

Comparison of Goodness-of-Fit Tests using Grouping Strategies for Multinomial Logit Regression Model

Mi Kyung Song^a · Inkyung Jung^{a,1}

^aDepartment of Biostatistics, Yonsei University College of Medicine

(Received July 11, 2013; Revised October 2, 2013; Accepted October 15, 2013)

Abstract

Several goodness-of-fit test statistics have been proposed for a multinomial logit regression model; however, the properties of the proposed tests were not adequately studied. This paper evaluates three different goodness-of-fit tests using grouping strategies, proposed by Fagerland *et al.* (2008), Bull (1994), and Pigeon and Heyse (1999). In addition, Pearson (1900)'s method is also examined as a reference. Simulation studies were conducted to evaluate the four methods in terms of null distribution and power. A real data example is presented to illustrate the methods.

Keywords: Multinomial logit regression, goodness-of-fit test, grouping strategy, null distribution.

1. 서론

로짓 회귀모형은 범주형 자료 형태의 종속변수와 범주형 또는 연속형 독립변수들 간의 관계를 나타내어 주는 방법으로 여러 다양한 연구 분야에서 사용된다. 특히 의학 연구에서는 종속변수가 범주형인 자료가 많이 존재하며, 때로는 연속형 자료를 범주화하여 결과를 해석하고자 하는 노력도 이루어지고 있다. 과거에 비해 점차 다항 종속변수에 대한 관심이 늘어나고 있는 추세인데, 이러한 자료의 분석을 가능하게 해주는 것이 다항 로짓 회귀모형이다. 그러나 다항 로짓 회귀모형의 적합 정도를 판단하는데 사용되는 분석 방법은 아직 구체적으로 연구되어 있지 않다.

본 연구에서는 다항 로짓 회귀모형의 적합성을 평가하기 위해 기존에 제시되어 있는 적합도(goodness-of-fit) 검정 방법들 중 그룹화 전략(grouping strategy)을 이용한 방법들에 대하여 확인, 비교, 평가하고자 한다. 기준으로써 로짓 회귀모형의 적합성을 관별하기 위해 일반적으로 사용되는 피어슨 χ^2 통계량, Hosmer와 Lemeshow (1980)의 방법과 피어슨 χ^2 통계량을 이용한 Fagerland 등 (2008)의 C_g 통계량, Hosmer와 Lemeshow (1980)의 예측 확률로부터 그룹화하는 전략을 확장하여 얻어진 \hat{C} 통계량 (Bull, 1994), 피어슨 χ^2 통계량의 과소산포(underdispersion)를 보정한 J^2 통계량 (Pigeon과 Heyse, 1999)을 고려한다. 이때 종속변수의 범주 수와 그룹의 수, 독립변수의 수를 다르게 설정한 후, 자료에 맞는 적절한 모형을 적합 시킬 때 각 통계량들은 자신의 귀무분포를 따르게 된다는 성질을 이용하여, 모의실험으로 얻어지는 통계량 값들의 귀무분포와 저자들이 제시한 분포를 비교하고자 한다. 또한 같은 상황 속에서 적절하지 않은 회귀모형을 적합 시켰을 때, 각 방법들이 이를 얼마나 잘 관별하는지 알아보하고자 한다.

¹Corresponding author: Assistant Professor, Department of Biostatistics, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, Korea. E-mail: ijung@yuhs.ac

2절에서는 본 연구에 들어가기에 앞서 다항 로짓 회귀모형에 대한 간단한 이론을 언급한다. 또한 본 연구에서 사용되는 네 가지 검정 통계량에 대한 이론적 배경 및 각 통계량들에 대한 식을 제시한다. 3절에서는 2절에서 소개한 네 가지 방법들을 비교, 평가하기 위한 모의실험 설계방법 및 결과를 제시한다. 4절에서는 대뇌 미세 출혈 자료에 대하여 그룹화 전략을 이용한 세 가지 방법을 적용한 결과를 비교, 평가한다. 마지막으로 5절에서는 결론 및 고찰을 제시한다.

2. 이론적 배경

2.1. 다항 로짓 회귀모형

로지스틱 회귀모형이란 이항자료의 종속변수와 이를 설명해주는 독립변수 사이의 선형적인 관계를 보기 위해 사용되는 방법이다. 이러한 로지스틱 회귀모형에서 종속변수가 이항이 아닌 다항으로 확장된 경우를 다항 로짓 회귀모형이라고 한다.

종속변수 Y 가 명목형 형태인 0에서 $c-1$ 까지의 가능한 c 개의 범주를 가지며 p 개의 독립적인 예측 변수 $\mathbf{x} = (x_1, x_2, \dots, x_p)$ 가 있다고 하자. 이때 기준 범주를 Y 가 0일 때로 놓으면, 기준 범주에 대한 j 범주의 로짓 함수는 식 (2.1)과 같다.

$$\begin{aligned} g_j(\mathbf{x}) &= \ln \left[\frac{P(Y = j | \mathbf{x})}{P(Y = 0 | \mathbf{x})} \right] \\ &= \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p, \quad j = 1, 2, \dots, c-1. \end{aligned} \quad (2.1)$$

식 (2.1)처럼 명목형 종속변수에 대한 로짓 모형은 임의로 한 기준 범주에 대한 나머지 각 범주와의 짝을 지어 로짓을 정의하며, 이는 기준 범주에 비해 j 범주에 속할 확률을 로짓을 통해 선형함수로 연결한다. 이러한 로짓 함수를 이용하여 조건부 \mathbf{x} 에 대한 각 종속변수의 범주에 속할 다항 확률은 식 (2.2)와 같이 정의된다.

$$\begin{aligned} P(Y = 0 | \mathbf{x}) &= \frac{1}{1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})} + \dots + e^{g_{c-1}(\mathbf{x})}} \\ P(Y = 1 | \mathbf{x}) &= \frac{e^{g_1(\mathbf{x})}}{1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})} + \dots + e^{g_{c-1}(\mathbf{x})}} \\ &\vdots \\ P(Y = c-1 | \mathbf{x}) &= \frac{e^{g_{c-1}(\mathbf{x})}}{1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})} + \dots + e^{g_{c-1}(\mathbf{x})}}. \end{aligned} \quad (2.2)$$

이를 크기가 n 인 독립적인 표본에 대하여 고려해보면, 위에서 언급했던 표기를 $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, 2, \dots, n$ 과 같이 나타낼 수 있는데, 여기서 i 번째 개체가 j 번째 범주에 속하는 경우 관측치 y_{ij} 는 1의 값을 가지며, 그 외의 범주에 속할 경우 0의 값을 갖는다. 또한 i 번째 개체가 j 번째 범주일 예측 확률($\hat{\pi}_{ij}$)은 얻어진 각 범주에 대한 로짓 함수를 이용하여 구하게 된다. 이때, 실제 얻어지는 종속변수 및 예측 확률은 c 개의 범주에 대한 합이 모두 1이 되어야 하는 제약조건을 만족해야 하는데, 즉, $\sum_{j=0}^{c-1} y_{ij} = 1$ 과 $\sum_{j=0}^{c-1} \hat{\pi}_{ij} = 1$ 이 성립해야 함을 의미한다.

