creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

[UCI]I804:21031-200000179652

이 학 석 사 학 위 논 문

이중 다단계 일반화 선형모형의 평균과 산포의 결합 모델링을 통한 열화시험데이터 분석

2019 년 2 월

부 경 대 학 교 대 학 원

통 계 학 과

김 혁 진

이 학 석 사 학 위 논 문

이중 다단계 일반화 선형모형의 평균과 산포의 결합 모델링을 통한 열화시험데이터 분석

지도교수 노 맹 석

이 논문을 석사 학위논문으로 제출함.

2019 년 2 월

부 경 대 학 교 대 학 원

통 계 학 과

김 혁 진

김혁진의 이학석사 학위논문을 인준함.

2019 년 2 월 22 일

위 원 장 이학박사 이 성 백 (인)

위 원 이학박사 박 인 호 (인)

위 원 이학박사 노 맹 석 (인)

목 차

표 차례 ii

그림 차례 iii

요약 iv

제 1장 서론(Introduction) ····· 1

제 2장 모형개요(Model Descripion) ····· 3

2.1 다단계 일반화 선형모형(HGLMs) ····· 3

2.2 결합 일반화 선형모형(JGLMs) ····· 4

2.3 이중 다단계 일반화 선형모형(DHGLMs) ····· 5

제 3장 데이터 분석(Data Analysis) ····· 10

3.1 데이터 설명(Data Description) ····· 10

3.2 모형 적합(Model Fitting Preocedure) ····· 15

3.2.1 GLM ····· 17

3.2.2 JGLM ····· 20

3.2.3 HGLM ····· 23

3.2.4 HGLM2 ····· 25

3.2.5 DHGLM ····· 29

3.2.6 DHGLM2 ····· 32

3.3 기울기 차이 검정(Slope Difference Test) ····· 35

3.4 잔차도표(Residual Plot) ····· 37

3.5 요약 정리(Summary) ····· 40

제 4장 결론 ····· 41

참고문헌 ····· 42

부록 A.1 R코드 ····· 44

표 차례

[표 2.3.1] 다양한 켤레분포의 조합 ····· 7

[표 3.1.1] 상위 8개의 열화시험 데이터 ····· 11

[표 3.1.2] 열화시험 데이터의 변수 설명 ····· 12

[표 3.1.3] 열화시험 데이터의 재설정된 변수 설명 ····· 14

[표 3.2.1] 분석에 사용된 6가지 모형 ····· 16

[표 3.2.2] 6가지 모형의 Linear predictor ····· 16

[표 3.2.3] 6가지 모형의 분포가정 및 연결함수 ····· 17

[표 3.2.1.1] GLM 적합 결과 ····· 18

[표 3.2.2.1] JGLM 적합 결과 ····· 21

[표 3.2.3.1] HGLM 적합 결과 ····· 24

[표 3.2.4.1] HGLM2 적합 결과 ····· 27

[표 3.2.5.1] DHGLM 적합 결과 ····· 30

[표 3.2.6.1] DHGLM2 적합 결과 ····· 33

[표 3.3.1] 각 모형에서의 기울기차이 검정결과 ····· 36

[표 3.5.1] 각 모형에서의 평균 및 산포모형 기울기 회귀계수 ····· 40

그림 차례

[그림 3.1.1] 타이어마모의 발생과정 ····· 13

[그림 3.1.2] 타이어의 마모도 정의 ····· 13

[그림 3.1.3] 설명변수와 반응변수의 산점도 ····· 14

[그림 3.4.1] 잔차도표(GLM,JGLM) ····· 38

[그림 3.4.2] 잔차도표(HGLM,HGLM2) ····· 39

[그림 3.4.3] 잔차도표(DHGLM,DHGLM2) ····· 49

PUKYONG NATIONAL UNIVERSITY 부경대학교

Analysis of Degradation Test Data via the Joint Modelling of Mean and Dispersion of the Double Hierarchical Generalized Linear Model

Hyuck Jin Kim

Department of Statistics, The Graduate School,
Pukyong National University

Abstract

Regression analysis was extended to include GLMs, LMMs and GLMMs as the random effect was added and the model became generalized. HGLMs have been proposed, which can expand the distribution of the random effect to a distribution other than the normal distribution (Lee and Nelder, 1996). DHGLMs is an extension of HGLMs, which also takes into account the random effect of the dispersion model (Lee and Nelder, 2006). In this study, we apply the fixed effect to the dispersion model through joint modeling (Lee and Nelder, 1991). The data used in the analysis are the degradation test data, which has the tire crack length and distance of driving as variables, and analyzed the influence of tire crack length on driving once. R package dhglm () was used for model fit of each model (Lee and Noh, 2018).

keywords : Degradation test data, Random effect, Double Hierarchical Generalized Linear Models, Joint modeling, Dispersion model

제 1장 서론(Introduction)

타이어의 마모는 타이어의 수명과 직접적인 관련이 있는 성능으로 운전자의 주요 관심사항 중 하나이다. 타이어 마모에 영향을 주는 타이어 부위 중 트레드(tread) 배합물 층은 노면과 직접 접촉하는 부위로써 타이어 원주의 최외각에 위치하고 있을 뿐만 아니라 전체 부피의 70%를 점유하고 있기 때문에 타이어에서 매우 큰 비중을 차지하고 있다. 이 타이어 트레드부와 노면 사이에 작용하는 마찰력이 트레드 고무의 강도보다 클 때 표면의 고무가 이탈하는 현상을 타이어마모라 정의할 수 있다. 이런 타이어 마모의 발생요인은 타이어의 구조, 공기압, 고무성질 등의 내부적인 요인과 차량의 사양, 노면구조, 사용조건 등의 외부적인 요인으로 나누어진다.

다단계 일반화 선형모형(Hierarchical Generalized Linear Models, HGLMs)은 일반화 선형 혼합모형(Generalized Linear Mixed Models, GLMMs)으로부터 확장된 모형으로써 산포를 모형화 할 수 있고, 다단계 가능도 접근법(Hierarchical likelihood approach)을 통하여 더욱 정밀한 추정이 가능하다. 또한 이중 다단계 일반화 선형모형은 다단계 일반화 선형모형의 확장으로써 평균모형 뿐만 아니라 산포모형에도 변량효과를 고려한 모형이다(Lee and Nelder, 1996).

본 논문에서는 여러 가지 모형의 적합을 통하여 한 번 주행 시 타이어의 마모도에 주는 영향을 분석하고자 한다. 고려한 모형은 일반화 선형

모형(Generalized Linear Models, GLMs), 다단계 일반화 선형 모형(Hierarchical Generalized Linear Models, HGLMs), 이중 다단계 일반화 선형모형(Double Hierarchical Generalized Linear Models, DHGLMs)의 3가지 평균모형과 결합 일반화 선형모형(Joint Generalized Linear Models, JGLMs)의 산포모형을 결합한 총 6가지의 모형이다. 각 모형의 비교를 위하여 모형 품질 평가의 척도로써 cAIC(Conditional Akaike Information Criterion)를 사용하였다(Lee and Noh, 2012). 각 모형의 모형 적합을 위하여 R package dhglm()을 사용하였다(Lee and Noh, 2018).

2장에서는 다단계 일반화 선형모형과 이중 다단계 일반화 선형모형, 결합 일반화 선형모형에 대하여 상세하게 설명하고, 3장에서는 분석에 사용된 열화시험데이터(Degradation test data)와 분석에 고려된 6가지의 모형과 이의 적합결과를 제시한다. 마지막으로 4장에서는 6가지 모형에 대하여 모형 선택(Model Selection)을 제시한다.

제 2장 모형 개요(Model Description)

2.1 다단계 일반화 선형모형(HGLMs)

일반화 선형 혼합모형(Generalized Linear Mixed Models, GLMMs)은 오차의 정규성 가정을 필요로 하는 선형혼합모형(Linear Mixed Models, LMMs)의 확장으로써 반응변수가 지수족 분포에서도 추론할 수 있도록 한 모형이다. 이를 더 확장하여 반응변수의 평균의 변량효과(random effect)에 정규분포 이외에 다양한 분포를 가정할 수 있도록 확장한 모형이 다단계 일반화 선형모형(Hierarchical Generalized Linear Models, HGLMs)이다(Lee and Nelder, 1996). 또한 이는 다단계 가능도 접근법(hierarchical likelihood approach)을 통하여 더욱 정밀한 추정을 가능하게 하고 모수에 대한 사전분포의 가정 없이도 관측되지 않는 변량효과를 추론하는 방법이 제시되었다.

다단계 일반화 선형모형의 간단한 구조는 평균모수에 대한 변량효과 $U^{(\mu)}=u^{(\mu)}$가 주어졌을 때 반응변수 Y의 모형으로 평균과 분산을 나타내는 두 조건부 적률은 다음과 같이 정의된다.

$$E(Y|v^{(\mu)})=\mu,\ \ Var(Y|v^{(\mu)})=\phi V(\mu).$$

여기서, ϕ는 산포모수, $V(\cdot\ )$은 분산함수이다. 평균 μ에 대한 linear

predictor를 통하여 다음과 같이 모형화된다.

$$\eta^{(\mu)} = h^{(\mu)}(\mu) = X^{(\mu)}\beta^{(\mu)} + Z^{(\mu)}v^{(\mu)}.$$

이 때 $h^{(\mu)}(\cdot)$는 μ를 고정효과와 변량효과의 선형결합과 이어주는 연결함수이다. $v^{(\mu)} = g^{(\mu)}(u^{(\mu)})$는 $u^{(\mu)}$의 함수이고 $g^{(\mu)}(\cdot)$는 단조함수이다. $\beta^{(\mu)}$와 $v^{(\mu)}$는 각각 평균모수에 대한 고정효과(Fixed effects)와 변량효과(Random effects)이고 $X^{(\mu)}$와 $Z^{(\mu)}$는 각각 이에 해당하는 모형행렬을 나타낸다. 더욱 확장된 다단계 일반화 선형모형의 구조는 3장에서 다루고자 한다.

2.2 결합 일반화 선형모형(JGLMs)

결합 일반화 선형모형(Joint Generalized Linear Models, JGLMs)는 평균모수 μ에 대해서만 고정효과를 모델링하는 일반화 선형모형의 확장으로써, 평균모수 μ 이외에도 산포모수 ϕ에 대해서도 고정효과를 모델링하는 방법론이다. linear predictor를 통하여 다음과 같이 모형화 할 수 있다.

$$\eta^{(\mu)} = h^{(\mu)}(\mu) = X^{(\mu)}\beta^{(\mu)},$$
$$\eta^{(\phi)} = h^{(\phi)}(\phi) = X^{(\phi)}\beta^{(\phi)}.$$

이 때 $h^{(\mu)}(\cdot)$와 $h^{(\phi)}(\cdot)$는 μ와 ϕ의 고정효과이고 $X^{(\mu)}$와 $X^{(\phi)}$는 이에 해당하는 모형행렬을 나타낸다.

2.3 이중 다단계 일반화 선형모형(DHGLMs)

이중 다단계 일반화 선형모형(Double Hierarchical Generalized Linear Models, DHGLMs)는 다단계 일반화 선형모형의 확장으로써 산포모형에도 변량효과를 고려한 모형이다(Lee and Nelder, 2006). 확장된 이중 다단계 일반화 선형모형의 구조를 위하여 $\lambda = var(u^{(\mu)}), \alpha = var(u^{(\lambda)})$를 고려하자. 이중 다단계 일반화 선형모형의 구조는 변량효과 $u^{(\mu,\lambda,\phi)} = (u^{(\mu)}, u^{(\lambda)}, u^{(\phi)})^T$가 주어졌을 때 반응변수 Y의 모형으로 평균과 분산을 나타내는 두 조건부 적률은 다음과 같이 정의된다.

$$E(Y|u^{(\mu,\lambda,\phi)}) = \mu, \quad Var(Y|u^{(\mu,\lambda,\phi)}) = \phi V(\mu).$$

다단계 일반화 선형모형은 이중 다단계 일반화 선형모형의 부속모형(Submodel)으로써 $\tau = var(u^{(\phi)}) = 0$, 즉 모든 산포모수의 변량효과 $u_i^{(\phi)}$가 1인 이중 다단계 일반화 선형모형과 동일하다.

