[6-7] # 가입전화 트래픽 예측 방법에 관한 연구

正會員 林 成 植* 正會員 鄭 永 在*

(A Study On The Subscriber Traffic Forecasting Method)

Seong Sik LIM*, Yeong Jae JEONG* Regular Members

要 約 본 논문에서는 보다 정확하게 가입전화 트래픽을 예측하기 위한 회귀진단 기법을 제안하였다. 기존 방법은 다만 회귀계수만 추정하여 트래픽을 예측하였지만, 새로운 방법에서는, 각각의 측정치에서 트래픽 예측에 영향을 미치는 측정치와 영향을 미치지 않는 즉, 이용할 수 없는 자료를 식별하는 방법을 보였으며, 이 방법를 이용하여 예측하는 것이 기존 예측 방법보다 타당성이 있음이 시뮬레이션에 의하여 입증되었다.

ABSTRACT This paper proposed the method of regression diagnostics with a view to forecasting the subscriber traffic with more accuracy. Up to now, the traffic forecasting method is only used the estimated regression coefficient.
However, the new method showed how to discriminate between influential measuring data and unavailable data. And this method proved out more reasonable for forecasting than any other method through simulation.

I. 서론

(x1i,x2i,...,xki,yi), i=1,2,...,n가 있을 때, 일반적으로 중회귀 모형은

$$Y = X\beta + \varepsilon \qquad (1.1)$$

여기서 X는 n x (k+1) 행렬이고

$$\varepsilon \sim N(0, I\sigma^2)$$

이라 가정할 때, b를 β 의 추정벡타라 하면 최소자승법에 의하여

$$b = (X'X)^{-1} X'Y \qquad (1.2)$$

가 된다.그러나 위와 같은 방법으로 회귀계수(b)를 추정하여 가입전화 트래픽을 예측한다면 주어진 자료에서 보다 유용한 정보를 얻을 수 가 없으므로,좀더 유용한 정보를 얻기위하여 여러가지 진단기법이 필요하게 되었다.

각 진단기법의 종류는 다음장에서 설명을 하기로 하고, 먼저 기본적으로 알아야 할 사항을 보기로 하자. 우선, 잔차행렬을

$$e = Y - Xb = (I - X(X'X)^{-1} X')Y \qquad (1.3)$$

으로 정의할 때, e의 기대치와 분산공분산행렬은

$$E(e) = 0$$

$$V(e) = (I - X(X'X)^{-1}X')\sigma^2 \qquad (1.4)$$

이 됨을 알 수 있다. 만일 σ^2의 추정벡타를 S^2이라 하면, $S^2 = e'e/(n-k-1)$이 되며, 이때 분산 공분산 행렬의 추정벡타는

$$\hat{V}(e) = (I - X(X'X)^{-1}X')S^2 \qquad (1.5)$$

이 되며, 식 (1.5)에 포함되어 있는 $X(X'X)^{-1}X'$는 "Hat matrix"라 부르며 회귀진단에서는 매우 중요한 행렬로서

$$H = X(X'X)^{-1}X' \qquad (1.6)$$

로 정의하며, i번째 대각선상의 값을 hii, (i,j)번째 값은 hij라고 한다면

$$hii = xi(X'X)^{-1}xi'$$

$$hij = xi(X'X)^{-1}xj' \qquad (1.7)$$

의 관계가 있다.

여기서 우리는 모든 가입전화 트래픽의 예측 경우가 단순회귀이므로 독립변수가 하나인 경우만을 실예로 분석하고자 한다. 이때 간단한 한 예로 단순회귀모형은 다음과 같고,

$$Yi = \beta_1 + \beta_2 Xi + \varepsilon, \quad i = 1,2,\ldots,n \qquad (1.8)$$

β_1, β_2의 추정값을 b_1, b_2라 할 때, 잔차는 $ei = Yi - b_1 - b_2 Xi$가 되며, hat matrix는

$$hij = \frac{1}{n} + \frac{(xi-\bar{x})(xj-\bar{x})}{\Sigma(xi-\bar{x})^2}$$

$$hii = \frac{1}{n} + \frac{(xi-\bar{x})(xi-\bar{x})}{\Sigma(xi-\bar{x})^2} \qquad (1.9)$$

이 된다.

잔차 ei의 기대치와 분산은

$$E(ei) = 0$$

$$V(ei) = (1 - hii)\sigma^2 \qquad (1.10)$$

임을 알수 있고, V(ei)의 추정치는

$$\hat{V}(ei) = (1 - hii)S^2 \qquad (1.11)$$

이 되며, 여기서 $S^2 = \Sigma ei^2/(n-2)$ 이다.

기존 가입 전화 트래픽 예측방식은 회귀계수(b)를 추정하여 각 모형별로 트래픽을 예측한 결과가 예측 자료로 이용할 수 없는 측정치에 의해 영향을 받는 경우가 발생할 수 있기 때문에, 이와 같이 영향을 미치지 않는 측정치와 영향을 미치는 측정치를 식별함으로서 예측자료의 특성을 파악할 수 있으며, 결과를 분석하는데 도움을 얻을 수 있다.

다음장에서는 회귀진단에 필요한 기법들을 소개하고, 3장에서는 실제 가입전화 트래픽 자료를 회귀진단 기법에 적용하여 얻은 회귀분석 결과와 기법을 이용하지 않고 분석한 결과를 비교 검토하였다.

Ⅱ. 회귀진단 기법의 종류

분석 결과에 따라 그 결과에 영향을 주는 측정치들중에는 이용가치가 없는 자료도 있고 이용가치가 있는 자료도 있다. 전자의 경우를 이상치라 하고 후자의 경우를 영향을 크게 주는 측정치라고 말한다.

다음에서 이상치와 크게 영향을 주는 측정치를 찾는 방법을 보기로 하자.

Ⅱ.1 이상치 검출

1) 표준화 잔차

Y가 $N(X\beta, I\sigma^2)$의 분포를 따르므로 e는 $N(0,(I-H)\sigma^2)$의 분포를 따름을 알 수 있다. 또한 i번째 잔차(ei)의 분포는

$$ei \sim N(0,(1-hii)\sigma^2) \qquad (2.1)$$

이 되므로 따라서, ei를 표준 정규분포화 하면

$$Zi = \frac{ei}{\sigma\sqrt{1-hii}} \sim N(0,1) \qquad (2.2)$$

와 같으며, σ 대신 측정치 S를 이용하면 표준화 잔차

$$\tau i = \frac{ei}{S\sqrt{1-hii}} \qquad (2.3)$$

가 얻어진다. 표준화 잔차 τi의 절대값 $|\tau i|$가 과다하게 크면 이상치라 생각할 수 있다. 이상치 판단기준은 [별첨 1]에 있다.