2.2. 모형 적합도 검정 통계량

관심 있는 종속변수와 이에 영향을 줄 것이라 예상하는 독립변수들 간의 모형 식을 구축하였을 때, 그 회귀모형 식이 자료를 반영하는 적절한 모형인지를 판단하는 것이 모형의 적합도 검정이다. 이 절에서

Table 2.1. Observed (O_{kj}) and estimated (E_{kj}) frequencies using grouping strategy

그룹	$Y = 0$		$Y = 1$...	$Y = c - 1$	
	관측	기대	관측	기대		관측	기대
1	O_{10}	E_{10}	O_{11}	E_{11}	...	$O_{1,c-1}$	$E_{1,c-1}$
2	O_{20}	E_{20}	O_{21}	E_{21}	...	$O_{2,c-1}$	$E_{2,c-1}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
g	O_{g0}	E_{g0}	O_{g1}	E_{g1}	...	$O_{g,c-1}$	$E_{g,c-1}$

는 다항 로짓 회귀모형에 관한 기존 검정 통계량들 중 그룹화 전략을 이용한 방법들을 소개하고자 한다. 여기서 그룹화 전략이란 Hosmer와 Lemeshow (1980)가 고안한 방법으로, 각 개체에 대한 효과를 보는 것이 아니라 여러 개체들을 그룹으로 묶어 그 그룹들의 효과를 보는 것이다. 그룹화 전략을 통해 형성된 자료 구조는 Table 2.1에 나타내었다. 종속변수가 다항인 경우에는 특정한 기준이 따로 존재하지 않으며 다양한 그룹화 전략이 존재한다. 이때, 각 칸에 들어갈 관측 빈도수와 기대 빈도수는 식 (2.3)과 같이 나타낼 수 있다.

$$O_{kj} = \sum_{l \in \Omega_k} y_{lj}, \quad E_{kj} = \sum_{l \in \Omega_k} \hat{\pi}_{lj} \tag{2.3}$$

여기서 $k = 1, 2, \dots, g, j = 1, \dots, c - 1$ 이며, Ω_k 는 k 그룹에 속한 n/g 개의 개체 자료를 의미한다.

2.2.1. Pearson의 방법 피어슨 χ^2 통계량은 피어슨의 곱적률상관계수(product-moment correlation)로 알려져 있으며, Karl Pearson이 1900년에 제안한 것이다. 이는 대부분의 적합도 검정 예를 들어, 한 모형을 다른 추가적인 효과를 가지는 모형과 비교하거나, 가지고 있는 자료에 맞는 모형을 선정하는 경우에 주로 사용되는 방법이다. 이 통계량 역시 실제 관측 빈도수와 기대 빈도수를 이용한 방법이며, 이에 대한 식은 (2.4)와 같다.

$$X^2 = \sum_{i=1}^n \sum_{j=0}^{c-1} \frac{(y_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}}. \tag{2.4}$$

이는 보통 자유도 $n(c - 1)$ 을 갖는 카이제곱 분포를 따르며, 산출되는 기대 빈도수가 너무 작을 경우, 언급된 카이제곱 분포를 잘 따르지 않는 경향이 있다.

2.2.2. Fagerland와 Hosemer, Bofin의 방법 여러 그룹화 방법들 중 Fagerland 등 (2008)이 제시한 것은 예측 확률들의 합 즉, $\sum_{j=1}^{c-1} \hat{\pi}_{ij} = 1 - \hat{\pi}_{i0}$ 을 이용하여 그룹을 형성하는 방법이다. 예를 들어, 만일 g 개의 그룹이 있다면 첫 번째 그룹은 가장 작은 합의 값을 갖는 n/g 만큼의 개체 데이터가 포함되고, 두 번째 그룹은 그 다음 작은 합의 값을 갖는 n/g 만큼이 포함되게 되는 형태로, 여러 범주를 모두 이용하여 그룹을 나누는 방법이다. 이와 같은 그룹화 방법과 식 (2.3)의 관측 빈도수, 기대 빈도수에 대한 피어슨 χ^2 통계량 식을 이용하면 식 (2.5)와 같은 식을 얻을 수 있다.

$$C_g = \sum_{k=1}^g \sum_{j=0}^{c-1} \frac{(O_{kj} - E_{kj})^2}{E_{kj}}. \tag{2.5}$$

이 통계량의 분포는 근사적으로 자유도 $(g - 2) \times (c - 1)$ 을 갖는 카이제곱 분포를 따르는 것이 모의 실험을 통하여 확인되었다. 이 검정 통계량은 종속변수가 이항 자료인 것에 대하여 분석할 경우 Hosmer와 Lemeshow (1980)의 방법에 비해 검정력이 비교적 높지 않다는 단점을 가지고 있으며, 또한 검

정에 사용되는 그룹의 수에 다소 의존하는 경향이 있다는 점이 제기되어있다 (Hosmer와 Hjort, 2002; Fagerland 등, 2008).

2.2.3. Bull의 방법 Bull (1994)은 Hosmer와 Lemeshow (1980)의 그룹화 전략을 확장한 것을 다항 로짓 회귀모형에 대한 총괄적인 적합도 검정 방법으로써 제시하였다. 여기서 Hosmer와 Lemeshow (1980)의 그룹화 전략의 주된 방법은 예측 확률을 이용하여 개체들을 그룹화하는 것으로 이를 통해 통계량 값을 얻는다. 이러한 그룹화 전략은 두 가지로 나뉘는데, 하나는 순서대로 해당그룹에 $n_k = n/g$ (k : 그룹)만큼 할당시키는 것으로, 한 범주의 예측 확률을 정렬시켜 그 순서대로 각 그룹에 할당시키는 방법이다. 다른 하나는 $k/10$ (k : 그룹) 질사점(cutpoints)을 이용하는 것으로, 역시 한 범주의 예측 확률을 이용하여 그 값을 질사점으로 나누어 그룹을 할당시키는 방법이다. 이러한 그룹화 전략을 통해 식 (2.3)에 나와 있는 관측 빈도수와 기대 빈도수를 이용하여 얻어지는 검정 통계량은 식 (2.6)과 같이 정의한다.

$$\hat{C} = \sum_{k=1}^g \sum_{j=0}^{c-1} \frac{(O_{kj} - E_{kj})^2}{E_{kj}}. \quad (2.6)$$

위의 식은 근사적으로 자유도 $g(c-1) - 2$ 를 갖는 카이제곱 분포를 따르게 된다고 언급되어 있다. 그러나 이 통계량의 경우는 이항 로짓 회귀모형에서의 모의실험에 대한 결과만이 제시되어 있고, 다항 로짓 회귀모형의 상황에 대해서는 확인되지 않았다. 이 방법 역시 종속변수가 다항일 경우에는 그룹화 방법에 대한 명확한 기준을 제시하고 있지 않으며, 어떤 기준을 세우느냐에 따라 결과 값에 약간의 차이를 보일 수 있다는 맹점이 있다. 기존의 연구결과에서 종속변수의 각 범주에 속할 예측 확률을 기준으로 삼을 수도 있으며, 관심있는 공변량을 기준으로 세우는 것 또한 유용한 방법이라고 언급되어 있다 (Pigeon과 Heyse, 1999). 따라서 본 연구에서는 그룹화의 기준을 독립변수로 설정하고자 한다.