결합 일반화 선형모형과 결합하여 $\mu, \phi, \lambda, \alpha, \tau$ 모두에 대하여 모델링함으로써 다음과 같이 5가지 모수에 대한 linear predictor를 통하여 확장할 수 있다.

$$\eta^{(\mu)} = h^{(\mu)}(\mu) = X^{(\mu)}\beta^{(\mu)} + Z^{(\mu)}v^{(\mu)},$$
$$\eta^{(\lambda)} = h^{(\lambda)}(\lambda) = X^{(\lambda)}\beta^{(\lambda)} + Z^{(\lambda)}v^{(\lambda)},$$
$$\eta^{(\tau)} = h^{(\tau)}(\tau) = X^{(\tau)}\beta^{(\tau)},$$
$$\eta^{(\phi)} = h^{(\phi)}(\phi) = X^{(\phi)}\beta^{(\phi)} + Z^{(\phi)}v^{(\phi)},$$
$$\eta^{(\alpha)} = h^{(\alpha)}(\alpha) = X^{(\alpha)}\beta^{(\alpha)}.$$

이 때 $h^{(\theta)}$는 해당 모수 θ와 이의 고정효과(또는 변량효과를 포함하여)의 선형결합을 이어주는 연결함수이다. $v^{(\theta)} = g^{(\theta)}(u^{(\theta)})$는 $u^{(\theta)}$의 함수이고 $g^{(\theta)}$는 단조함수이다. 그리고 $\beta^{(\theta)}$와 $v^{(\theta)}$는 각각 θ의 고정효과 및 변량효과이고 $X^{(\theta)}$와 $Z^{(\theta)}$는 각각 이에 해당하는 모형행렬을 나타낸다.

반응변수 $Y|w$는 지수족 분포이고 변량효과 $u^{(\mu)}, u^{(\lambda)}, u^{(\phi)}$는 $Y|w$의 켤레분포(Conjugate Distribution)로 정규분포 이외의 분파의 가정이 가능하며 다양한 조합을 생성할 수 있다. 표 2.3.1은 이러한 켤레분포의 다양한 조합을 나타낸다.

Lee와 Nelder(1996)은 이중 다단계 일반화 선형모형에서 주변가능도(Marginal Likelihood, ML)에서는 다루기 힘든 적분을 보완한 다단계 가능도(Hierarchical Likelihood)를 사용하였고, $(y, v^{(\mu)}, v^{(\lambda)}, v^{(\phi)})^T$에 대한 로그 결합 밀도함수를 다음과 같이 정의하였다.

$$h = \log f_{\beta^{(\mu)},\beta^{(\phi)}}(y|v^{(\mu)}, v^{(\phi)}) + \log f_{\beta^{(\lambda)}}(v^{(\mu)}|v^{(\lambda)}) + \log f_{\beta^{(\alpha)}}(v^{(\lambda)}) + \log f_{\beta^{(\tau)}}(v^{(\phi)}).$$

표 2.3.1 다양한 켤레분포의 조합

Model name	$Y\|u^{(\mu)}$ Dist.	Link $h^{(\mu)}(\mu)$	$u^{(\mu)}$ dist.	Link $g^{(\mu)}(u^{(\mu)})$
LMMs	Gaussian	Identity	Gaussian	Identity
Binomial conjugate	Binomial	Logit	Beta	Logit
Binomial GLMMs	Binomial	Logit	Gaussian	Identity
Binomial frailty	Binomial	CLL[1)]	Gamma	Log
Poisson GLMMs	Poisson	Log	Gaussian	Identity
Poisson conjugate	Poisson	Log	Gamma	Log
Gamma GLMMs	Gamma	Log	Gaussian	Identity
Gamma conjugate	Gamma	Inverse	Inv. Gamma[2)]	Inverse
Gamma-Gamma	Gamma	Log	Gamma	Log

여기서 $f_{\beta^{(\mu)},\beta^{(\phi)}}(y|v^{(\mu)},v^{(\phi)})$, $f_{\beta^{(\lambda)}}(v^{(\mu)}|v^{(\lambda)})$, $f_{\beta^{(\alpha)}}(v^{(\lambda)})$, $f_{\beta^{(\tau)}}(v^{(\phi)})$은 각각 $(y|v^{(\mu)},v^{(\phi)})$, $(v^{(\mu)}|v^{(\lambda)})$, $(v^{(\lambda)})$, $(v^{(\phi)})$의 조건부 밀도함수이다.

다단계 가능도 추정법(Maximum Hierarchical Likelihood Estimate, MHLE)은 다단계 가능도를 최대로 하는 추정법이다. 최대가능도추정법으로 구해진 $\hat{\beta}$을 이용하여 $\hat{v}$은 경험적 베이즈 추정법(Empirical Bayes E

1) Complementary-Log-Log

2) Inverse Gamma

stimate)으로 $\hat{E}(v|y)$의 형태를 가지며 적분함으로써 구해진다. 따라서 주변가능도는 h를 적분함으로써 얻어지는 $m=\log\int\exp(h)dv$이며 이를 최대화하는 v를 추정한다. Fisher Likelihood를 얻기 위한 적분은 일반적으로 수치계산이 쉽지 않기 때문에 근사적인 방법으로써 1차 라플라스 근사에 해당하는 다음의 수정된 단면 다단계 가능도(Adjusted Profile Hierarchical Likelihoods, APHLs)를 제안하였다(Lee and Nelder, 2001a).

$$p_{\xi}(\ell)=\left[\ell-\frac{1}{2}log\det\{H(\ell,\xi)/(2\pi)\}\right]\Big|_{\xi=\tilde{\xi}}$$

여기서, ℓ은 다단계 가능도를 의미하고 $H(\ell,\xi)=-\partial^2\ell/\partial\xi\partial\xi^T$이고 $\hat{\xi}$는 $\partial\ell/\partial\xi=0$의 해이다. 또한 더욱 정밀한 근사를 위하여 다음의 2차 라플라스근사를 이용하였다.

$$p^{s}_{v^{(\mu,\lambda,\phi)}}(h)=p_v(\mu,\lambda,\phi)(h)-tr(F)/24.$$

여기서, $F=\left[-\{3h_4+5h_3H(h,v)^{-1}h_3\}H(h,v)^{-2}\right]\Big|_{v^{(\mu,\lambda\phi)}=\tilde{v}^{(\mu,\lambda\phi)}}$이고,
$h_3=\partial^3h/\partial(v^{(\mu,\lambda,\phi)})^3, h_4=\partial^4h/\partial(v^{(\mu,\lambda,\phi)})^4$이다.

또한 변량효과가 포함된 모형에서 사용되는 모형 품질 평가의 척도로써 cAIC(conditional Akaike Information Criterion)를 사용하였고, 다음과 같이 정의된다(Lee and Noh, 2012).

$$cAIC = D + 2p_D,$$

이 때 $D = D(\hat{\mu}) = -2\ell(\hat{\mu};y|v^{(\mu,\lambda,\phi)})$, $\ell(\mu;y|v^{(\mu,\lambda,\phi)}) = \log f_{\beta^{(\mu)},\beta^{(\phi)}}(y|v^{(\mu,\lambda,\phi)})$, $p_D = tr(H^{-1}H^*)$, $H = -2\partial^2 h/\partial(\beta^{(\mu)}, v^{(\mu,\lambda,\phi)})^2$, $H^* = -2\partial^2\ell/\partial(\beta^{(\mu)}, v^{(\mu,\lambda,\phi)})^2$이다. cAIC는 변량효과가 없는 모형에서 AIC(Akaike Information Criterion)와 동일한 값을 가진다.

PUKYONG NATIONAL UNIVERSITY 부경대학교

제 3장 데이터 분석(Data Analysis)

3.1 데이터 설명(Data Description)

제품의 수명 예측을 위한 수명시험에서 제품의 수명이 다할 때 까지 시험을 하는 것은 경제적, 시간적으로 큰 부담이 된다. 산업기술 발전 도가 빨라지고 제품의 수명주기와 개발시간이 단축되었고, 따라서 결과적으로 수반되는 설계상 문제의 해결과 신기술 및 부품적용에 대한 신뢰성 평가를 위해 사용조건보다 가혹한 조건으로 시험 시간을 단축하여 실시하는 가속수명시험 또는 열화시험 등이 시행되기도 한다. 본 논문에서는 열화시험 데이터인 수명이 다하지 않은 타이어의 데이터를 이용하여 타이어의 마모도에 주는 영향을 차량별 변량효과를 고려함으로써 분석하고자 한다.

데이터는 다음과 같이 수집되었다. 임의로 선택된 다섯 대의 물류 차량의 앞바퀴 양쪽에 같은 종류의 새 타이어를 장학한 후 각각의 차량에 화물을 적재하고 정해진 같은 구간을 반복 주행하였다. 일정기간이 지난 후 차량들의 주행거리와 타이어의 홈의 깊이를 각각 측정하였다. 주행거리는 주행 후 계기판의 값과 타이어 장착 시 계기판의 값의 차이이고 타이어의 서로 다른 2군데에서 측정한 홈의 깊이의 평균값이다. 총 4회의 주행시험이 시행되었다.

3.1.1 상위 8개의 열화시험 데이터

No.	ID	K	T	Y	X_1	X_2
1	1	1	1	1.515	1.593	0
2	1	1	2	3.700	4.110	0
3	1	1	3	5.500	5.705	0
4	1	1	4	9.538	8.445	0
5	1	2	1	1.150	0	1.593
6	1	2	2	3.000	0	4.110
7	1	2	3	3.925	0	5.705
8	1	2	4	6.738	0	8.445

표 3.1.1은 본 논문에서 사용된 열화시험 데이터의 상위 8개를 정리표한 표이고, 표 3.1.2는 데이터의 변수를 설명한 표이다. 설명변수는 왼쪽 및 오른쪽 주행거리로 구성된다. 표 3.1.1에서 K는 왼쪽바퀴와 오른쪽바퀴의 위치를 의미하는 변수로써 $K=1$은 왼쪽바퀴, $K=2$는 오른쪽바퀴를 나타낸다. 따라서 $K=1$에 해당하는 X_2의 관측값과 $K=2$에 해당하는 X_1의 관측값은 모두 0이다. 또한 T는 시행횟수로써 총 4회의 주행시험이 시행되었으므로 1,2,3,4의 값을 가진다. 4번째 시험에서의 5번째 차량의 왼쪽바퀴 마모도는 결측되었다. 따라서 마모도의 총 관측수는 5×4×2(=차량의수×반복측정횟수×바퀴수)에서 결측치 1건을 제외한 39회이다. ID는 차량번호로써 1,2,3,4,5의 5대의 차량을 나타낸다. Y는 반응변수로써 타이어의 마모도이며 단위는 mm이다. 설명변수는 X_1, X_2로 각각 왼쪽 및 오른쪽 주행거리를 나타내며 단위는 $10{,}000km$이다.

표 3.1.2 열화시험 데이터의 변수 설명

변수명	설명	값	단위
Y	타이어 마모도(17−홈의 깊이)	$0 \le Y \le 17$	mm
ID	차량번호	1,2,3,4,5	
X_1	왼쪽 주행거리	$X_1 \ge 0$	$10{,}000km$
X_2	오른쪽 주행거리	$X_2 \ge 0$	$10{,}000km$

타이어마모의 발생과정과 타이어의 마모도에 대한 정의는 그림 3.1.1과 그림 3.1.2를 통하여 설명되어있다. 타이어마모는 차량 주행 시 노면에 닿는 타이어 트레드부의 마찰이 발생함으로써 표면의 고무가 이탈하여 발생한다. 또한 이탈된 표면의 고무는 타이어의 홈(groove)으로 밀려들어가게 되고 따라서 타이어의 홈의 깊이는 줄어들게 된다. 따라서 타이어의 마모도를 초기 홈의 깊이인 $17mm$에서 주행시험 이후에 측정된 홈의 깊이를 뺀 값으로 정의하였다.

