

■ 연구논문

국내 G3 팩시밀리 화상품질에 관한 통계 분석

이성덕

충북대학교 통계학과

권세혁

한남대학교 응용통계학과

Statistical Analysis on the Measurement of the Image Quality of G3 facsimile

Sung Duck Lee

Dept. of statistics, Chungbuk National University

Sehyg Kwon

Dept. of Applied Statistics, Hannam University

Abstract

Two user groups, expert and non-expert, are sampled to measure the image quality of G3 facsimile. A ITU-TS tset chart No. 2 has been transmitted among some selected cities and evaluated by user groups. Their subjective evaluation to the image quality is quantified by Mean Opinion Score method. There is highly significant difference in the image quality between expert and non-expert. From modified logit model, it is concluded that there is no significance in two considered factors, the effects of the number of links and transmission time. The derived percent curves show that 80% of non-experts(90% of expert) is considering the image quality of G3 facsimile "fair, good, or excellent".

1. 서론

그룹 3 팩시밀리를 사용하여 문서를 보내는 경우 아날로그 신호를 사용하고 있으나, 보내고자 하는 문자나 그림 등의 화상이 팩시밀리에 입력되면 디지털 신호로 읽혀진다. 그러므로 팩시밀리 안에 내장된 모뎀에 의해 디지털 신호는 아날로그 신호로 변조된 후 전화

망에 의해 송신되며, 송신된 신호는 다시 디지털 신호로 복조되어 송신된 화상이 감열방식에 의해 인쇄된다. 팩시밀리는 입력된 정보를 프레임 단위로 보내지지만 프레임에 포함된 정보의 내용에 따라 각 프레임은 서로 다른 수의 비트로 인식되어 전송된다. 왜냐하면 주사선과 부주사선으로 등분된 한 펄은 하나의 정보로 인식되어 디지털 부호화 하지만 같은 정보가 반복되면 송신시간 절약을 위해 독자적 부호 방식을 가지므로 같은 길이의 프레임이더라도 디지털 부호화한 후 비트의 갯수는 상이하게 된다.

전송된 비트에 에러가 발생하더라도 수신된 화상이나 문자가 손상되지 않을 수 있으므로 팩시밀리에 있어서 화상품질에 영향을 미치는 주요 인자는 비트에러보다 프레임에러를 고려한다. 그러나 프레임 내에서 하나의 비트에 에러가 생겨도 그 프레임은 에러가 발생한 것으로 간주하게 되어 실제 이용자가 느끼는 화상품질을 측정하는데는 어려움이 있다. 한 페이지에서 연속적으로 4개 이상 프레임 에러들이 한 번 이상 발생한 경우, 12개 이상의 프레임 에러가 발생한 경우, 그리고 연속적으로 2개나 3개의 프레임 에러들의 발생이 3번 이상인 경우가 아니면 화상을 판독하는데 큰 어려움이 없다는 것이 ITU-TS (1993)에 연구 발표되었다. 그리고 전송된 화상에 있어서 프레임에러를 측정하는데 어려움이 있으며 프레임에러의 형태에 따라 화상품질이 달라지므로, 본 연구에서는 팩시밀리 화상품질을 측정하기 위해 프레임에러보다는 주관적 평가 방법의 하나인 MOS (Mean Opinion Score) 방법을 사용하여 화상품질을 측정하였다.

MOS 방법은 5점 척도를 사용하여 응답자의 주관적인 성향을 계량화시키는 평가방법이다. 측정하기를 원하는 항목을 설정하고, 각 설문 문항의 보기를 “매우 좋다”, “좋다”, “보통이다”, “나쁘다”, “매우 나쁘다”의 순으로 배열하여 응답자의 평가를 조사한 후, 각 각을 5부터 1까지의 점수로 환산하여 측정한다. i 를 평가점수, P_i 를 i 점수의 상대뒀수라고 하면 각 문항의 평가점수의 평균 (MOS)과 분산 (VAROS)은 아래와 같이 정의된다.

$$\begin{aligned} \text{MOS} &= \sum_{i=1}^5 i P_i \\ \text{VAROS} &= \sum_{i=1}^5 (i - \text{MOS})^2 P_i \end{aligned} \quad (1)$$

