

Firework plot for evaluating the impact of outliers in statistical inference

Sungho Moon^{a,1}

^aDepartment of Data Management, Busan University of Foreign Studies

(Received February 5, 2018; Revised February 8, 2018; Accepted February 8, 2018)

Abstract

Outliers and influential observations often distort many numerical measures for data analysis. Jang and Anderson-Cook (*Quality and Reliability Engineering International*, **30**, 1409–1425, 2014) proposed a graphical firework plot method for exploratory analysis purpose to provide a possible visualization of the trace of the impact of the possible outlying and influential observations on the univariate/bivariate data analysis and regression. They developed 3-D plot as well as pairwise plot for the appropriate measures of interest. We use firework plots as a graphical exploratory data analysis tool to detect outliers and evaluate the impact of outliers in statistical inference.

Keywords: outliers, influential observations, firework plot, three-dimensional firework, firework plot matrix

1. 서론

특이점(이상점이라고도 함) 및 영향점(영향력이 큰 관측값이라고도 함)은 자료분석시 많은 수치적 측도들(measures)을 왜곡한다. 특이점 및 영향점의 평가는 주로 회귀분석에서 다루어졌다 (Fox, 2008; Chatterjee와 Hadi, 2012). 우리는 Hampel (1974)이 제안한 민감도 곡선(sensitivity curve)이나 Cook (1986)이 제안한 영향력 함수(influence function)를 사용하여 이러한 특이점 및 영향점을 탐지할 수 있다. 그러나 이러한 방법은 이론적 전개가 필요한 방법들이다. Jang과 Anderson-Cook (2014)은 불꽃그림이란 이름을 붙인 탐색적 그림도구를 발표하였는데 특이점이나 영향점이 일변량/이변량 자료분석 및 회귀분석에 어떠한 영향을 미치는지 알기 위하여 불꽃그림, 3-D 불꽃그림 및 짝진 불꽃그림 행렬을 제시하였다. 관측값에 부여된 가중치가 1에서 0으로 변화함에 따라 특이점이 일변량/이변량 자료분석시의 여러 수치적 측도들에 어떠한 영향을 미치는지 불꽃그림, 3-D 불꽃그림 및 불꽃그림 행렬을 통하여 살펴보았다. 또한, 특이점이나 영향점이 회귀계수 및 잔차제곱합(residual sum of squares; SSE)에 어떠한 영향을 미치는지 3차원 그림에 추적곡선을 그려 보았을 뿐만 아니라 쌍으로 대비시켜 봄으로써 분석의 시각적인 효과를 증대시켰다. Jang과 Anderson-Cook (2015, 2017)은 일반화선형모형이나 능형회귀모형하에서 특이점이나 영향점이 회귀계수 및 잔차제곱합에 어떠한 영향을 미치는지 불꽃그림, 3-D 불꽃그림 및 불꽃그림 행렬을 통하여 살펴보았다. Moon (2016)은 불꽃그림이 일변량 자료의 왜도

This work was supported by the research grant of Busan University of Foreign Studies in 2017.

¹Department of Data Management, Busan University of Foreign Studies, Pusan 46234, Korea.

E-mail: shmoon@bufs.ac.kr

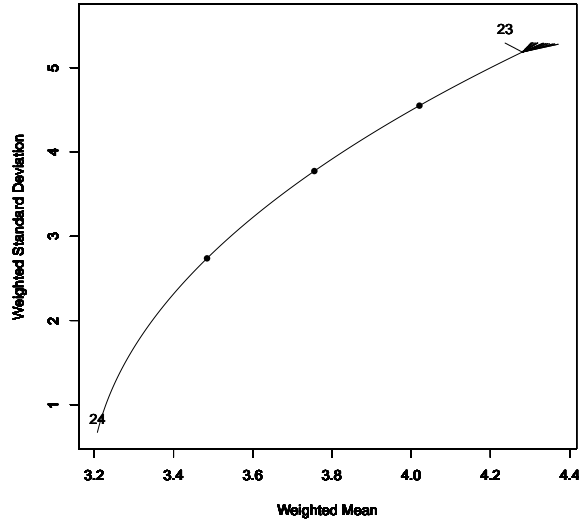


Figure 2.1. Mean-standard deviation firework plot for copper data.

와 침도에서 특이점의 영향을 평가하기 위한 탐색적 자료분석 그림도구로 사용될 수 있음을 보였다. 본 연구에서는 이러한 불꽃그림이 통계적 추론에서 특이점의 영향을 평가하기 위한 탐색적 자료분석 그림 도구로 사용될 수 있음을 보이고자 한다. 제 2절에서는 불꽃그림의 작성방법을 평균-표준편차 불꽃그림을 중심으로 서술하고 제 3절에서는 일변량 자료에 대한 통계적 추론에서 특이점의 영향을 평가하기 위한 탐색적 자료분석 그림도구로서의 불꽃그림을 제시하고 실제 사례를 통하여 불꽃그림의 유용성을 살펴보았다. 제 4절에서 결론으로 마무리하였다.