본 논문에서는 "*한번 주행시*" 주행거리에 따른 타이어의 마모도에 미치는 영향을 분석하고자 한다. 따라서 반응변수 X_1과 X_2 및 설명변수 Y에 대하여 다음과 같이 변수를 재설정 하였다.

$$d_{ikt} = D_{ikt} - D_{ik,t-1}.$$

여기서 D_{ikt}는 i번째 차량의 시행횟수 t에서의 왼쪽바퀴($k=1$) 또는 오른쪽바퀴($k=2$)의 관측값이고, d_{ikt}는 i번째 차량의 왼쪽 또는 오른쪽바퀴

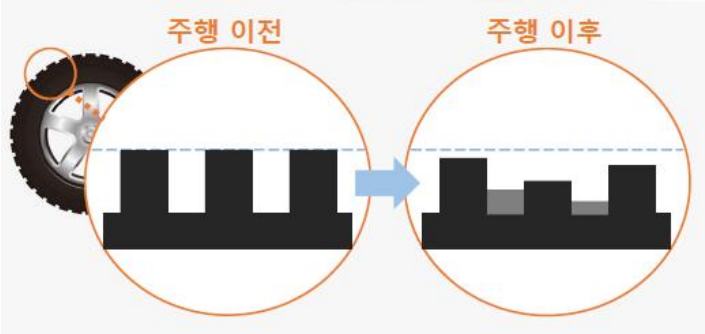

그림 3.1.1 타이어마모의 발생과정

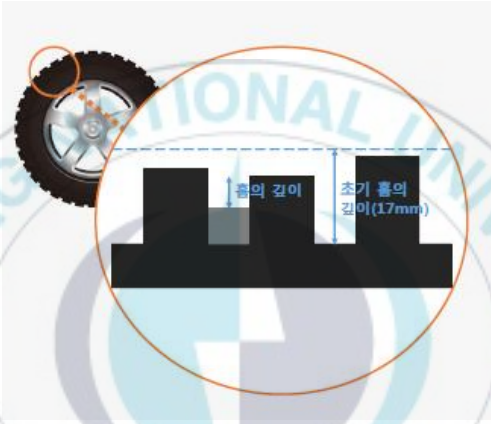

그림 3.1.2 타이어의 마모도 정의

퀴의 관측값의 시행횟수 t시점과 $t-1$시점의 차이값이다. 또한 새로운 변수 x_3: 주행시험 이전의 타이어 누적마모도를 추가하였다. 따라서 한 번 주행 시 타이어의 마모도에 주는 영향을 분석하고자 한다. 새로운 변수에 대한 설명은 표 3.1.3에 제시되어있다.

설명변수와 반응변수의 산점도는 그림 3.1.3에 제시되어있다. 이를 통하여 대체적으로 왼쪽 및 오른쪽의 해당 주행시험 이후의 주행거리(x_1, x_2)

표 3.1.3 열화시험 데이터의 재설정된 변수 설명

변수명	설명	값	단위
Y	해당 주행시험에서의 타이어 마모도	$0 \leq Y \leq 17$	mm
ID	차량번호	$1,2,3,4,5$	
x_1	해당 주행시험에서의 왼쪽 주행거리	$x_1 \geq 0$	$10,000km$
x_2	해당 주행시험에서의 오른쪽 주행거리	$x_2 \geq 0$	$10,000km$
x_3	해당 주행시험 이전의 타이어 누적마모도	$x_3 \geq 0$	mm

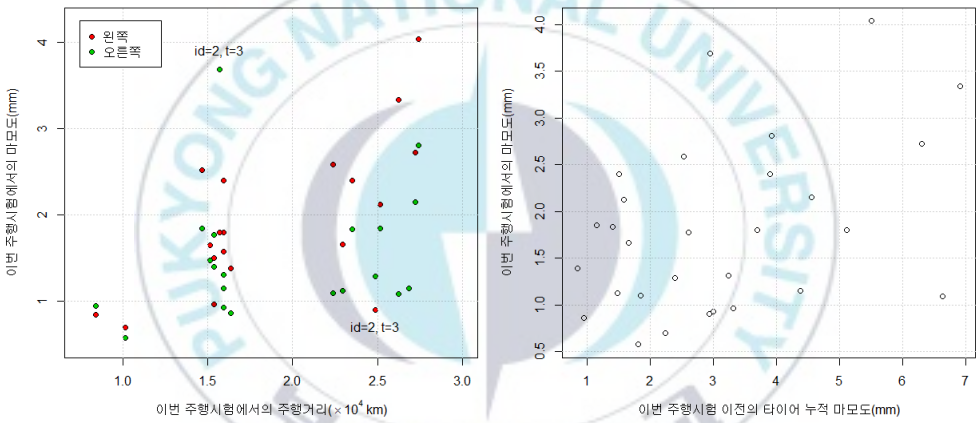

그림 3.1.3 설명변수와 반응변수의

산점도(좌 : $x_1, x_2 vs\ y$, 우 : $x_3 vs\ y$)

가 증가할수록 마모도가 각각 선형적으로 증가하고 있음을 알 수 있다. 또한 해당 주행시험 이전의 타이어 누적마모도가 증가할수록 마모도가 선형적으로 증가하고 있음을 알 수 있다.

또한 본 논문에서는 왼쪽 및 오른쪽의 해당 주행시험 이후의 주행거리 (x_1, x_2)가 0이고 해당 주행시험 이전의 타이어 누적마모도(x_3)가 0일 때, 해당 주행시험에서의 타이어 마모도(y)가 0을 가지는 절편이 없는 선형 모형을 고려하고자 한다.

3.2 모형 적합(Model Fitting Procedure)

모형은 평균모수 μ에 대한 모형과 산포모수 ϕ에 대한 모형에 대하여 여러 가지의 linear predictor 구성요소 조합으로 이루어진다. 더 나아가서, 표 3.2.1과 같이 각 모형을 평균모수 및 산포모수에 대하여 모형화 하였는지, 고정효과 또는 변량효과가 고려되었는지에 따라 상세하게 분류할 수 있다.

표 3.2.2는 6가지 모형에 대하여 μ, ϕ에 대한 2가지의 linear predictor를 나타낸 표이다. 평균모형 μ에서 변량효과가 포함된 모형에서는 왼쪽과 오른쪽바퀴 주행거리의 차량별 변량효과를 고려하였다. 또한 산포모형 ϕ에서 변량효과가 포함된 모형에서는 차량별 변량효과를 고려하였다.

표 3.2.3은 6가지 모형의 반응변수와 변량효과의 분포가정, 평균모형 및 산포모형의 연결함수를 나타낸 표이다.

표 3.2.1 분석에 사용된 6가지 모형

Random effects \ Fixed effects	Mean model	Mean, Dispersion model
Not included	GLM	JGLM
Mean model	HGLM	HGLM2
Mean, Dispersion model	DHGLM	DHGLM2

표 3.2.2 6가지 모형의 Linear predictor

Model	Linear Predictor	
	μ	ϕ
GLM	$\log\mu = X^{(\mu)}\beta^{(\mu)}$	$\log\phi = \beta_0^{(\phi)}$
HGLM	$\log\mu = X^{(\mu)}\beta^{(\mu)} + X^{(\mu)}v^{(\mu)}$	$\log\phi = \beta_0^{(\phi)}$
DHGLM	$\log\mu = X^{(\mu)}\beta^{(\mu)} + X^{(\mu)}v^{(\mu)}$	$\log\phi = \beta_0^{(\phi)} + v^{(\phi)}$
JGLM	$\log\mu = X^{(\mu)}\beta^{(\mu)}$	$\log\phi = X^{(\phi)}\beta^{(\phi)}$
HGLM2	$\log\mu = X^{(\mu)}\beta^{(\mu)} + X^{(\mu)}v^{(\mu)}$	$\log\phi = X^{(\phi)}\beta^{(\phi)} + v^{(\phi)}$
DHGLM2	$\log\mu = X^{(\mu)}\beta^{(\mu)} + X^{(\mu)}v^{(\mu)}$	$\log\phi = X^{(\phi)}\beta^{(\phi)} + v^{(\phi)}$

표 3.2.3 6가지 모형의 분포가정 및 연결함수

Model	Dist. of response	Link for mean	Dist. of random effects for mean para.[3]	Dist of random effect for disp. para.[4]	Link for dispersion
GLM	Gamma	Log			
HGLM	Gamma	Log	Gaussian		
DHGLM	Gamma	Log	Gaussian	Gaussian	Log
JGLM	Gamma	Log			Log
HGLM2	Gamma	Log	Gaussian		Log
DHGLM2	Gamma	Log	Gaussian	Gaussian	Log

3.2.1 GLM

첫 번째 모형인 GLM은 6가지 모형 중에서 가장 기본적인 모형으로써, 다음과 같은 linear predictor로 이루어진 모형이다. 수식에서 지시함수 표현의 단순화를 위하여 $x_{jikt} = x_{jikt} I(K = j), j = 1,2$로 표현하였다.

$$\eta_{ikt}^{(\mu)} = \log \mu_{ikt} = \sum_{j=1}^{3} \beta_j^{(\mu)} x_{jikt},$$

$$\eta_{ikt}^{(\phi)} = \log \phi_{ikt} = \beta_0^{(\phi)}.$$

3) Random slope for each level of ID : (x1|id)+(x2|id)

4) Random intercept for each level of ID : (1|id)

표 3.2.1.1 GLM 적합 결과

	Estimate	Std. Error	t value	Pr(>|t|)	95% CI
$\hat{\beta}_1^{(\mu)}$	0.298	0.073	4.099	0.000***	0.298 ± 0.145 (0.155 ~ 0.440)
$\hat{\beta}_2^{(\mu)}$	0.174	0.064	2.707	0.007**	0.174 ± 0.129 (0.048 ~ 0.300)
$\hat{\beta}_3^{(\mu)}$	0.034	0.039	0.862	0.388	0.034 ± 0.078 (−0.043 ~ 0.110)
$\hat{\beta}_0^{(\phi)}$	−1.847	0.252	−7.343	0.000***	−1.847 ± 0.503 (−2.340 ~ −1.354)
cAIC	77.076				

여기서, 반응변수의 분포 Y_{ikt}는 감마분포를 가정하였다. μ_{ikt}는 반응변수에 대한 기댓값으로, 해당 주행시험에서의 i번째 차량의 시행횟수 t에서의 왼쪽바퀴($k=1$) 또는 오른쪽바퀴($k=2$)의 평균 타이어 마모도이다. ϕ_{ikt}는 반응변수에 대한 잔차의 분산(residual variance)으로, 해당 주행시험에서의 i번째 차량의 시행횟수 t에서의 왼쪽바퀴($k=1$) 또는 오른쪽바퀴($k=2$)의 타이어의 마모도에 대한 잔차의 분산이다. 모든 모수에 대한 연결함수 $h^{(\mu)}, h^{(\phi)}$는 로그연결함수(log link function)을 고려하였다. 그리고 x_{jikt}는 i번째 차량의 시행횟수 t에서의 왼쪽바퀴($k=1$) 또는 오른쪽바퀴($k=2$)의 j번째 설명변수이고, $\beta_j^{(\mu)}$는 이에 해당하는 고정효과이다. 평균모형인 μ의 linear predictor에서는 절편(intercept)이 제외된 고정효과가 있는 가법모형(additive model)을 고려하였고, 산포모형인 ϕ

의 linear predictor에서는 절편만 포함된 모형을 고려하였다. $I(k=k_0)$는 왼쪽 또는 오른쪽바퀴의 위치를 나타내는 지시함수(indicator function)이다. 참고로 모형행렬(design matrix)은 왼쪽과 오른쪽바퀴에 대한 위치를 지시함수로 나타냄으로써 x_1과 x_2는 직교(orthogonal)하게 나타난다.

GLM에서의 추정모수는 평균모형 μ에서의 전체 차량에 대한 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리의 기울기 $\beta_1^{(\mu)}, \beta_2^{(\mu)}$와 이전 누적마모도의 기울기 $\beta_3^{(\mu)}$의 3개의 고정효과로 이루어지고, 산포모형 ϕ에서의 절편 $\beta_0^{(\phi)}$으로 이루어진다. 적합결과는 cAIC와 함께 표 3.2.1.1에 제시되어 있다.

평균모형에서의 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리에 대한 기울기의 추정치 $\hat{\beta}_1^{(\mu)}, \hat{\beta}_2^{(\mu)}$은 각각 $0.298, 0.174$로, 다른 설명변수가 고정되었을 때 왼쪽 및 오른쪽 주행거리 $10{,}000km$당 각각 평균 타이어 마모가 $1.347, 1.190$배 증가하는 것으로 나타났다. 그리고 평균모형에서의 해당 주행시험 이전의 타이어 누적마모도에 대한 기울기의 추정치 $\hat{\beta}_3^{(\mu)}$은 0.034로, 다른 설명변수가 고정되었을 때 이전 누적마모도 $1mm$당 평균 타이어 마모가 1.034배 증가하는 것으로 나타났다. 귀무가설 $H_0 : \beta_j^{(\mu)} = 0, j = 1,2,3$에 대하여 $p-value$는 각각 $0.000, 0.007, 0.388$로 유의수준 5%에서 해당 주행시험 이전의 타이어 누적마모도(x_3)가 통계적으로 유의하지 않은 것으로 나타났다.

모형의 cAIC는 변량효과가 없으므로 AIC와 동일한 $-2h+2df=77.076$으로 나타났다. 여기서 h는 다단계 가능도이고, df는 자유도이다.