비트에러는 화상품질에 직접적인 영향을 미치지 않으나, 일반적으로 비트에러가 커질수록 화상품질은 저하되므로 비트에러에 영향을 주는 요인들을 팩시밀리 화상품질에 영향을 미치는 요인들로 고려할 수 있다. 한국전자통신연구소 품질공학연구실(1992)의 연구에 의하면 송 수신 간에 거쳐야 하는 교환기의 수가 많을수록, 시간대별 통화량이 많을수록 비트에러가 커지므로 이 두 요인들이 화상품질에 영향을 미치는지를 로짓모형을 이용하여 분석하고자 한다. 우리나라의 경우 통화량이 폭주하여 우회 선로를 사용하지 않는다면 시내 전화망의 교환기의 수는 1이나 2이며 시외 전화망에서는 3 혹은 4이다. 그리고 통화량이 많은 시간은 오전 9시부터 오전 12시 그리고 오후 6시부터 오후 9시까지로 간주되므로, 같은 방법으로 시간대를 통화량의 효과로 분석하였다. 다른 요인들로는 교환기종과 전송 선로의 길이나 종류 등을 고려할 수 있으나 영향이 극히 미미하다고 판단되거나 혹은 측정상 어려움이 있어서 팩시밀리 화상품질에 영향을 주는 요인으로는 위의 두가지만 고려하였다.

시험 화상으로 사용된 길이 297mm이고 폭 210mm인 ITU-TS 표준 시험 도표 No. 2는 문자의 판독성에 대한 일반적인 평가용으로 ITU-TSC(1988)에 권고되고 있는데, 프레임 왜곡 정도, 최소 허용 해상도, 수평 수직 해상도, 시험패턴 표면의 균일성, 그리고 문자 판독의 한계 등을 측정할 수 있는 도형이나 문자로 만들어져 있다. 그러므로 종합적인 화상 품질을 측정하기 위해서는 각 항목들을 하나의 5점 척도 설문 문항으로 만들어 평가하고, 각 항목에 가중치를 고려하여 측정한다. 그러나 권세혁과 황건(1994)은 상관분석을 통해 종합적인 품질만을 묻는 하나의 문항만으로 ITU-TS 표준 시험 도표 No. 2가 측정하는 여러 문항을 포함한 설문지를 이용한 측정 결과를 대치할 수 있다는 것을 보였고, 그러므로 국내 G3 팩시밀리 화상 품질을 측정하기 위해 각 항목의 평가 점수에 가중치를 고려하여 하나의 평가점수를 산출하는 대신 전체적인 느낌을 묻는 항목의 평가점수를 전송된 시험화상의 화상 품질을 측정하는 계량지표로 사용하였다.

본 연구는 전자통신 분야에 있어서 팩시밀리 화상 품질과 같이 주관적인 평가로만 측정이 가능한 항목들에 대하여 주관적인 평가를 객량화하거나 측정값의 해석방법, 그리고 요인분석방법을 제시하고자 하는데 또 하나의 목적이 있다.

전자통신 분야에 종사하는 전문가 그룹과 일반 팩시밀리 사용자인 비전문가 그룹으로 나누어 두 그룹간의 화상 품질 평가의 차이에 대한 유의성 검정, 로짓모형을 이용하여 화상 품질에 영향을 미치는 요인들의 유의성 검정, 그리고 누적퍼센트 곡선을 이용하여 두 그룹이 느끼는 국내 G3 팩시밀리 화상 품질을 측정 비교하였다.

2. 통계적 분석방법

두 그룹간의 평가 점수에 대한 유의성검정은 일원분산분석을 이용하되 종속변수는 평가점수의 평균을 사용하며, 화상 품질에 영향을 미친다고 고려된 두 요인들의 결합조건들을 반복수로 사용하였다. 그러므로 일원분산분석에서 총 자료 수는 8개이다. 로짓모형을 이용하여 요인들의 유의성을 검정한 결과, 요인들은 유의하지 않았으므로 요인들의 결합조건을 반복수로 사용하였다.