2.2.4. Pigeon과 Heyse의 방법 Pigeon과 Heyse (1999)의 J^2 통계량은 로짓 회귀모형에서 일반적으로 사용되는 적합도 검정 방법인 피어슨 χ^2 통계량을 보정한 것이다. 피어슨 χ^2 통계량은 각 개체들이 각 종속변수의 범주에 속할 서로 다른 확률을 가지는 경우 카이제곱 분포를 잘 따르지 않는다. 또한 피어슨 χ^2 통계량의 경우 각 범주의 관측 빈도수 O_j 의 분산인 $\sum_{i=1}^n \hat{\pi}_{ij}(1 - \hat{\pi}_{ij})$ 이 다항 분포를 가정하였을 때 얻어지는 분산인 $n\pi_j(1 - \pi_j)$ 보다 더 작기 때문에($\pi_j = \sum_{i=1}^n \hat{\pi}_{ij}/n$), 다항 분포의 상황에서 산포가 더 작게 추정되는 경향이 있다. 따라서 J^2 통계량은 개체마다 종속변수에 속할 확률이 다를 수도 있는 상황에 대하여 확장하였으며, 피어슨 χ^2 통계량의 분모에 과소산포를 보정해줄 수 있는 가중치를 부여하여 다항 분포 상황에 대한 문제를 해결하였다. 이 통계량은 각 그룹의 변동을 보정해주는 문제를 해결해주어, 다항 로짓 회귀모형 상황에서 정확하지 않았던 그룹을 나누는 기준에 대한 효과를 줄일 수 있다는 장점을 가지고 있다. 이 통계량을 다항 로짓 회귀모형의 적합도 검정 방법으로 확장하기 위하여 Hosmer와 Lemeshow (1980)의 그룹화 전략을 적용한 식은 다음과 같다.

$$J^2 = \sum_{k=1}^g \sum_{j=0}^{c-1} \frac{(O_{kj} - E_{kj})^2}{\phi_{kj} E_{kj}}, \quad \phi_{kj} = \frac{\sum_{i=1}^{n_g} \hat{\pi}_{ij}(1 - \hat{\pi}_{ij})}{n_g \pi_{kj}(1 - \pi_{kj})},$$

여기서 π_{kj} 는 k 그룹 내의 개체들에 대한 $\hat{\pi}_{ij}$ 의 평균이다. 이 통계량은 근사적으로 자유도 $(g-1) \times (c-1)$ 을 갖는 카이제곱 분포를 따른다고 제시되어 있다. 그러나 이 통계량은 종속변수가 이항인 경우에 대한 결과뿐 아니라 다항인 경우에 대한 결과도 기준에 언급된 바 없으며, 단지 예제를 통하여 \hat{C} 통

Table 3.1. Four settings for simulation of the null distribution

상황1	범주 수가 4이고 독립변수의 수가 1인 경우
상황2	범주 수가 4이고 독립변수의 수가 2인 경우
상황3	범주 수가 5이고 독립변수의 수가 1인 경우
상황4	범주 수가 5이고 독립변수의 수가 2인 경우

계량과의 결과 비교만이 제시되어있다 (Pigeon과 Heyse, 1999). 이 통계량 역시 그룹화 방법에 여러 기준을 이용할 수 있으며, 본 연구에서는 독립변수를 이용하여 그룹을 나누고자 한다.

3. 모의실험을 통한 비교

3.1. 귀무분포

3.1.1. 모의실험 방법 올바른 모형을 적절 시킬 때, 각 통계량의 분포는 각각 자신의 귀무분포를 따르게 된다. Fagerland 등 (2008)의 방법인 C_g 는 $\chi^2_{(g-2)(c-1)}$ 를 따른다고 제시되어 있다. Bull (1994)의 방법인 \hat{C} 은 $\chi^2_{g(c-1)-2}$ 를 따른다고 언급하고 있으며 마지막으로 Pigeon과 Heyse (1999)의 방법인 J^2 은 $\chi^2_{(g-1)(c-1)}$ 를 따른다고 한다. 따라서 올바른 모형을 적합한 후, 각 방법들을 이용한 분석을 1000번 반복하여 얻어진 통계량 값들의 분포가 저자들이 제시한 각 통계량의 분포와 유사한지를 알아보려고 한다. 이때, Table 3.1에 나타나있는 상황에 대하여 그룹의 수가 5와 10일 때로 나누어 모의실험을 진행하였다.

모의실험의 자료를 얻기 위해 로짓 함수를 이용하는데, 종속변수의 범주 수가 4와 5인 경우와 독립변수의 개수가 1과 2인 경우의 로짓 함수는 다음과 같다.

$$\begin{aligned} \text{범주 수} = 5 & \begin{cases} \text{범주 수} = 4 & \begin{cases} g_1(x) = -2.1 - 0.2x, \\ g_2(x) = -1.1 - 0.35x, \\ g_3(x) = -1.9 - 0.21x, \\ g_4(x) = x, \end{cases} \\ \text{범주 수} = 5 & \begin{cases} g_1(x) = -2.1 - 0.2x + z, \\ g_2(x) = -1.1 - 0.35x + z, \\ g_3(x) = -1.9 - 0.21x + z, \\ g_4(x) = x + z. \end{cases} \end{cases} \end{aligned} \quad (3.1)$$

식 (3.1)에서 범주 수가 4일 때 사용되는 로짓 함수와 범주 수가 5일 때 사용되는 세 개의 로짓 함수의 회귀계수가 동일한데, 이는 범주 수가 하나 늘었을 때 어떠한 경향을 가지는지를 보고자 하기 위함이다. 마찬가지로 이유로 범주 수가 5가 되었을 때 추가되는 로짓 함수는 상수항의 효과를 배제하고 오로지 독립변수 하나만의 효과를 추가하고자 그 회귀계수를 1로 설정하였다. 이를 통해 단지 순수한 독립변수 효과만큼의 영향을 받는 범주 수가 하나 더 늘어난 다항 로짓 회귀모형 상황에서 각 통계량들은 제시된 분포를 얼마나 잘 따르는지를 알아보려고 한다. 위의 로짓 함수의 회귀계수는 기존 연구결과를 참고하였다 (Hosmer와 Hjort, 2002; Fagerland 등, 2008). 또한 독립변수 x 와 z 를 생성하기 위해 $U(-6, 6)$, $U(-1, 1)$ 두 가지 분포를 이용하였는데, 그 이유는 분포의 범위에 따라 생성되는 독립변수의 범위가 달라지기 때문이다. 마찬가지로 분포에 대한 선정은 기존 연구결과들에서 사용하였던 분포들을 참고하였다 (Hosmer와 Hjort, 2002; Fagerland 등, 2008). 다른 회귀계수들에 대한 설정과 정규 분포에 대한 경우에 있어서도 모의실험을 진행하였으나 유사한 결과가 도출되었기 때문에, 이를 따로 실지는 않았다.

Table 3.2. Simulated per cent rejection rates at 5 per cent significant level and simulated mean values