2. 불꽃그림의 작성 방법

일변량 데이터 x_1, x_2, \dots, x_n 이 주어졌을 때 이 데이터에 대한 가중산술평균과 가중표준편차는 다음과 같이 정의된다.

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \quad s_w = \frac{\sum_{i=1}^n w_i (x_i - \bar{x}_w)^2}{\sum_{i=1}^n w_i}, \quad (2.1)$$

여기서 w_i ($i = 1, 2, \dots, n$)는 x_i 에 대응되는 가중값이다.

i 번째 데이터에 부여된 가중치의 값 w_i 을 1에서 0으로 연속적으로 변화하게 하고, 나머지 가중치인 경우 w_j ($j \neq i, j = 1, 2, \dots, n$)로 고정된 다음 가중산술평균과 가중표준편차를 구한 후 이 값들을 연결하여 곡선을 구한다. 이러한 작업을 모든 데이터에 대하여 적용하면 평균-표준편차 불꽃그림이 그려진다. 이 불꽃그림을 통하여 산술평균 및 표준편차 입장에서 어떤 데이터가 특이점인지를 탐색적으로 알아낼 수 있다.

다음 자료는 통밀 속에 있는 구리의 양(단위: parts per million; PPM)을 나타내는 24개 자료 (Maronna 등, 2006)이다.

2.20, 2.20, 2.40, 2.40, 2.50, 2.70, 2.80, 2.90, 3.03, 3.03, 3.10, 3.37, 3.40, 3.40, 3.40, 3.50, 3.60, 3.70, 3.70, 3.70, 3.70, 3.77, 5.28, 28.95.

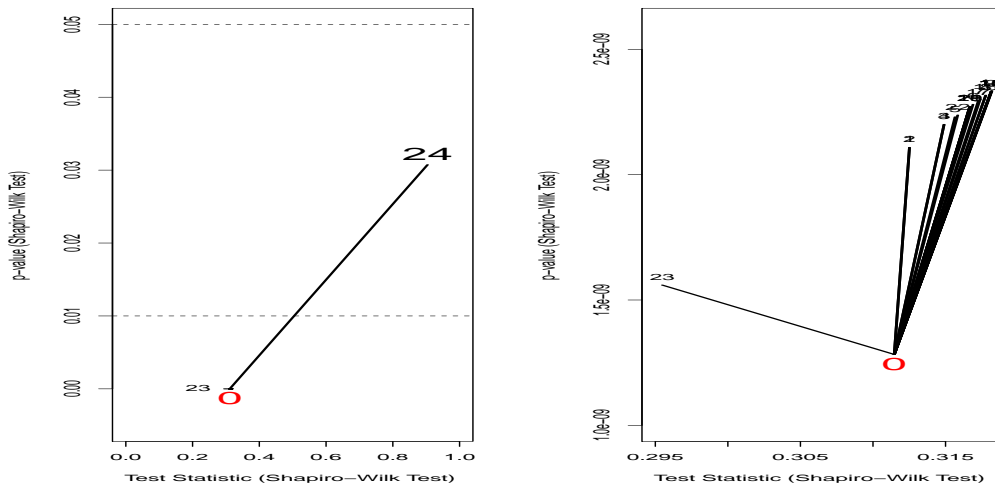


Figure 3.1. Normality test firework plot for copper data.

이 자료를 대상으로 평균-표준편차 불꽃그림을 그리면 Figure 2.1과 같다. 24번째 관측값에 대응되는 곡선이 다른 관측값들에 대응되는 곡선과는 반대방향으로 매우 길게 뻗어 있음을 알 수 있다. 24번째 관측값이 특이점임을 알 수 있다. 이 곡선의 길이는 민감도곡선과 관련이 있다 (Jang과 Anderson-Cook, 2014). 24번째 관측값에 대응하는 곡선에서 3개의 점은, 24번째 관측값에 대한 가중치를 0.25, 0.50, 0.75로 부여하고 다른 관측값들에 대한 가중치는 1로 두어 구한 가중산술평균값과 가중표준편차값을 가리킨다.

3. 통계적 추론에서 특이점의 영향을 평가하기 위한 탐색적 자료분석 그림도구로서의 불꽃그림

통계적 추론에서 특이점의 영향을 평가하기 위한 탐색적 자료분석 그림도구로서의 불꽃그림을 제시하기 위하여 일표본문제(특정한 값과 하나의 모평균의 비교)와 이표본문제(두 모집단의 비교)로 나누어 살펴보고자 한다.

