Chapter 8. Generalized linear model (I)

8.1 Data and Statistical Models

(1) Types of data

data : numerical (quantitative) and categorical (qualitative)

- numerical : continuous and discrete
- categorical : nominal and ordinal

(2) Types of categorical data

- response : binary (dichotomous) and polytomous (polychotomous)
- categorical covariate : factor and level

(3) Types of models

- categorical response : logistic and log-linear
- continuous response : multiple linear regression and ANOVA

8.2 Exponential Family

(1) Definition

Let Y be r.v. with pdf

$$f(y;\theta,\,\phi) = \exp\left[\,\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\,\phi)\,\right],$$

then Y belongs to an $exponential\ family$ with $natural$ $(canonical)$ parameter θ if ϕ is known. Also, we assume that $a(\ \cdot\ ), b(\ \cdot\ ), c(\ \cdot,\ \cdot\ )$ are known functions.

(i) $N(\mu,\sigma^2)$ case

$$\begin{aligned} f(y;\theta,\,\phi) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{\frac{1}{\sigma^2}\left(y\mu - \frac{\mu^2}{2}\right) - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + log(2\pi\sigma^2)\right)\right\} \end{aligned}$$

so that $\theta = \mu,\ \phi = \sigma^2$ and

$$a(\phi) = \phi,\ b(\theta) = \frac{\theta^2}{2},\ c(y,\,\phi) = -\frac{1}{2}\left\{\frac{y^2}{\sigma^2} + log(2\pi\sigma^2)\right\}$$

(ii) $P(\lambda)$ case

$$\begin{aligned} f(y;\theta,\,\phi) &= \lambda^y e^{-\lambda} y!, \quad y = 0,\,1,\,2,\,\cdots \\ &= \exp(ylog\ \lambda - \lambda - log\ y!) \end{aligned}$$

so that $\theta = log\lambda,\ \phi = 1$ and

$$a(\phi) = 1,\ b(\theta) = e^{\theta},\ c(y,\,\phi) = -log\ y!$$

(iii) $B(n, \pi)$ case

$$\begin{aligned} f(y;\theta,\ \phi) &= \binom{n}{y}\pi^y(1-\pi)^{n-y}, \quad y = 0,\ 1,\ \cdots,\ n \\ &= \exp\left\{ \frac{\frac{1}{n} y log\left(\frac{\pi}{1-\pi}\right) + log(1-\pi)}{1/n} + log\binom{n}{y} \right\} \end{aligned}$$

so that $\theta = log\left(\frac{\pi}{1-\pi}\right)$, $\phi = 1/n$ and

$$a(\phi) = \phi,\ b(\theta) = log(1+e^{\theta}),\ c(y,\ \phi) = log\binom{1/\phi}{y}$$

(2) Properties

Let

$$l(\theta;\phi,\ y) = logf(y;\theta,\ \phi)$$

be the log-likelihood function, then we have the following theorem, called *Bartlett identity*.

Theorem 8.1

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0,\ \ E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left[\left(\frac{\partial l}{\partial \theta}\right)^2\right] = 0$$

(Proof)

$$E\left(\frac{\partial l}{\partial \theta}\right) = \int \frac{1}{f(y;\theta)}\left(\frac{\partial}{\partial \theta}\right) f(y;\theta)dy = \int \frac{\partial}{\partial \theta} f(y;\theta)dy = \frac{\partial}{\partial \theta}\int f(y;\theta)dy = 0$$

$$\frac{\partial^2 l}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta}f(y;\theta)}{f(y;\theta)}\right) = \frac{\frac{\partial^2}{\partial\theta^2}f(y;\theta)}{f(y;\theta)} - \left\{\frac{\frac{\partial}{\partial\theta}f(y;\theta)}{f(y;\theta)}\right\}^2$$

$$E\left(\frac{\partial^2 l}{\partial\theta^2}\right) = \int \frac{\partial^2}{\partial\theta^2}f(y;\theta)dy - E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right] = \frac{\partial^2}{\partial\theta^2}\int f(y;\theta)dy - E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right] = -E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right]$$

which completes the proof.