2) 스트던트화 잔차

이 잔차는 표준화잔차에서 S, b대신에 i번째 측정치 Yi를 제외시키고 n-1개의 측정치로부터 얻어지는 S,b (S,b를 각각 S(t), b(i) 로 표기하자)를 사용 한다면, 그때 S^2은

$$S^2(i) = \frac{1}{n-k-2}\sum_{k\neq i}[Yk - Xk\,b(i)]^2 \qquad (2.4)$$

이 되며, 위 식을 간단한 형태로 변형하면

$$(n-k-2)S^2(i) = (n-k-1)S^2 - \frac{ei^2}{1-hii} \qquad (2.5)$$

이 된다.

스트던트화 잔차는 표준화잔차에서 S대신 S(i)로 대체시킨 식으로 다음과 같이

$$\tau i^* = \frac{ei}{S(i)\sqrt{1-hii}} \qquad (2.6)$$

로 정의되며, 자유도 n-k-2 를 가진 t분포 형태를 따른다.

식 (2.6)를 좀더 간략하게 정리하면

$$\tau i^* = \frac{ei}{S(i)\sqrt{1-hii}} = \tau i \cdot \left[\frac{n-k-2}{n-k-1-\tau i^2}\right]^{1/2} \qquad (2.7)$$

이 된다.

여기서 $|\tau i^*|$가 어느 정도 커야만 Yi를 이상치로 볼 수 있는지를 결정하기 위한 검정방법은

$$|\tau i^*| \geq t(n-k-2;\alpha/2) \qquad (2.8)$$

이면, 유의수준 α에서 Yi 를 이상치라 판정할 수 있다.

Ⅱ.2. 영향을 크게 주는 측정치 검출

1) 행렬 H의 대각선 원소

최소자승법에 의하여 적합된 중회귀 모형은

$$\hat{Y} = Xb = X(X'X)^{-1}X'Y = HY \qquad (2.9)$$

가 얻어지며, 예측치 $\hat{Y}i$를 Yi로 편미분하면

$$\frac{\delta\hat{Y}i}{\delta Yi} = hii$$

로 행렬 H의 i번째 대각선 원소가 된다.

따라서 hii의 값이 1 에 가까울 수록 Yi는 $\hat{Y}i$에 영향을 크게 준다고 판단할 수 있다. 즉 이 말은 hii=1 이면 Yi=$\hat{Y}i$, 다시 말하면 ei=0을 의미한다.

여기서 H의 성질을 몇가지 살펴보기로 하자.

- HH=H 가 성립하므로 멱등행렬이고, H의 계수는

$$\text{rank}(H) = \sum_{i=1}^{n} hii = k+1$$

이 된다.

- 행렬 H는 양반정치(Positive Semi-definite)행렬이므로 $hii \geq 0$ 이고, $(1-hii)\sigma^2 \geq 0$ 이어야 하므로 $hii \leq 1$이 되어야 한다. 따라서

$$0 \leq hii \leq 1$$

의 관계가 있다. 그리고 hii의 평균치는 (k+1)/n 이다.

hii가 어느 정도 커야만 Yi가 영향을 크게 주는 측정치인가를 판단하는 기준은

$$\frac{n-k-1}{k}\cdot\frac{hii-(1/n)}{1-hii} \geq F(k,n-k-1;\alpha) \quad (2.10)$$

이면, Yi 가 크게 영향을 받는 측정치라 말할 수 있다. 그런데 k>10 이고 n-k-1>50 이상일 때 95%의 신뢰도에서 F 분포의 값은 2보다 작은 값이 나타난다. 그러므로 좀더 간단히 하면 다음과 같은 식으로 판단기준을 설정하여 Yi가 영향을 주는 측 측정치인지 아닌지를 판단할 수 있다.

$$hii \geq (2k+1)/n \quad (2.11)$$

2) DFFITS

이 경우는 b와 b(i)의 차이 값이 가장 큰 $\hat{Y}i$ 를 찾는 방법을 말한다. i번째 Yi를 제외시킨 상태에서의 적합된 모형의 차이 값은

$$DFFIT = \hat{Y}i - \hat{Y}i(i) = Xi(b-b(i)) = \frac{hii\cdot ei}{1-hii} \quad (2.12)$$

이되며, 위 식을 S(i) hii로 나누면 DFFITS(i)는

$$DFFITS(i) = \left[\frac{hii}{1-hii}\right]^{1/2}\cdot\frac{ei}{S(i)\sqrt{1-hii}} \quad (2.13)$$

으로 정의되며, 위 식은 스트던트화 잔차의 식을 이용하여

$$DFFITS(i) = \left[\frac{hii}{1-hii}\right]^{1/2}\tau i^{*} \quad (2.14)$$

으로 바꿀 수 있다.

DFFITS(i)가 어느 정도 커야만 Yi 가 영향을 크게 주는 측정치로 판단할 수 있는가를 보기 위하여, 먼저 $|\tau i^{*}| \geq 2$ 이고, $hii \geq (k+1)/n$ 이므로, 식(2.14)에서

$$|DFFiTS(i)| \geq \left[\frac{(k+1)/n}{1-(k+1)/n}\right]^{1/2}\cdot 2 = 2\left[\frac{k+1}{n-k-1}\right]^{1/2} \geq 2\left[\frac{k+1}{n}\right]^{1/2} \quad (2.15)$$

으로 되므로 위식을 이용하여 판단할 수 있다.

3) Cook의 통계량

일반적으로 모든 b^{*}에 대하여 $100(1-\alpha)\%$ 신뢰구간은

$$\frac{(b-b^{*})'X'X(b-b^{*})}{(k+1)S^{2}} \leq F(k+1,n-k-1;1-\alpha) \quad (2.16)$$

로 정의되는데 Cook의 통계량은 b^{*} 대신에 b(i) 를 대입시킨 즉,

$$D(i) = \frac{(b-b(i))'X'X(b-b(i))}{(k+1)S^{2}} \quad (2.17)$$

으로 정의되며, 이는 F(k+1, n-k-1; 1-α)의 분포를 가지며, 이 값을 크게 하는 것을 영향을 크게 주는 측정치로 판단할 수 있다.

D(i)을 간단히 하기 위하여 표준화잔차를 이용하면

$$D(i) = \frac{1}{k+1}\cdot\frac{hii\,ei^{2}}{S^{2}(1-hii)^{2}} = \frac{hii}{(k+1)(1-hii)}\cdot\tau i^{2} \quad (2.18)$$

와 같다. 위식에서 τi^{2}의 값이 커지면 D(i)도 값이 커지고, hii값이 커지면 D(i)의 값이 커짐을 알 수 있다.

