

확률분포를 이용한 포털들의 응답시간 품질에 관한 연구

류귀열 *

* 서경대학교 컴퓨터과학과

A Study on Quality of Portals Based on Probability Distributions of Response Time

Ryu Gui-Yeol *

* Department of Computer Science, Seokyeong University

ABSTRACT

Purpose: The purpose of this paper is estimate response quality of three major portal in Korea based on the response time. In addition to response time, the response time by 1Kbyte will be analysed.

Methods: Data was collected from July 2010 to November 2013 using Firebug. For comparing averages, ANOVA will be used. For comparing distributions, Chisquare test and Kolmogov-Smirnov test will be used for parametric and non parametric test respectively.

Results: For response quality based on response time, Daum gets the first place, Naver the second place, and Nate the third place. But the order of the response time per 1Kbyte is different. The order is Naver, Daum and Nate.

Conclusion: The response quality may be estimated using various factors. Response time is the most important factor. Daum provides the shortest response time. We could say Daum provides the best response quality. But Naver provides the shortest response time per 1Kbyte. From these results, we know reducing packets is very important thing in response time.

Key Words: Analysis of Variance, Chisquare Test, Kolmogov-Smirnov Test, Portal, Response Time

• Received 11 December 2013, 1st revised 3 January 2014, 2nd revised 7 January 2014, accepted 10 January 2014

† Corresponding Author(gryu@skuniv.ac.kr)

© 2013, The Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

* 본 연구는 2012년도 서경대학교 교내연구비 지원에 의하여 이루어졌음

1. 서론

우리나라의 인터넷은 1982년 시작되었다. Korea Internet & Security Agency (2013)에 의하면 2001년 다운로드 속도가 초당 1 Mbps (Mbyte per second)를 넘는 초고속망 네트워크에서 세계 1위를 차지하였고, 인터넷 이용자도 2,000만 명을 돌파하였다. 2012년 현재 인터넷이용자 수는 3,812만 여명 이용률은 78.4%를 보이고 있다. 인터넷 네트워크 분야에서도 초고속 인터넷 네트워크인 VDSL (very high-bit rate digital subscriber line), FTTH (fiber to the home), LAN (local area network) 등을 적극적으로 개발하여 도입함으로써 세계에서 유례를 찾아볼 수 없을 정도로 빠르게 성장했다.

인터넷에서 사용되는 HTTP (hypertext transfer protocol) 프로토콜은 FTP (file transfer protocol)에서의 파일 전송의 비효율성을 줄이기 위해 제안된 것으로 지금은 널리 사용되고 있는 프로토콜이다. Touch et al. (1998)은 HTTP에서 데이터 수신 구조를 설명하였다. 웹서비스의 응답시간은 서버의 성능과 웹페이지의 성능, 네트워크의 성능, 서버의 위치, 이용자의 컴퓨팅 성능 등에 영향을 받으므로 매우 복잡한 구조에 의해 결정된다. 따라서 응답시간 측정은 인터넷 서비스 사업자나 네트워크 사업자 등 개별 사업자에 의해 수행되기 쉽지 않다. 응답시간에 관한 연구는 주로 서버의 성능에 관한 연구로 수행되었고, Chiew (2009)는 웹페이지의 성능을 연구하였다. 그러나 이용자 관점에서의 응답시간에 관한 연구는 미미한 실정이다.

Hoxmier and DiCesare (2000)는 웹페이지의 응답시간은 사용자 만족도를 측정하는 가장 솔직한 방법이라고 주장하였다. Kim (2012)은 “웹서비스는 빠를수록 좋다. 비즈니스 측면에서 보면 시간은 바로 금이다. 웹서비스가 1초 느려지면 매출은 3% 떨어진다는 마이크로소프트(Bing)의 통계도 있다. 웹서비스는 0.1초라도 빠를수록 좋다. 수천만 이용자가 모이면 이러한 시간이 모여 이용자에게도 사업자에게도 낭비가 된다.”라고 주장하였다. 응답시간에 대한 사용자들의 심리를 분석한 Seow (2008)는 2초가 넘어가면 시스템이 멈추지 않았으므로 주기적으로 알려 줄 필요가 있고, 5초가 넘어가면 반드시 피드백을 주어야 하고, 7초 이상이면 사용자들이 떠난다고 주장하였다. 상세 내용은 Table 1에 나와 있다.