We call $U = \frac{\partial l}{\partial\theta}$ $score\ function$ and

$$-E\left(\frac{\partial^2 l}{\partial\theta^2}\right) = E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right]$$

is called $Fisher\ information\ number$. Now, we apply the Bartlett identity in Theorem 8.1 to the exponential family. The log-likelihood function is

$$l(\theta;y) = \frac{\{y\theta - b(\theta)\}}{a(\phi)} + c(y,\ \phi)$$

and we have

$$\frac{\partial l}{\partial\theta} = \frac{\{y - b'(\theta)\}}{a(\phi)}, \quad \frac{\partial^2 l}{\partial\theta^2} = -\frac{b''(\theta)}{a(\phi)}$$

hence, we have

$$E(Y) = \mu = b'(\theta).$$

Also, from $-\dfrac{b''(\theta)}{a(\phi)} + \dfrac{Var(Y)}{a^2(\phi)} = 0$, we have

$$Var(Y) = b''(\theta)a(\phi).$$

We call $b''(\theta)$ *variance function*, and ϕ *dispersion parameter*. Further, we can express θ in terms of $\mu = E(Y)$, and we denote the variance function as $V(\mu)$.

8.3 Construction of GLMs

The *GLM* (*Generalized Linear Models*), suggested by Nelder and Wedderburn (1972), consists of 3 parts;

1. $Y_1, \cdots, Y_n$ are independent and belongs to an exponential family.
2. $\eta = \sum_{j=0}^{p-1} X_j\beta_j$ is called *linear predictor*, where $X_0 \equiv 1$.
3. There exists a function g, called a *link function*, s.t. $g(\mu_i) = \eta_i$, where η_i is the linear predictor and $\mu_i = E(Y_i)$. Also, g is assumed to be monotone and differentiable

The classical multiple linear model can be regarded as a special case of GLM with

$$\mu_i = \sum_{j=0}^{p-1} x_{ij}\,\beta_j = \eta_i$$

with identity link function.

For the binomial response, 3 popular link functions are as follows;

1. logit : $\eta = log\{\mu/(1-\mu)\},\ \ 0 < \mu < 1$
2. probit : $\eta = \Phi^{-1}(\mu),\ \ \Phi(\cdot)$
3. complementary log-log : $\eta = log\{-log(1-\mu)\}$

A link function satisfying $\theta = \eta$ is called *natural* (*canonical*) *link*, and natural link functions for each distribution is as follows;

$$
\begin{array}{cl}
\text{normal} & : \eta = \mu \\
\text{Poisson} & : \eta = log\mu \\
\text{binomial} & : \eta = log\{\mu/(1-\mu)\} \\
\text{gamma} & : \eta = \mu^{-1} \\
\text{inverse Gaussian} & : \eta = \mu^{-2}
\end{array}
$$

Also, properties of exponential families are summarized in the followings;

distribution	Normal	Poisson	Binomial	Gamma	Inverse Gaussian
notation	$N(\mu,\ \sigma^2)$	$P(\mu)$	$B(m,\ \pi)$	$G(\mu,\ \nu)$	$IG(\mu,\ \sigma^2)$
dispersion	σ^2	1	$1/m$	$s\nu^{-1}$	σ^2
$b(\theta)$	$\theta^2/2$	e^θ	$log(1+e^\theta)$	$-log(-\theta)$	$-(-2\theta)^{1/2}$
$\mu(\theta)$	θ	e^θ	$e^\theta/(1+e^\theta)$	$-1/\theta$	$(-2\theta)^{-}1/2$
natural link	identity	log	logit	inverse	$1/\mu^2$
$V(\mu)$	1	μ	$\mu(1-\mu)$	μ^2	μ^3

8.4 Estimation of Regression Coefficients

In the GLMs, estimation of $\boldsymbol{\beta}$ is done by the maximum likelihood method. But, the score function