3.1. 일표본문제

특이점은 자료분석상의 많은 수치적 척도들, 예를 들면 산술평균, 분산, 표준편차, 왜도, 첨도, 공분산, 상관계수 등을 왜곡시킨다. 또한 통계적 추론에서도 많은 왜곡이 일어나는 데 특이점의 존재 여부에 따라 신뢰구간이나 검정의 결과가 다르게 나타나게 된다. 그러므로 통계적 추론에서 이러한 특이점이 존재하는지의 여부와 이러한 특이점이 신뢰구간이나 검정에 어떤 영향을 주는지를 살펴보아야한다.

Figure 3.1은 구리의 양 자료에서 각 관측값 제거 전과 후의 Shapiro-Wilk 정규성검정에 대한 불꽃그림이다. 각 관측값 제거 전과 후의 Shapiro-Wilk 정규성검정시 계산되는 검정통계량 값과 그에 대응하는 p -값을 구한 후 직선으로 연결하여 그림을 작성한다. Figure 3.1에서 '0'은 모든 관측값을 다 사용한다는 의미이다. 왼쪽 불꽃그림은 각 관측값 제거 전과 후의 Shapiro-Wilk 정규성검정에 대한 불꽃그림이고 오른쪽 불꽃그림은 1-23번째 관측값 제거 전과 후의 Shapiro-Wilk 정규성검정에 대한 불꽃그림이다. 1-23번째 관측값 제거시에는 모든 경우 24번째 관측값이 자료에 포함되어 있어 p -값이 0.01보다 작

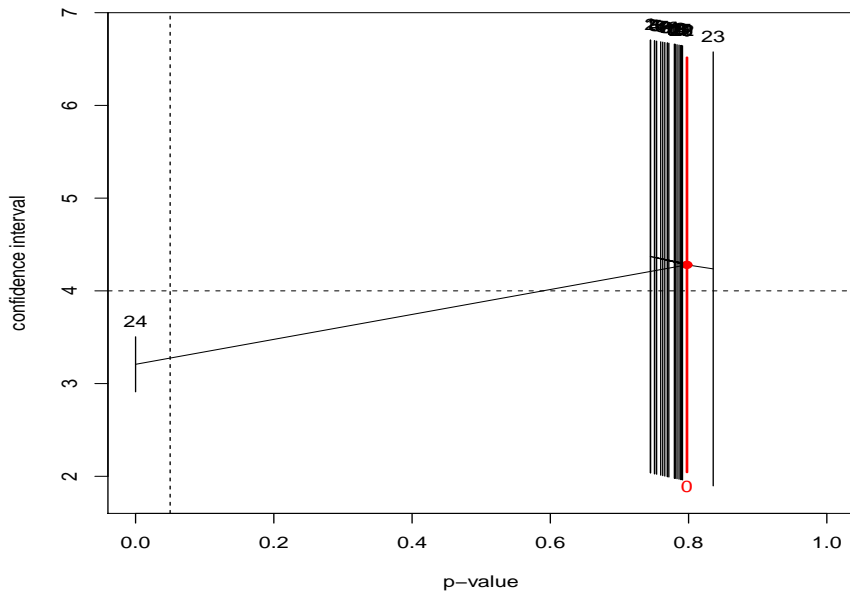


Figure 3.2. Statistical inference firework plot for copper data ($H_0 : \mu = 4.0$, $H_1 : \mu \neq 4.0$).

이 자료가 정규분포를 따른다고 볼 수 없으나 24번째 관측값 제거시에는 p -값이 0.01보다 커져 유의수준 $\alpha = 0.01$ 에서 자료가 정규분포를 따른다고 볼 수 있다. 24번째 관측값에 대응되는 직선이 다른 관측값들에 대응되는 직선과는 반대방향으로 매우 길게 뻗어 있음을 알 수 있다. 24번째 관측값이 특이점임을 알 수 있고 특이점인 24번째 관측값의 영향을 확실하게 확인할 수 있다.