Table 1. Seow's category for response time

Category	Response time	Guidelines for software design
Instantaneous	0.2 Second	Response time for input button: 5~100ms Time for displaying menu: 200ms
Immediate	0.5~1 Second	Time for turning page Expected time for page up or down
Continuous	2~5 Second	Need alert information of system periodically Must give feedback if response time is over 5 second
Captive	7~10 Second	Users leave the web pages

포털 서비스는 이용자들이 브라우저를 통해 서버에 서비스를 요청하면, 서버는 HTTP 프로토콜을 이용하여 포털의 정보를 패킷화하여 인터넷 네트워크를 통해 이용자들에게 전달하는 방식으로 이루어지고 있다. 이용자들은 복잡한 응답구조를 이해하지 않고, 서비스가 얼마나 빨리 제공되는가에 대해서만 평가하는 경향이 있다.

응답품질 평가에는 응답시간, 콘텐츠, 디자인 등 여러 가지 요인들이 있으나, 응답시간이외의 요인들은 주관적이며 계량화하기 어려운 점이 있다. 따라서 응답시간은 응답품질을 평가하는 객관적이고 매우 중요한 요소이다. 본 논문은 포털 들의 응답품질을 응답시간으로 평가할 것이며, 연구 대상 사이트는 텍스트 기반의 단순 소개 페이지를 제

공하는 Google을 제외하고, 방문자 수 기준으로 상위 3대 포털 사이트인 Naver와 Daum, Nate로 선정하였다. 본 논문의 결과는 포털 사이트들이 응답품질을 높이기 위해 응답시간을 줄이는 기술개발을 촉진시킬 수 있을 것이다.

2. 연구 설계 및 결과

응답시간을 측정하기 위해 사용된 도구는 이용자 환경을 유지하고, 객관성을 높이기 위해 상용 브라우저에서 제공하는 도구를 사용하였다. 네스케이프 사에서 개발한 Firefox는 네트워크 관리 툴인 Firebug를 제공하며, Firebug는 응답시간을 측정한다. 또한 연구의 일관성을 유지하기 위해 서경대학교 연구실에서만 응답시간을 측정하였다. Figure 1은 Firebug가 응답시간을 측정하는 화면이다. 이 방법은 Ryu (2012, 2013a)에서도 사용하였다. 본 연구는 포털들의 매출에도 영향을 줄 수 있는 중요한 문제이기 때문에 긴 기간 동안 많은 실험을 통해 객관성을 높이는 것을 요구한다. 따라서 우리는 확률분포를 이용해 비교연구할 것이며, 이를 위해 실험기간은 2010년 7월 6일에서 2013년 11월 11일까지 3년 4개월 실시하였으며, 실험횟수는 2,430회 실시하였다.

응답시간을 비교하기 위해 평균을 이용한 비교와 분포를 이용한 비교가 바람직하다. 평균의 동일성 검정을 위해서 Ree (2013), Kim and Kim(2013), Han et al. (2012), Han and Kang (2012)과 Kim et al. (2004) 등에서 설명되어 있는 바와 같이 분산분석을 사용할 것이다. 분포를 비교하는 방법에는 모수적 방법과 비모수적 방법이 있다. 모수적 방법은 Press et al. (2002)에 설명되어 있는 것과 같이, 카이스퀘어 검정을 실시할 것이다. 비모수적 검정은 Kim and Oh (2003), Hong and Kim (2003) 등에서 설명되어 있는 것과 같이 2 표본 콜모고로프-스미르노프 (Kolmogov-Smirnov) 검정을 사용할 것이다.