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \left(\frac{\partial l}{\partial \beta_j}\right)_{j=0,\ \cdots,\ p-1} = \mathbf{0}$$

does not give an explicit solution for $\boldsymbol{\beta}$. Therefore, we use iteration methods such as $Newton - Raphson\ method$ or $Fisher's\ scoring\ method$. In fact, they are based on the Taylor expansion. The 1st Taylor expansion of $f(y)$ about $y = \mu$ is

$$f(y) \ \simeq \ f(\mu) + (y - \mu) f'(\mu)$$

and we apply it to $\partial l \ / \ \partial \beta_j$, then we have

$$\begin{aligned} 0 &= \frac{\partial l}{\partial \beta_j} \\ &\simeq \frac{\partial l}{\partial \beta_j}\Big]_{\beta_j = \hat{\beta}_j^{(0)}} + (\beta_j - \hat{\beta}_j^{(0)}) \cdot \frac{\partial^2 l}{\partial \beta_j^2}\Big]_{\beta_j = \hat{\beta}_j^{(0)}}, \quad j = 0,\ \cdots,\ p-1, \end{aligned}$$

where $\hat{\beta}_j^{(0)}$ is an initial value. In matrix notation for $\beta_j \ \ j = 0,\ 1,\ \cdots,\ p-1$,

$$\begin{aligned} \mathbf{0} &= \frac{\partial l}{\partial \boldsymbol{\beta}} \\ &\simeq \frac{\partial l}{\partial \boldsymbol{\beta}}\Big]_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(0)}} + \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\Big]_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(0)}} \cdot (\ \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(0)}) \end{aligned}$$

and let the solution of $\boldsymbol{\beta}$ in the above equation be $\hat{\boldsymbol{\beta}}^{(1)}$, then

$$\hat{\boldsymbol{\beta}}^{(1)} = \hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{A}^{-1}(\hat{\boldsymbol{\beta}}^{(0)}) \cdot \boldsymbol{U}(\hat{\boldsymbol{\beta}}^{(0)}),$$

where $A^{-1}(\hat{\boldsymbol{\beta}}^{(0)})$ is the inverse matrix of $p \times p$ matrix

$$\boldsymbol{A}(\hat{\boldsymbol{\beta}}^{(0)}) = -\left[\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(0)}}\right]$$

and $\boldsymbol{U}(\hat{\boldsymbol{\beta}}^{(0)})$ is given by

$$\boldsymbol{U}(\hat{\boldsymbol{\beta}}^{(0)}) = \left[\frac{\partial l}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(0)}}\right]$$

Next, we iterate the above process using $\hat{\boldsymbol{\beta}}^{(1)}$ as an initial vector, and we have

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + \boldsymbol{A}^{-1}(\hat{\boldsymbol{\beta}}^{(t)}) \cdot \boldsymbol{U}(\hat{\boldsymbol{\beta}}^{(t)}), \quad t = 0,\ 1,\ \cdots$$

We continue this process until $\hat{\boldsymbol{\beta}}^{(t+1)}$ is close enough to$\hat{\boldsymbol{\beta}}^{(t)}$. If we use

$$\boldsymbol{I} = -E\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right)$$

instead of $\boldsymbol{A}$, then it is called Fisher's scoring method.

Theorem 8.2 We can express

$$\boldsymbol{U} = \boldsymbol{X}'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{\mu})\frac{\partial \eta}{\partial \mu},$$

where $\boldsymbol{W} = \mathrm{diag}(w_{ii}) = \boldsymbol{V}^{-1}\ \left(\frac{\partial \mu}{\partial \eta}\right)^2$ and $\boldsymbol{V} = a(\phi)\mathrm{diag}(\nu_{ii})$. Also,

$$\boldsymbol{I} = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}$$

(Proof) First, we compute $\boldsymbol{U} = (U_0,\ \cdots,\ U_{p-1})'$, where

$$U_j = \frac{\partial l}{\partial \beta_j} = \sum \frac{\partial l_i}{\partial \beta_j} = \sum \frac{\partial l_i}{\partial \theta}\frac{\partial \theta}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}$$