3.2.2 JGLM

두 번째 모형인 JGLM은 첫 번째 모형 GLM에서 산포모형의 고정효과가 추가된 모형으로써, 다음과 같은 linear predictor로 이루어진 모형이다.

$$\eta_{ikt}^{(\mu)} = \log\mu_{ikt} = \sum_{j=1}^{3}\beta_j^{(\mu)}x_{jikt},$$

$$\eta_{ikt}^{(\phi)} = \log\phi_{ikt} = \beta_0^{(\phi)} + \sum_{j=1}^{3}\beta_j^{(\phi)}x_{jikt}.$$

GLM과 마찬가지로, 반응변수의 분포 Y_{ikt}는 감마분포를 가정하였고 $h^{(\mu)}, h^{(\phi)}$은 로그연결함수를 고려하였다. 평균모형인 μ의 linear predictor에서는 절편이 제외된 고정효과가 있는 가법모형을 고려하였고, 산포모형인 ϕ의 linear predictor에서는 절편을 포함하는 고정효과가 포함된 가법모형을 고려하였다.

JGLM에서의 추정모수는 평균모형 μ에서의 전체 차량에 대한 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리의 기울기 $\beta_1^{(\mu)}, \beta_2^{(\mu)}$와 이전 누적마모도의 기울기 $\beta_3^{(\mu)}$의 3개의 고정효과로 이루어지고, 산포모형 ϕ

표 3.2.2.1 JGLM 적합 결과

	Estimate	Std. Error	t value	$Pr(>\|t\|)$	95% CI
$\hat{\beta}_1^{(\mu)}$	0.300	0.055	5.473	0.000***	0.300 ± 0.107 (0.193 ~ 0.407)
$\hat{\beta}_2^{(\mu)}$	0.185	0.052	3.564	0.000***	0.185 ± 0.102 (0.083 ~ 0.287)
$\hat{\beta}_3^{(\mu)}$	0.025	0.040	0.617	0.537	0.025 ± 0.079 (−0.054 ~ 0.103)
$\hat{\beta}_0^{(\phi)}$	−1.306	0.864	−1.512	0.131	−1.306 ± 1.694 (−3.000 ~ 0.387)
$\hat{\beta}_1^{(\phi)}$	−0.807	0.548	−1.474	0.140	−0.807 ± 1.073 (−1.881 ~ 0.266)
$\hat{\beta}_2^{(\phi)}$	−0.744	0.516	−1.441	0.149	−0.744 ± 1.012 (−1.755 ~ 0.268)
$\hat{\beta}_3^{(\phi)}$	−0.352	0.150	2.358	0.018*	−0.352 ± 0.293 (−0.059 ~ 0.645)
cAIC	72.398				

에서의 절편 $\beta_0^{(\phi)}$와 전체 차량에 대한 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리의 기울기 $\beta_1^{(\phi)}, \beta_2^{(\phi)}$와 이전 누적마모도의 기울기 $\beta_3^{(\phi)}$의 3개의 고정효과로 이루어진다. 적합결과는 cAIC와 함께 표 3.2.2.1에 제시되어 있다.

평균모형에서의 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리에 대한 기울기의 추정치 $\hat{\beta}_1^{(\mu)}, \hat{\beta}_2^{(\mu)}$은 각각 $0.300, 0.185$로, 다른 설명변수가 고정되었을 때 왼쪽 및 오른쪽 주행거리 $10{,}000km$당 각각 평균 타이어 마모가 $1.350, 1.203$배 증가하는 것으로 나타났다. 그리고 평균모형에서의 해당 주행시험 이전의 타이어 누적마모도에 대한 기울기의 추정치 $\hat{\beta}_3^{(\mu)}$은 0.025로, 다른 설명변수가 고정되었을 때 이전 누적마모도 $1mm$당 평균 타이어 마모가 1.025배 증가하는 것으로 나타났다. 귀무가설 $H_0 : \beta_j^{(\mu)} = 0, j = 1,2,3$에 대하여 $p-value$는 각각 $0.000, 0.000, 0.537$로 유의수준 5%에서 해당 주행시험 이전의 타이어 누적마모도(x_3)가 통계적으로 유의하지 않은 것으로 나타났다.

산포모형에서의 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리에 대한 기울기의 추정치 $\hat{\beta}_1^{(\phi)}, \hat{\beta}_2^{(\phi)}$는 각각 $-0.807, -0.744$로, 다른 설명변수가 고정되었을 때 왼쪽 및 오른쪽 주행거리 $10{,}000km$당 타이어 마모의 잔차 분산이 각각 $0.446, 0.475$배 감소하는 것으로 나타났다. 또한, 산포모형에서의 해당 주행시험 이전의 타이어 누적마모도에 대한 기울기의 추정치 $\hat{\beta}_3^{(\phi)}$은 0.352로, 다른 설명변수가 고정되었을 때 이전 누적마모도 $1mm$당 타이어 마모의 잔차 분산이 1.422배 증가하는 것으로 나타났다. 귀무가설 $H_0 : \beta_j^{(\phi)} = 0, j = 1,2,3$에 대하여 $p-value$는 각각 $0.140, 0.149$,0.018로 유의수준 5%에서 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리(x_1, x_2)가 통계적으로 유의하지 않은 것으로 나타났다.

모형의 cAIC는 변량효과가 없으므로 AIC와 동일한 $-2h + 2df = 72.398$

로 나타났다.

3.2.3 HGLM

세 번째 모형인 HGLM은 GLM에서 평균모형에 변량효과가 추가된 모형이다. 변량효과는 random slope로써 왼쪽과 오른쪽바퀴 주행거리의 차량별 변량효과를 고려하였으며, 다음과 같은 linear predictor로 이루어진 모형이다.

$$\eta_{ikt}^{(\mu)} = \log\mu_{ikt} = \sum_{j=1}^{3}\beta_j^{(\mu)}x_{jikt} + \sum_{j=1}^{2}v_{ji}^{(\mu)}x_{jikt},$$

$$\eta_{ikt}^{(\phi)} = \log\phi_{ikt} = \beta_0^{(\phi)}.$$

여기서 변량효과 $v_i^{(\mu)} = (v_{1i}^{(\mu)}, v_{2i}^{(\mu)})^T$가 주어졌을 때 반응변수 $Y_{ikt}|v_i^{(\mu)}$는 감마분포이고, 평균모형에 대한 변량효과의 분포는 $v_{ji}^{(\mu)} \sim N(0,\lambda_j), j=1,2$를 가정하였다. 평균 및 산포모수에 대한 연결함수 $h^{(\mu)}, h^{(\phi)}$는 모두 로그연결함수를 고려하였다. 평균모형인 μ의 linear predictor에서는 절편이 없고 고정효과와 변량효과가 모두 있는 가법모형을 고려하였다. 또한, 산포모형인 ϕ의 linear predictor에서는 절편만 포함된 모형을 고려하였다.

HGLM에서의 추정모수는 평균모형 μ에서의 전체 차량에 대한 왼쪽과 오른쪽바퀴의 해당누적마모도의 기울기 $\beta_3^{(\mu)}$의 3개의 고정효과로 이루어지고, 관측되지 않는 주행시험에서의 주행거리의 기울기 $\beta_1^{(\mu)}, \beta_2^{(\mu)}$와 이전

표 3.2.3.1 HGLM 적합 결과

	Estimate	Std. Error	t value	$\Pr(>\|t\|)$	95% CI
$\hat{\beta}_1^{(\mu)}$	0.283	0.091	3.096	0.002**	0.283 ± 0.179 (0.104 ~ 0.462)
$\hat{\beta}_2^{(\mu)}$	0.193	0.058	3.318	0.000***	0.193 ± 0.114 (0.079 ~ 0.307)
$\hat{\beta}_3^{(\mu)}$	0.016	0.036	0.461	0.645	0.016 ± 0.070 (−0.053 ~ 0.086)
$\hat{\beta}_1^{(\lambda)}$	−3.879	1.212	−3.200	0.001**	−3.879 ± 2.376 (−6.256 ~ −1.503)
$\hat{\beta}_2^{(\lambda)}$	−8.449	6.167	−1.370	0.171	−8.449 ± 12.087 (−20.536 ~ 3.638)
$\hat{\beta}_0^{(\phi)}$	−2.019	0.285	−7.087	0.000***	−2.019 ± 0.558 (−2.577 ~ −1.460)
cAIC	73.050				

왼쪽과 오른쪽바퀴의 주행거리의 차량별 변량효과 $v_{ji}^{(\mu)}, j=1,2$와 이의 변량효과의 분산인 $\lambda_j, j=1,2$와 산포모형 ϕ에서의 절편 $\beta_0^{(\phi)}$도 포함된다. 적합결과는 cAIC와 함께 표 3.2.3.1에 제시되어 있다.

평균모형에서의 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리에 대한 기울기의 추정치 $\hat{\beta}_1^{(\mu)}, \hat{\beta}_2^{(\mu)}$은 각각 $0.283, 0.193$으로, 다른 설명변수가 고정되었을 때 왼쪽 및 오른쪽 주행거리 $10{,}000km$당 각각 평균 타이어

마모가 1.327, 1.213배 증가하는 것으로 나타났다. 그리고 평균모형에서의 해당 주행시험 이전의 타이어 누적마모도에 대한 기울기의 추정치 $\hat{\beta}_3^{(\mu)}$은 0.016으로, 다른 설명변수가 고정되었을 때 이전 누적마모도 $1mm$당 평균 타이어 마모가 1.017배 증가하는 것으로 나타났다. 귀무가설 $H_0 : \beta_j^{(\mu)} = 0, j = 1,2,3$에 대하여 $p-value$는 각각 0.002, 0.000, 0.645로 유의수준 5%에서 해당 주행시험 이전의 타이어 누적마모도(x_3)가 통계적으로 유의하지 않은 것으로 나타났다.

또한 변량효과의 분산에 대한 추정치인 $\beta_j^{(\lambda)}, j = 1,2$는 각각 $-3.879, -8.449$로 평균모형에 추가된 왼쪽과 오른쪽바퀴 주행거리의 차량별 변량효과의 분산은 각각 $0.021mm^2, 0.0002mm^2$로 나타났다. 귀무가설 $H_0 : \beta_j^{(\lambda)} = 0, j = 1,2$에 대하여 $p-value$는 각각 0.001, 0.171로 유의수준 5%에서 오른쪽바퀴 주행거리의 차량별 변량효과가 통계적으로 유의하지 않은 것으로 나타났다. 그리고 모형의 cAIC는 73.050으로 나타났다.

3.2.4 HGLM2

네 번째 모형인 HGLM2는 HGLM에서 산포모형에 고정효과가 추가된 모형이며, 다음과 같은 linear predictor로 이루어진 모형이다.

$$\eta_{ikt}^{(\mu)} = \log\mu_{ikt} = \sum_{j=1}^{3}\beta_j^{(\mu)}x_{jikt} + \sum_{j=1}^{2}v_{ji}^{(\mu)}x_{jikt},$$

$$\eta_{ikt}^{(\phi)} = \log\phi_{ikt} = \beta_0^{(\phi)} + \sum_{j=1}^{3}\beta_j^{(\phi)}x_{jikt}.$$

HGLM과 마찬가지로 변량효과 $v_i^{(\mu)} = (v_{1i}^{(\mu)}, v_{2i}^{(\mu)})^T$가 주어졌을 때 반응변수 $Y_{ikt}|v_i^{(\mu)}$는 감마분포이고, 평균모형에 대한 변량효과의 분포는 $v_{ji}^{(\mu)} \sim N(0,\lambda_j), j=1,2$를 가정하였다. 평균 및 산포모수에 대한 연결함수 $h^{(\mu)}, h^{(\phi)}$는 모두 로그연결함수를 고려하였다. 평균모형인 μ의 linear predictor에서는 절편이 없고 고정효과와 변량효과가 모두 있는 가법모형을 고려하였다. 또한, 산포모형인 ϕ의 linear predictor에서는 절편을 포함하는 고정효과가 포함된 가법모형을 고려하였다.

HGLM2에서의 추정모수는 평균모형 μ에서의 전체 차량에 대한 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리의 기울기 $\beta_1^{(\mu)}, \beta_2^{(\mu)}$와 이전 누적마모도의 기울기 $\beta_3^{(\mu)}$의 3개의 고정효과로 이루어지고, 관측되지 않는 왼쪽과 오른쪽바퀴의 주행거리의 차량별 변량효과 $v_{ji}^{(\mu)}, j=1,2$와 이의 변량효과의 분산인 $\lambda_j, j=1,2$가 포함된다. 또한 산포모형 ϕ에서의 절편 $\beta_0^{(\phi)}$과 전체 차량에 대한 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리의 기울기 $\beta_1^{(\phi)}, \beta_2^{(\phi)}$와 이전 누적마모도의 기울기 $\beta_3^{(\phi)}$의 3개의 고정효과도 포함된다. 적합결과는 cAIC와 함께 표 3.2.4.1에 제시되어 있다.