표본의 크기가 충분히 크면 각 점수에 대해 평가 점수의 상대 도수를 구할 수 있으므로 Aldrich와 Nelson(1984)의 기본적인 로짓모형을 다음과 같이 변형된 로짓모형을 제시하고 그 결과를 이용하여 화상 품질에 영향을 미치는 요인들의 영향을 분석한다. Y 를 종속 변수(평가점수), N 을 교환기수 효과, T 를 통화량 효과, NT 를 교호작용이라 하면 변형된 로짓 회귀 모형은 아래와 같다.

$$\begin{aligned} & \text{Ln}\left(\frac{\text{Pr}(Y=s)}{\text{Pr}(Y=5)}\right) \\ & = \mu + L_i + T_j + LT_{ij} + \varepsilon_{sij}. \end{aligned} \quad (2)$$

여기서 i 는 1, 2, 3, 그리고 4(링크 수)이고, j 는 0(통화량이 많은 시간대) 또는 1(통화량이 적은 시간대), 그리고 s 는 1, 2, 3, 그리고 4의 평가점수이다. 그러므로 위의 변형된 로

짓 회귀 모형의 독립변수는 각 결합 조건에서 1점부터 4점까지의 상대도수(relative frequency)로 측정된다. 종속 변수가 연속변수로 변형이 되었으므로 요인들에 대한 유의성 검정은 일반적인 분산분석을 이용할 수 있다.

두 요인의 결합 조건에 따라 실제 전송된 화상에 대해 응답자의 응답 결과 얻은 평가 점수에 의해 점수 히스토그램을 얻을 수 있다. 이 점수 히스토그램들을 정규 분포 곡선에 근사시킨 후 각 정규분포의 평균과 분산을 이용하여 누적 퍼센트 곡선을 구하는데 그 순서를 정리하면 다음과 같다.

순서 1: 표본의 평균과 표준편차 계산

두 요인들의 결합 조건에 의해 전송된 시험화상의 평가점수의 평균과 표준편차를 식 (1)을 이용하여 구한다.

순서 2: 점수 히스토그램을 정규분포로 근사

측정 결과 각 결합 조건에서 얻은 점수 히스토그램을 정규 분포 곡선으로 근사시킨다. 이산인 점수 영역을 점수 1을 $-\infty$ 부터 1.5까지, 점수 2를 1.5부터 2.5까지, 점수 3을 2.5부터 3.5까지, 점수 4를 3.5부터 4.5까지, 그리고 점수 5를 4.5부터 ∞ 까지 구간으로 나눈다. 그리고 근사된 정규 분포 곡선의 평균 (μ)과 표준편차 (σ)는 다음 연립 방정식의 근으로 구한다.

$$\begin{aligned} \text{MOS} &= 5 - \sum_{i=1}^4 \Phi\left(\frac{i+0.5-\mu}{\sigma}\right) \\ \text{VAROS} &= 25 - \sum_{i=1}^4 (2i+1) \Phi\left(\frac{i+0.5-\mu}{\sigma}\right) \\ &\quad - (\text{MOS})^2. \end{aligned} \quad (3)$$

순서 3: 통합 분산 (pooled variance) 추정

두 요인들의 결합조건 내에 근사된 정규 분포 곡선의 분산들을 이용하여 통합 분산을 아래와 같이 구한다. n_i 및 σ_i^2 는 각 결합조건에 대한 자료의 크기 및 분산이며 측정 자료의 크기이다.

$$\hat{\sigma}^2 = \sum_{i=1}^8 \frac{n_i \sigma_i^2}{n} \quad (4)$$

순서 4: 정규 분포 곡선의 평균들을 조정

순서 3에서 구한 통합 분산의 제곱근 (표준편차)을 식 (3)에 대입하여 정규 분포의 평균들을 다시 구한다. 통합 분산과 평균을 구한 이유는 누적퍼센트 곡선의 근사를 위한 자료들의 분포 모수를 같게 하는데 있다.

순서 5: 누적 퍼센트 곡선

순서 4에서 다시 구한 정규 분포의 평균들을 독립변수로 하고 각 그룹의 “매우 좋다”, “좋다 혹은 매우 좋다”, “보통이다, 좋다 혹은 매우 좋다”, “나쁘다, 보통이다, 좋다 혹은 매우 좋다”, “매우 나쁘다, 나쁘다, 보통이다, 좋다 혹은 매우 좋다” 5개의 누적 퍼센트를 종속 변수로 하여 누적 퍼센트 곡선을 구한다. 누적 퍼센트 곡선은 다항 함수에 의해 근사된다. 근사를 위해 사용되는 다항 함수의 차수는 요인들의 결합조건의 수에 의존한다.