and

$$l = \sum l_i = \sum [y_i\theta - b_i(\theta)a(\phi) + c(y_i,\ \phi)] = \frac{1}{a(\phi)}\sum\{y_i\theta - b(\theta)\} + \sum c(y_i,\ \phi)$$

and hence

$$\begin{aligned}
\frac{\partial l_i}{\partial \theta} &= \frac{1}{a(\phi)}(y_i - b'(\theta)) = \frac{1}{a(\phi)}(y_i - \mu_i) \\
\frac{\partial \theta}{\partial \mu_i} &= \left(\frac{\partial \mu_i}{\partial \theta}\right)^{-1} = \left(\frac{\partial {b_i}'(\theta)}{\partial \theta}\right)^{-1} = \frac{1}{\nu_{ii}} \\
\frac{\partial \eta_i}{\partial \beta_j} &= x_{ij}
\end{aligned}$$

so that

$$U_j = \sum_i \frac{(y_i - \mu_i)}{a(\phi)}\frac{1}{\nu_{ii}}\frac{\partial \mu_i}{\partial \eta_i}x_{ij} = \sum_i (y_i - \mu_i)w_{ii}\frac{\partial \eta_i}{\partial \mu_i}x_{ij},$$

where $w_{ii} = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\ /\ a(\phi)\nu_{ii}$.

On the other hand, we compute $\boldsymbol{I} = (I_{jk})\ ,\ \ j,\ k = 0,\ \cdots\ ,\ p-1$, then

$$\begin{aligned} I_{jk} &= -E\left(\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\right) = -E\left(\frac{\partial U_j}{\partial\beta_k}\right) \\ &= -E\left[\frac{\partial}{\partial\beta_k}\left\{\sum(y_i-\mu_i)w_{ii}\frac{\partial\eta_i}{\partial\mu_i}x_{ij}\right\}\right] = \sum\frac{\partial\mu_i}{\partial\beta_k}w_{ii}\frac{\partial\eta_i}{\partial\mu_i}x_{ij} \\ &= \sum\frac{\partial\eta_i}{\partial\beta_k}w_{ii}x_{ij} = \sum x_{ik}w_{ii}x_{ij} \end{aligned}$$

which completes the proof.

Now, the Fisher's scoring method is

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + \boldsymbol{I}^{-1}(\hat{\boldsymbol{\beta}}^{(t)})\boldsymbol{U}(\hat{\boldsymbol{\beta}}^{(t)})$$

and if we multiply $I(\hat{\boldsymbol{\beta}}^{(t)})$ on both sides, we have

$$\boldsymbol{I}(\hat{\boldsymbol{\beta}}^{(t)})\hat{\boldsymbol{\beta}}^{(t+1)} = \boldsymbol{I}(\hat{\boldsymbol{\beta}}^{(t)})\hat{\boldsymbol{\beta}}^{(t)} + \boldsymbol{U}(\hat{\boldsymbol{\beta}}^{(t)})$$

Now, if we use the result of Theorem 8.2, then the left hand side becomes

$$\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}}^{(t+1)}$$

and the right hand side becomes

$$\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}}^{(t)} + \boldsymbol{X}'\boldsymbol{W}(\boldsymbol{y}-\hat{\boldsymbol{\mu}})\frac{\partial\eta}{\partial\mu} = \boldsymbol{X}'\boldsymbol{W}\left\{\hat{\boldsymbol{\eta}}^{(t)} + (\boldsymbol{y}-\hat{\boldsymbol{\mu}}^{(t)})\frac{\partial\eta}{\partial\mu}\right\} = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{z}^{(t)},$$

where $\hat{\boldsymbol{\eta}}^{(t)} = \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(t)}$, $\hat{\boldsymbol{\mu}}^{(t)} = g^{-1}(\hat{\boldsymbol{\eta}}^{(t)})$ and

$$\boldsymbol{z}^{(t)} = \hat{\boldsymbol{\eta}}^{(t)} + (\boldsymbol{y}-\hat{\boldsymbol{\mu}}^{(t)})\frac{\partial\eta}{\partial\mu}$$

which is called *adjusted dependent variable*. Finally, we have

$$\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}}^{(t+1)} = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{z}^{(t)}, \quad t = 0,\ 1,\ \cdots$$

and it has the same form as the weighted LSE, so that it is called *IRLS* (*Iterative Reweighted Least Squares*).