평균모형에서의 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리에 대한 기울기의 추정치 $\hat{\beta}_1^{(\mu)}, \hat{\beta}_2^{(\mu)}$은 각각 $0.276, 0.190$으로, 다른 설명변수가 고정되었을 때 왼쪽 및 오른쪽 주행거리 $10{,}000km$당 각각 평균 타이어 마모가 $1.318, 1.209$배 증가하는 것으로 나타났다. 그리고 평균모형에서의 해당 주행시험 이전의 타이어 누적마모도에 대한 기울기의 추정치 $\hat{\beta}_3^{(\mu)}$

표 3.2.4.1 HGLM2 적합 결과

	Estimate	Std. Error	t value	$Pr(>\|t\|)$	95% CI
$\hat{\beta}_1^{(\mu)}$	0.276	0.081	3.394	0.001***	0.276 ± 0.160 (0.117 ~ 0.436)
$\hat{\beta}_2^{(\mu)}$	0.190	0.046	4.114	0.000***	0.190 ± 0.090 (0.099 ~ 0.280)
$\hat{\beta}_3^{(\mu)}$	0.017	0.030	0.554	0.580	0.017 ± 0.060 (−0.043 ~ 0.076)
$\hat{\beta}_1^{(\lambda)}$	−3.661	1.065	−3.439	0.001**	−3.661 ± 2.087 (−5.748 ~ −1.574)
$\hat{\beta}_2^{(\lambda)}$	−7.380	3.473	−2.125	0.171	−7.380 ± 6.807 (−14.187 ~ −0.573)
$\hat{\beta}_0^{(\phi)}$	−1.383	0.706	−1.958	0.050	−1.383 ± 1.384 (−2.767 ~ 0.001)
$\hat{\beta}_1^{(\phi)}$	−1.240	0.482	−2.570	0.010*	−1.240 ± 0.945 (−2.185 ~ −0.294)
$\hat{\beta}_2^{(\phi)}$	−0.777	0.426	−1.823	0.068	−0.777 ± 0.835 (−1.611 ~ 0.058)
$\hat{\beta}_3^{(\phi)}$	−0.396	0.125	3.160	0.002**	−0.396 ± 0.246 (0.150 ~ 0.641)
cAIC	60.133				

은 0.017으로, 다른 설명변수가 고정되었을 때 이전 누적마모도 $1mm$당 평균 타이어 마모가 1.017배 증가하는 것으로 나타났다. 귀무가설 $H_0 : \beta_j^{(\mu)} = 0, j = 1,2,3$에 대하여 $p-value$는 각각 $0.001, 0.000, 0.580$으로 유의수준 5%에서 해당 주행시험 이전의 타이어 누적마모도(x_3)가 통계적으로 유의하지 않은 것으로 나타났다.

또한 변량효과의 분산에 대한 추정치인 $\beta_j^{(\lambda)}, j = 1,2$는 각각 $-3.661, -7.380$로 평균모형에 추가된 왼쪽과 오른쪽바퀴 주행거리의 차량별 변량효과의 분산은 각각 $0.026mm^2, 0.001mm^2$로 나타났다. 귀무가설 $H_0 : \beta_j^{(\lambda)} = 0, j = 1,2$에 대하여 $p-value$는 각각 $0.001, 0.171$로 유의수준 5%에서 오른쪽바퀴 주행거리의 차량별 변량효과가 통계적으로 유의하지 않은 것으로 나타났다.

산포모형에서의 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리에 대한 기울기의 추정치 $\hat{\beta}_1^{(\phi)}, \hat{\beta}_2^{(\phi)}$는 각각 $-1.240, -0.777$로, 다른 설명변수가 고정되었을 때 왼쪽 및 오른쪽 주행거리 $10,000km$당 타이어 마모의 잔차 분산이 각각 $0.289, 0.460$배 감소하는 것으로 나타났다. 또한, 산포모형에서의 해당 주행시험 이전의 타이어 누적마모도에 대한 기울기의 추정치 $\hat{\beta}_3^{(\phi)}$은 0.396로, 다른 설명변수가 고정되었을 때 이전 누적마모도 $1mm$당 타이어 마모의 잔차 분산이 1.486배 증가하는 것으로 나타났다. 귀무가설 $H_0 : \beta_j^{(\phi)} = 0, j = 1,2,3$에 대하여 $p-value$는 각각 $0.010, 0.068, 0.002$로 유의수준 5%에서 오른쪽바퀴의 해당 주행시험에서의 주행거리(x_2)가 통계적으로 유의하지 않은 것으로 나타났다. 그리고 cAIC는 60.133으로 나타났다.

3.2.5 DHGLM

다섯 번째 모형인 DHGLM은 HGLM에서 산포모형에 변량효과가 추가된 모형이다. 변량효과는 random intercept로써 각 차량별 변량효과를 고려하였으며, 다음과 같은 linear predictor로 이루어진 모형이다.

$$\eta_{ikt}^{(\mu)} = \log\mu_{ikt} = \sum_{j=1}^{3}\beta_j^{(\mu)}x_{jikt} + \sum_{j=1}^{2}v_{ji}^{(\mu)}x_{jikt},$$

$$\eta_{ikt}^{(\phi)} = \log\phi_{ikt} = \beta_0^{(\phi)} + v_i^{(\phi)}.$$

여기서 변량효과 $v_i^{(\mu)} = (v_{1i}^{(\mu)}, v_{2i}^{(\mu)}, v_i^{(\phi)})^T$가 주어졌을 때 반응변수 $Y_{ikt}|v_i^{(\mu,\phi)}$는 감마분포이다. 평균모형에 대한 변량효과의 분포는 $v_{ji}^{(\mu)} \sim N(0,\lambda_j)$,$j=1,2$를 가정하였고, 산포모형에 대한 변량효과의 분포는 $v_i^{(\phi)} \sim N(0,\tau)$를 가정하였다. 평균 및 산포모수에 대한 연결함수 $h^{(\mu)}, h^{(\phi)}$는 모두 로그연결함수를 고려하였다. 평균모형인 μ의 linear predictor에서는 절편이 없고 고정효과와 변량효과가 모두 있는 가법모형을 고려하였다. 또한, 산포모형인 ϕ의 linear predictor에서는 절편과 각 차량별 변량효과가 포함된 모형을 고려하였다.

DHGLM에서의 추정모수는 평균모형 μ에서의 전체 차량에 대한 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리의 기울기 $\beta_1^{(\mu)}, \beta_2^{(\mu)}$와 이전 누적마모도의 기울기 $\beta_3^{(\mu)}$의 3개의 고정효과로 이루어지고, 관측되지 않는 왼쪽과 오른쪽바퀴의 주행거리의 차량별 변량효과 $v_{ji}^{(\mu)}, j=1,2$와 이의

표 3.2.5.1 DHGLM 적합 결과

	Estimate	Std. Error	t value	$\Pr(>\|t\|)$	95% CI
$\hat{\beta}_1^{(\mu)}$	0.283	0.069	4.083	0.000***	0.283 ± 0.136 (0.147 ~ 0.419)
$\hat{\beta}_2^{(\mu)}$	0.192	0.045	4.282	0.000***	0.192 ± 0.088 (0.104 ~ 0.279)
$\hat{\beta}_3^{(\mu)}$	0.032	0.028	1.129	0.259	0.032 ± 0.055 (−0.023 ~ 0.086)
$\hat{\beta}_1^{(\lambda)}$	−4.681	1.368	−3.423	0.001***	−4.681 ± 2.681 (−7.362 ~ −2.001)
$\hat{\beta}_2^{(\lambda)}$	−19.783	1287.371	−0.015	0.988	−19.783 ± 2523.201 (−2542.984 ~ 2503.418)
$\hat{\beta}_0^{(\phi)}$	−2.282	0.326	−7.008	0.000***	−2.282 ± 0.638 (−2.577 ~ −1.460)
$\hat{\tau}$	−0.763	0.580	−1.315	0.188	−0.763 ± 1.136 (−1.899 ~ 0.374)
cAIC	63.235				

변량효과의 분산인 $\lambda_j, j=1,2$가 포함된다. 또한 산포모형 ϕ에서의 절편 $\beta_0^{(\phi)}$와 관측되지 않는 각 차량별 변량효과 $v_i^{(\phi)}$가 포함된다. 적합결과는 cAIC와 함께 표 3.2.5.1에 제시되어 있다.

평균모형에서의 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리에 대한 기울기의 추정치 $\hat{\beta}_1^{(\mu)}, \hat{\beta}_2^{(\mu)}$은 각각 $0.283, 0.192$으로, 다른 설명변수가 고정되었을 때 왼쪽 및 오른쪽 주행거리 $10{,}000km$당 각각 평균 타이어 마모가 $1.327, 1.211$배 증가하는 것으로 나타났다. 그리고 평균모형에서의 해당 주행시험 이전의 타이어 누적마모도에 대한 기울기의 추정치 $\hat{\beta}_3^{(\mu)}$은 0.032으로, 다른 설명변수가 고정되었을 때 이전 누적마모도 $1mm$당 평균 타이어 마모가 1.032배 증가하는 것으로 나타났다. 귀무가설 $H_0 : \beta_j^{(\mu)} = 0, j = 1,2,3$에 대하여 $p-value$는 각각 $0.000, 0.000, 0.259$로 유의수준 5%에서 해당 주행시험 이전의 타이어 누적마모도(x_3)가 통계적으로 유의하지 않은 것으로 나타났다.

또한 변량효과의 분산에 대한 추정치인 $\beta_j^{(\lambda)}, j = 1,2$는 각각 $-4.681, -19.783$으로 평균모형에 추가된 왼쪽과 오른쪽바퀴 주행거리의 차량별 변량효과의 분산은 각각 $0.009mm^2, 2.56 \times 10^{-9} mm^2$로 나타났다. 귀무가설 $H_0 : \beta_j^{(\lambda)} = 0, j = 1,2$에 대하여 $p-value$는 각각 $0.001, 0.988$으로 유의수준 5%에서 오른쪽바퀴 주행거리의 차량별 변량효과가 통계적으로 유의하지 않은 것으로 나타났다.

그리고 산포모형에서의 변량효과의 분산에 대한 추정치인 $\beta^{(\tau)}$는 -0.763으로 산포모형에 추가된 차량별 변량효과의 분산은 $0.466mm^2$로 나타났다. 귀무가설 $H_0 : \beta^{(\tau)} = 0$에 대하여 $p-value$는 0.188으로 유의수준 5%에서 각 차량별 변량효과가 통계적으로 유의하지 않은 것으로 나타났다. 그리고 모형의 cAIC는 63.235로 나타났다.

3.2.6 DHGLM2

여섯 번째 모형인 DHGLM2는 DHGLM에서 산포모형에 고정효과가 추가된 모형이며, 다음과 같은 linear predictor로 이루어진 모형이다.

$$\eta_{ikt}^{(\mu)} = \log\mu_{ikt} = \sum_{j=1}^{3}\beta_j^{(\mu)}x_{jikt} + \sum_{j=1}^{2}v_{ji}^{(\mu)}x_{jikt},$$

$$\eta_{ikt}^{(\phi)} = \log\phi_{ikt} = \beta_0^{(\phi)} + \sum_{j=1}^{3}\beta_j^{(\phi)}x_{jikt} + v_i^{(\phi)}.$$

DHGLM과 마찬가지로 변량효과 $v_i^{(\mu)} = (v_{1i}^{(\mu)}, v_{2i}^{(\mu)}, v_i^{(\phi)})^T$가 주어졌을 때 반응변수 $Y_{ikt}|v_i^{(\mu,\phi)}$는 감마분포이다. 평균모형에 대한 변량효과의 분포는 $v_{ji}^{(\mu)} \sim N(0,\lambda_j), j=1,2$를 가정하였고, 산포모형에 대한 변량효과의 분포는 $v_i^{(\phi)} \sim N(0,\tau)$를 가정하였다. 평균 및 산포모수에 대한 연결함수 $h^{(\mu)}, h^{(\phi)}$는 모두 로그연결함수를 고려하였다. 평균모형인 μ의 linear predictor에서는 절편이 없고 고정효과와 변량효과가 모두 있는 가법모형을 고려하였다. 또한, 산포모형인 ϕ의 linear predictor에서는 절편과 각 차량별 변량효과가 포함된 모형을 고려하였다.