Figure 3.2는 구리의 양 자료에서 각 관측값 제거 전과 후의 95% 신뢰구간과 양측검정 $H_0 : \mu = 4.0$, $H_1 : \mu \neq 4.0$ 에 대응하는 p -값을 나타내는 불꽃그림이다. Figure 3.2에서 각 관측값 제거 전과 후의 95% 신뢰구간의 하한값과 상한값을 두개의 y -값으로 하고 p -값을 x -값으로 하여 직선으로 연결하고 신뢰구간의 중앙에 위치하는 산술평균을 y -값으로 하고 p -값을 x -값으로 하여 직선으로 연결한다. Figure 3.2에서 '0'은 모든 관측값을 다 사용한다는 의미이다. 1-23번째 관측값 제거시에는 신뢰구간이 모두 4.0을 포함하나 24번째 관측값 제거시에는 신뢰구간이 4.0을 포함하지 않게 된다. 또한 1-23번째 관측값 제거시에는 모든 경우 24번째 관측값이 자료에 포함되어 있어 신뢰구간의 폭이 매우 넓고 신뢰구간의 한가운데 값인 산술평균이 4.0보다 크나 24번째 관측값 제거시에는 신뢰구간의 폭이 매우 좁게 되고 신뢰구간의 한가운데 값인 산술평균이 4.0보다 작게 된다. 24번째 관측값에 대응되는 직선이 다른 관측값들에 대응되는 직선과는 반대방향으로 매우 길게 뻗어 있음을 알 수 있다. 24번째 관측값이 특이점임을 알 수 있고 특이점인 24번째 관측값의 영향을 확실하게 확인할 수 있다.

미국 국립공원인 Yellow Stone Park 내에 있는 간헐온천(old faithful geyser)에서 관심사항은 온천물이 나오는 지속시간(duration)과 온천물이 분출한 후 다음 온천물이 쏟아질 때까지의 간격시간(interval)이다. 다음 자료는 온천물의 지속시간을 분 단위로 켤 자료(107개)를 나타낸다.

4.37, 3.87, 4.00, 4.03, 3.50, 4.08, 2.25, 4.70, 1.73, 4.93, 1.73, 4.62, 3.43, 4.25, 1.68, 3.92, 3.68, 3.10, 4.03, 1.77, 4.08, 1.75, 3.20, 1.85, 4.62, 1.97, 4.50, 3.92, 4.35, 2.33, 3.83, 1.88, 4.60, 1.80, 4.73, 1.77, 4.57, 1.85, 3.52, 4.00, 3.70, 3.72, 4.25, 3.58, 3.80, 3.77, 3.75, 2.50, 4.50, 4.10, 3.70, 3.80, 3.43, 4.00,

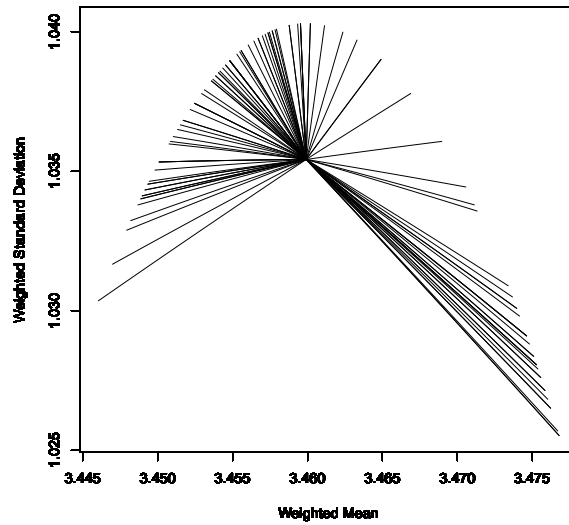


Figure 3.3. Mean-standard deviation firework plot for eruption length data.

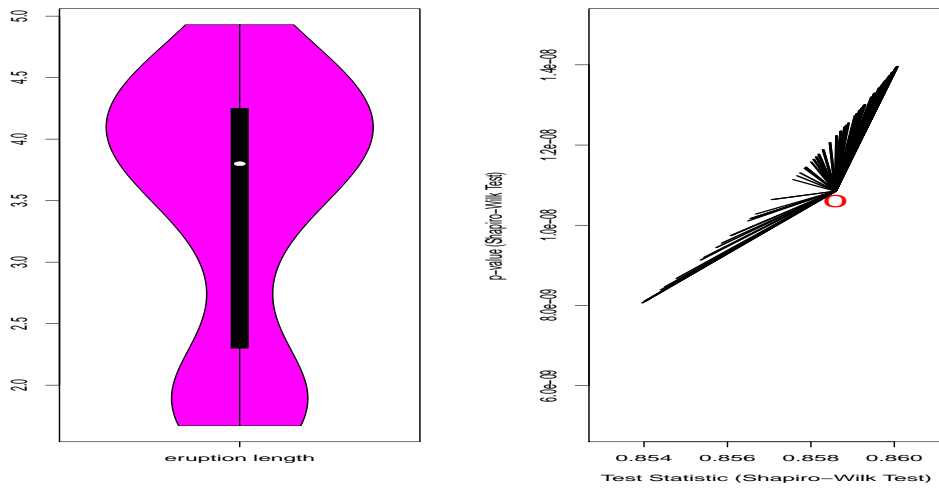


Figure 3.4. Violin plot and normality test firework plot for eruption length data.