8.5 Goodness-of-Fit Measures for GLMs

There are two types of goodness-of-fit measures for GLMs : deviance and Pearson's chi-square statistic

(1) Deviance

Let $\hat{\boldsymbol{\beta}}_{max}$ and $\hat{\boldsymbol{\beta}}$ be estimators of $\boldsymbol{\beta}$ under the maximal model and the current model, respectively. Also, let L be the likelihood function. Then, the likelihood ratio test statistic is

$$\lambda = \frac{L(\hat{\boldsymbol{\beta}}_{max}; y)}{L(\hat{\boldsymbol{\beta}}\,; y)}.$$

If the current model is good, then λ will be close to 1, and if not, then it will be very large. Based on this idea, Nelder and Wedderburn (1972 suggested the *deviance*, defined as

$$D = 2log\ \lambda = 2[\ l(\hat{\boldsymbol{\beta}}_{max}; y) - l(\hat{\boldsymbol{\beta}}\,; y)\ ]$$

It can be shown that the deviance D is asymptotically χ^2 distribution with d.f. $n-p$ if the model is good (i.e., under the null hypothesis). Hence, we conclude (reject the null hypothesis) that the current model is not good if $D > \chi^2_\alpha(n-p)$.

Ex.8.5 (p.313) (Normal dist.) Let $Y_1, \cdots, Y_n$ be independent $N(\mu_i, \sigma^2)$, then the log-likelihood becomes

$$l(\boldsymbol{\beta}\,;\, \boldsymbol{y}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2 - \frac{n}{2}log(2\pi\sigma^2)$$

Now, under the maximal model, we have $E(Y_i) = \mu_i,\ i = 1, \cdots, n$, then the MLE of μ_i is $\hat{\mu}_i = y_i$. Therefore, the log-likelihood under the maximal model is

$$l(\hat{\boldsymbol{\beta}}_{max};\, \boldsymbol{y}) = -\frac{n}{2}log(2\pi\sigma^2)$$

On the other hand, if we assume that the current model is $E(Y_i) = \mu$ (i.e., one parameter), then the MLE of μ is $\hat{\mu} = \bar{y}$. Hence, the log-likelihood under the current model becomes

$$l(\hat{\boldsymbol{\beta}}\,;\, \boldsymbol{y}) = -\frac{1}{2\sigma^2}\sum(y_i-\bar{y})^2 - \frac{n}{2}log(2\pi\sigma^2).$$

Therefore, the deviance is

$$D = 2\left[l(\hat{\boldsymbol{\beta}}_{max};\boldsymbol{y}) - l(\hat{\boldsymbol{\beta}}\,;\, \boldsymbol{y})\right] = \frac{1}{\sigma^2}\sum(y_i-\bar{y})^2$$

which follows exactly χ^2 distribution with d.f. $n-1$.

Ex.8.6 (p.314) (Poisson dist.) Let $Y_1, \cdots, Y_n$ be independent $P(\lambda_i)$, then the log-likelihood function is

$$l(\boldsymbol{\beta}\,;\,\boldsymbol{y}) = \sum y_i log\ \lambda_i - \sum \lambda_i - \sum log\ y_i!$$

and under the maximal model $E(Y_i) = \lambda_i$, the MLE is $\hat{\lambda}_i = y_i$, so that

$$l(\hat{\boldsymbol{\beta}}_{max};\boldsymbol{y}) = \sum y_i log y_i - \sum y_i - \sum log y_i!$$

On the other hand, if we assume that the current model is $E(Y_i) = \lambda$ (i.e., one parameter), then the MLE of λ is $\hat{\lambda} = \bar{y}$. Hence, the log-likelihood under the current model becomes

$$l(\hat{\boldsymbol{\beta}}\,;\,\boldsymbol{y}) = \sum y_i log\bar{y} - n\bar{y} - \sum log y_i!$$

Hence, the deviance is

$$D = 2\left[\ \sum y_i log y_i - \sum y_i log\ \bar{y}\ \right] = 2\sum y_i log(y_i\ /\ \bar{y})$$

which follows approximately $\chi^2(n-1)$.