순서 6: 팩시밀리 화상품질 추정

전문가 그룹과 비전문가 그룹에 있어서 평가점수의 차이가 유의하면 각각 근사화된 누적 퍼센트 곡선을 구하여 전문가 그룹이 느끼는 팩시밀리 화상품질과 비전문가의 화상 품질을 측정하고, 유의하지 않으면 두 그룹을 합쳐 하나의 누적퍼센트 곡선을 구하여 국내 전체의 화상 품질을 측정할 수 있다.

3. 자료분석

교환기 수와 통화량의 결합조건에 따라 전송 지역 (서울, 대구, 대전)과 시간을 설정하여 실제 선로에서 국내 팩시밀리 기종을 이용하여 ITU-TS 표준 시험 도표를 전송하여 시험 화상들을 수집하였다. 통화량의 차이는 언급한대로 시간대를 나누어 전송함으로써 교환기 수의 효과의 수준과 통화량의 효과의 수준에 의한 8개 결합 조건하에서 시험 화상을 수집하였다. 수집된 시험 화상 중 각 그룹 내에서 6장씩 총 48장의 표본 시험화상을 추출하였다. 전자통신 분야에 재직하는 연구원 중 8명을 선택하여 전문가 그룹으로 하고, 대학생 중 65명을 선택하여 비전문가 그룹으로 하였다. 각 응답자에게 추출된 시험 화상을 보고 ITU-TS가 권고하는 측정 항목에 따라 작성된 설문지를 작성하게 하였다. 전문가 그룹을 먼저하고 설문 조사한 결과 48장을 모두 평가하는데 상당한 시간이 소요되어 응답의 신뢰성이 떨어질 것이라고 판단되어 비전문가 그룹에게는 48장 중 24장만 평가하였다. 그러므로 전문가 그룹으로부터 384장의 설문지가 수집되었고 반면에 비전문가 그룹으로부터 1560장의 설문지가 수집되었다. 종합적인 화상 품질을 묻는 문항의 평가점수 평균과 표준편차를 구하여 <표 1>에 정리하였다.

응답자들의 신뢰성을 조사하기 위하여 Cronbach 계수를 이용하였는데, 예상대로 전문가 그룹의 계수는 0.8515로 비전문가 그룹의 0.7673보다 높았다. 일반적으로 Cronbach 계수가 0.6이상이면 신뢰성이 있다고 판단하므로 설문에 응답한 응답자들의 신뢰도는 높다고 할수있다. 표본의 크기가 16인 일원분산분석 결과, 전문가 그룹과 비전문가 그룹의 평가점수 평균의 차이는 99% 유의 ($F_{값} = 44.99$)하였으므로 전문가 그룹이 비전문가 그룹보다 화상 품질을 높게 평가하고 있음을 알 수 있다. 그리고 비전문가 그룹과는 달리 전문가 그룹의 응답자 중 누구도 화상 품질을 1점이라고 평가한 경우는 없었다. 이처럼 전문가 그룹의 평가가 높은 것은 팩시밀리에 대해 잘 알고 있으므로 어느 정도의 판독이 힘든 화상에 대해 관대하기 때문일 것이라고 사료된다.

(표 1) 화상 품질 평가 점수의 평균과 표준편차
(Mean and Standard deviation of Opinion Score)