x, z	표본 수 = 100							표본 수 = 400						
	X^2	그룹 = 5			그룹 = 10			X^2	그룹 = 5			그룹 = 10		
		C_g	\hat{C}	J^2	C_g	\hat{C}	J^2		C_g	\hat{C}	J^2	C_g	\hat{C}	J^2
	300*	9*	13*	12*	24*	28*	27*	1200*	9*	13*	12*	24*	28*	27*
$U(-6, 6)$														
상황1	3.00	4.90	0.40	0.90	4.40	1.30	1.90	0.10	5.70	1.20	1.90	5.50	2.00	2.70
	294.61	9.19	9.18	9.21	24.10	24.08	24.11	1197.26	9.18	9.18	9.20	24.13	24.13	24.15
상황2	9.70	7.80	0.10	1.30	7.50	0.40	1.40	21.00	8.90	0.20	1.10	9.10	0.20	2.80
	255.90	8.63	7.19	9.38	21.16	19.03	24.63	1181.66	9.25	7.23	9.29	23.12	18.95	24.36
상황3	22.60	3.80	2.60	3.30	4.00	4.10	4.80	58.40	9.00	3.90	5.30	8.20	6.50	7.80
	335.56	10.70	9.93	10.07	29.40	28.85	28.97	1557.77	12.57	11.79	11.91	31.30	31.29	31.38
상황4	9.70	4.60	1.40	2.30	4.00	1.50	2.80	37.00	5.30	2.90	3.80	6.60	4.30	6.50
	225.72	9.61	7.18	9.04	23.06	18.83	23.93	1423.98	11.20	9.66	11.35	28.38	25.06	29.78
$U(-1, 1)$														
상황1	0.30	5.60	1.00	1.50	4.30	1.30	1.90	0.00	4.40	1.40	1.60	5.40	1.40	2.30
	297.66	9.61	9.39	9.39	24.58	24.46	24.46	1199.63	9.09	9.03	9.04	24.21	24.16	24.16
상황2	2.00	3.50	0.40	0.90	4.20	1.40	1.70	0.10	5.40	0.30	0.60	4.50	0.60	1.10
	294.11	9.23	9.19	9.50	24.04	23.69	24.45	1199.43	9.05	8.76	9.00	23.87	23.46	24.08
상황3	93.70	6.20	0.20	0.60	4.20	0.60	1.10	100.00	5.00	0.60	1.30	4.90	1.50	2.50
	387.08	12.75	12.06	12.08	32.52	31.60	31.62	1597.99	12.26	12.14	12.15	32.43	32.30	32.30
상황4	84.20	4.40	0.50	1.10	3.30	0.80	1.80	100.00	4.60	0.20	1.50	4.40	0.40	1.00
	374.71	12.30	12.12	12.47	31.10	31.18	32.04	1595.70	12.05	11.91	12.15	31.71	31.46	32.08

* : 자유도(degrees of freedom)

이렇게 설정된 상황을 기반으로 독립변수들과 미리 설정된 로짓 함수를 이용하여 다항 로짓 예측 확률을 구한 후, $U(0, 1)$ 에서 임의로 생성한 u 값을 다항 로짓 예측 확률과 비교하여 종속변수의 값을 결정한다. 이때 범주 수가 4와 5인 경우는 식 (3.2)의 기준을 통해 종속변수의 값을 결정한다.

$$\begin{cases} y = 3; u > \pi_0 + \pi_1 + \pi_2, \\ y = 2; \pi_0 + \pi_1 < u \leq \pi_0 + \pi_1 + \pi_2, \\ y = 1; \pi_0 < u \leq \pi_0 + \pi_1, \\ y = 0; u \leq \pi_0, \end{cases} \quad \begin{cases} y = 4; u > \pi_0 + \pi_1 + \pi_2 + \pi_3, \\ y = 3; \pi_0 + \pi_1 + \pi_2 < u \leq \pi_0 + \pi_1 + \pi_2 + \pi_3, \\ y = 2; \pi_0 + \pi_1 < u \leq \pi_0 + \pi_1 + \pi_2, \\ y = 1; \pi_0 < u \leq \pi_0 + \pi_1, \\ y = 0; u \leq \pi_0. \end{cases} \quad (3.2)$$

각 상황에 대한 표본 크기는 100과 400으로 나누어 진행하였다.

3.1.2. 결과 네 가지 상황에 대하여 그룹의 수가 5와 10일 때, 각 통계량들의 귀무분포를 알아보고자 시행한 모의실험의 결과는 Table 3.2에 제시되어 있는데, 이는 각 상황에서 유의수준 5%에 해당하는 분율을 나타낸 값이다. 우선 피어슨 χ^2 통계량은 기존 연구결과 (Fagerland 등, 2008)에서 제시한 것처럼 대부분의 경우 유의수준 5%를 만족하지 않는 것을 볼 수 있다. 각 상황에서 얻어진 통계량의 평균값 역시 대부분 큰 차이를 두고 제시된 분포의 평균에서 벗어나는 양상을 보인다.

범주 수가 3이고 독립변수의 수가 1인 상황에 대해서 확인되었던 C_g 통계량의 경우, 본 연구결과에서는 종속변수의 범주 수와 독립변수의 개수가 늘어남에 따라 유의수준 5%에 만족하지 않는 적절치 못한 값들이 다소 증가하였다. 평균의 경우에는 상황1일 때, 대체로 제안된 분포의 평균값과 유사한 결과를

Table 3.3. Three settings for simulation of power

상황1	주효과 항 생략 $g_j(x) = \beta_{j0} + \beta_{j1}x, j = 1, 2, 3, \beta_{j1} : 0.01, 0.5, 1.0$
상황2	제곱 항 생략 $g_j(x) = \beta_{j0} + \beta_{j1}x + \beta_{j2}x^2, j = 1, 2, 3, \beta_{j2} : 0.01, 0.05, 0.1, 0.5$
상황3	교호작용 항 생략 $g_j(x) = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \beta_{j3}x_1x_2, j = 1, 2, 3, \beta_{j3} : 0.2, 0.6, 1.0$

보이며, 범주 수나 독립변수의 수가 늘어나는 상황에서도 몇몇 경우를 제외하고는 대부분 제안된 분포의 평균과 유사한 것을 알 수 있었다. 이전 연구결과 (Fagerland 등, 2008)에 비해 벗어남이 보이지만 C_g 의 경우는 범주 수와 독립변수의 수의 증가에 있어서도 근사적으로 제안된 분포를 따른다고 볼 수 있다.

\hat{C} 통계량과 J^2 통계량의 경우는 기존에 제시된 결과 없이 이항 범주에 대한 결과를 기반으로 다항 범주 일 경우에 근사적으로 카이제곱 분포를 따를 것이라고 언급되었기 때문에, C_g 통계량의 결과와 비교하여 귀무분포에 대한 결론을 해석하고자 한다. J^2 통계량이 \hat{C} 통계량에 비해 다소 높은 값을 갖지만, 두 통계량의 값들이 대부분 C_g 통계량들에 비해 작은 값으로 추정되는 것을 볼 수 있으며, 종속변수의 범주 수에 상관없이 모두 유의수준 5%에 만족하는 경우가 매우 드물게 관찰되었다. 평균의 경우에도 역시 J^2 통계량의 값이 더 크지만, 두 통계량 모두 제안된 분포의 평균 근처에 도달하는 경우가 거의 없으며, 대부분의 결과가 C_g 통계량의 평균 결과와 유사한 것을 볼 수 있다. 카이제곱 분포는 자유도가 평균을 나타내므로 두 통계량에 대한 제시된 분포가 적절하지 않다고 할 수 있다. 다시 말해, \hat{C} 통계량과 J^2 통계량은 자유도 $g(c - 1) - 2$ 와 자유도 $(g - 1) \times (c - 1)$ 을 갖는 카이제곱 분포를 따른다고 보기 어렵다.