실험 자료를 확률분포를 표현하기 위해 수평축에는 0초와 20초 사이를 0.1초 간격으로 구분하였으며 20초 이상의 범주의 자료는 넓게 퍼져 있지만 빈도가 매우 낮아 분포로 표현하기 어렵기 때문이다. 수직축은 세 분포의 통일성을 위해 0%~8%로 표현하였다. Naver와 Daum, Nate의 응답시간의 분포가 Figure 2에 나타나 있다. Naver의 응답시간의 분포는 봉우리가 두 개 있는 쌍봉 형태로 나타났으며, 이는 Ryu (2013b)에 나타난 응답패킷의 분포와 유사하다. 쌍봉 형태는 응답시간이 짧은 분포와 긴 분포가 혼재되어 있음을 보여 준다. Daum과 Nate의 응답시간의 분포는 오른쪽 꼬리가 긴 형태를 보이고 있다.

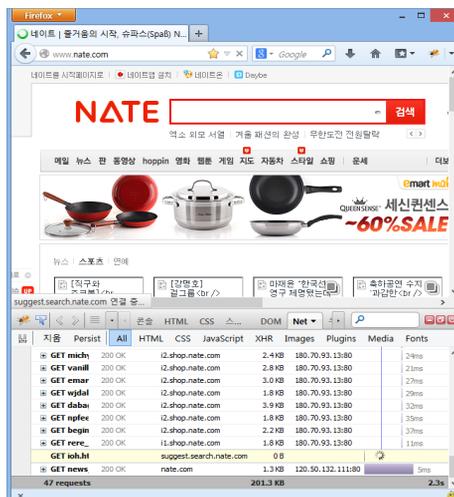


Figure 1. Response time using firebug

응답시간의 평균은 Naver가 3.74초, Daum 2.90초, Nate 3.94초로 Daum이 가장 빨랐으며 Nate가 가장 느렸다. 중앙값으로는 Naver가 3.40초, Daum 2.43초, Nate 3.18초로 Daum이 가장 빨랐으며 Naver가 가장 느렸다. 평균의 순서와 중앙값의 순서가 다른 이유는 Nate의 응답시간 분포는 오른쪽 꼬리가 긴 분포이나, Naver의 응답시간 분포는 쌍봉 형태이기 때문이라 추정된다. 응답시간의 표준편차는 Naver가 2.49초, Daum 2.36초, Nate 3.06초로 Daum이 가장 작으며 Nate가 가장 크게 나타났다. 결과는 Table 2에 나와 있다.

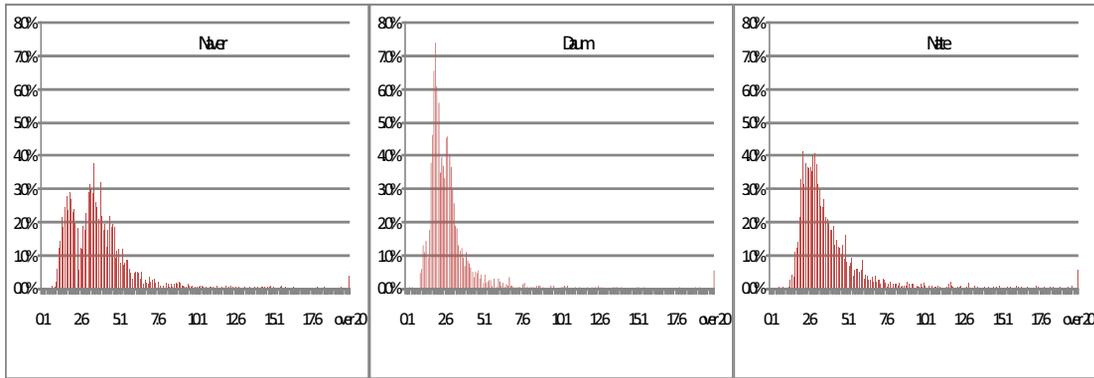


Figure 2a Naver

Figure 2b Daum

Figure 2c Nate

Figure 2. Response Time Distribution of Naver

Table 2. Descriptive Statistics of Naver, Duam, Nate

(unit:second)	Naver	Daum	Nate
Average	3.74	2.90	3.94
Median	3.40	2.43	3.18
Standard Deviation	2.49	2.36	3.06
Range	34.07	35.99	61.36