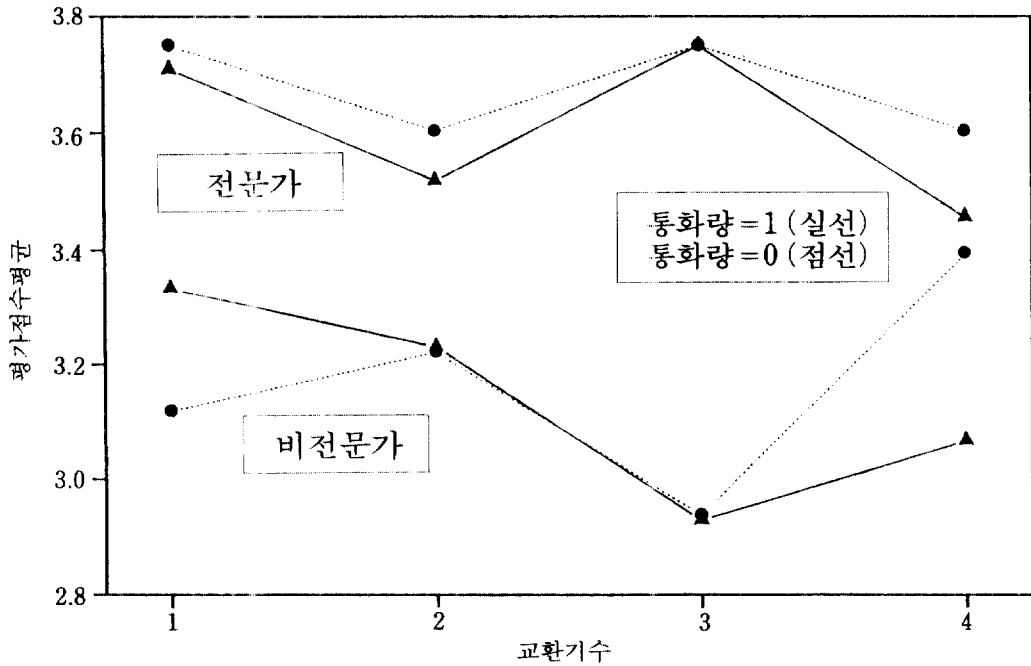
| 교환기수 | 통화량 | 전문가 그룹 | | 비전문가 그룹 | |
|------|-----|--------|-------|---------|-------|
| | | 평균 | 표준편차 | 평균 | 표준편차 |
| 1 | 0 | 3.708 | .6510 | 3.333 | .8165 |
| 1 | 1 | 3.750 | .7293 | 3.119 | .8403 |
| 2 | 0 | 3.521 | .7143 | 3.231 | .7819 |
| 2 | 1 | 3.604 | .7920 | 3.222 | .7251 |
| 3 | 0 | 3.750 | .6358 | 2.928 | .8547 |
| 3 | 1 | 3.750 | .6995 | 2.939 | .7641 |
| 4 | 0 | 3.458 | .8241 | 3.067 | .7976 |
| 4 | 1 | 3.604 | .8184 | 3.395 | .8141 |

〈표 2〉는 화상품질에 있어서 교환기 수의 효과와 통화량의 효과의 유의성을 검정하기 위해 각 그룹에 있어서 두수분포표를 작성한 것이다. 〈표 2〉의 자료를 이용하여 식 (2)의 로짓모형을 교호작용이 있는 경우와 없는 경우 분석하였는데, $Lm(0)$ 는 정의되지 않으므로 전문가 그룹에 있어서 자료의 크기는 24이고 비전문가의 경우는 30이었다. 분석결과 전문가 그룹이나 비전문가 그룹에 있어서 교호작용은 물론 두 주요인도 유의하지 않았다. 이 결론은 각 그룹의 결합조건의 평가점수의 평균을 나타낸 〈그림 1〉에서도 알 수 있다. 그러므로 교환기 수나 통화량은 화상품질에 영향을 미치지 않음을 알 수 있다. 앞에서 언급한 선로 상의 여러 특성들을 수량화하여 위의 로짓모형을 이용하여 화상품질에 영향을 미치는 인자를 찾는 연구가 앞으로 가능할 것이다.

전문가 그룹과 비전문가 그룹간에 화상품질 평가점수의 차이가 유의하므로 각 그룹에 있어서 누적퍼센트 곡선을 근사하여 화상품을 측정하였다. 주요인 (교환기 수, 통화량)들이 유의하지 않으므로 누적퍼센트 곡선의 범위도 작게 된다. 왜냐하면 평가점수의 평균의 범위가 작기 때문이다. 누적퍼센트 곡선을 근사하기 위하여 앞장에 기술된 순서대로 계산과정을 정리하면 〈표 3〉과 같으며, 두 그룹에 있어서 누적 퍼센트 곡선들은 〈그림 2〉에 있다. 그리고 전문가 그룹의 통합분산은 0.4796이고 비전문가 그룹의 통합분산은 0.5917이다.