영향을 크게 주는 측정치의 판단기준은

$$D(i) \geq F(k+1;\ n-k-1;\alpha) \qquad (2.19)$$

이 되는 Yi를 찾는 방법으로 Cook는 대략적으로

$$D(i) \geq F(k+1,\ n-k-1;0.5) \qquad (2.20)$$

이면 영향을 크게 주는 측정치로 판단해도 좋다고 제안하였다.

4) Andrew-Pregibon 의 통계량

행렬X와 벡타Y를 같이 고려하여, 다음의 통계량

$$AP(i) = \frac{|\overset{*}{X}(i)'\overset{*}{X}(i)|}{|\overset{*}{X}'\overset{*}{X}|} \qquad (2.21)$$

을 제안하였다. 여기서 $\overset{*}{X}(i)=(X(i):Y(i))$, $\overset{*}{X}=(X:Y)$을 의미하는데, $\overset{*}{X}(i)$는 $\overset{*}{X}$ 행렬에서 i 번째 행을 제거시킨 것이다.

위식을 다시 표현하면

$$AP(i) = (1 - \frac{ei^2}{(1-hii)\ SSE})(1-hii)$$

$$= 1-hii - \frac{ei^2}{(n-k-1)S^2} \qquad (2.22)$$

이 된다. 여기서 SSE는 잔차 자승합을 의미하며, $SSE=(n-k-1)S^2$으로 이미 알고 있다. 식(2.22)에서 hii가 크거나 또는 ei^2이 크면 AP(i)값은 작아지며, 이들 값 중 가장 작은 AP(i)의 값을 Yi가 영향을 크게 주는 측정치로 판단하면 된다.

5) COVRATIO

회귀계수 추정치의 분산공분산 행렬은

$$V(b) = (X'X)^{-1}\sigma^2$$

$$V(b(i)) = [X(i)'X(i)]^{-1}\sigma^2 \qquad (2.23)$$

이다. σ^2 대신에 각각 S^2, $S^2(i)$를 대입시킨 후, 그 행렬식의 비율을 COVRATIO라 정의하고, 그 식은

$$COVRATIO(i) = \frac{|S^2(i)[X(i)'X(i)]^{-1}|}{|S^2(X'X)^{-1}|} \qquad (2.24)$$

와 같다. 위 식(2.16)을 좀더 간략하게 하기 위해 먼저 다음과 같은 식

$$|X'(i)X(i)| = (1-hii)|X'X|$$

을 이용하여 계산하면

$$COVRATIO(i) = \frac{[S^2(i)]^{k+1}}{(S^2)^{k+1}} \cdot \frac{1}{1-hii} \qquad (2.25)$$

되는데, 식(2.5)와 (2.6)를 사용하면 식(2.25)는

$$COVRATIO(i) = \frac{1}{[\frac{n-k-2}{n-k-1} + \frac{\varepsilon i^{*2}}{n-k-1}]^{k+1}(1-hii)} \qquad (2.26)$$

이 된다.

COVRATIO(i)의 값이 1에 가까우면 Yi는 별로 영향을 못주며, 1에서 멀어질수록 영향을 크게 주는 측정치라 말할 수 있다.

Yi가 영향을 크게 주는 측정치라고 판단하는 기준이 다음과 같으면

$$|\ COVRATIO(i)-1\ | \geq 3(k+1)/n \qquad (2.27)$$

Yi를 영향을 크게 주는 측정치라고 한다.

6) FVARATIO

하나의 측정치 Yi가 제외되었을 때, Yi의 분산이 어떻게 변화하는 가에 대한 것으로서, 이것을 하기위해 먼저 $\hat{Y}i$ 와 $\hat{Y}i(i)$ 의 분산공분산 행렬은

$$V(\hat{Y}i) = hii \cdot S^2$$

$$V(\hat{Y}i(i)) = \frac{hii}{1-hii} S^2 \qquad (2.28)$$

이 되며, 이를 비율로 구하면

$$FVARATIO(i) = \frac{hii \cdot S^2(i)/(1-hii)}{hii \cdot S^2} = \frac{S^2(i)}{S^2(1-hii)} \qquad (2.29)$$

얻게 된다.

Yi가 영향을 크게 주는 측정치인지 아닌지의 판단기준은

$$FVARATIO(i) \leq 1 - 3/n \quad \text{이거나,}$$
$$FVARATIO(i) \geq 1 + (2k+3)/n \qquad (2.30)$$

이면 영향을 주는 측정치로 볼 수 있다.

Ⅱ.3. 측정치 측도의 비교

2절에서 이상치를 검출하는 방법으로 2 가지를, 영향을 크게 주는 측정치를 검출해내는 방법으로 6 가지를 다루었으며, 이들 간에는 상호 연관성이 있음을 알 수 있다. 이를 한번 고찰해 보기로 하자.

이상치를 찾는 2가지 방법 즉, 표준화 잔차 및 스트던트화 잔차는 모두 잔차(ei)의 함수이며, |ei|가 커지면 이들 값들은 커진다. 일반적으로 잔차가 매우 크면 이상치라 볼 수 있다. 또한 이들 측도들은 "hat matrix" hii 의 함수로서 hii 가 커지면 이들 값들도 커진다. 따라서 hii 가 크고 ei도 크면 Yi는 이상치라 판단할 수 다.

영향을 크게 주는 측정치를 찾는 방법 중에서, hii 값은 종속변수 Y의 값과는 직접 관련이 없고 X에 대한 유관한 측도 이므로, hii의 값이 크면 i 번째가 다른 측정치로부터 멀리 떨어져 있다고 다고 판단할 수 있다. 따라서 hii의 값이 크면 그 측정치 Yi는 영향을 크게 주는 측정치가 될 가능성이 높다.

그리고 영향을 크게 주는 측도로서 DFFITS(i), D(i), AP(i), COVRATIO(i), FVARATIO(i) 들은 모두 측정된 회귀계수들의 변화를 검출해 내는 방법으로 만들어져 있어서 영향을 크게주는 측정치를 찾는 측도로 쓰이나, 이들은 모두 잔차ei의 함수로 나타난다.

따라서, 이들은 영향을 크게 주는 측정치로 판단될 때에는 동시에 이상치가 될 가능성도 있다.

만일, 이상치로 판정되었다고 하여 이것을 무조건 버리고 나머지 만으로 분석을 실시하는 것은 옳은 방법이 아니다. 이상치로 판정되면 원인을 규명하여야 한다. 원인이 규명되면 다시 측정하여 대체시키는 경우가 있고, 측정이 불가능한 경우는 측정치를 제거시킬 수도 있다.