평균비교를 위해 분산분석을 실시하였다. 분산분석 표는 Table 3에 나와 있다. 분산분석 결과 유의확률이 2.45×10^{-45} 이므로 평균이 동일하다는 귀무가설을 기각할 수 있다. 따라서 평균이 동일하다는 귀무가설을 기각할 수 있으므로 평균은 동일하지 않다고 결론을 내릴 수 있다. 다중비교를 위해 분산의 동일성 검정을 실시하였다. Olkin et al. (1960)은 분산의 동일성 검정통계량을 제시하였다. Levene의 검정통계량에 의한 결과는 Table 4에 나와 있다. 유의확률이 0.001보다 작으므로 분산이 동일하다는 귀무가설을 기각할 수 있다. Tamhane (1977)은 이분산일 때 다중비교 통계량을 제안하였다. Tamhane의 다중비교 결과는 Table 5에 나와 있다. 유의수준 5% 하에서 모두 유의하게 차이가 났으며, 응답 평균 시간이 Daum이 가장 빨랐고, 다음으로 Naver, 마지막으로 Nate로 나타났다.

응답시간 분포를 직접 비교하기 위해, 먼저 모수적 방법으로 카이제곱 검정을 실시하였다. 카이제곱 검정을 위해 구간을 51개로 분할하였다. 검정결과는 Table 6에 나와 있다. Table 2.5에 의하면 유의수준 1% 하에서 Naver와 Daum, Nate의 응답시간의 분포는 모두 동일하지 않았다. 따라서 모수적 분포 비교에 의해서도 Daum이 가장 빠르고 다음으로 Naver, 그리고 Nate가 가장 느리다는 사실을 알 수 있다.

비모수적 검정을 위해 2 표본 콜모고로프-스미르노프 검정을 실시하였다. 검정결과는 Table 7에 나와 있다. 결과

를 보면 유의수준 1% 하에서 세 분포 모두 동일하지 않다는 사실을 알 수 있다. 지금까지의 결과는 분산분석 결과에서 응답시간의 평균을 차이가 있었으며, 분포를 비교하는 모수적 비모수적 방법에 의해서도 유의한 차이가 있었고 응답시간은 Daum이 가장 빠르고, 다음으로 Naver, 마지막으로 Nate임을 말해준다.

Table 3. Analysis of Variance

Factor	SS	DF	MS	F ratio	P-value
Treatment	1469	2	734.75	104.2	<0.001
Error	51393	7287	7.053		
Total	52863	7289			

Table 4. Test for Equal Variances

Levene Statistic	df1	df2	Sig.
35.25674344	2	7287	<0.001

Table 5. Multiple Comparison of Tamhane

I	J	I-J	Standard Error	Sig.
Naver	Daum	0.84	0.07	<0.001
	Nate	-0.19	0.08	0.05
Daum	Naver	-0.84	0.07	<0.001
	Nate	-1.03	0.08	<0.001

Table 6. Chisquare Test

	Daum	Nate
Naver	683.46**	373.19**
Daum		696.58**

** means the null hypothesis can be rejectes under $\alpha = 0.01$.

Table 7. Two Sample Kolmogorov-Smirnov Test

	Test statistic	Sig.
Naver vs Daum	0.438	<0.001
Naver vs Nate	0.253	<0.001
Daum vs Nate	0.201	<0.001

다음으로 응답시간에 영향을 많이 줄 것으로 추정되는 패킷량 효과를 제외하기 위해 1Kbyte당 응답시간을 비교하였다. 1Kbyte당 응답시간의 분포는 Figure 3에 있다. 응답시간 분포와는 다르게 Naver의 분포는 쌍봉 분포가 아니었다. 세 분포는 오른쪽 꼬리가 긴 비슷한 형태를 띠고 있다. Table 8에 의하면 1Kbyte당 응답시간의 평균은

Naver가 0.0185초, Daum 0.0233초, Nate 0.0294초로 Naver가 가장 빨랐으며 Nate가 가장 느렸다. 중앙값으로는 Naver가 0.0126초, Daum 0.0186초, Nate 0.0194로 Naver가 가장 빨랐으며 Nate가 가장 느렸다. 이러한 현상은 응답시간과는 다른 경향을 보이고 있다.