〈 표 2 〉 평가점수 횡수분포표
(Frequency Table)

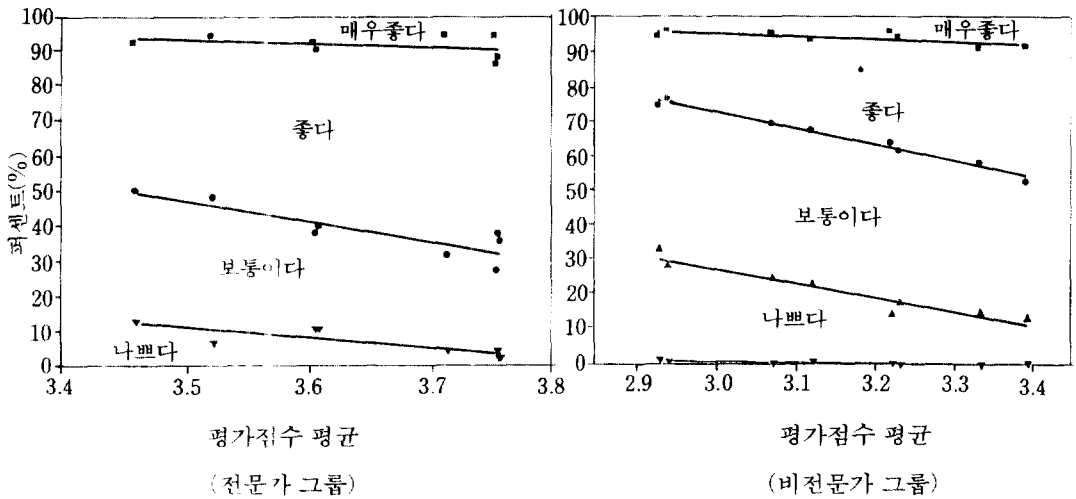
| 링크수 | 통화량 | 전문가 그룹 | | | | | 비전문가 그룹 | | | | |
|-----|-----|--------|------|------|------|------|---------|------|------|------|-----|
| | | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 1 | 0 | 0.0 | 4.2 | 27.1 | 62.5 | 6.3 | 0.0 | 14.9 | 44.1 | 33.8 | 7.2 |
| 1 | 1 | 0.0 | 2.1 | 35.4 | 47.9 | 14.6 | 1.0 | 22.2 | 45.4 | 26.8 | 4.6 |
| 2 | 0 | 0.0 | 6.3 | 41.7 | 45.8 | 6.3 | 0.0 | 17.9 | 44.6 | 33.8 | 3.6 |
| 2 | 1 | 0.0 | 10.4 | 27.1 | 54.2 | 8.3 | 0.5 | 13.9 | 50.5 | 33.0 | 2.1 |
| 3 | 0 | 0.0 | 4.2 | 22.9 | 66.7 | 6.3 | 1.5 | 32.0 | 42.3 | 20.6 | 3.6 |
| 3 | 1 | 0.0 | 2.1 | 33.3 | 52.1 | 12.5 | 1.0 | 27.7 | 49.2 | 20.5 | 1.5 |
| 4 | 0 | 0.0 | 12.5 | 37.5 | 41.7 | 8.3 | 0.5 | 24.4 | 45.6 | 26.9 | 2.6 |
| 4 | 1 | 0.0 | 10.4 | 29.2 | 50.0 | 10.4 | 0.5 | 12.8 | 40.0 | 40.0 | 6.7 |



〈 그림 1 〉 평가점수의 평균
(Mean of Opinion Score)