For the normal distribution with p parameters $\beta_0, \beta_1, \cdots, \beta_{p-1}$ in the current model, let the MLE of μ_i, $i = 1, \cdots, n$ be $\hat{\mu}_i$, then the log-likelihood is

$$l(\hat{\boldsymbol{\beta}}\,;\,\boldsymbol{y}) = -\frac{1}{2\sigma^2}\sum(y_i - \hat{\mu}_i)^2 - \frac{n}{2}log(2\pi\sigma^2)$$

and, therefore, the deviance is

$$D = 2[\ l(\hat{\boldsymbol{\beta}}_{max}; \boldsymbol{y}) - l(\hat{\boldsymbol{\beta}}\ ;\ \boldsymbol{y})\ ] = \sum (y_i - \hat{\mu}_i)^2\ /\ \sigma^2$$

which is exactly the same as SSE/σ^2 in the multiple regression model, and follows $\chi^2(n-p)$.

Here are deviance for the distributions in the exponential family.

Normal : $\sum (y_i - \hat{\mu}_i)^2\ /\ \sigma^2$

Poisson : $2\sum\{y_i log(y_i\ /\ \hat{\mu}_i) - (y_i - \hat{\mu}_i)\}$

Binomial : $2\sum\{y_i log(y_i\ /\ \hat{\mu}_i) + (m_i - y_i) log[(m_i - y_i)\ /\ (m_i - \hat{\mu}_i)\}$

Gamma : $2\sum\{-log(y_i\ /\ \hat{\mu}_i) + (y_i - \hat{\mu}_i)\ /\ \hat{\mu}_i\}$

Inverse gaussian : $\sum (y_i - \hat{\mu}_i)^2\ /\ (\hat{\mu}_i^2 y_i)$

(2) Pearson's χ^2 statistic

Another goodness-of-fit measure in the GLMs is the Pearson's χ^2 defined as

$$X^2 = \sum (y_i - \hat{\mu}_i)^2\ /\ V(\hat{\mu}_i)$$

where $V(\hat{\mu}_i)$ is the variance function. It can be shown that X^2 is asymptotically $\chi^2(n-p)$. Under the Gaussian distribution, $D = X^2$ and X^2 follows exactly $\chi^2(n-p)$.

8.6 Testing and Residuals

(1) Testing

Assume that the current model consists of p-parameters $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_{q-1}, \beta_q, \cdots, \beta_{p-1})'$, and consider

$$H_0 \ : \beta_q = \ \cdots = \beta_{p-1} = 0$$

Hence, under H_0, the model consists of q parameters. Now, the test statistics is the difference between (D_1) (deviance under the current model) and (D_0) (deviance under the null model) Let $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ be estimates of $\boldsymbol{\beta}$ under the null model and the current model, respectively. Then,

$$\begin{aligned} \Delta D &= D_0 - D_1 \\ &= 2[\, l(\hat{\boldsymbol{\beta}}_{max}; \boldsymbol{y}) - l(\hat{\boldsymbol{\beta}}_0\, ;\, \boldsymbol{y})\,] - 2[\, l(\hat{\boldsymbol{\beta}}_{max}; \boldsymbol{y}) - l(\hat{\boldsymbol{\beta}}_1\, ;\, \boldsymbol{y})\,] \\ &= 2[\, l(\hat{\boldsymbol{\beta}}_1\, ;\, \boldsymbol{y}) - l(\hat{\boldsymbol{\beta}}_0\, ;\, \boldsymbol{y})\,]. \end{aligned}$$

Since D_1 is asymptotically $\chi^2(n-p)$, and D_0 is asymptotically $\chi^2(n-q)$, it can be shown that ΔD is asymptotically $\chi^2(p-q)$. Therefore, we reject H_0 if $\Delta D > \chi^2_\alpha(p-q)$.