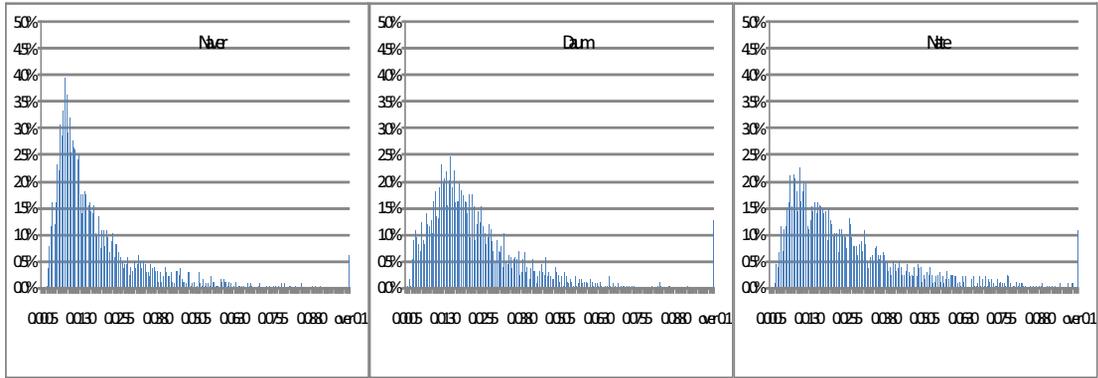


Figure 3a Naver

Figure 3b Daum

Figure 3c Nate

Figure 3. Response Time Distribution of Naver per 1Kbyte

Table 8. Descriptive Statistics of Naver, Duam, Nate per 1Kbyte

(Unit: Second)	Naver	Daum	Nate
Average	0.0185	0.0233	0.0294
Median	0.0126	0.0186	0.0194
Standard Deviation	0.0303	0.0229	0.1887
Range	1.2611	0.5767	9.2556

평균비교를 위해 분산분석을 실시하였다. 분산분석 표는 Table 9에 나와 있다. 분산분석 결과 유의확률이 0.003이므로 평균이 동일하다는 귀무가설을 유의수준 1% 하에서 기각할 수 있다. 따라서 평균이 동일하지 않다는 결론을 내릴 수 있다. 다중 비교를 위해 분산의 동일성 검정을 실시하였다. Levene의 검정통계량에 의한 결과가 Table 10에 나와 있다. 유의확률이 0.009이므로 유의수준 1% 하에서 분산이 동일하지 않다는 결론을 내릴 수 있다. 분산이 동일하지 않기 때문에 이분산에서 사용하는 Tamhane의 다중비교를 실시하였다. 결과는 Table 11에 나와 있다. 다중비교 결과 유의수준 5% 하에서 1Kbyte당 평균 응답시간에서 Naver는 Daum과 Nate에 비해 유의하게 빨랐으나, Daum과 Nate는 유의하게 차이하지 않았다.

Table 9. Analysis of Variance per 1Kbyte

Factor	SS	DF	MS	F ratio	P-value
Treatment	0.145	2	0.072	5.859	0.003
Error	90.003	7287	0.012		
Total	90.148	7289			

Table 10. Test for Equal Variances per 1Kbyte

Levene Statistic	df1	df2	Sig.
4.729	2	7287	0.009

Table 11. Multiple Comparison of Tamhane per 1Kbyte

I	J	I-J	표준오차	Sig.
Naver	Daum	-0.0048	0.00077	<0.001
	Nate	-0.0108	0.00388	0.015
Daum	Naver	0.0048	0.00077	<0.001
	Nate	-0.0061	0.00386	0.301

분포를 직접 비교하기 위해, 모수적 방법으로 카이제곱 검정을 실시하였다. 카이제곱 검정을 위해 구간을 51개로 분할하였다. 검정결과는 Table 12에 나와 있다. 표에 의하면 Naver와 Daum, Nate의 응답시간의 분포는 유의수준 1% 하에서 모두 동일하지 않다는 사실을 알 수 있다. 따라서 모수적 분포 비교에 의해서도 1Kbyte당 응답시간에서 Naver가 가장 빠르고, 다음으로 Daum, 그리고 Nate가 가장 느리다는 사실을 알 수 있다.