또한 영향을 크게 주는 측정치로 판정이 되면, 일단 이 측정치가 어떻게 해서 가장 영향을 크게 주는가를 검토해 볼 필요가 있다. 만일 다른 측정치들로 부터 멀리 떨어져 있는 것이 원인이라면, 그 중간에 몇개의 측정치를 추가하여 진단을 실시하고 분석을 하는 것이 바람직하다. 만일, 다른 측정치들로 부터 멀리 떨어져 있지 않는데도 불구하고 잔차가 커서 영향을 크게주는 측정치로 판정되었다면, 이것은 이상치이므로 이상치와 동일하게 취급하면 된다.

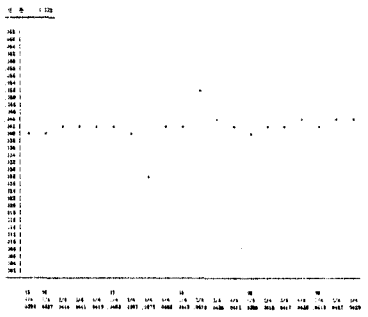
Ⅲ. 트래픽 예측 방법의 비교

이 장에서는 서울 신촌국의 NO1A 가입 전화 실측 트래픽 자료를 이용하여 트래픽 예측 결과를 비교 설명하고자 한다.

먼저, 가입 전화 실측 트래픽 자료를 수집, 분석한 후 계절지수를 고려하여 보정트래픽을 산정한다. 신촌국의 년도별 보정 트래픽은 [표 3-1] 과 같으며, 이에 대한 분기별 보정트래픽 산점도는 [그림 3-1] 과 같다.

[표 3-1] 신촌국 보정 트래픽

년 도	분 기	보정트래픽
'84	4/4	0.0387
'85	1/4	0.0390
	2/4	0.0403
	3/4	0.0385
	4/4	0.0398
'86	1/4	0.0387
	2/4	0.0414
	3/4	0.0411
	4/4	0.0419
'87	1/4	0.0403
	2/4	0.0397
	3/4	0.0275
	4/4	0.0408
'88	1/4	0.0419
	2/4	0.0510
	3/4	0.0436
	4/4	0.0412
'89	1/4	0.0396
	2/4	0.0416
	3/4	0.0417
	4/4	0.0436
'90	1/4	0.0419
	2/4	0.0427
	3/4	0.0429



[그림 3-1] 신촌국 보정트래픽 그래프

또한 가입 전화 트래픽 예측에 사용되는 추세모형은 [표 3-2]와 같다.

[표 3-2] 추세모형

모 형	산 출 식
직 선	Yt = a + bt
지 수	Yt = a · exp(bt)
수정지수	Yt = K - a · exp(bt)
성장곡선	Yt = K / (1 + a · exp(bt))

주) Yt : 가입전화 보정트래픽
t : 시간
a.b : 회귀계수
K : 상한치
exp : 자연대수

Ⅲ.1 기존 트래픽 예측 방법

기존 트래픽 예측 방식은 회귀계수를 추정하여 적합된 회귀모형을 이용하여 익년도 가입 전화 트래픽을 예측하기 때문에 각 모형에 대한 타당성 검토에 관계없이 F-통계량 값이 가장 큰 모형을 선정하고, 이때 선정된 모형에 의해 예측된 값을 최적 예측치로 선정하므로서 실제 타당성이 없는 모형인데도 그모형의 예측치를 전화교환시설 설계에 적용하므로서 신뢰성이 떨어진다고 볼 수 있다.

Ⅲ.2 회귀진단에 의한 트래픽 예측

1절의 문제점을 보완하기 위해서 가입전화 트래픽 예측에 영향을 미치는 측정치와 영향을 미치지 않는 측정치를 검출하는 분석과 함께 영향을 미치는 측정치이면서 이상치인 경우의 자료도 예를 통하여 분석하여 보자.

[표 3-3]은 각 모형별 회귀진단방법을 이용하여 트래픽 예측 결과에 영향을 크게주는 측정치와 이상치를 검출하기 위한 표이며,유의수준을 α=0.05로 하자.

이상치를 검출하는 방법으로서 표준화잔차(r_i)에 의한 이상치 검출방법은 [표 3-1]에서 k=1이고 n=24이므로 [별첨 1]의 표에는 기각치가 없으므로 보간법을 이용하여 구하면 n=24 에서 기각치가 2.868 이 된다. 따라서 각 모형별 $|r_i|$의 값이 2.868보다 큰 분기는 87년 3/4분기 자료이므로,이 분기의 트래픽 0.0275 Erl은 이상치임을 알수있다.

다음으로 스트먼트화잔차(r_i^*)에의한 이상치를 검출하는 방법으로서, 자유도가 21이므로, 이상치로 판단하는 기준인 기각치는 t(21;0.025)=2.080 이므로 $|r_i^*|$가 이값 보다는 큰 것들은 이상치

라 판단할 수 있는데 87년 3/4분기 및 88년 2/4분기의 트래픽 자료가 2.080 보다 큰 값이므로 이상치 임을 알 수 있다.