(2) Residuals

(i) Pearson residual

The Pearson residual is defined as

$$r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

and note that

$$X^2 = \sum_{i=1}^{n} r_{Pi}^2$$

(ii) Anscombe residual

Since the Pearson residual do not follow the normal distribution, Anscombe suggested a transformation $A(\ \cdot\ )$, defined as

$$A(\ \cdot\ ) = \int V^{-1/3}(\mu) d\mu$$

For example, in the Poisson distribution, $V(\mu) = \mu$, so that

$$\int \mu^{-1/3} d\mu = \frac{3}{2}\mu^{2/3}$$

and by the delta method, we have $\sqrt{V(A(Y))} \approx A'(\mu)\sqrt{V(\mu)} = \mu^{-1/3}\mu^{1/2} = \mu^{1/6}$. Hence,

$$r_{Ai} = \frac{\frac{3}{2}(y_i^{2/3} - \hat{\mu}_i^{2/3})}{\hat{\mu}_i^{1/6}}$$

(iii) deviance residual

Recall that $D = \sum d_i$, and the deviance residual is defined as

$$r_{Di} = \text{sign}(y_i - \hat{\mu}_i)\sqrt{d_i}$$

For example, in the Poisson case,

$$r_{Di} = \text{sign}(y_i - \hat{\mu}_i)\left\{2(y_i log\left(\frac{y_i}{\hat{\mu}_i}\right) - y_i + \hat{\mu}_i)\right\}^{1/2}$$

8.7 ANOVA models

(1) ANOVA models

analysis of variance model : continuous response and categorical covariates

one covariate case : one-factor experiment (one-way classification)

two covariates case : two-factor experiment (two-way classification)

(2) Estimation

We can write the ANOVA model as a multiple linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, however, $\boldsymbol{X}$ is not a full-rank matrix, so that $\boldsymbol{X}'\boldsymbol{X}$ is singular. Therefore, we cannot have the unique solution for $\boldsymbol{\beta}$ in the normal equation;

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$$

In this case, we have to assign some constraints. Assume that $\boldsymbol{X}$ is $n \times p$ $(p < n)$ matrix. If q column vectors are linearly independent among p column vectors, then we have to assign $p - q$ constraints to $\boldsymbol{\beta}$. There are two methods of assigning constraints; $sum - to - zero\ constraint$ and $corner - point\ constraint$.

First, we consider oneway ANOVA, and data structure for oneway ANOVA with J levels are given in Table 8.5.

Data structure for oneway ANOVA with J levels

level	response				total
A_1	Y_{11}	Y_{12}	$\cdots$	Y_{1n_1}	$Y_{1.}$
A_2	Y_{21}	Y_{22}	$\cdots$	Y_{2n_2}	$Y_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_J	Y_{J1}	Y_{J2}	$\cdots$	Y_{Jn_J}	$Y_{J.}$

We assume that $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0},\ \sigma^2 \boldsymbol{I})$, and let $N = \sum_{j=1}^{J} n_j$ and

$$\boldsymbol{y} = (Y_{11},\ \cdots,\ Y_{n1},\ \cdots,\ Y_{J1},\ \cdots,\ Y_{Jn_J})'$$

Further, we assume that $n_j \equiv K$ for all $j = 1, \cdots, J$, so that $N = JK$.