본 연구에서 \hat{C} 통계량과 J^2 통계량은 기존에 제시되었던 분포에 잘 맞지 않는다는 결론이 도출되었다. 그 주된 이유는 C_g 통계량과 그룹을 묶는 방법이나 보정하는 측면에서 조금 다를 뿐 유사한 통계량을 가졌으나, 검정 통계량의 기각 여부를 결정하는 기각역은 세 통계량이 서로 다르기 때문이다. 즉, 얻어진 자료에 비해 자유도가 큰 카이제곱 분포를 따르는 \hat{C} 통계량이나 J^2 통계량의 경우는 큰 기각값을 가지기 때문에 유의수준 5%에서 모형이 적합하지 않다고 결론 내릴 가능성이 매우 낮은 것이고, 이는 적절치 않은 모형임에도 불구하고 적절하다고 잘못 판단할 가능성이 높아지는 것을 의미한다. Figure 3.1은 이를 뒷받침해주는 자료로 1000번의 모의실험 결과 얻어지는 각 통계량 값들의 히스토그램과 제안된 분포를 나타내고 있다. 이때, $U(-1, 1)$ 에서 독립변수를 생성하고 첫 번째 상황을 적용한 결과 도출된 모의실험 자료를 이용하였다. 위에서 언급한 것과 마찬가지로 세 통계량들의 히스토그램은 거의 일치하나, 자유도에 따른 각각의 통계량 분포는 상이한 것을 볼 수 있다. 또한 각 상황에 따라 분석된 결과에서 C_g 통계량의 귀무분포가 기존에 제시된 분포와 가장 유사한 것을 볼 수 있다.

3.2. 검정력

3.2.1. 모의실험 방법 자료를 잘 설명하는 적절한 모형과 자료를 반영하지 못하는 부적절한 모형 사이의 차이를 판별하는 것은 적합도 검정 통계량의 중요한 역할이다. 따라서 이 절에서는 인위적으로 만든 부적절한 상황에 대하여 검정 통계량들이 이를 얼마나 잘 판단하는지를 알아보하고자 한다. Table 3.3은 모의실험을 진행하기 위해 설정된 상황을 나타낸 것으로, 각 상황에 해당하는 항을 생략함으로써 발생하는 모형의 적합성 결여 정도를 증가시키기 위해 해당하는 항의 회귀계수 값을 미리 설정하였다.

검정력에 대한 모의실험에서는 종속변수의 범주 수가 4인 경우로 한정 지었다. 그 이유는 귀무분포의

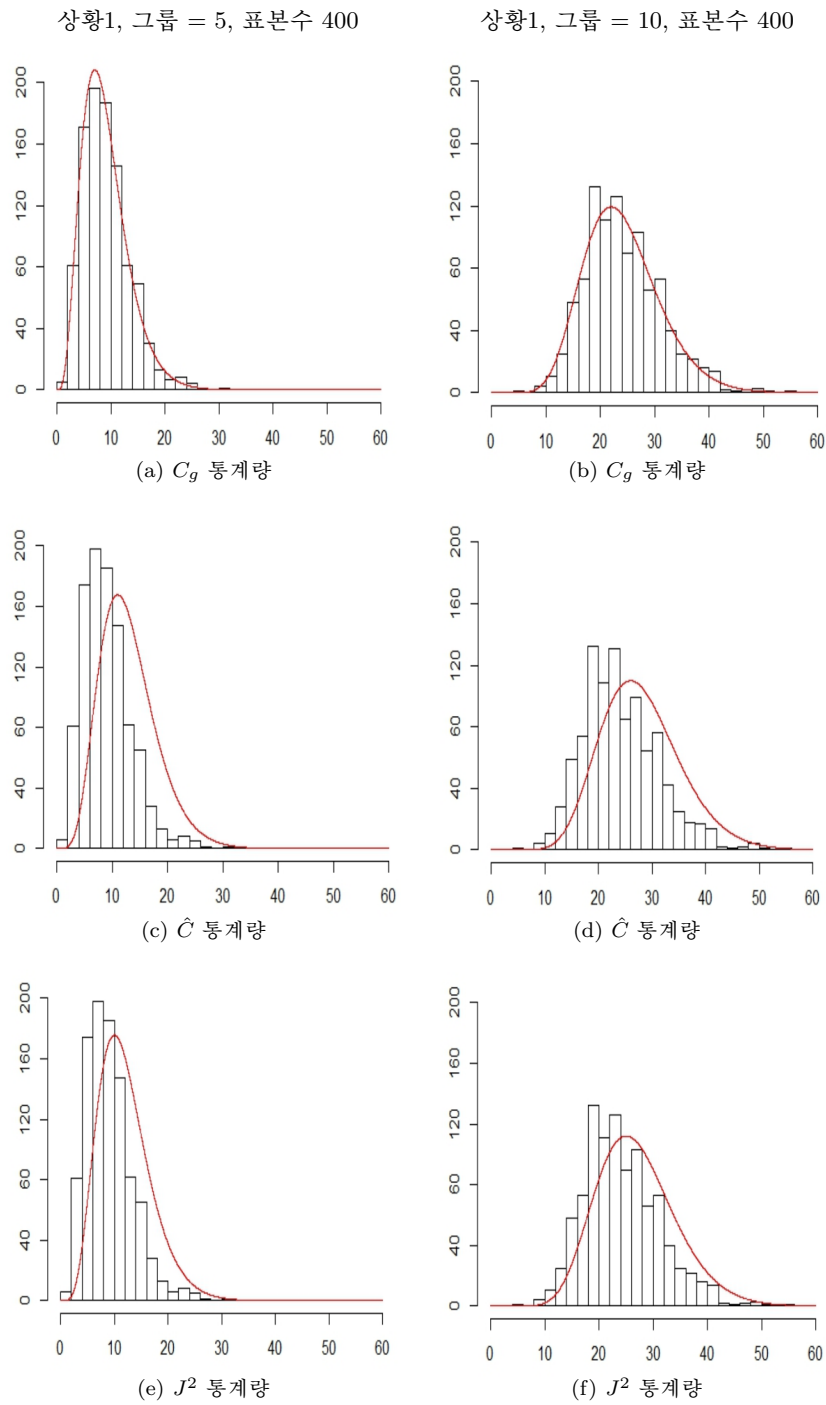


Figure 3.1. Histograms of simulated values and proposed null distributions for the three test statistics