2.27, 4.40, 4.05, 4.25, 3.33, 2.00, 4.33, 2.93, 4.58, 1.90, 3.58, 3.73, 3.73, 1.82, 4.63, 3.50, 4.00, 3.67, 1.67, 4.60, 1.67, 4.00, 1.80, 4.42, 1.90, 4.63, 2.93, 3.50, 1.97, 4.28, 1.83, 4.13, 1.83, 4.65, 4.20, 3.93, 4.33, 1.83, 4.53, 2.03, 4.18, 4.43, 4.07, 4.13, 3.95, 4.10, 2.72, 4.58, 1.90, 4.50, 1.95, 4.83, 4.12

이 자료를 대상으로 평균-표준편차 불꽃그림을 그리면 Figure 3.3과 같다. 어떤 곡선도 특이한 차이가 없으므로 특이점이 없음을 알 수 있다.

Figure 3.4는 온천물의 지속시간 자료 107개 데이터 모두를 사용하여 그린 바이올린그림 및 각 관측값 제거 전과 후의 Shapiro-Wilk 정규성검정에 대한 불꽃그림이다. Figure 3.4에서 '0'은 모든 관측값을 다 사용한다는 의미이다. 왼쪽 바이올린그림은 107개 데이터 모두를 사용하여 그린 바이올린그림이다.

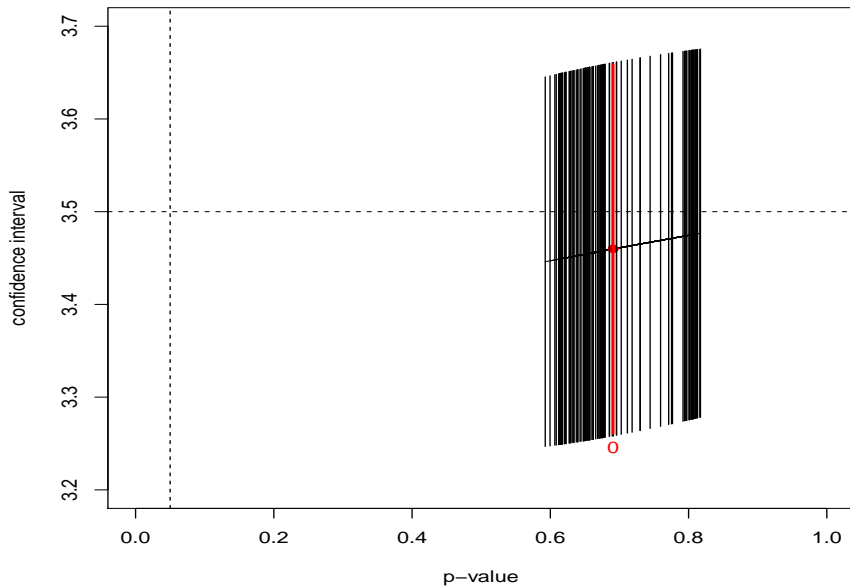


Figure 3.5. Statistical inference firework plot for eruption length data ($H_0 : \mu = 3.5$, $H_1 : \mu \neq 3.5$).

쌍봉(bimodal) 분포임을 알 수 있다. 오른쪽 불꽃그림은 1-107번째 관측값 제거 전과 후의 Shapiro-Wilk 정규성검정에 대한 불꽃그림이다. 1-107번째 관측값 제거시 모든 경우 유의수준 $\alpha = 0.05$ 에서 정규분포를 따른다고 볼 수 없다. 원 자료가 쌍봉분포이다 보니 어떤 관측값을 제거하더라도 검정통계량 값이나 p -값에서 변화가 거의 없음을 알 수 있다. 특이점이 없음을 알 수 있다.

Figure 3.5는 온천물의 지속시간 자료 각 관측값 제거 전과 후의 95% 신뢰구간과 양측검정 $H_0 : \mu = 3.5$, $H_1 : \mu \neq 3.5$ 에 대응하는 p -값을 나타내는 불꽃그림이다. Figure 3.5에서 '0'은 모든 관측값을 다 사용한다는 의미이다. 1-107번째 관측값 제거시 신뢰구간이 모두 3.5를 포함하는 것을 알 수 있고 모든 관측값에 대응되는 직선에 차이가 큰 차이가 없다. 특이점이 없음을 확인할 수 있다.

3.2. 이표본문제

특이점은 두 모집단의 비교에 대한 통계적 추론에서도 많은 왜곡이 일어나는 데 특이점의 존재 여부에 따라 신뢰구간이나 검정의 결과가 다르게 나타나게 된다. 그러므로 통계적 추론에서 이러한 특이점이 존재하는지의 여부와 이러한 특이점이 신뢰구간이나 검정에 어떤 영향을 주는지를 살펴보아야한다.

다음 Table 3.1은 Montgomery (2012)에 나타나는 각 12마리 실험용 쥐의 신경조직과 근육조직에 대한 2시간 후의 표준화 형광성을 나타내는 자료이다.