국명 : 신 촌 1332

REGRESSION DIAGNOSTICS

모형 : 직 선 형

NO	QURT	실측 T	예측 T	RESIDUAL	STAD RES	STUD RES	DFFITS(I)	D(I)	AP(I)	COVRATIO	FVARATIO
1	84 4/4	0.0387	0.0387	0.0000	0.0036	0.0035	0.0015	0.0000	0.8433	1.3014	1.2422
2	85 1/4	0.0390	0.0389	0.0001	0.0378	0.0369	0.0143	0.0001	0.8624	1.2724	1.2146
3	85 2/4	0.0403	0.0391	0.0012	0.3635	0.3562	0.1316	0.0090	0.8746	1.2324	1.1835
4	85 3/4	0.0385	0.0392	-0.0007	-0.2143	-0.2096	0.0716	0.0027	0.8934	1.2205	1.1674
5	85 4/4	0.0398	0.0394	0.0004	0.1081	0.1056	0.0333	0.0006	0.9049	1.2055	1.1514
6	86 1/4	0.0387	0.0396	-0.0009	-0.2593	-0.2537	-0.0740	0.0029	0.9140	1.1834	1.1333
7	86 2/4	0.0414	0.0398	0.0016	0.4565	0.4482	0.1210	0.0076	0.9232	1.1553	1.1134
8	86 3/4	0.0411	0.0400	0.0011	0.3176	0.3111	0.0781	0.0032	0.9364	1.1560	1.1085
9	86 4/4	0.0419	0.0402	0.0017	0.4699	0.4813	0.1131	0.0066	0.9313	1.1130	1.0934
10	87 1/4	0.0403	0.0403	0.0000	0.0124	-0.0121	0.0027	0.0000	0.9529	1.1517	1.0994
11	87 2/4	0.0397	0.0405	-0.0008	-0.2321	-0.2210	0.0485	0.0012	0.9540	1.1420	1.0927
12	87 3/4	0.0275	0.0407	-0.0132	-3.6997	-5.4807	-1.2295	0.2992	0.3620	0.1635	0.4131
13	87 4/4	0.0408	0.0409	-0.0001	-0.0269	-0.0263	0.0055	0.0000	0.9531	1.1454	1.0934
14	88 1/4	0.0419	0.0411	0.0008	0.2298	0.2248	0.0480	0.0012	0.9541	1.1431	1.0928
15	88 2/4	0.0510	0.0413	0.0097	2.7338	3.2870	0.7308	0.1647	0.6292	0.5021	0.7259
16	88 3/4	0.0436	0.0414	0.0022	0.6059	0.5970	0.1403	0.0101	0.9319	1.1198	1.0870
17	88 4/4	0.0412	0.0416	-0.0004	-0.1221	-0.1193	-0.0300	0.0005	0.9401	1.1651	1.1129
18	89 1/4	0.0396	0.0418	-0.0022	-0.6292	-0.6203	0.1675	0.0144	0.9153	1.1356	1.1038
19	89 2/4	0.0416	0.0420	-0.0004	-0.1142	-0.1116	-0.0324	0.0006	0.9210	1.1495	1.1361
20	89 3/4	0.0417	0.0422	-0.0005	0.1391	-0.1360	-0.0429	0.0010	0.9086	1.2047	1.1510
21	89 4/4	0.0436	0.0424	0.0012	0.3569	0.3497	0.1154	0.0074	0.8903	1.2114	1.1631
22	90 1/4	0.0419	0.0426	-0.0007	-0.1905	-0.1863	-0.0688	0.0025	0.8784	1.2433	1.1887
23	90 2/4	0.0427	0.0427	0.0000	-0.0106	0.0104	-0.0041	0.0000	0.8625	1.2725	1.2147
24	90 3/4	0.0429	0.0429	0.0000	-0.0060	-0.0058	-0.0025	0.0000	0.8433	1.3014	1.2422

(a) 직 선 형

국명 : 신 촌 1332

REGRESSION DIAGNOSTICS

모형 : 지 수 형

NO	QURT	실측 T	예측 T	RESIDUAL	STAD RES	STUD RES	DFFITS(I)	D(I)	AP(I)	COVRATIO	FVARATIO
1	84 4/4	0.0387	0.0386	0.0001	0.0359	0.0351	0.0151	0.0001	0.8433	1.3012	1.2422
2	85 1/4	0.0390	0.0388	0.0002	0.0729	0.0712	0.0284	0.0004	0.8623	1.2719	1.2144
3	85 2/4	0.0403	0.0389	0.0014	0.4009	0.3931	0.1453	0.0110	0.8734	1.2292	1.1820
4	85 3/4	0.0385	0.0391	-0.0006	-0.1742	-0.1703	-0.0582	0.0018	0.8943	1.2222	1.1682
5	85 4/4	0.0398	0.0393	0.0005	0.1500	0.1467	0.0463	0.0011	0.9085	1.2044	1.1508
6	86 1/4	0.0387	0.0395	0.0008	0.2152	-0.2104	-0.0614	0.0020	0.9197	1.1859	1.1344
7	86 2/4	0.0414	0.0396	0.0018	0.5018	0.4931	0.1332	0.0092	0.9216	1.1557	1.1112
8	86 3/4	0.0411	0.0398	0.0013	0.3644	0.3571	0.0897	0.0042	0.9350	1.1528	1.1069
9	86 4/4	0.0419	0.0400	0.0019	0.5379	0.5290	0.1243	0.0080	0.9352	1.1274	1.0909
10	87 1/4	0.0403	0.0402	0.0001	0.0370	0.0362	0.0080	0.0000	0.9525	1.1516	1.0993
11	87 2/4	0.0397	0.0403	-0.0006	0.1816	-0.1776	-0.0379	0.0008	0.9549	1.1441	1.0938
12	87 3/4	0.0275	0.0405	-0.0130	-3.6460	-5.4622	-1.2339	0.2906	0.3792	0.1794	0.4327
13	87 4/4	0.0408	0.0407	0.0001	0.0247	0.0241	0.0050	0.0000	0.9581	1.1454	1.0934
14	88 1/4	0.0419	0.0409	0.0010	0.2815	0.2756	0.0589	0.0018	0.9529	1.1393	1.0915
15	88 2/4	0.0510	0.0411	0.0099	2.7838	3.3794	0.7513	0.1915	0.6172	0.4832	0.7121
16	88 3/4	0.0436	0.0413	0.0023	0.6575	0.6488	0.1524	0.0119	0.9291	1.1130	1.0837
17	88 4/4	0.0412	0.0414	-0.0002	-0.0702	-0.0686	-0.0172	0.0002	0.9405	1.1661	1.1134
18	89 1/4	0.0396	0.0416	-0.0020	-0.5773	-0.5683	0.1535	0.0122	0.9179	1.1421	1.1070
19	89 2/4	0.0416	0.0418	0.0002	-0.0633	-0.0618	-0.0180	0.0002	0.9214	1.1906	1.1365
20	89 3/4	0.0417	0.0420	-0.0003	0.0889	-0.0869	-0.0274	0.0004	0.9091	1.2060	1.1515
21	89 4/4	0.0436	0.0422	0.0014	0.4057	0.3979	0.1359	0.0096	0.8888	1.2073	1.1611
22	90 1/4	0.0419	0.0424	-0.0005	-0.1425	-0.1393	0.0515	0.0014	0.8790	1.2451	1.1896
23	90 2/4	0.0427	0.0426	0.0001	0.0358	0.0350	0.0160	0.0001	0.8624	1.2734	1.2146
24	90 3/4	0.0429	0.0428	0.0001	0.0388	0.0379	0.0164	0.0001	0.8433	1.3012	1.2422