Table 3.4. Simulated power (%) at 5 per cent nominal level

	표본 수 = 100							표본 수 = 400									
	X^2	그룹 = 5			그룹 = 10			X^2	그룹 = 5			그룹 = 10					
		C_g	\hat{C}	J^2	C_g	\hat{C}	J^2		C_g	\hat{C}	J^2	C_g	\hat{C}	J^2			
상황1: $U(-6, 6)$																	
0.01	0.00	14.30	3.20	4.60	9.30	3.50	5.10	0.00	17.20	3.00	4.60	12.00	3.50	5.40			
0.5	0.00	14.60	99.70	99.80	10.20	97.70	98.60	0.00	15.60	100.00	100.00	10.70	100.00	100.00			
1.0	0.00	15.40	100.00	100.00	10.30	100.00	100.00	0.00	14.00	100.00	100.00	10.30	100.00	100.00			
상황1: $U(-1, 1)$																	
0.01	0.00	13.80	2.90	4.80	9.20	4.20	5.50	0.00	16.60	2.60	3.80	11.30	3.30	4.50			
0.5	0.00	15.20	6.60	9.40	8.00	5.40	6.80	0.00	13.40	29.60	35.30	10.80	20.00	23.60			
1.0	0.00	13.30	24.30	30.60	9.30	16.80	20.10	0.00	13.30	94.90	96.50	10.20	86.10	88.50			
상황2: $U(-6, 6)$																	
0.01	4.50	5.00	0.60	0.60	4.30	1.70	2.30	0.40	7.60	1.20	2.10	5.60	1.90	2.50			
0.05	6.50	19.80	4.30	6.00	13.20	5.30	6.80	2.30	76.00	51.20	57.70	64.10	45.40	49.40			
0.1	3.60	61.50	36.60	43.20	48.40	29.00	33.60	0.30	100.00	99.90	99.90	99.80	99.50	99.50			
0.5	0.50	92.00	92.90	95.20	83.10	89.00	91.20	0.00	100.00	100.00	100.00	100.00	100.00	100.00			
상황2: $U(-1, 1)$																	
0.01	0.30	5.70	1.00	1.50	4.30	1.20	1.80	0.00	4.70	1.20	1.60	5.20	1.40	2.30			
0.05	0.30	6.00	1.10	1.60	4.40	1.20	1.70	0.00	5.10	1.20	1.40	5.60	1.40	2.20			
0.1	0.30	5.20	1.00	1.40	4.30	1.20	1.40	0.00	4.80	0.90	1.80	5.70	1.40	1.70			
0.5	0.50	6.00	0.80	1.10	4.50	0.90	1.40	0.00	10.30	2.30	3.10	9.40	2.90	4.10			
상황3: $U(-6, 6)$																	
0.2	21.90	7.80	0.30	1.70	11.80	0.40	3.20	36.70	45.60	1.70	10.90	53.90	1.80	8.70			
0.6	26.10	65.00	0.40	0.90	56.20	1.10	2.30	45.30	83.50	0.60	1.00	87.70	1.50	2.60			
1.0	32.00	75.50	0.60	1.20	67.80	1.10	1.60	54.20	91.00	0.90	1.60	94.20	1.70	2.60			
상황3: $U(-1, 1)$																	
0.2	1.80	4.50	0.60	1.40	3.20	1.00	2.00	0.10	4.80	0.50	0.80	3.60	1.00	1.50			
0.6	1.90	4.20	0.60	1.00	4.30	1.80	3.10	0.30	7.90	0.30	0.90	6.30	0.80	1.40			
1.0	2.30	6.50	0.60	1.10	4.50	0.80	2.40	0.30	9.80	0.40	1.10	9.90	0.60	1.10			

모의실험 결과 범주 수가 5인 경우에 비해 4인 경우에 더 적합함을 보였기 때문이다. 이때 각 상황에서 사용되는 상수항과 주효과 항의 회귀계수는 귀무분포의 모의실험에서 사용한 값과 동일하다. 각 상황에 대하여 표본 크기를 100과 400으로 나누어 진행하였으며, 1000번의 반복을 통하여 얻어진 통계량 값들 중 명목수준 5%에서 모형이 적합하지 않다고 결론 내린 분율을 검정력으로 평가하였다.

3.2.2. 결과 각 세 상황에 대한 검정력 모의실험 결과를 Table 3.4에 나타내었다. 우선 주효과 항을 생략하는 상황1을 보면, 피어슨 χ^2 통계량의 검정력은 모두 0인 것을 알 수 있다. 따라서 모형에 상수항만을 적합하였을 때, 피어슨 χ^2 통계량은 주효과 항에 대한 효과를 감지하는 능력이 없다고 할 수 있다. C_g 통계량의 경우는 주효과에 대한 회귀계수가 증가하거나 표본 수가 증가하는 것에 상관없이, 그리고 공변량의 분포에 상관없이 대체로 비슷한 결과를 나타내고 있다. 이 통계량은 대체로 그룹의 수가 작은 경우에 검정력이 더 높은 것으로 보이며, 상수항만을 적합 시킨 모형에 대해서는 낮은 검정력을 보인다. \hat{C} , J^2 통계량은 주효과의 회귀계수가 증가함에 따라 검정력이 높아지는 것을 알 수 있다. 공변량의 분포가 좁은 것에 비해 넓은 것에서 검정력이 높고, 그룹 수가 5인 것에 비해 10일 때, 더 좋은 결과를 나타내고 있다. 두 통계량 중에서는 \hat{C} 에 비해 J^2 이 좀 더 높은 검정력을 갖는 것으로 보인다. 상황2의 경우, 피어슨 χ^2 통계량의 결과가 상황1의 결과에 비해 다소 높아진 것을 볼 수 있으나 대체적으로 낮은 검정력을 가진다. C_g 통계량의 경우 좁은 공변량의 분포에서는 일부를 제외하고 회귀계수 값의 증가에도 유사한 결과를 나타내고 있으나, 넓은 분포에서는 회귀계수가 증가함에 따라 검정력이 모두 증가하는 것을 볼 수 있다. 또한 그룹이 10일 때에 비해 5일 때 더 나은 결과를 나타내고 있다. \hat{C} , J^2 통

Table 4.1. A real data of cerebral microbleeds

변수	변수 유형	변수 값
CMB분류	명목형	0 = 미세출혈이 없음(787(69.22%))
		1 = 비뇌엽성 영역(248(21.81%))
		2 = 뇌엽성 영역(36(3.17%))
		3 = 비뇌엽성 및 뇌엽성 복합영역(66(5.80%))
나이	연속형	64.88±11.62 (19-98)
체질량지수(BMI)	연속형	24.08± 3.11 (14.47-39.00)
맥과전과속도/표준편차(PWVSD)	연속형	3.81± 1.00 (1.56-10.42)
수축기혈압(systolic)	연속형	150.84±23.19 (95.00-236.00)

계량은 좁은 공변량의 분포에서 C_g 통계량에 비해 낮은 검정력을 보이거나 넓은 분포에서는 C_g 통계량과 유사한 결과를 보인다. 특히 그룹이 10인 경우에 그 차이가 더 명확하게 나타난다. 좁은 분포에서는 표본 수가 큰 영향을 미치지 못하는 반면, 넓은 분포에서는 표본 수가 증가할 경우 더 좋은 결과가 도출되었다. 마지막으로 상황3을 보면, 피어슨 χ^2 통계량의 결과가 이전의 다른 결과들에 비해 상당히 안정되고 좋아졌음을 볼 수 있다. 또한 공변량의 분포에 상관없이 회귀계수가 증가함에 따라 검정력 역시 증가하였다. 그러나 상대적으로 다른 세 통계량의 결과에 비해 더 낮은 검정력을 가진다. C_g 통계량의 경우는 회귀계수가 증가함에 따라 대체로 검정력이 높아지는 것을 볼 수 있다. 제곱항의 상황과 마찬가지로 C_g 통계량에서는 좁은 공변량의 분포에 비해 넓은 분포에서 검정력이 더 높은 것을 확인하였다. 또한 대체로 그룹 수가 작은 쪽에서 더 높은 검정력을 보인다. \hat{C} , J^2 통계량은 제곱항의 상황에 비해 낮은 검정력이 도출되었다. 두 통계량 모두 공변량의 분포에 상관없이 낮은 검정력을 유지하는 것을 알 수 있으며, 회귀계수 증가에도 상관없이 유사한 검정력을 보이고 있다. 표본 수에 큰 영향을 받지 않으나, 그룹의 수가 5일 때에 비해 10일 때 다소 높은 검정력을 갖는 것을 볼 수 있다.