이 자료를 대상으로 신경조직과 근육조직에 대응되는 평균-표준편차 불꽃그림들을 같이 그리면 Figure 3.6과 같다. Figure 3.6에서 'N'은 신경조직, 'M'은 근육조직을 의미한다. 산술평균과 표준편차에 입장에서 10, 11 번째 관측값들에 대응하는 곡선이 다른 곡선과는 다르게 나타나므로 10, 11번째 관측값들(신경조직에서 작은 값 2개)이 특이점임을 알 수 있다.

Figure 3.7은 신경조직과 근육조직 각각에 대하여 각 관측값 제거 전과 후의 Shapiro-Wilk 정규성검정에 대한 불꽃그림이다. 왼쪽 그림은 신경조직에 대하여 각 관측값 제거 전과 후의 Shapiro-Wilk 정규

Table 3.1. The mean normalized fluorescence for nerve and muscle tissue

Observation no.	Tissue	The mean normalized fluorescence
1	Nerve	6625
2	Nerve	6000
3	Nerve	5450
4	Nerve	5200
5	Nerve	5175
6	Nerve	4900
7	Nerve	4750
8	Nerve	4500
9	Nerve	3985
10	Nerve	900
11	Nerve	450
12	Nerve	2800
13	Muscle	3900
14	Muscle	3500
15	Muscle	3450
16	Muscle	3200
17	Muscle	2980
18	Muscle	2800
19	Muscle	2500
20	Muscle	2400
21	Muscle	2200
22	Muscle	1200
23	Muscle	1150
24	Muscle	1130

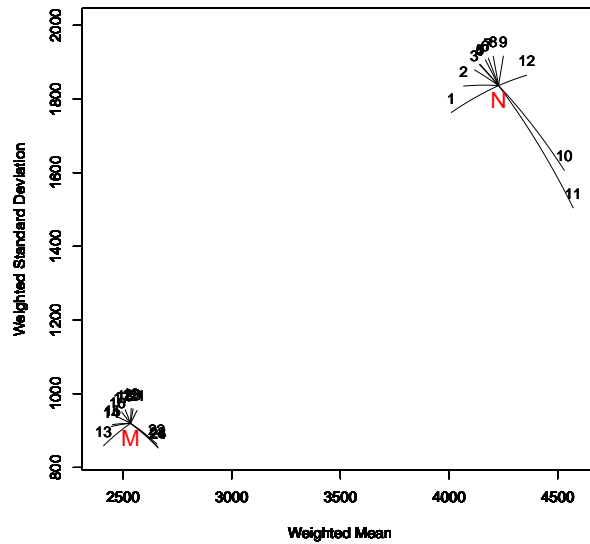


Figure 3.6. Mean-standard deviation firework plot for fluorescence data.

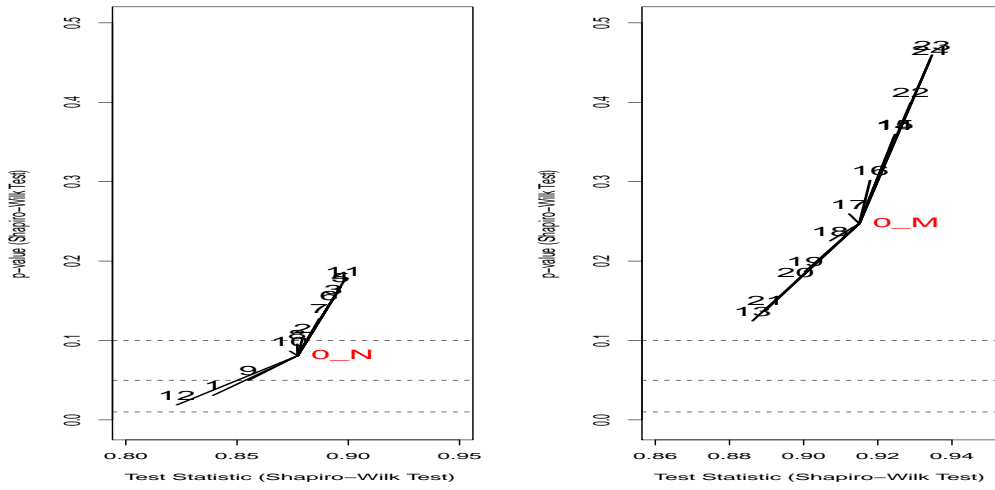


Figure 3.7. Normality test firework plot for fluorescence data.