Table 12. Chisquare Test per 1Kbyte

	Daum	Nate
Naver	102.09**	202.49**
Daum		126.53**

** means the null hypothesis can be rejectes under $\alpha = 0.01$.

다음으로 비모수적 검정을 위해 2 표본 콜모고로프-스미르노프 검정을 실시하였다. 검정결과는 Table 13에 나와 있다. 결과를 보면 유의수준 1% 하에서 세 분포 모두 동일하지 않다는 사실을 알 수 있다.

Table 13. Two Sample Kolmogorov-Smirnov Test per 1Kbyte

	Test statistic	Sig.
Naver vs Daum	0.233	<0.001
Naver vs Nate	0.210	<0.001
Daum vs Nate	0.074	<0.001

3. 요약 및 결론

포털의 응답품질 평가에는 응답시간, 콘텐츠, 디자인 등 여러 가지 요인들이 있으나, 응답시간이외의 요인들은 주관적이며 계량화하기 어려운 점이 있다. 응답시간은 사용자 만족도를 평가는 가장 솔직한 방법이다. 또한 포털 이용자들은 요청한 서비스가 얼마나 빨리 제공되는가에 대해서만 평가하는 경향이 있다. 따라서 응답시간은 응답품질을 평가하는 객관적이고 매우 중요한 요소이다. 응답시간은 서비스의 패킷량과 서버의 성능, 페이지의 성능, 네트워크

트래픽, 서버의 위치, 이용자 컴퓨팅 성능 등 여러 가지 요소들에 영향을 받는다. 본 논문은 3년 4개월에 걸쳐 조사한 데이터를 바탕으로 대한민국의 상위 3개 포털들의 응답시간을 이용하여 응답품질을 평가하였다. 연구 방법으로는 평균비교를 위해 분산분석을 사용하였으며, 분포비교를 위해 모수적 방법으로는 카이스퀘어 검정, 비모수적 방법으로는 콜모고로프-스미르노프 검정을 실시하였다.

세 가지 검정 결과 모두 유의수준 1% 하에서 유의한 차이가 있는 것으로 나타났다. 다중비교에서도 유의한 차이가 나타났다. 응답시간은 Daum이 가장 빨랐으며, 다음으로 Naver, 마지막으로 Nate 순이었다. 평균응답시간은 Daum이 2.9초, Naver가 3.74초, Nate가 3.94초였다. 이 결과는 웹서비스가 1초 느려지면 매출은 3% 떨어진다는 마이크로소프트(Bing)의 통계를 기준으로 보면 매우 중요한 시사점을 가지고 있다.

패킷량 효과를 제외하기 위한 비교를 위하여 1Kbyte당 응답시간을 분석하였다. 1Kbyte당 응답시간에 대해서는 세 가지 검정 결과 유의수준 1% 하에서 유의한 차이가 있는 것으로 나타났다. 다중비교에서 Naver가 Daum과 Nate에 대해서는 유의한 차이가 났지만 Daum과 Nate는 유의한 차이를 발견하지 못하였다. 1Kbyte당 평균 응답시간은 Naver가 0.0185초, Daum이 0.0233초, Nate가 0.0294초였다.