4. 실제 자료 분석

이 절에서는 그룹화 전략을 이용한 세 적합도 검정 통계량 방법을 대뇌 미세출혈(cerebral microbleeds) 자료에 적용하여 결과를 비교하고자 한다. 피어슨 χ^2 통계량은 앞서 언급되었던 것과 같이 다항 로짓 회귀모형의 적합도 검정 방법으로는 적절치 않다고 판단되었으므로 실제 예제에 대해 분석하는 과정에서는 제외하였다. 나머지 세 통계량에 대해서는 크게 두 가지 방향으로 결과를 보고자 하였는데, 하나는 적절한 모형이 적합 되었을 때 모형이 적합하다고 잘 판단하는지에 관한 것이고, 다른 하나는 적절하지 않은 모형이 적합 되었을 때 모형이 적합하지 않다고 잘 판단하는지에 대한 것이다. 이때, 그룹화 전략 방법으로는 종속변수의 범주에서 기준범주를 제외한 나머지 각 범주에 속할 예측 확률을 이용하였다. 모의실험에서 사용된 그룹화 전략과 다른 방법을 적용한 이유는 모의실험의 경우 데이터를 얻는 과정에서 모든 범주의 결과 값이 얻어지지 않아 일부 범주에 대한 값의 손실이 발생하였고, 이로 인해 추정이 잘 되지 않는 경향이 관찰되었다. 또한, 이 절에서는 추후 임의로 적절하지 않다고 생각되는 모형을 적합 시켜 이를 잘 판별하는지를 알아보기 위해 예제에서 관심 있어 하는 독립변수를 제외하고 분석하게 되는데, 이 경우에는 모의실험에서 사용된 그룹화 전략을 이용할 수 없으므로 예제에서 확인하고자 하는 두 상황을 비교하는데 적합하지 않다고 판단하였기 때문이다.

4.1. 자료 설명

본 자료는 2006년 6월부터 2012년 1월까지 증상이 발병한 후 7일 이내에 뇌경색(cerebral infarction)이

Table 4.2. Goodness-of-fit test statistics and p -values of the three methods for cerebral microbleeds data

그룹화 전략	통계량	그룹 수 = 5		그룹 수 = 10	
		통계량 값	p 값	통계량 값	p 값
상황1					
$1 - \hat{\pi}_{i0}$	C_g	8.0070	0.5334	30.0163	0.1842
$\hat{\pi}_{i1}$	\hat{C}	20.2159	0.0900	35.6647	0.1514
	J^2	20.3004	0.0616	35.9211	0.1170
$\hat{\pi}_{i2}$	\hat{C}	8.9748	0.7748	23.6928	0.6976
	J^2	9.1504	0.6900	24.0125	0.6296
$\hat{\pi}_{i3}$	\hat{C}	11.8112	0.5432	25.6888	0.5901
	J^2	11.8846	0.4550	25.8355	0.5277
상황2					
$1 - \hat{\pi}_{i0}$	C_g	3.9651	0.9137	20.5611	0.6645
$\hat{\pi}_{i1}$	\hat{C}	3.6689	0.9942	20.5611	0.8432
	J^2	3.6769	0.9886	20.6353	0.8031
$\hat{\pi}_{i2}$	\hat{C}	3.6689	0.9942	20.5611	0.8432
	J^2	3.6769	0.9886	20.6353	0.8031
$\hat{\pi}_{i3}$	\hat{C}	3.6689	0.9942	20.5611	0.8432
	J^2	3.6769	0.9886	20.6353	0.8031

나 일과성허혈발작(transient ischemic attack)을 진단받은 환자 1,137명을 대상으로 뇌 영상(brain imaging studies), 혈관 영상(vascular imaging studies), 가슴 x선 촬영, 심전도법(electrocardiography)과 기본적인 혈액 검사를 실시한 결과이다. 본 예제의 연구 목적은 환자들에게 얻어진 결과들을 토대로 여러 요인들을 보정하였을 때, 동맥 경직도를 나타내는 맥파전파속도가 대뇌 미세출혈의 발생 위치에 어떠한 영향을 미치는지를 알아보려고 하는 것이다. 이때 대뇌 미세출혈의 위치는 세 가지로 세분화하였다. Table 4.1은 예제 자료들 중 실제 다항 로짓 모형에 사용된 변수와 해당 변수의 설명 및 기술통계량 값을 나타낸 것으로 본 연구에서는 연속형 독립변수만을 고려하였기 때문에, 예제 자료 중 종속변수에 유의한 영향을 미치는 연속형 변수 및 임상적으로 의미 있는 연속형 변수만을 다항 로짓 회귀모형의 독립변수로써 고려하였다.

4.2. 결과

예제 자료에 적절할 것으로 예상되는 다항 로짓 회귀모형을 적합 시켰을 때, 그 적합성 여부를 판단하기 위하여 본 연구에서 평가하고 있는 그룹화 전략을 이용한 세 가지 통계량을 적용한 결과(상황1)와 단변량 분석 결과 종속변수인 대뇌 미세출혈 발생위치와 유의한 관계가 없는 변수만을 모형에 적합 시켜 얻은 결과(상황2)를 Table 4.2에 제시하였다. 상황1의 결과, 전반적으로 모든 통계량이 모형이 적합하다는 결론을 내리는 것을 볼 수 있다. 그러나 \hat{C} 통계량과 J^2 통계량의 일부 결과 값이 다른 결과에 비해 변동이 큰 것을 볼 수 있으며 일부를 제외하고는 대체로 그룹의 수에 상관없이 C_g 통계량에 비해 더 큰 p 값을 보이는 것을 알 수 있다. 이는 통계량 값이 그들의 제안된 분포에 비해 작은 값으로 추정된 것을 의미하는데, 그 이유는 3절에서 언급한 것과 같이 자유도가 상대적으로 큰 \hat{C} 통계량과 J^2 통계량은 원래 그들의 통계량 값들이 가지는 귀무분포에 비해 오른쪽으로 치우친 형태이기 때문이다. 따라서 모형이 적합하다는 결론을 내릴 가능성이 크며, 이로 인해 얻어진 검정 통계량 값에 해당하는 p 값이 당연히 커질 수밖에 없는 결과가 발생하는 것이다. 상황2의 결과를 보면 모든 모형이 적합하다고 결론내리고 있음을 알 수 있으며, 오히려 4개의 공변량이 포함된 모형에 비해 세 통계량 모두 높은 p 값을 보이

는 것을 알 수 있었다. 무엇보다도 \hat{C} 통계량과 J^2 통계량의 결과가 각 예측 확률에 대해 모두 동일하였고, 이전 모형에 비해 매우 안정적인 추세를 보이는 것을 확인하였다. 이는 세 통계량 모두 부적절한 모형 적합으로 인해 발생하는 적합성 결여를 잘 판단하지 못한다고 할 수 있으며, 따라서 전반적으로 낮은 검정력을 가진다고 할 수 있다.

5. 결론 및 고찰

이항 로짓 회귀모형의 경우 자주 사용되는 적합성 평가의 방법들 (Agresti, 2007)이 존재하나 다항 로짓 회귀모형의 적합성 평가에 대한 방법들은 구체화되지 않은 실정이다. 최근 이항 로짓 회귀모형의 적합도 검정 방법들을 확장하거나 변형시킨 방법들이 제안되고 있으나, 이 방법들에 대한 구체적인 평가가 제대로 이루어지지 않았다는 점에 착안하여 본 논문은 기존에 제시된 방법들 중 일반적으로 사용되는 피어슨 χ^2 방법과 그룹화 전략을 이용한 일부 통계량을 선정하여 비교 및 평가하였다.