DHGLM2에서의 추정모수는 평균모형 μ에서의 전체 차량에 대한 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리의 기울기 $\beta_1^{(\mu)}, \beta_2^{(\mu)}$와 이전 누적마모도의 기울기 $\beta_3^{(\mu)}$의 3개의 고정효과로 이루어지고, 관측되지 않는 왼쪽과 오른쪽바퀴의 주행거리의 차량별 변량효과 $v_{ji}^{(\mu)}, j=1,2$와 이

표 3.2.6.1 DHGLM2 적합 결과

	Estimate	Std. Error	t value	$Pr(>\|t\|)$	95% CI
$\hat{\beta}_1^{(\mu)}$	0.272	0.071	3.839	0.000***	0.272 ± 0.139 (0.133 ~ 0.412)
$\hat{\beta}_2^{(\mu)}$	0.197	0.036	5.497	0.000***	0.197 ± 0.070 (0.127 ~ 0.268)
$\hat{\beta}_3^{(\mu)}$	0.033	0.025	1.284	0.199	0.033 ± 0.050 (−0.017 ~ 0.082)
$\hat{\beta}_1^{(\lambda)}$	−3.986	1.108	−3.596	0.000***	−3.986 ± 2.172 (−6.158 ~ −1.813)
$\hat{\beta}_2^{(\lambda)}$	−11.596	20.015	−0.579	0.562	−11.596± 39.228 (−50.825 ~ 27.632)
$\hat{\beta}_0^{(\phi)}$	−2.067	0.724	−2.855	0.004***	−2.067 ± 1.419 (−3.486 ~ −0.648)
$\hat{\beta}_1^{(\phi)}$	−0.913	0.417	−2.190	0.029*	−0.913 ± 0.817 (−1.730 ~ −0.096)
$\hat{\beta}_2^{(\phi)}$	−0.562	0.395	−1.423	0.155	−0.562 ± 0.774 (−1.335 ~ −0.212)
$\hat{\beta}_3^{(\phi)}$	0.402	0.112	3.609	0.000***	0.402 ± 0.218 (0.184 ~ 0.621)
$\hat{\tau}$	−0.744	0.566	0.566	0.189	−0.744 ± 1.110 (−1.899 ~ 0.374)
cAIC	53.580				

의 변량효과의 분산인 $\lambda_j, j=1,2$가 포함된다. 또한 산포모형 ϕ에서의 절편 $\beta_0^{(\phi)}$과 전체 차량에 대한 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리의 기울기 $\beta_1^{(\mu)}, \beta_2^{(\mu)}$와 이전 누적마모도의 기울기 $\beta_3^{(\mu)}$의 3개의 고정효과, 관측되지 않는 각 차량별 변량효과 $v_i^{(\phi)}$가 포함된다. 적합결과는 cAIC와 함께 표 3.2.6.1에 제시되어 있다.

평균모형에서의 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리에 대한 기울기의 추정치 $\hat{\beta}_1^{(\mu)}, \hat{\beta}_2^{(\mu)}$은 각각 $0.272, 0.197$으로, 다른 설명변수가 고정되었을 때 왼쪽 및 오른쪽 주행거리 $10{,}000km$당 각각 평균 타이어 마모가 $1.313, 1.218$배 증가하는 것으로 나타났다. 그리고 평균모형에서의 해당 주행시험 이전의 타이어 누적마모도에 대한 기울기의 추정치 $\hat{\beta}_3^{(\mu)}$은 0.033으로, 다른 설명변수가 고정되었을 때 이전 누적마모도 $1mm$당 평균 타이어 마모가 1.033배 증가하는 것으로 나타났다. 귀무가설 $H_0 : \beta_j^{(\mu)} = 0, j=1,2,3$에 대하여 $p-value$는 각각 $0.000, 0.000, 0.199$로 유의수준 5%에서 해당 주행시험 이전의 타이어 누적마모도(x_3)가 통계적으로 유의하지 않은 것으로 나타났다.

또한 변량효과의 분산에 대한 추정치인 $\beta_j^{(\lambda)}, j=1,2$는 각각 $-3.986, -11.596$으로 평균모형에 추가된 왼쪽과 오른쪽바퀴 주행거리의 차량별 변량효과의 분산은 각각 $0.019mm^2, 9.19\times10^{-6}mm^2$로 나타났다. 귀무가설 $H_0 : \beta_j^{(\lambda)} = 0, j=1,2$에 대하여 $p-value$는 각각 $0.000, 0.562$로 유의수준 5%에서 오른쪽바퀴 주행거리의 차량별 변량효과가 통계적으로 유의하지 않은 것으로 나타났다.

산포모형에서의 왼쪽과 오른쪽바퀴의 해당 주행시험에서의 주행거리에 대한 기울기의 추정치 $\hat{\beta}_1^{(\phi)}, \hat{\beta}_2^{(\phi)}$는 각각 $-0.913, -0.562$로, 다른 설명변수가 고정되었을 때 왼쪽 및 오른쪽 주행거리 $10{,}000km$당 타이어 마모의 잔차 분산이 각각 $0.401, 0.570$배 감소하는 것으로 나타났다. 또한, 산포모형에서의 해당 주행시험 이전의 타이어 누적마모도에 대한 기울기의 추정치 $\hat{\beta}_3^{(\phi)}$은 0.402로, 다른 설명변수가 고정되었을 때 이전 누적마모도 $1mm$당 타이어 마모의 잔차 분산이 1.495배 증가하는 것으로 나타났다. 귀무가설 $H_0 : \beta_j^{(\phi)} = 0, j = 1,2,3$에 대하여 $p-value$는 각각 $0.029, 0.155, 0.000$로 유의수준 5%에서 오른쪽바퀴의 해당 주행시험에서의 주행거리(x_2)가 통계적으로 유의하지 않은 것으로 나타났다.

그리고 산포모형에서의 변량효과의 분산에 대한 추정치인 $\beta^{(\tau)}$는 -0.744로 산포모형에 추가된 차량별 변량효과의 분산은 $0.475mm^2$으로 나타났다. 귀무가설 $H_0 : \beta^{(\tau)} = 0$에 대하여 $p-value$는 0.189로 유의수준 5%에서 각 차량별 변량효과가 통계적으로 유의하지 않은 것으로 나타났다. 그리고 모형의 cAIC는 53.580으로 나타났다.

3.3 기울기차이 검정(Slope Difference Test)

평균모형에서 왼쪽바퀴와 오른쪽바퀴의 기울기가 동일한가에 대한 가설검정 $H_0 : \beta_1^{(\mu)} = \beta_2^{(\mu)}$는 다음 식을 통하여 알 수 있다. 6가지 모형 중에서 평균모형이 가장 확장된 모형인 DHGLM 또는 DHGLM2에서의 μ의 linear predictor를 고려해보자.

표 3.3.1 각 모형에서의 기울기차이 검정결과

$\hat{\beta}_2^{(\mu)} - \hat{\beta}_1^{(\mu)}$	Estimate	Std. Error	t value	$\Pr(>\|t\|)$	95% CI
GLM	−0.124	0.067	−1.855	0.064	−0.124 ± 0.131 (−0.255 ~ 0.007)
JGLM	−0.115	0.056	−2.040	0.041*	−0.002 ± 0.110 (−0.225 ~ −0.004)
HGLM	−0.105	0.055	−1.924	0.054	−0.105 ± 0.107 (−0.213 ~ 0.002)
HGLM2	−0.101	0.058	−1.737	0.082	−0.101 ± 0.114 (−0.214 ~ 0.013)
DHGLM	−0.099	0.042	−2.373	0.018*	−0.099 ± 0.082 (−0.181 ~ −0.017)
DHGLM2	−0.092	0.047	−1.956	0.051	−0.092 ± 0.092 (−0.183 ~ 0.000)

$$\log\mu_{ikt} = \beta_1^{(\mu)} x_{1ikt} + \beta_2^{(\mu)} x_{2ikt} + \beta_3^{(\mu)} x_{3ikt} + v_{1i}^{(\mu)} x_{1ikt} + v_{2i}^{(\mu)} x_{1ikt},$$

$$\log\mu_{ikt} = \beta_1^{(\mu)} x_{1ikt} + \beta_2^{(\mu)} x_{2ikt} + \beta_3^{(\mu)} x_{3ikt} + v_{1i}^{(\mu)} x_{1ikt} + v_{2i}^{(\mu)} x_{1ikt} + \{\beta_1^{(\mu)} x_{2ikt} - \beta_1^{(\mu)} x_{2ikt}\},$$

$$= \beta_1^{(\mu)} \{x_{1ikt} + x_{2ikt}\} + (\beta_2^{(\mu)} - \beta_1^{(\mu)}) x_{2ikt} + \beta_3^{(\mu)} x_{3ikt} + v_{1i}^{(\mu)} x_{1ikt} + v_{2i}^{(\mu)} x_{1ikt}.$$

따라서, $x_4 = x_1 + x_2, x_2, x_3$의 3가지 설명변수의 적합을 통하여 왼쪽과 오른쪽 바퀴의 주행거리의 기울기가 동일한가에 대한 검정을 할 수 있다. 변량효과가 포함되지 않은 축소된 모형은 $v_{1i}^{(\mu)} = v_{2i}^{(\mu)} = 0$을 대입함으로써 동일한 결과를 도출할 수 있다. 평균모형에서의 왼쪽바퀴와 오른쪽바퀴

의 기울기가 동일한가에 대한 가설검정결과는 표 3.3.1에 제시되어있다. $\hat{\beta}_2^{(\mu)} - \hat{\beta}_1^{(\mu)}$의 기울기 검정 결과 $p-value$는 각각 $0.064, 0.041, 0.054, 0.082, 0.051, 0.018$로써 유의수준 5%에서 JGLM, DHGLM2의 2가지 모형에서 양쪽 타이어의 기울기는 통계적으로 유의하게 나타났다.

3.4 잔차 도표(Residual Plot)

GLM과 JGLM에 대한 잔차도표는 그림 3.4.1에 제시되어있다. 좌측의 Q-Q 도표(Quantile-Quantile plot)를 통하여 두 모형 모두 대부분의 값들이 정규분포의 quantile을 나타내는 직선 주위에 나타나고 있다. 하지만 산포모형에 고정효과를 포함하지 않는 모형인 GLM은 우측 상단에 이상값이 나타나는 것을 볼 수 있는데, JGLM은 산포모형에 고정효과를 추가함으로써 이러한 이상값을 제어해주는 것을 확인할 수 있다.

HGLM과 HGLM2에 대한 잔차도표는 그림 3.4.2에 제시되어있다. 좌측의 Q-Q 도표(Quantile-Quantile plot)를 통하여 두 모형 모두 대부분의 값들이 정규분포의 quantile을 나타내는 직선 주위에 나타나고 있다. 하지만 산포모형에 고정효과를 포함하지 않는 모형인 HGLM은 우측 상단에 이상값이 나타나는 것을 볼 수 있는데, HGLM2는 산포모형에 고정효과를 추가함으로써 이러한 이상값을 제어해주는 것을 확인할 수 있다.

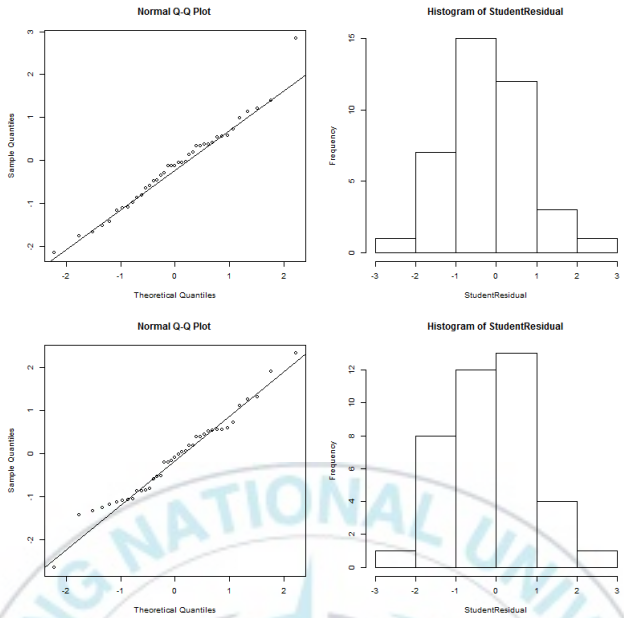

그림 3.4.1 잔차도표(상측2개:GLM, 하측2개:JGLM)

DHGLM과 DHGLM2에 대한 잔차도표는 그림 3.4.3에 제시되어있다. 좌측의 Q-Q 도표(Quantile-Quantile plot)를 통하여 두 모형 모두 대부분의 값들이 정규분포의 quantile을 나타내는 직선 주위에 나타나고 있다. 하지만 DHGLM과 DHGLM2에서 각각 우측상단과 좌측하단에 무시할만한 수준의 이상값이 나타나는 것을 볼 수 있는데, 이는 다른 4가지의 모형과 비교하였을 때 산포모형에서 변량효과가 추가되었기 때문에 나타난 것으로 예측된다.