(c) 수정 지수형

국명 : 신 촌 1332

REGRESSION DIAGNOSTICS

모형 : 수정지수형

NO	QURT	실측 T	예측 T	RESIDUAL	STAD RES	STUD RES	DFFITS(I)	D(I)	AP(I)	COVRATIO	FVARATIO
1	84 4/4	0.0387	0.0388	-0.0001	-0.0435	-0.0425	-0.0183	0.0002	0.8433	1.3012	1.2421
2	85 1/4	0.0390	0.0391	-0.0001	-0.0167	-0.0163	-0.0065	0.0000	0.8625	1.2725	1.2147
3	85 2/4	0.0403	0.0393	0.0010	0.3011	0.2948	0.1089	0.0062	0.8762	1.2311	1.1858
4	85 3/4	0.0385	0.0395	-0.0010	0.2803	-0.2744	-0.0937	0.0046	0.8923	1.2158	1.1657
5	85 4/4	0.0398	0.0397	0.0001	0.0354	0.0346	0.0107	0.0001	0.9034	1.2067	1.1519
6	86 1/4	0.0387	0.0399	-0.0012	-0.3352	-0.3284	-0.0958	0.0048	0.9169	1.1757	1.1309
7	86 2/4	0.0414	0.0401	0.0013	0.3733	0.3659	0.0988	0.0051	0.9261	1.1627	1.1169
8	86 3/4	0.0411	0.0403	0.0008	0.2312	0.2262	0.0568	0.0017	0.9384	1.1610	1.1109
9	86 4/4	0.0419	0.0405	0.0014	0.3995	0.3917	0.0920	0.0044	0.9408	1.1414	1.0974
10	87 1/4	0.0403	0.0407	-0.0004	-0.1036	-0.1013	-0.0225	0.0003	0.9524	1.1506	1.0989
11	87 2/4	0.0397	0.0409	-0.0011	-0.3248	-0.3181	-0.0679	0.0024	0.9518	1.1366	1.0902
12	87 3/4	0.0275	0.0411	-0.0136	-3.7802	-6.2389	-1.3044	0.3123	0.3358	0.1407	0.3832
13	87 4/4	0.0408	0.0412	-0.0004	-0.1240	-0.1212	0.0253	0.0003	0.9574	1.1439	1.0927
14	88 1/4	0.0419	0.0414	0.0005	0.1305	0.1275	0.0272	0.0004	0.9556	1.1468	1.0946
15	88 2/4	0.0510	0.0418	0.0094	2.6234	3.0919	0.6874	0.1701	0.6548	0.5439	0.7555
16	88 3/4	0.0436	0.0418	0.0018	0.5039	0.4952	0.1163	0.0070	0.9367	1.1315	1.0927
17	88 4/4	0.0412	0.0420	-0.0008	-0.2251	0.2143	-0.0543	0.0015	0.9386	1.1615	1.1112
18	89 1/4	0.0396	0.0422	-0.0026	-0.7257	-0.7116	-0.1938	0.0192	0.9097	1.1218	1.0971
19	89 2/4	0.0416	0.0423	-0.0007	-0.2120	-0.2074	-0.0605	0.0019	0.9197	1.1860	1.1344
20	89 3/4	0.0417	0.0425	-0.0008	-0.2356	-0.2305	-0.0727	0.0028	0.9071	1.2007	1.1491
21	89 4/4	0.0436	0.0427	0.0009	0.2599	0.2543	0.0889	0.0039	0.8928	1.2181	1.1663
22	90 1/4	0.0419	0.0429	-0.0010	-0.2832	-0.2772	-0.1024	0.0055	0.8766	1.2383	1.1863
23	90 2/4	0.0427	0.0430	-0.0003	-0.1016	-0.0993	-0.0396	0.0008	0.8621	1.2713	1.2141
24	90 3/4	0.0429	0.0432	-0.0003	-0.0941	-0.0919	-0.0396	0.0008	0.8430	1.3003	1.2417

(c) 수정 지수형

국명 : 신 촌 1332

REGRESSION DIAGNOSTICS

모형 : 성장곡선형

NO	QURT	실측 T	예측 T	RESIDUAL	STAD RES	STUD RES	DFFITS(I)	D(I)	AP(I)	COVRATIO	FVARATIO
1	84 4/4	0.0387	0.0387	0.0000	-0.0144	-0.0141	-0.0061	0.0000	0.8433	1.3014	1.2422
2	85 1/4	0.0390	0.0389	0.0001	0.0159	0.0155	0.0062	0.0000	0.8625	1.2725	1.2147
3	85 2/4	0.0403	0.0391	0.0012	0.3376	0.3307	0.1222	0.0078	0.8753	1.2346	1.1845
4	85 3/4	0.0385	0.0393	-0.0008	-0.2429	-0.2376	-0.0812	0.0034	0.8931	1.2190	1.1667
5	85 4/4	0.0398	0.0395	0.0003	0.0763	0.0743	0.0235	0.0003	0.9098	1.2062	1.1517
6	86 1/4	0.0387	0.0397	-0.0010	-0.2937	-0.2875	-0.0839	0.0037	0.9180	1.1816	1.1323
7	86 2/4	0.0414	0.0399	0.0015	0.4187	0.4107	0.1109	0.0064	0.9246	1.1589	1.1161
8	86 3/4	0.0411	0.0401	0.0010	0.2775	0.2716	0.0682	0.0024	0.9374	1.1585	1.1097
9	86 4/4	0.0419	0.0402	0.0016	0.4474	0.4392	0.1032	0.0055	0.9391	1.1371	1.0954
10	87 1/4	0.0403	0.0405	0.0002	-0.0562	-0.0549	-0.0122	0.0001	0.9528	1.1514	1.0992
11	87 2/4	0.0397	0.0407	-0.0010	-0.2775	-0.2716	-0.0580	0.0018	0.9530	1.1395	1.0916
12	87 3/4	0.0275	0.0409	-0.0134	-3.7428	-6.0611	-1.2685	0.3062	0.3480	0.1512	0.3972
13	87 4/4	0.0408	0.0411	0.0003	-0.0754	0.0737	-0.0154	0.0001	0.9579	1.1449	1.0931
14	88 1/4	0.0419	0.0413	0.0006	0.1799	0.1759	0.0376	0.0007	0.9550	1.1442	1.0938
15	88 2/4	0.0510	0.0414	0.0096	2.6881	3.1907	0.7094	0.1775	0.6418	0.5224	0.7404
16	88 3/4	0.0436	0.0416	0.0020	0.5531	0.5450	0.1281	0.0085	0.9345	1.1260	1.0900
17	88 4/4	0.0412	0.0418	-0.0006	-0.1728	-0.1699	-0.0426	0.0010	0.9394	1.1635	1.1123
18	89 1/4	0.0396	0.0420	-0.0024	0.6806	-0.6720	-0.1815	0.0169	0.9124	1.1285	1.1004
19	89 2/4	0.0416	0.0422	0.0006	-0.1663	-0.1625	-0.0474	0.0012	0.9204	1.1875	1.1353
20	89 3/4	0.0417	0.0424	-0.0007	-0.1910	-0.1868	-0.0589	0.0018	0.9079	1.2028	1.1501
21	89 4/4	0.0436	0.0425	0.0011	0.3048	0.2984	0.1019	0.0054	0.8917	1.2152	1.1649
22	90 1/4	0.0419	0.0427	0.0008	-0.2413	-0.2361	-0.0872	0.0040	0.8775	1.2408	1.1875
23	90 2/4	0.0427	0.0429	-0.0007	-0.0608	-0.0594	-0.0237	0.0003	0.8623	1.2721	1.2146
24	90 3/4	0.0429	0.0431	0.0002	-0.0558	-0.0537	-0.0232	0.0003	0.8432	1.3010	1.2422

