에서는 S_{QQ} 가 S_{PP} 보다 우수한 것으로 나타났으며, 표본의 크기 60에서는 Power와 정확도에서는 S_{PP} 가 우수하고 AUC에서는 S_{QQ} 가 우수하였다. 그러나 표본의 크기가 100 이상의 경우에는 세 통계량(W , S_{QQ} , S_{PP}) 모두 우수한 것으로 나타났다.

특히, ROC 분석을 실시하면 그림 1에서와 같은 두 검정 통계량의 정확도에 대한 비교가 그림으로도 나타나 쉽게 알 수 있다. 그림 1은 크기가 15인 표본에 대하여 S_{PP} 와 다른 두 통계량의 비교를 실시한 ROC 곡선 그래프이다. (a)에서는 S_{PP} (AUC=0.714)가 S_{QQ} (AUC=0.630)보다 곡선 아래의 면적이 넓어 우수함을 보여주고 있으며(Difference=0.084; Std of Difference=0.006; P값<0.001), (b)에서는 W (AUC=0.711)와 S_{PP} 의 면적이 거의 비슷함을 보여주고 있다(Difference=-0.003; Std of Difference=0.005; P값=0.545).

5. 결 론

정규성 검정에 사용되는 대표적인 그래프를 이용하여 S_{PP} 와 S_{QQ} 의 검정 통계량을 제시하였고, 이것을 균일분포에 대하여 ROC 분석을 실시하고, 그 결과를 통하여 Power와 정확도 측면에서 고찰하였고 AUC를 이용하여 어느 통계량이 우수한가를 검정하였다. 그 결과, 표본의 크기가 약 15 미만에서는 P-P 플롯에서 얻은 S_{PP} 통계량이 Shapiro-Wilk의 검정 통계량 W 와 Q-Q 플롯에서 얻은 S_{QQ} 통계량보다 우수함을 알 수 있었고, 크기가 18이상의 표본에서는 Shapiro-Wilk의 검정 통계량 W 가 S_{PP} 와 S_{QQ} 통계량보다 우수함을 알 수 있었다. 또한 S_{PP} 와 S_{QQ} 통계량의 비교에서는 크기가 50 이하의 표본에서는 S_{PP} 가 우수하고 70 이상의 표본에서는 S_{QQ} 가 우수함을 알 수 있었다. 또한 크기가 100 이상의 표본에서는 어느 통계량이나 우수하였다. 이것을 통하여 정규성 검정을 위한 그래프를 적용할 경우에, 비교되어 고려해야 할 분포가 균일분포라면 크기가 50 이하의 표본에서 P-P 플롯이, 70 이상의 표본에 대하여는 Q-Q 플롯이 더 우수한 정규성 검정의 그래프라는 추론을 할 수 있었다.