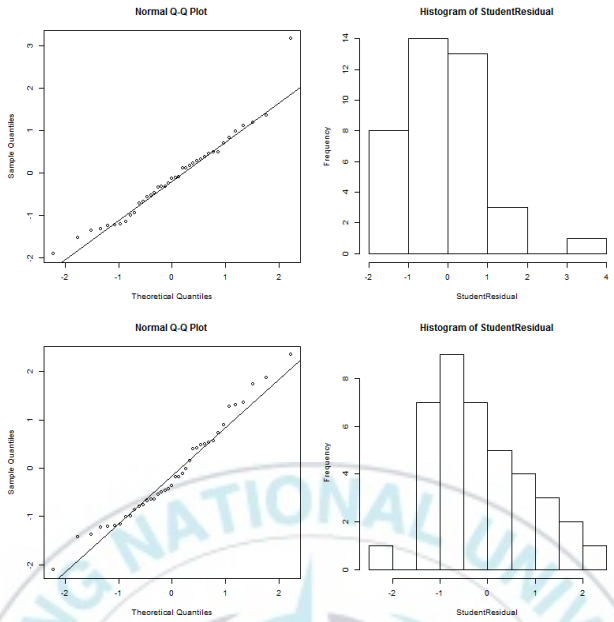

그림 3.4.2 잔차도표(상측2개:HGLM, 하측2개:HGLM2)

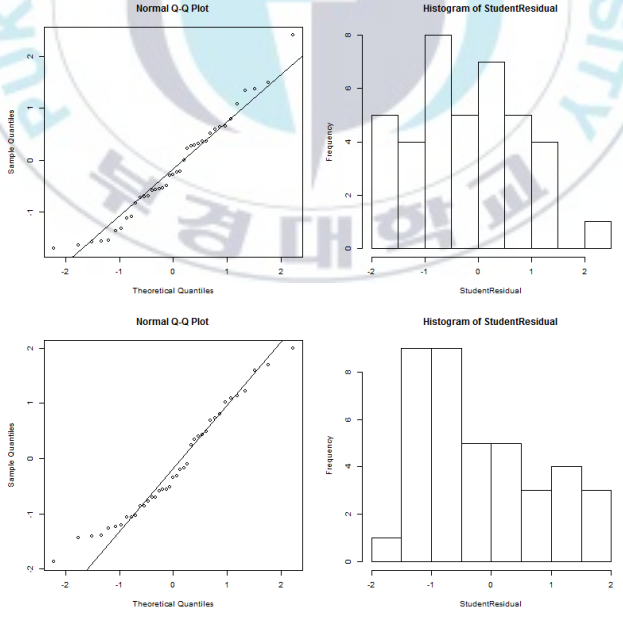

그림 3.4.3 잔차도표(상측2개:DHGLM, 하측2개:DHGLM2)

3.5 요약 정리(Summary)

6가지 모형들의 비교를 위하여 표 3.5.1에서 평균 및 산포모형에서의 기울기에 해당하는 회귀계수를 나타내었다. 평균모형에서는 모든 모형에 $\beta_1^{(\mu)}, \beta_2^{(\mu)}$가 유의하고, $\beta_3^{(\mu)}$는 유의하지 않은 동일한 결과가 나타났다. 산포모형에서는 JGLM에서 $\beta_3^{(\phi)}$, HGLM2와 DHGLM2에서 $\beta_1^{(\phi)}, \beta_3^{(\phi)}$가 유의하게 나타났다. 평균모형에서는 새로운 변수인 이전 누적마모도 x_3를 추가하였지만, 6가지 모형 모두 유의하게 나타나지 않았다. 그러나, 산포모형에서 고정효과를 추가함으로써 JGLM, HGLM2, DHGLM2에서 모두 유의하게 나타났다. 그리고 모형이 더 확장될수록, 또는 산포모형에 고정효과를 추가할 때 cAIC가 낮아지는 것을 확인할 수 있다.

표 3.5.1 모든 모형에 대한 평균 및 산포모형에서의 기울기 회귀계수

Model	$\hat{\beta}_1^{(\mu)}$	$\hat{\beta}_2^{(\mu)}$	$\hat{\beta}_3^{(\mu)}$	$\hat{\beta}_1^{(\phi)}$	$\hat{\beta}_2^{(\phi)}$	$\hat{\beta}_3^{(\phi)}$	cAIC
GLM	0.298***	0.174**	0.034				77.076
HGLM	0.283***	0.173***	−0.016				73.050
DHGLM	0.283***	0.192***	0.032				63.235
JGLM	0.300***	0.185***	0.025	−0.807	−0.744	0.352*	72.400
HGLM2	0.276***	0.190***	−0.017	−1.240*	−0.777	0.396**	60.133
DHGLM2	0.272***	0.197***	0.033	−0.913*	−0.562	0.402***	53.580

제 4장 결론(Conclusion)

HGLM2와 DHGLM2는 산포모형에 고정효과를 추가함으로써 평균모형에서의 $\beta_3^{(\mu)}$와 산포모형에서의 $\beta_2^{(\phi)}$를 제외한 모든 회귀계수가 유의하게 나타났다. 또한 Q-Q 도표를 통하여 살펴볼 수 있었듯이, 산포모형에서 고정효과를 포함하는 모형이 이상값을 제어해주는 것을 확인할 수 있었다. 즉, 산포모형에 고정효과를 포함하지 않는 모형에서 나타나는 이상값을 산포모형에 고정효과를 추가함으로써 이를 제어해주었기 때문이다. 따라서 모형선택(Model Selection)을 수행을 6가지의 모형 중에서 산포모형에서 고정효과를 포함하는 모형을 고려하여 선택하고자 한다. 그리고 산포모형에서 고정효과를 포함하는 모형 JGLM, HGLM2, DHGLM2의 cAIC는 각각 72.400, 60.133, 53.580으로 가장 낮은 cAIC를 가지는 모형은 DHGLM2이며, 다른 모형의 cAIC와 비교하였을 때 어느 정도 유의한 값의 차이를 가진다. 따라서 6가지의 모형 중에서 최종적으로 선택된 모형은 DHGLM2이다.

DHGLM2 모형에서, 해당 주행시험에서의 왼쪽과 오른쪽 주행거리 (x_1,x_2)는 해당 주행시험에서의 평균 타이어 마모도에 영향을 주고, 해당 주행시험에서의 왼쪽 주행거리와 해당 주행시험 이전의 타이어 누적마모도(x_1,x_3)는 해당 주행시험에서의 타이어 마모도의 잔차의 분산에 영향을 주는 것으로 나타났다.

참고문헌

[1] Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 619–678.

[2] Lee, Y. and Nelder, J. A. (2001a). Hierarchical generalized linear models: a synthesis of generalised linear modelss, random–effect models and structured dispersion. *Biometrika*, 88(4):987–1006

[3] Lee, Y. and Nelder, J. A. (2006). Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 55(2):139–185

[4] Lee, Y. and Noh, M. (2012). Modelling random effect variance with double hierarchical generalized linear models. *Statistical Modelling*, 12(6):487–502

[5] Lee, Y. and Noh, M. (2018). dhglm: Double Hierarchical General ized Linear Models. Available at *https://CRAN.R–project.org/pa ckage=dhglm* version 2.0

[6] Lee, Y. G., Lee, K. W., Baek, S. Y., Kam, D. U., Kim, D. W., and Yoo, S. G. (2016). Method and apparatus for measuring tire tread abrasion. US Patent APP. 14/941,332.

[7] Nelder, J. A. and Lee, Y. (1991). Generalized linear models for the analysis of tagushi-type experiments. *Applied stochastic models and data analysis*, 7(1):107-120

[8] Nelder, J. A., Pawitan, Y., and Lee, H. J. (2006). *Generalized linear models with random effects: unified analysis via H-likelihood*. Ghapman and Hall/CRC.

[9] Noh, M., Ok, Y. J., Na, M. H., and Ng, C. T. (2018). Analysis of degradation data using double hierarchical generalized linear model. *The Korean Data & Information Science Society*, 29(1):217-228

[10] Vaida, F. and Blanchard, S. (2005). Conditional akaike information for mixed-effects models., *Biometrika* 92 351-370.

부 록

A.1 R코드

```
# 각 path 지정. org_path만 고치면 됨.
org_path = "K:/MyFile/Thesis/Code(HJ)"
data_path = paste0(org_path, "/data", sep="")
code_path = paste0(org_path, "/code", sep="")
graph_path = paste0(org_path, "/graph", sep="")
#*********************************************************
# 다른파일 함수 로딩
# outputCharGraph.R                    데이터특성 그래프
# outputSlopeCIGraph.R                 기울기의 신뢰구간 그래프
# outputRanefQQPlot.R                  변량효과 q-q plot
# dhglmfit_mod_20150917_Faster.R       DHGLM 패키지 소스
#*********************************************************
setwd(code_path)
source("outputCharGraph.R", encoding="UTF-8")
source("outputSlopeCIGraph.R", encoding="UTF-8")
source("outputOutlierGraph.R", encoding="UTF-8")
source("outputRanefQQPlot.R", encoding="UTF-8")
source("dhglmfit_mod_20150917_Faster.R")
#*********************************************************
```

```
## 1. CI, 1.96SE구하는 function
ci=function(x){
  lower=upper=vector(); result=list()
  for ( i in 1:dim(x)[1] ){
    lower[i]=x[i,1]-x[i,2]*qnorm(0.975)
    upper[i]=x[i,1]+x[i,2]*qnorm(0.975)
  }
  result=list( CI=round(rbind(lower,upper),3),
               "1.96SE"=round(x[,2]*qnorm(0.975),3) )
  return( result )
}

## 2. residual plot 저장하는 function
# 저장할 path 지정.
setwd(paste0(graph_path,"/model_check"))
name=c("GLM","HGLM","DHGLM","JGLM","HGLM2","DHGLM2")

fig=function(x,i){
  png(filename=paste0(i,"_",name,"_plot_check.png")[i],
      width=800, height=800, units="px")
  plotdhglm(x); dev.off()
}
```

```
## 3. p-value 및 ***표시하는 function
pvalue=function(p) {
  result=result.t=vector()
  list.t=list()

  for ( i in 1:dim(p)[1] ) {
    if( 2*pnorm(p[i,3])>1 ) result[i]=2-2*pnorm(p[i,3])
    else result[i]=2*pnorm(p[i,3])

    if ( result[i] < 0.001 ) result.t[i]="***"
    else if ( result[i] < 0.01 ) result.t[i]="**"
    else if ( result[i] < 0.05 ) result.t[i]="*" else result.t[i]="No"
    }
}

  list.t[[1]]=round(result,3)
  list.t[[2]]=result.t

  print(list.t)

  if ( sum(list.t[[2]]=="No") !=0 ) warning('"No" means that coefficients
are not significant')
}
```