〈 표 3 〉 누적 퍼센트 곡선 근사 과정
(Procedure of CPC)

| 링크수 | 통화량 | 전문가 그룹 | | | | | | 비전문가 그룹 | | | | | |
|-----|-----|--------|-------|-------|-------|-------|-------|---------|-------|-------|-------|--|--|
| | | 순서 1 | | 순서 2 | | 순서 4 | 순서 1 | | 순서 2 | | 순서 4 | | |
| | | 평균 | STD | 평균 | STD | 평균 | 평균 | STD | 평균 | STD | 평균 | | |
| 1 | 0 | 3.708 | .6510 | 3.712 | .5951 | 3.715 | 3.332 | .8165 | 3.333 | .7712 | 3.334 | | |
| 1 | 1 | 3.750 | .7294 | 3.753 | .6801 | 3.757 | 3.117 | .8403 | 3.119 | .7949 | 3.121 | | |
| 2 | 0 | 3.521 | .7143 | 3.523 | .7395 | 3.524 | 3.232 | .7819 | 3.231 | .7300 | 3.232 | | |
| 2 | 1 | 3.604 | .7920 | 3.602 | .7361 | 3.607 | 3.221 | .7251 | 3.222 | .6736 | 3.222 | | |
| 3 | 0 | 3.750 | .6358 | 3.751 | .5533 | 3.756 | 2.929 | .8547 | 2.928 | .8172 | 2.929 | | |
| 3 | 1 | 3.750 | .6995 | 3.755 | .6434 | 3.759 | 2.939 | .7641 | 2.939 | .7084 | 2.940 | | |
| 4 | 0 | 3.458 | .8241 | 3.459 | .7796 | 3.461 | 3.068 | .7976 | 3.067 | .8106 | 3.071 | | |
| 4 | 1 | 3.604 | .8184 | 3.607 | .7770 | 3.610 | 3.395 | .8141 | 3.395 | .8326 | 3.394 | | |



〈 그림 2 〉 누적 퍼센트 곡선
(Cumulative percent Curve)

평가 점수의 평균의 범위가 좁으므로 평균누적 퍼센트 곡선은 평균의 선형 함수로 근사된 선형함수는 누적퍼센트를 나타내는데, 예를들어 위에서 두번째 직선 위의 평가자들이 화상품질을 “좋다” 혹은 “매우 좋다”로 평가할 확률이 된다.

4. 결론

본 연구에서는 G3 팩시밀리 국내 화상품을 측정하기 위해 전자 통신 분야에 종사하는 전문가 그룹과 일반 팩시밀리 사용자인 비전문가 그룹으로 나누어 설문지를 통한 주관적인 평가 방법인 MOS 방법을 이용하여 계량화하고 분석하였다.

일원 분산분석 결과 전문가 그룹과 비전문가 그룹간에 평가점수 평균의 차이는 유의하지 않았으므로 두 그룹을 나누어 각 그룹이 느끼는 화상품을 측정하였다. 화상품질에 있어서 교환기 수와 통화량의 효과를 로짓모형을 이용하여 분석한 결과 두 요인들은 화상 품질에 영향을 미치지 않았으나 다음과 같은 결론을 얻을 수 있다.

1. 팩시밀리 사용자의 30%는 팩시밀리 화상품질이 “좋다 혹은 매우 좋다”로 평가하며, 전자 통신 분야에 종사하는 전문가들의 40%는 팩시밀리 화상품질이 “좋다 혹은 매우 좋다”로 평가한다.
2. 팩시밀리 사용자의 80%나 전문가의 90%는 팩시밀리 화상품질이 보통 이상으로 평가하고 있다.
3. 기준을 만드는 기관이 국내 팩시밀리 사업자에게 팩시밀리 사용자의 50% 정도가 화상품질이 “좋다 혹은 매우 좋다”는 평가를 얻게 하려면 팩시밀리 사용자 중 표본을 추출하여 설문 조사 결과 평가 점수의 평균은 3.48이 되게 규정하여야 한다.

참고문헌

- [1] 권세혁, 황건 (1994), 팩시밀리화상품질 측정에 관한 연구, 전자통신연구소 동향분석 Vol. 30.
- [2] 품질공학연구실 (1992), 전기통신 서비스품질에 관한 연구, 한국 전자통신 연구소 2KR410002650F.
- [3] J. H. Aldrich and F. D. Nelson (1984), Linear Probability, Logit and Probit Models, Sage Publication.
- [4] ITU-TS (1988), Terminal Equipment and Protocols for Telematic Services, Series T. 21.
- [5] ITU-TS(1993), Quality as corrupted by transmission-induced scan errors, Draft Recommendation E. 45y.