(d) 성장곡선

[표 3-3] 모형별 각 기법에 대한 측도들의 값

다음은 영향을 크게 주는 측정치를 검출하는 방법으로서 DFFITS(i)를 검토해보면 $2\sqrt{(k+1)/n} = 2\sqrt{2/24} = 0.5574$ 이므로, | DFFITS(i) | 가 이 값보다 큰 것들은 스트던트화잔차에서와 같은 분기의 값들이 영향을 크게주는 측정치로 판정이 되며, D(i)의 경우는 F(2,22;0.5)=0.716이므로, 이 값보다 큰 D(i)값은 없으나 일반적으로 볼때 87년 3/4분기가 그중에서 제일 큰 값이므로 영향을 준다고 볼 수 있으며, 또한 AP(i)의 값은 가장 작은 값이 영향을 크게 주는 측정치인데, AP(i)값들 중 가장 작은 값은 87년 3/4분기 값이며, 그 다음이 88년 2/4분기임을 알수 있다. COVRATIO는 COVRATIO $\geq 1+3(k+1)/n = 1+3(2)/24 = 1.25$ 이거나 COVRATIO $< 1+3(k+1)/n = 1 - 6/24 = 0.75$인 경우는 영향을 주는 측정치로 볼 수 있는데 1-2번째, 12,14번째, 23-24 번째가 모두 영향을 주는 측정치라 판단 할 수 있으며, 특히 12 번째 값이 가장 크게 영향을 주는 측정치임을 알수 있다.

다음은 FVARATIO 로서 FVARATIO $\geq 1 + (2k+3)/n = 1 + 5/24 = 0.8750$이거나 FVARATIO $\leq 1 - 3/n = 1 - 3/24 = 1.2083$인 경우는 영향을 주는 측정판단할 수 있는데 COVRATIO 와 동일하다.

위에서 검토한 모든 기법의 측도들을 분석해 보

면 12,14번째 측정치 자료는 이상치이면서 영향을 크게 주는 측정치로 판단이 되며, 1,2,23,24 번째 측정치는 영향을 크게주는 측정치로 판단할 수 있으며, 위와같이 12,14번째 이상치 자료는 보다 세밀한 분석.검토를 요한다.

그러므로 보다 정확한 트래픽을 예측하기 위하여, 이상치로 판정된 측정치는 제거하거나 수정하여야 하며 그 결과는 기존 트래픽 예측 방법 보다 정확한 값을 갖는 다고 볼 수 있다. 그 이유는 F-통계량 값에 의해 모형의 타당성이 입증되기 때문인대 모형의 타당성에 대한 논의는 다음절에서 하기로하자.

Ⅲ.3 예측 결과 비교

추정된 회귀계수를 이용하여 예측한 결과와 회귀진단을 이용하여 이상치를 제거한 후 적합된 회귀모형을 이용하여 예측한 결과에 대해 비교하고자 한다.

모형에 대한 타당성 검정은 F-통계량 값을 이용하는데, 먼저 이상치를 제거하기 전과 이상치를 제거한 후의 모형에 대한 타당성을 검토하기 위하여 다음 [표 3-4]의 F-통계량 값을 비교하여 보자.

[표 3-4] F-통계량 값 비교

모 형	F - 통계량 값	
	이상치 제거 전	이상치 제거 후
직 선	2.9369	25.7884
지 수	2.8872	25.6442
수정지수	2.7292	26.0946
성장곡선	2.8817	26.0332

유의수준 α =0.05에서 모형의 타당성 검정에 대한 기각치는 F(1,22;0.05) = 4.28 이므로 이상치 제거전의 F-통계량 값은 기각치 보다 작은 값이기 때문에 모형이 타당하지 않으며, 이상치를 제거한 후의 F-통계량값은 기각치 보다 크기 때문에 모형이 타당하다는 것을 [표 3-4]에서 알 수 있다.

이와 같이 이상치를 제거하고 가입 전화 트래픽을 예측하는 것이 보다 정확하고 타당성이 있는것이므로, 회귀진단을 이용하여 불필요한 측정치를 검출해 내고 가입전화 트래픽을 예측하는 것이 유의하다고 본다

Ⅳ.결 론

본 논문에서는 가입 전화 트래픽을 예측하는대 있어서 이상치를 제거하기 전에 예측한 결과와 이상치를 제거한 후 예측결과를 비교할 때 제거후는 모형의 타당성이 유의하지만 제거전은 타당성이 없는 것으로 나타났다.

그러므로 모든 자료는 회귀진단기법을 이용하여 먼저 이상치와 영향을 크게 주는 자료를 분석검토하여 이상치는 제거를 하고 가입전화 트래픽을 예측하는 것이 신뢰성이 있으므로 이 방법을 이용하는 것이 좋다고 생각된다.

또한 기존 방법에서 탈피하여 보다 나은 시계열 분석방법을 연구하여 정확하게 트래픽을 예측함으로서 가입자에게 양질의 서비스를 제공할 수 있고 경제적인 통신투자사업을 할 수 있게 하기 위하여 계속 연구가 되어야 할 것으로 사료된다.