적절한 모형을 적합 시켰을 때, 각 통계량들의 귀무분포와 제시되어 있는 그들의 분포가 유사할 것을 토대로 귀무분포에 대한 모의실험을 진행하였다. 피어슨 χ^2 통계량의 경우는 기존 연구결과 (Fagerland 등, 2008)에서 언급했던 것과 같이 유의수준 5%에서 상당히 벗어남을 보였으며, 평균에서도 불일치하는 것을 볼 수 있었다. C_g 통계량의 경우는 종속변수의 범주 수와 독립변수의 개수가 늘어남에 따라 적절치 않은 값들의 발생이 증가함에도 불구하고 대부분이 유의수준 5%를 만족하였고, 평균 역시 저자들이 제시한 각 통계량 분포의 평균 근처에 놓여있는 것을 확인하였다. \hat{C} 과 J^2 통계량은 어떤 상황에 대해서도 모두 작게 추정되는 것을 보았다. 유의수준 5%에 만족하는 경우가 매우 드물며, 평균 역시 제시된 분포의 평균에 비해 매우 작은 값을 가지는 것을 볼 수 있었다. 따라서 \hat{C} 통계량과 J^2 통계량은 저자들이 제시한 분포를 유지하는 능력이 매우 떨어지며, 잘못된 결론을 도출할 가능성이 크기 때문에 C_g 통계량이 다른 통계량들에 비해 그나마 유용하다는 결론을 내렸다.

비록 정확한 모형을 적합 시켰을 때, 통계량들이 제시된 분포에 비교적 잘 따르지 않는 것을 보았으나 각 방법에 대한 더 정확한 평가를 위해 검정력에 대한 모의실험을 진행하였다. 검정력에 대한 평가를 위해 세 가지 상황에 대하여 진행하였는데 피어슨 χ^2 통계량은 세 가지 상황에서 모두 낮은 검정력을 보였다. 다른 세 통계량의 경우, 상황1에서 예상외로 \hat{C} 과 J^2 통계량이 높은 검정력을 보였다. 이는 상수항만을 적합 시켰을 경우 각 종속변수에 속할 확률이 개체별로 같은 값을 갖게 되는데, C_g 통계량은 이를 잘 판단하지 못하였기 때문이다. 상황2의 결과에서는 독립변수의 값을 얻은 분포에 따라 각 통계량들의 검정력이 달라짐을 볼 수 있었다. 좁은 분포에서는 세 통계량 모두 낮은 검정력을 보였으며, 넓은 분포에서는 회귀계수가 증가함에 따라 세 통계량 모두 다소 높은 검정력을 보였다. 상황3의 결과에서는 상황2의 결과에 비해 낮은 검정력을 갖는 것을 볼 수 있으며, 넓은 분포에서 비교적 C_g 통계량이 적절한 검정력을 갖는 것을 볼 수 있었다. 검정력 평가에서도 C_g 통계량이 그나마 유용하다고 할 수 있으며, 나머지 다른 통계량들은 다항 로짓 회귀분석의 적합도 검정 방법으로써 사용하는 것은 무리가 있다고 생각된다.

또한 세 가지 적합도 검정 통계량들을 실제 자료에 적용하여 그 결과를 확인하였다. 우선 예제 자료를 잘 설명할 것이라 예상되는 적절한 모형을 적합 시켰을 경우, \hat{C} 통계량과 J^2 통계량의 p 값들에 변동이 있는 것을 볼 수 있었으며, 전반적으로 C_g 통계량에 비해 p 값들이 큰 것을 알 수 있었다. 이는 실제 통계량의 분포에 비해 가정된 분포가 오른쪽으로 치우친 경향이 있기 때문에 발생하는 결과이다. 하나의 독립변수만을 포함한 부적절한 모형의 결과, 세 통계량에서 오히려 이전 상황에 비해 더 안정적이고 큰 p 값을 도출하였는데, 이는 세 통계량 모두 부적절한 모형에 의해 발생하는 적합성 결여를 잘 판단하지 못한다는 것을 의미한다. 따라서 이 예제에서도 모의실험의 결과와 유사하게 \hat{C} 통계량과 J^2 통계량의

분포가 저자들에 의해 제시된 분포와 다름을 나타내고 있으며, 특히 부적절한 모형을 판별해내는 검정력이 낮다고 할 수 있다.

본 연구에서는 다항 로짓 회귀모형의 적합도 검정 통계량을 여러 상황 하에 비교 분석하였다. 이 방법들은 기존에 잘 알려진 Hosmer와 Lemeshow (1980)의 방법과 피어슨 χ^2 방법을 보정한 방법들임에도 불구하고 저자들이 제시한 통계량의 분포가 통계량들의 귀무분포와 차이가 있음을 알 수 있었으며, 검정력 또한 낮았다. 모의실험 결과는 설정한 상황에 한한 것이기 때문에 다른 상황에 대한 분석의 여지가 남아 있다. 본 연구에서의 확장 즉, 더 많은 독립변수가 포함되는 경우나 다른 유형의 독립변수가 동시에 포함되는 경우, 그리고 독립변수를 생성하는 다양한 분포에 대하여 생각해볼 수 있으며, 여러 분포의 조합도 고려할 수 있다. 실제 자료에 적용하는데 있어 하나의 통계량만을 이용하는 것보다는 여러 통계량을 동시에 제시하거나 잔차도와 같이 적합도 검정에 기본적으로 사용할 수 있는 그림을 추가하여 적합도 검정 여부를 판단하는 것이 더 적절할 것이라 여겨진다.

References

- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd ed. Wiley, New Jersey.
- Bull, S. (1994). Analysis of attitudes toward workplace smoking restrictions. In: Lange, N., Ryan, L., Billard, D., Conquest, L. and Greenhouse, J. (1994). *Case Studies in Biometry*. Wiley, New York, 249–271.
- Fagerland, M. W., Hosmer, W. H. and Bofin, A. M. (2008). Multinomial goodness-of-fit tests for logistic regression models, *Statistics in Medicine*, **27**, 4238–4253.
- Hosmer, D. W. and Hjort, N. L. (2002). Goodness-of-fit processes for logistic regression: Simulation results, *Statistics in Medicine*, **21**, 2723–2738.
- Hosmer, D. H. and Lemeshow, S. (1980). Goodness-of-fit tests for the multiple logistic regression model, *Communications in Statistics, Part A, Theory and Methods*, **9**, 1043–1069.
- Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*, 2nd ed, Wiley, New York.
- Pigeon, J. G. and Heyse, J. F. (1999). An improved goodness of fit statistic for probability prediction models, *Biometrical Journal*, **41**, 71–82.

다항 로짓 회귀모형에서의 그룹화 전략을 이용한 적합도 검정 방법 비교

송미경^a · 정인경^{a,1}

^a연세대학교 의과대학 의학통계학과

(2013년 7월 11일 접수, 2013년 10월 2일 수정, 2013년 10월 15일 채택)

요약

지금까지 제안되어 있는 다항 로짓 회귀모형의 적합도 검정 방법들에 대하여 저자들이 제안한 방법들이 타당한지를 확인하고자 본 연구를 진행하였다. 여러 검정 통계량들 중 그룹화 전략을 이용한 통계량들 (Fagerland 등, 2008; Bull, 1994; Pigeon과 Heyse, 1999)을 선정하였고, 이러한 통계량의 기반이 되는 피어슨 χ^2 통계량 또한 같이 비교하였다. 제안된 분포가 모의실험의 상황 하에 얻어지는 귀무분포와 유사한지, 그리고 부적절한 모형의 판별을 적절히 수행하는지에 대하여 확인하였으며, 실제 자료에 세 가지 방법을 적용한 결과를 비교, 평가하였다.

주요용어: 다항 로짓 회귀모형, 적합도 검정, 그룹화 전략, 귀무분포.

¹교신저자: (120-752) 서울특별시 서대문구 신촌동 134번지, 연세대학교 의과대학 의학통계학과, 조교수.
E-mail: ijung@yuhs.ac