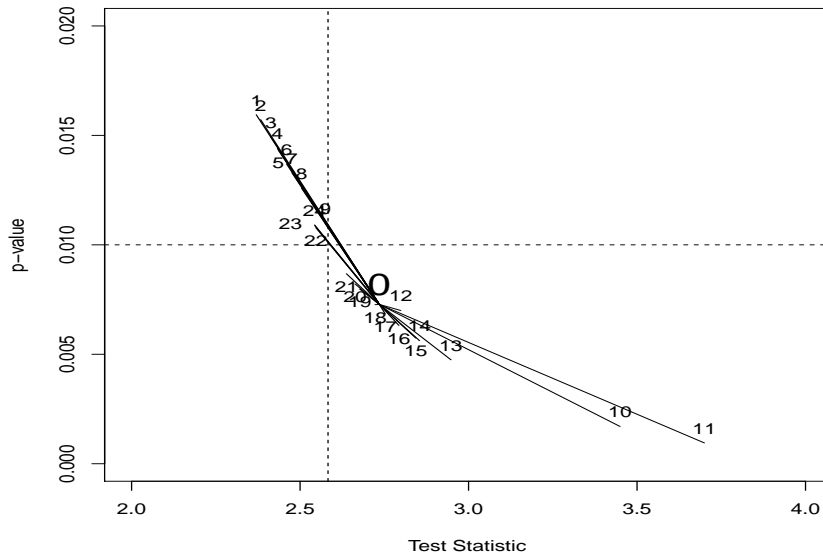


Figure 3.8. Statistical inference firework plot for fluorescence data ($H_0 : \mu_1 \leq \mu_2, H_1 : \mu_1 > \mu_2$).

성검정에 대한 불꽃그림이고 오른쪽 그림은 근육조직에 대하여 각 관측값 제거 전과 후의 Shapiro-Wilk 정규성검정에 대한 불꽃그림이다. 신경조직에서 1번째 관측값(신경조직에서 큰 값 10개 중 제일 큰 값)이나 12번째 관측값(신경조직에서 큰 값 10개 중 제일 작은 값) 제거시에는 p -값이 0.01과 0.05 사이에 놓이게 되고 1번째 관측값이나 12번째 관측값에 대응되는 직선이 다른 관측값 제거시 대응되는 직선과는 다른 방향에 위치하므로 정규성검정 입장에서 신경조직에서는 1번째 관측값과 12번째 관측값이 특이점을 알 수 있다. 산술평균과 표준편차에 입장에서 10, 11번째 관측값들이 특이점인 것과 대조가 된다. 근육조직에서는 어떤 관측값을 제거하여도 p -값이 0.1 보다 큼을 알 수 있다. 정규성검정 입장에서 근육조직에서는 특이점이 없음을 알 수 있다.

Figure 3.8은 표준화 형광성 자료에서 각 관측값 제거 전과 후의 양측검정 $H_0 : \mu_1 \leq \mu_2$, $H_1 : \mu_1 > \mu_2$ 에 대응하는 검정통계량 값과 p -값을 나타내는 불꽃그림이다. 각 관측값 제거 전과 후의 검정통계량 값과 p -값을 연결하여 직선을 구한다. Figure 3.8에서 '0'은 모든 관측값을 다 사용한다는 의미이다. 모든 자료를 다 사용하는 경우는 p -값이 0.01보다 작으나 1-9번째 관측값(신경조직에서 큰 값 9개)이 제거되거나 22-24번째 관측값(근육조직에서 작은 값 3개)이 제거되면 p -값이 0.01와 0.05 사이에 놓이게 된다. 평균-표준편차 불꽃그림이나 정규성검정에 대한 불꽃그림에서 특이점을 찾은 결과하고는 다른 양상을 나타낸다.

4. 결론

일변량 자료의 통계적 추론에서 불꽃그림을 통하여 특이점을 찾아내는 방법을 제시하고 이러한 특이점이 통계적 추론에 어떤 영향을 주는지를 불꽃그림을 통하여 살펴보았다. 실제 자료를 통하여 불꽃그림이 일변량 자료의 통계적 추론에서 특이점의 영향을 평가하기 위한 탐색적 자료분석 그림도구로 사용될 수 있음을 살펴보았다. 이 불꽃그림을 사용하여 다양한 자료를 대상으로 탐색적 방법으로 특이점을 찾아낼 수 있다. 또한 이러한 특이점이 통계적 추론에 어떤 영향을 주는지를 평가할 수 있는 탐색적 자료 분석 그림도구로서 불꽃그림이 유용하게 쓰일 수 있다.