PUKYONG NATIONAL UNIVERSITY 부경대학교

```
#*********************************************************
# Modify the Data
#*********************************************************

# Raw Data
setwd(data_path)
data0=read.csv('TireData.csv',header=T); head(data0)

# Make the new variable
dat=data.frame( matrix(nrow=dim(data0)[1],ncol=3) )
colnames(dat)=c('y','x1','x2')

## Model the difference
# z : new variable ( dat ), x : origin variable ( data0 )
for ( i in 1:dim(data0)[1] ){
  if ( data0$t[i]==1 ) dat[i,1:3]=data0[i,4:6]     # when time==1
  else dat[i,1:3]=data0[i,4:6]-data0[(i-1),4:6] # when time!=1
}
x3=vector()
for ( i in 1:dim(data0)[1] ){
  if ( data0[i,]$t==1 ) x3[i]=0 else x3[i]=data0$y[i-1]
}
x3; data.frame(dat$y,x3)

# Combine
# id : 차량번호, k : 1-왼쪽, 2:오른쪽, t : 시행횟수, y : 해당 주행 마모도
# x1: 해당 주행시험 왼쪽 주행거리, x2: 해당 주행시험 오른쪽 주행거리
# x3: 이전 주행 누적 마모도
data=cbind(id=data0$id,k=data0$k,t=data0$t,dat,x3)
head(data); attach(data)
```

```
#*********************************************************
# 기초통계
#*********************************************************

# x1,x2 vs y 산점도
# x1, x2 = 0인 것들 삭제하고 산점도( x1,x2 위치 indicator 제거 )
plot(x1[x1!=0],y[x1!=0],ylab="해당 주행시험에서의 마모도(mm)",
xlab=expression( "해당 주행시험에서의 주행거리("%*%10**4 ~ "km)" ),
xlim=c(0.75,3),ylim=c(0.5,4.25),pch=21,bg=2  )
points(x2[x2!=0],y[x2!=0],pch=21,bg=3)
grid()

# n: x2에서 이상값으로보이는 y값
text(data[15,]$x2,3.9,"id=2, t=3"); text(data[20,]$x1,0.7,"id=3, t=4")
legend(0.75,4.25,legend=c("왼쪽","오른쪽"),pch=21,pt.bg=c(2,3))

# (참고) 상관검정, x2에 이상값있어서 유의하지 않게 나옴.
cor.test(x1[x1!=0],y[x1!=0]); cor.test(x2[x2!=0],y[x2!=0])

# x2 이상값 제거 후 유의하게 나타남.
dd=data[-c(15,20),]
cor.test( dd$x2[dd$x2!=0], dd$y[dd$x2!=0] )

# x3
# 이전누적마모도가 이번주행마모도에 영향을 끼치기는 하는데 작은 선형적관
계를 가짐.
plot(x3[x3!=0],y[x3!=0]); cor.test(x3,y)
plot(data$x3[data$x3!=0],data$y[data$x3!=0],
xlab=expression( "해당 주행시험 이전의 타이어 누적 마모도(mm)" ),
ylab=expression( "해당 주행시험에서의 마모도(mm)" ))
grid()
which(data$x3==6.6375)
cor.test(data$x3[data$x3!=0],data$y[data$x3!=0])
```

```
#*********************************************************
# 1.1 Generalized Linear Model ( family : Gamma )
#*********************************************************

# h-likelihood로 표현하기위해서 모두 dhglmfit으로 적합.
glm.mu = DHGLMMODELING(Model="mean",
                Link="log",LinPred=y~-1+x1+x2+x3)
glm.phi = DHGLMMODELING(Model="dispersion")
glm.res = dhglmfit(RespDist="gamma",DataMain=data,
                MeanModel=glm.mu,DispersionModel=glm.phi)

pvalue( glm.res$beta_coeff ); pvalue( glm.res$phi_coeff )
ci( glm.res$beta_coeff ); ci( glm.res$phi_coeff )

# AIC, 모형체크도표
cat("AIC:", glm.res$caic); fig(glm.res,1)

#*********************************************************
# 1.2 Generalized Linear Model ( family : Gamma )
# H0: b1 = b2, 평균모형에서 왼쪽 오른쪽 기울기 같은가
#*********************************************************
data$x4 = data$x1 + data$x2

glm.mu.d = DHGLMMODELING(Model="mean",Link="log",
                LinPred=y~-1+x4+x2+x3)
glm.phi.d = DHGLMMODELING(Model="dispersion")
glm.res.d = dhglmfit(RespDist="gamma",DataMain=data,
                MeanModel=glm.mu.d, DispersionModel=glm.phi.d)

# H0:b1=b2 채택, 기울기는 같다.
glm.res.d$beta_coeff[2,]; pvalue(glm.res.d$beta_coeff);
ci(glm.res.d$beta_coeff)
```

```
#*********************************************************
# 2.1 Hierarchical Generalized Linear Model ( family : Gamma )
#*********************************************************

hglm.mu = DHGLMMODELING(Model="mean",Link="log",
                LinPred=y~-1+x1+x2+x3+(x1|id)+(x2|id),
                RandDist=c("gaussian","gaussian"))
hglm.phi = DHGLMMODELING(Model="dispersion")
hglm.res = dhglmfit(RespDist="gamma",DataMain=data,
                MeanModel=hglm.mu, DispersionModel=hglm.phi)

pvalue( hglm.res$beta_coeff ); ci( hglm.res$beta_coeff )
pvalue( hglm.res$lambda_coeff ); ci( hglm.res$lambda_coeff )
pvalue( hglm.res$phi_coeff ); ci( hglm.res$phi_coeff )

# AIC, 모형체크도표
cat("AIC:", hglm.res$caic); fig(hglm.res,2)

#*********************************************************
# 2.2 Hierarchical Generalized Linear Model ( family : Gamma )
# H0: b1 = b2, 평균모형에서 왼쪽 오른쪽 기울기 같은가
#*********************************************************

hglm.mu.d = DHGLMMODELING(Model="mean", Link="log",
        LinPred=y~-1+x4+x2+x3+(x1|id)+(x2|id)+(x3|id),
        RandDist=c("gaussian","gaussian","gaussian"))
hglm.phi.d = DHGLMMODELING(Model="dispersion")
hglm.res.d = dhglmfit(RespDist="gamma",DataMain=data,
        MeanModel=hglm.mu.d, DispersionModel=hglm.phi.d)

# H0:b1=b2 채택, 기울기는 같다.
hglm.res.d$beta_coeff[2,]; pvalue(hglm.res.d$beta_coeff)
ci(hglm.res.d$beta_coeff)
```

```
#*********************************************************
# 3.1 DHGLM ( family : Gamma )
#*********************************************************
dhglm.mu = DHGLMMODELING(Model="mean", Link="log",
                LinPred=y~-1+x1+x2+x3+(x1|id)+(x2|id),
                RandDist=c("gaussian","gaussian"))
dhglm.phi = DHGLMMODELING(Model="dispersion",Link="log",
                LinPred=phi~(1|id),RandDist="gaussian")
dhglm.res = dhglmfit(RespDist="gamma",DataMain=data,
                MeanModel=dhglm.mu, DispersionModel=dhglm.phi)
pvalue( dhglm.res$beta_coeff ); ci( dhglm.res$beta_coeff )
pvalue( dhglm.res$lambda_coeff ); ci( dhglm.res$lambda_coeff )
pvalue( dhglm.res$phi_coeff ); ci( dhglm.res$phi_coeff )
pvalue( dhglm.res$tau_coeff ); ci( dhglm.res$tau_coeff )

# AIC, 모형체크도표
cat("AIC:", dhglm.res$caic); fig(dhglm.res,3)

#*********************************************************
# 3.2 DHGLM ( family : Gamma )
# H0: b1 = b2, 평균모형에서 왼쪽 오른쪽 기울기 같은가
#*********************************************************
dhglm.mu.d = DHGLMMODELING(Model="mean", Link="log",
        LinPred=y~-1+x4+x2+x3+(x1|id)+(x2|id)+(x3|id),
        RandDist=c("gaussian","gaussian","gaussian"))
dhglm.phi.d = DHGLMMODELING(Model="dispersion",Link="log",
        LinPred=phi~(1|id),RandDist="gaussian")
dhglm.res.d = dhglmfit(RespDist="gamma",DataMain=data,
        MeanModel=dhglm.mu.d, DispersionModel=dhglm.phi.d)

# H0:b1=b2 기각, 기울기는 다르다.
dhglm.res.d$beta_coeff[2,]; pvalue(dhglm.res.d$beta_coeff);
ci(dhglm.res.d$beta_coeff)
```

```
#####################################
# 새로추가한 x3가 유의하게 나오지 않음: joint modeling을 추가.
# glm  : mu가 증가하고 감소하는지, jglm : phi가 증가하고 감소하는지
#*******************************************************
# 4.1 Joint Generalized Linear Model ( family : Gamma )
#*******************************************************
jglm.mu = DHGLMMODELING(Model="mean", Link="log",
                LinPred=y~-1+x1+x2+x3)
jglm.phi = DHGLMMODELING(Model="dispersion",Link="log",
                LinPred=phi~x1+x2+x3)
jglm.res = dhglmfit(RespDist="gamma",DataMain=data,
                MeanModel=jglm.mu, DispersionModel=jglm.phi)

pvalue( jglm.res$beta_coeff ); ci( jglm.res$beta_coeff )
pvalue( jglm.res$phi_coeff ); ci( jglm.res$phi_coeff )

# AIC, 모형체크도표
cat("AIC:", jglm.res$caic); fig(jglm.res,4)

#*******************************************************
# 4.2 JGLM ( family : Gamma ) : GLM + JGLM
# H0: b1 = b2, 평균모형에서 왼쪽 오른쪽 기울기 같은가
#*******************************************************
jglm.mu.d = DHGLMMODELING(Model="mean", Link="log",
                LinPred=y~-1+x4+x2+x3)
jglm.phi.d = DHGLMMODELING(Model="dispersion",Link="log",
                LinPred=phi~x4+x2+x3)
jglm.res.d = dhglmfit(RespDist="gamma",DataMain=data,
                MeanModel=jglm.mu.d, DispersionModel=jglm.phi.d)

# H0:b1=b2 기각, 기울기는 다르다.
jglm.res.d$beta_coeff[2,];pvalue(jglm.res.d$beta_coeff);
ci(jglm.res.d$beta_coeff)
```

```
#*********************************************************
# 5.1 HGLM2 ( family : Gamma ) : HGLM + JGLM
#*********************************************************
hglm2.mu = DHGLMMODELING(Model="mean", Link="log",
                LinPred=y~-1+x1+x2+x3+(x1|id)+(x2|id),
                RandDist=c("gaussian","gaussian"))
hglm2.phi = DHGLMMODELING(Model="dispersion",Link="log",
                LinPred=phi~x1+x2+x3)
hglm2.res = dhglmfit(RespDist="gamma",DataMain=data,
                MeanModel=hglm2.mu, DispersionModel=hglm2.phi)
pvalue( hglm2.res$beta_coeff ); ci( hglm2.res$beta_coeff )
pvalue( hglm2.res$lambda_coeff ); ci( hglm2.res$lambda_coeff )
pvalue( hglm2.res$phi_coeff ); ci( hglm2.res$phi_coeff )

# AIC, 모형체크도표
cat("AIC:", hglm2.res$caic); fig(hglm2.res,5)

#*********************************************************
# 5.2 HGLM2 ( family : Gamma )
# H0: b1 = b2
# 평균모형에서 왼쪽 오른쪽 기울기 같은가
#*********************************************************
hglm2.mu.d = DHGLMMODELING(Model="mean", Link="log",
        LinPred=y~-1+x4+x2+x3+(x1|id)+(x2|id)+(x3|id),
        RandDist=c("gaussian","gaussian","gaussian"))
hglm2.phi.d = DHGLMMODELING(Model="dispersion",Link="log",
        LinPred=phi~x4+x2+x3)
hglm2.res.d = dhglmfit(RespDist="gamma",DataMain=data,
        MeanModel=hglm2.mu.d, DispersionModel=hglm2.phi.d)

# H0:b1=b2 채택, 기울기는 같다.
hglm2.res.d$beta_coeff[2,]; pvalue(hglm2.res.d$beta_coeff);
ci(hglm2.res.d$beta_coeff)
```

```
#*********************************************************
# 6.1 DHGLM2 ( family : Gamma ) : DHGLM + JGLM
#*********************************************************
dhglm2.mu = DHGLMMODELING(Model="mean", Link="log",
        LinPred=y~-1+x1+x2+x3+(x1|id)+(x2|id),
        RandDist=c("gaussian","gaussian"))
dhglm2.phi = DHGLMMODELING(Model="dispersion",Link="log",
        LinPred=phi~x1+x2+x3+(1|id), RandDist="gaussian")
dhglm2.res = dhglmfit(RespDist="gamma",DataMain=data,
        MeanModel=dhglm2.mu, DispersionModel=dhglm2.phi)

pvalue( dhglm2.res$beta_coeff ); ci( dhglm2.res$beta_coeff )
pvalue( dhglm2.res$lambda_coeff ); ci( dhglm2.res$lambda_coeff )
pvalue( dhglm2.res$phi_coeff ); ci( dhglm2.res$phi_coeff )
pvalue( dhglm2.res$tau_coeff ); ci( dhglm2.res$tau_coeff )

# AIC, 모형체크도표
cat("AIC:", dhglm2.res$caic); fig(dhglm2.res,6)

#*********************************************************
# 6.2 DHGLM2 ( family : Gamma )
# H0: b1 = b2, 평균모형에서 왼쪽 오른쪽 기울기 같은가
#*********************************************************
dhglm2.mu.d = DHGLMMODELING(Model="mean", Link="log",
        LinPred=y~-1+x4+x2+x3+(x1|id)+(x2|id)+(x3|id),
        RandDist=c("gaussian","gaussian","gaussian"))
dhglm2.phi.d = DHGLMMODELING(Model="dispersion",Link="log",
        LinPred=phi~x4+x2+x3+(1|id),RandDist="gaussian")
dhglm2.res.d = dhglmfit(RespDist="gamma",DataMain=data,
        MeanModel=dhglm2.mu.d, DispersionModel=dhglm2.phi.d)
# H0:b1=b2 기각, 기울기는 같다.
dhglm2.res.d$beta_coeff[2,]; pvalue(dhglm2.res.d$beta_coeff);
ci(dhglm2.res.d$beta_coeff)
```