본 논문의 결과는 응답시간은 Daum이 가장 빨랐지만 1Kbyte당 응답시간은 Naver가 가장 빨랐다. 이는 응답시간에 응답패킷이 큰 영향을 주고 있다는 사실을 알 수 있고, 응답시간을 줄이기 위해서 포털들의 홈페이지의 패킷을 줄이려는 노력이 매우 효과적인 방법임을 말하고 있다. 이 문제는 모바일 포털에서는 더욱더 중요한 문제인데, 그 이유는 모바일 서비스는 주로 패킷량에 비례하여 과금을 하는 종량제를 채택하고 있고, 모바일 네트워크가 협대역이기 때문이다. 본 논문의 한계는 동일한 장소에서 실험한 결과이기 때문에, 향후 좀 더 다양한 장소에서 실험하는 연구가 필요하다. 또한 응답시간에 영향을 주는 변인들에 대한 추가 연구도 필요하다. 그 외 모바일 포털들에 대한 응답품질에 관한 연구도 필요하다.

REFERENCES

- Chiew, Thiam Kian. 2009. "Web page performance analysis." PhD diss., University of Glasgow.
- Han, Dong-Wook, and Kang, Min-Chae. 2012. "Study on application of information and communication technology in special education." *Journal of the Korean Data & Information Science Society* 23:927-37.
- Han, Jin-Man, Lee, Kyeong-Jun, and Yang, Jeong-Ok. 2012. "The effects of the 16-weeks' combined exercise program on metabolic syndrome and autonomic nerve system of low-level physical strength group." *Journal of the Korean Data & Information Science Society* 23:895-904.
- Hong, Jong Sun, and Kim, Ji Hoon. 2009. "Nonparametric homogeneity tests of two distributions for credit rating model validation." *Journal of the Korean Data & Information Science Society* 20:261-72.
- Hoxmeier, John A., and DiCesare, Chris. 2000. "System response time and user satisfaction: an experimental study of browser-based applications." *Proceedings of the 4th COLLECTeR Conference on Electronic Commerce*, Breckenridge, April 11.
- Kim, Dae-Hak, and Oh, Kwang-Sik. 2003. "On the equality of two distributions based on nonparametric kernel density estimator." *Journal of the Korean Data & Information Science Society* 14:247-55.
- Kim, Il-Hwan. 2012. "Web service, the faster the better." *Proceedings of the DEVIEW 2012*.
- Kim, Sang-Cheol, Yun, Won-Young, and Chun, Young-Rok. 2004. "Optimal process condition for products with multi-categorical ordinal quality characteristic." *Journal of the Korean Society for Quality Management* 32:109-125.
- Kim, Tai-Kyoo, and Kim, Myung Joon. 2013. "A study on the model improvement of Korean industrial standards-Quality Excellence Index(KS-QEI)." *Journal of the Korean Society for Quality Management* 41:327-336.
- Korea Internet & Security Agency. 2013. 2013 Korea internet white paper. Kyeonggi: Korea Internet & Security

Agency.

- Olkin, Ingram, Ghurye, Sudhish G., Hoeffding, Wassily, Madow, William G., and Mann, Henry, B. 1960. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. California: Stanford University Press.
- Press, William H., Flannery, Brian P., Teukolsky, Saul A., and Vetterling, William T. 2002. *Numerical recipes in C*. Cambridge: Cambridge University Press.
- Ree, Sangbok. 2013. "Study on the result changes with the size of the variance in Taguchi method and factor experimental." *Journal of the Korean Society for Quality Management* 41:119-134.
- Ryu, Gui-Yeol. 2012. "A study on response time of WiBro depending on signal intensity." *Journal of the Korean Data Analysis Society* 14:1119-1128.
- Ryu, Gui-Yeol. 2013. "A study on comparing response times between Wibro and wired internet using portals." *Journal of the Korean Data & Information Science Society* 24:437-444.
- Ryu, Gui-Yeol. 2013. "A study on distribution comparison of response packets for major portal sites." *Journal of the Korean Data & Information Science Society* 24:23-32.
- Seow, Steven C. 2008. *Designing and engineering time*. Boston: Addison-Wesley.
- Tamhane, AC. 1977. "Multiple comparisons in model I one-way ANOVA with unequal variances." *Communications in Statistics B(9)*:167-178.
- Touch, Joe, Heidemann, John, and Obraczka, Katia. 1998. "Analysis of HTTP performance." USC/ISI Research Report 98(463):1-10.