부록 A: R 코드

A1. Figure 3.2를 구현하기 위한 R 코드

```
# 데이터의 입력
copper=c(2.20, 2.20, 2.40, 2.40, 2.50, 2.70, 2.80, 2.90, 3.03, 3.03, 3.10, 3.37,
3.40, 3.40, 3.40, 3.50, 3.60, 3.70, 3.70, 3.70, 3.70, 3.77, 5.28, 28.95)
# 데이터의 개수
n=length(copper)
mat1=c(rep(0,(n+1)*5))
wt=matrix(mat1,nrow=(n+1))
copper.test=t.test(copper,mu=4)
wt[1,1]=copper.test$p.value
wt[1,2]=copper.test$conf.int[1]
wt[1,3]=copper.test$conf.int[2]
wt[1,4]=copper.test$statistic
wt[1,5]=copper.test$estimate
for(i in 1:n)
{
copper.test=t.test(copper[-i],mu=4)
wt[i+1,1]=copper.test$p.value
wt[i+1,2]=copper.test$conf.int[1]
```

```

wt[i+1,3]=copper.test$conf.int[2]
wt[i+1,4]=copper.test$statistic
wt[i+1,5]=copper.test$estimate
}
# Figure 3.2
plot(1,2,type="n",xlim=c(0,1),ylim=c(1.8,6.8),xlab="p-value",ylab="confidence interval",main="")
for(i in 1:n)
{
arrows(wt[1,1],wt[1,5],wt[i+1,1],wt[i+1,5],length=0)
}
points(wt[1,1],wt[1,5],pch=19,col="red")
for(i in 1:n)
{
arrows(wt[i+1,1],wt[i+1,2],wt[i+1,1],wt[i+1,3],length=0)
text(wt[i+1,1],wt[i+1,3],i,pos=3)
}
arrows(wt[1,1],wt[1,2],wt[1,1],wt[1,3],length=0,col="red",lwd=2)
text(wt[1,1],wt[1,2],0,pos=1,col="red")
abline(v=0.05,lty=2)
abline(h=4,lty=2)

```

References

- Chatterjee, S. and Hadi, A. S. (2012). *Regression Analysis by Example* (5th ed), Wiley, Hoboken.
- Cook, R. D. (1986). Assessment of local influence, *Journal of Royal Statistical Society, Series B*, **48**, 133–169.
- Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models* (2nd ed), Sage, New York.
- Hampel, F. R. (1974). The influence curve and its role in robustness, *The Annals of Statistics*, **45**, 383–393.
- Jang, D. H. and Anderson-Cook C. M. (2014). Firework plot as a graphical exploratory data analysis tool for evaluating the impact of outliers in data exploration and regression, *Quality and Reliability Engineering International*, **30**, 1409–1425.
- Jang, D. H. and Anderson-Cook C. M. (2015). Firework plots for evaluating the impact of outliers and influential observations in generalized linear models, *Quality Technology & Quantitative Management*, **12**, 419–432.
- Jang, D. H. and Anderson-Cook, C. M. (2017). Firework plots for evaluating the impact of outliers and influential observations in ridge regression, *Quality and Reliability Engineering International*, **33**, 709–725.
- Maronna, R. A., Martin, D., and Yohai, V. J. (2006). *Robust Statistics*, John Wiley & Sons, New York.
- Montgomery, D. C. (2012). *Design and Analysis of Experiments* (8th ed), John Wiley & Sons, New York.
- Moon, S. (2016). Firework plots for evaluating the impact of outliers and influential observations in ridge regression, *The Korean Journal of Applied Statistics*, **29**, 355–368.

통계적 추론에서 특이점의 영향을 평가하기 위한 탐색적 자료분석 그림도구로서의 불꽃그림

문승호^{a,1}

^a부산외국어대학교 데이터경영학과

(2018년 2월 5일 접수, 2018년 2월 8일 수정, 2018년 2월 8일 채택)

요약

특이점 및 영향점은 자료분석을 하는 데 사용되는 수치적 측도들을 왜곡한다. 각종 자료분석에 있어서의 특이점 검색을 위한 검정 통계량이나 그림도구에 관한 연구는 꾸준히 전개되어 왔다. Jang과 Anderson-Cook (*Quality and Reliability Engineering International*, **30**, 1409–1425, 2014)은 불꽃그림이란 이름을 붙인 그림도구를 발표하였는데 특이점이나 영향점이 일변량/이변량 자료분석 및 회귀분석에 어떠한 영향을 미치는지 알기 위하여 3-D 불꽃그림 및 불꽃그림 행렬을 제시하였다. 본 연구에서는 이러한 불꽃그림이 통계적 추론에서 특이점을 찾아내고 이러한 특이점의 영향을 평가하기 위한 탐색적 자료분석 그림도구로서 사용될 수 있음을 보였다.

주요용어: 특이점, 영향점, 불꽃그림, 삼차원 불꽃그림, 짝진 불꽃그림행렬

이 논문은 2017학년도 부산외국어대학교 학술연구조성비에 의해 연구되었음.

¹(46234) 부산광역시 금정구 금샘로 485번길, 부산외국어대학교 데이터경영학과. E-mail: shmoon@bufs.ac.kr