[부록 1] 각 유의 수준에서의 기각치

유의수준 α = 0.10

표본의 크기 n	측정변수 k									
	1	2	3	4	5	6	8	10	15	25
5	1.07									
6	2.06	1.89								
7	2.10	2.02	1.90							
8	2.11	2.12	2.03	1.91						
9	2.14	2.20	2.13	2.05	1.93					
10	2.30	2.26	2.21	2.16	2.06	1.92				
12	2.39	2.37	2.33	2.29	2.24	2.17	1.93			
14	2.47	2.46	2.42	2.39	2.36	2.32	2.19	1.94		
16	2.53	2.51	2.50	2.47	2.45	2.42	2.34	2.20		
18	2.58	2.57	2.56	2.54	2.53	2.50	2.44	2.36		
20	2.63	2.62	2.61	2.59	2.58	2.56	2.52	2.46	2.31	
25	2.72	2.72	2.71	2.70	2.69	2.68	2.66	2.63	2.60	
30	2.80	2.79	2.79	2.78	2.77	2.77	2.75	2.73	2.68	2.13
35	2.86	2.86	2.85	2.85	2.84	2.84	2.82	2.81	2.77	2.55
40	2.91	2.91	2.90	2.90	2.90	2.89	2.88	2.87	2.84	2.72
45	2.95	2.95	2.95	2.95	2.94	2.94	2.93	2.93	2.90	2.82
50	2.99	2.98	2.99	2.98	2.98	2.98	2.98	2.97	2.96	2.88
60	3.04	3.04	3.05	3.05	3.05	3.06	3.06	3.06	3.03	3.00
70	3.11	3.11	3.11	3.11	3.11	3.11	3.10	3.10	3.09	3.07
80	3.16	3.16	3.16	3.16	3.16	3.16	3.16	3.15	3.14	3.12
90	3.19	3.20	3.19	3.19	3.19	3.19	3.19	3.19	3.18	3.17
100	3.23	3.23	3.23	3.23	3.23	3.23	3.23	3.22	3.22	3.21

유의수준 $\alpha = 0.05$

표본의 크기 n	독립변수 k									
	1	2	3	4	5	6	8	10	15	25
5	1.92									
6	2.07	1.93								
7	2.18	2.08	1.94							
8	2.28	2.20	2.10	1.94						
9	2.36	2.29	2.21	2.10	1.95					
10	2.42	2.37	2.31	2.22	2.11	1.95				
12	2.52	2.49	2.45	2.39	2.33	2.24	1.96			
14	2.61	2.58	2.55	2.51	2.47	2.41	2.25	1.98		
16	2.68	2.66	2.63	2.60	2.57	2.53	2.43	2.26		
18	2.75	2.72	2.70	2.68	2.65	2.62	2.55	2.44		
20	2.78	2.77	2.76	2.74	2.72	2.70	2.64	2.53	2.16	
25	2.89	2.88	2.87	2.86	2.84	2.83	2.80	2.76	2.60	
30	2.96	2.96	2.95	2.94	2.94	2.93	2.90	2.88	2.79	2.17
35	3.03	3.02	3.02	3.01	3.00	3.00	2.98	2.97	2.91	2.64
40	3.08	3.08	3.07	3.07	3.06	3.06	3.05	3.03	3.00	2.84
45	3.13	3.12	3.12	3.12	3.11	3.11	3.10	3.09	3.06	2.96
50	3.17	3.16	3.16	3.16	3.16	3.15	3.14	3.14	3.11	3.04
60	3.23	3.23	3.23	3.23	3.22	3.22	3.22	3.21	3.20	3.15
70	3.28	3.29	3.28	3.28	3.28	3.28	3.27	3.27	3.26	3.23
80	3.33	3.33	3.33	3.33	3.33	3.33	3.32	3.32	3.31	3.29
90	3.37	3.37	3.37	3.37	3.37	3.37	3.36	3.36	3.36	3.34
100	3.41	3.41	3.40	3.40	3.40	3.40	3.40	3.40	3.39	3.38

유의수준 $\alpha = 0.01$

표본의 크기 n	독립변수 k									
	1	2	3	4	5	6	8	10	15	25
5	1.98									
6	2.17	1.98								
7	2.32	2.17	1.98							
8	2.44	2.32	2.18	1.98						
9	2.54	2.44	2.33	2.18	1.99					
10	2.62	2.55	2.46	2.33	2.18	1.99				
12	2.76	2.70	2.64	2.56	2.46	2.34	1.99			
14	2.86	2.82	2.76	2.72	2.65	2.57	2.36	1.99		
16	2.95	2.90	2.86	2.84	2.79	2.73	2.58	2.35		
18	3.02	3.00	2.97	2.94	2.90	2.85	2.76	2.69		
20	3.08	3.06	3.04	3.01	2.98	2.95	2.87	2.76	2.20	
25	3.21	3.19	3.18	3.16	3.14	3.12	3.07	3.01	2.78	
30	3.30	3.29	3.28	3.26	3.25	3.24	3.21	3.17	3.04	2.21
35	3.37	3.36	3.35	3.34	3.34	3.33	3.30	3.28	3.19	2.81
40	3.43	3.42	3.42	3.41	3.40	3.40	3.38	3.36	3.30	3.05
45	3.48	3.47	3.47	3.46	3.46	3.45	3.44	3.43	3.38	3.22
50	3.52	3.52	3.51	3.51	3.50	3.49	3.48	3.48	3.45	3.34
60	3.60	3.59	3.59	3.59	3.58	3.58	3.57	3.56	3.54	3.48
70	3.65	3.65	3.65	3.65	3.64	3.64	3.64	3.63	3.61	3.57
80	3.70	3.70	3.70	3.70	3.69	3.69	3.69	3.68	3.67	3.64
90	3.74	3.74	3.74	3.74	3.74	3.74	3.73	3.73	3.72	3.70
100	3.78	3.78	3.78	3.77	3.77	3.77	3.77	3.77	3.76	3.74

참 고 문 헌

1. Andrew, D.F. "Significance tests based on residuals." Biometrica. 58, pp139-148. 1971.

2. Andrews, D.F. and Pregibon, D. "Finding the outliers that matter." Journal of the Royal Statistical Society. B40, pp85-93. 1978.

3. Behnken, D.W., and Draper, N.R. "Residuals and their variance pattern." Technometrics. 14, pp101-111. 1972.

4. Bersley, D.A., Kuh, E. and Welsch, R.E. "Regression diagnostics." New York. John Wiley. 1983.

5. Cook, R.D. "Detection of influential observation in linear regression." Technometrics. 19, pp15-18. 1977.

6. Draper, N.R. and John, J.A. "Infiuential observations and outliers in regression." Technometrics.23, pp21-26. 1981.

7. Hoaglin, D.C. and Welsch, R.E. "The hat matrix in regression and ANOVA." American Statistican.32, pp17-22. 1978.

8. Lund, R.E. "Tables for an approximate test for outliers in linear models."Technometrics. 17, pp473-476.1975.

9. Prescott, P. "An approximate test for outliers in linear models." Technometrics.17, pp 129-132. 1975.

10. Stefansky, W. "Rejecting outliers by maximum normed residual." Ann. Statist. 42, pp35-45. 1971.

11. Thompson, W.R. "On a criteria for the rejection of outliers and the distribution of the ratio of the deviation to the sample standard deviation." Ann. Math. Statist. 6, pp214-219. 1935.

12. Tietjen, G.L., Moore, R.H. and Beckmin, R.J. "Testing for a single outlier in simple linear regression." Technometrics.15, pp717-721. 1973.