표 5. 표본의 크기가 30 이상인 경우의 ROC 분석 결과

Statistic	Sample Size	ROC Area	Best Cut-off				5% Sig.		
			Criteria	Sens. (Power)	Spec.	Accuracy	Criteria	Sens. (Power)	Accuracy
<i>W</i>	30	0.889	0.9558	88.4	75.5	81.95	0.9291	43.8	69.40
	40	0.952	0.9509	87.4	89.8	88.60	0.9412	70.0	82.50
	50	0.981	0.9549	94.8	91.3	93.05	0.9493	88.2	91.60
	60	0.993	0.9519	95.4	96.8	96.10	0.9551	96.7	95.85
	70	0.996	0.9550	98.2	96.6	97.40	0.9575	98.8	96.90
	80	0.999	0.9526	98.4	99.0	98.70	0.9608	99.9	97.45
	90	0.999	0.9575	99.9	98.3	99.10	0.9628	100.0	97.50
	100	0.999	0.9510	99.2	99.7	99.45	0.9638	100.0	97.50
	200	1.000	0.9472	100.0	100.0	100.00	0.9733	100.0	97.50
	300	1.000	0.9459	100.0	100.0	100.00	0.9764	100.0	97.50
<i>S_{QQ}</i>	30	0.775	0.1859	77.4	67.5	72.45	0.2616	17.2	56.10
	40	0.856	0.1795	79.9	77.3	78.60	0.2304	36.2	65.60
	50	0.905	0.1696	86.3	81.1	83.60	0.2120	50.3	72.65
	60	0.943	0.1659	90.4	84.4	87.40	0.1937	66.9	80.95
	70	0.962	0.1668	90.8	90.4	90.60	0.1834	79.5	87.25
	80	0.976	0.1618	94.4	90.9	92.65	0.1776	85.0	90.00
	90	0.983	0.1628	94.3	93.7	94.00	0.1671	91.4	93.20
	100	0.990	0.1588	97.1	95.1	96.10	0.1588	97.1	96.05
	200	1.000	0.1454	100.0	100.0	100.00	0.1177	100.0	97.50
	300	1.000	0.1233	100.0	100.0	100.00	0.0975	100.0	97.50
<i>S_{PP}</i>	30	0.819	0.0456	81.5	70.0	75.75	0.0643	32.9	63.95
	40	0.877	0.0429	81.8	79.2	80.50	0.0534	50.7	72.85
	50	0.913	0.0393	84.9	81.1	83.00	0.0471	61.1	78.05
	60	0.935	0.0375	86.5	84.6	85.55	0.0433	70.6	82.80
	70	0.953	0.0362	89.6	87.8	88.60	0.0407	73.3	84.15
	80	0.965	0.0340	92.4	88.0	90.20	0.0384	79.4	87.20
	90	0.970	0.0343	90.5	91.9	91.20	0.0369	83.0	89.00
	100	0.978	0.0321	94.2	90.7	92.45	0.0351	87.5	91.25
	200	1.000	0.0278	99.5	99.2	99.35	0.0242	100.0	97.50
	300	1.000	0.0258	100.0	99.9	99.95	0.0201	100.0	97.50
	500	1.000	0.0208	100.0	100.0	100.00	0.0152	100.0	97.50

6. 참고문헌

- [1] Altman, D. G. (1992). *Practical statistics for medical research*, London : Chapman and Hall.
- [2] De Wet, T. and Venter, J. H. (1972). Asymptotic distributions of certain test criteria of normality, *South African Statistical Journal*, Vol. 6, 135-149.
- [3] Endrenyi, L. and Patel, M. (1991). A new, sensitive graphical method for detecting deviations from the normal distribution of drug responses: the NTV plot, *Br. J. clin. pharmac.*, Vol. 32, 159-166.
- [4] Fang, J. and Case, K. (1996). Using a correlation test for normality, *J. of Quality Tech.*, Vol. 28, 356-362.
- [5] Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, Vol. 143, 29-36.
- [6] Hanley, J. A. and McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases, *Radiology*, Vol. 148, 839-843.
- [7] Jackson, P. R., Tucker, G. T. and Woods, H. F. (1989). Testing for bimodality in frequency distributions of data suggesting polymorphisms of drug metabolism-histograms and probit plots, *Br. J. clin. pharmac.*, Vol. 28, 647-653.
- [8] Kim, N. (1997). 이변량 정규분포의 적합도 검정을 위한 통계량의 극한분포에 대한 연구, *한국통계학회논문집*, 4권 3호, 863-879.
- [9] Lee, J.-Y. and Rhee, S.-W. (1997). 특정분포에 따른 확률 Plot들의 정규성과 Bimodality 비교, *한국통계학회논문집*, 4권 1호, 243-254.
- [10] Lee, J.-Y., Woo, J. S., and Choi, D. W. (1998). Using a normal test variable (NTV) for clinical research, *응용통계연구*, 11권 1호, In print.
- [11] Lin, C.-C. and Mudholkar, G. (1980). A simple test for normality to against asymmetric alternatives, *Biometrika*, Vol. 67, 455-461.
- [12] Looney, S. W. (1995). How to use tests for univariate normality to assess multivariate normality, *The American Statistician*, Vol. 49, 64-70.
- [13] Metz, C. E. (1978). Basic principles of ROC analysis, *Seminars in Nuclear Medicine*, Vol. 8, 283-298.
- [14] Schoonjans, F., Zalata, A., Depuydt, C. E., and Comhaire, F. H. (1995). MedCalc: a new computer program for medical statistics, *Computer Methods and Programs in Biomedicine*, Vol. 48, 257-262.
- [15] Shapiro, S. S. and Francia, R. S. (1972). An approximation analysis of variance test for normality, *J. Am. Stat. Assoc.*, Vol. 67, 215-216.
- [16] Shapiro, S. S. and Wilk, M. B. (1965). An analysis-of-variance test for normality (complete sample), *Biometrika*, Vol. 52, 591-611.
- [17] Zweig, M. H. and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, Vol. 39, 561-577.