Also, note that the deviance can be written as

$$D = \frac{1}{\sigma^2} \sum (y_i - \hat{y}_i)^2 = \frac{1}{\sigma^2} (\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y})$$

(i) sum-to-zero constraint

Note that we may write

$$E(Y_{jk}) = \mu + \alpha_j, \quad j = 1, \cdots, J; \;\; k = 1, \;\cdots\;, K,$$

where μ is an overall effect of the response and α_j is the relative effect of jth level. When we write $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, then

$$\underset{JK \times (J+1)}{\boldsymbol{X}} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} \end{bmatrix} \quad , \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_J \end{bmatrix}$$

where $\mathbf{1} = (1, \cdots, 1)'$ and $\mathbf{0} = (0, \cdots, 0)'$ are $K-$vectors. Hence,

$$\underset{(J+1) \times (J+1)}{\boldsymbol{X}'\boldsymbol{X}} = \begin{bmatrix} N & K & \cdots & K & \\ K & K & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \\ K & 0 & 0 & \cdots & K \end{bmatrix} \quad , \quad \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} Y_{\cdot\cdot} \\ Y_{1\cdot} \\ \vdots \\ Y_{J\cdot} \end{bmatrix}$$

Note that the 1st column of $\boldsymbol{X}'\boldsymbol{X}$ is sum of the other J columns, so that the rank of $\boldsymbol{X}'\boldsymbol{X}$ is J, and therefore, we need one constraint on $\boldsymbol{\beta}$. Here, we assign sum-to-zero-constraint, i.e., $\sum_{j=1}^{J} \alpha_j = 0$. Under this constraint,

$$\hat{\mu} = \frac{Y_{\cdot\cdot}}{N} = \bar{y}_{\cdot\cdot}, \quad \hat{\alpha}_j = \frac{Y_{j\cdot}}{K} - \frac{Y_{\cdot\cdot}}{N} = \bar{y}_{j\cdot} - \bar{y}_{\cdot\cdot}$$

and

$$\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = \frac{Y_{\cdot\cdot}^2}{N} + \sum_{j=1}^{J} Y_{j\cdot} \left(\frac{Y_{j\cdot}}{K} - \frac{Y_{\cdot\cdot}}{N} \right) = \frac{1}{K} \sum_{j=1}^{J} Y_{j\cdot}^2$$

In this estimation, we have $\sum_{j=1}^{J} \hat{\alpha}_j = 0$ due to the constraint $\sum_{j=1}^{J} \alpha_j = 0$.

(ii) corner-point constraint

The corner point constraint assumes that any specific relative effect α_j is zero. For example, if we need one constraint, then we may assume $\alpha_1 = 0$. Under this constraint, we have

$$\underset{JK \times J}{\boldsymbol{X}} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} \end{bmatrix} , \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix}$$

and therefore,

$$\underset{J \times J}{\boldsymbol{X}'\boldsymbol{X}} = \begin{bmatrix} N & K & K & \cdots & K \\ K & K & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K & 0 & 0 & \cdots & K \end{bmatrix} , \quad \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} Y_{\cdot\cdot} \\ Y_{2\cdot} \\ \vdots \\ Y_{J\cdot} \end{bmatrix}$$

Now, $\boldsymbol{X}'\boldsymbol{X}$ becomes non-singular and the estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_J \end{bmatrix} = \frac{1}{K} \begin{bmatrix} Y_{1\cdot} \\ Y_{2\cdot} - Y_{1\cdot} \\ \vdots \\ Y_{J\cdot} - Y_{1\cdot} \end{bmatrix}$$

Also, we have

$$\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = \frac{1}{K} \left[Y_{\cdot\cdot}\, Y_{1\cdot} + \sum_{j=2}^{J} Y_{j\cdot}(Y_{j\cdot} - Y_{1\cdot}) \right] = \frac{1}{K} \sum_{j=1}^{J} Y_{j\cdot}^2$$

Note that $D = (\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y})/\sigma^2$, so that the deviance based on both methods (sum-to-zero constraint and the corner point constraint) are the same.

(3) Testing

To test whether there are treatment effects or not, consider

$$H_0 \,:\, \alpha_1 = \ \cdots = \alpha_J = 0$$

then, under H_0, we have $E(Y_{jk}) = \mu$, so that $\boldsymbol{X}'\boldsymbol{X} = N,\ \boldsymbol{X}'\boldsymbol{y} = Y_{\cdot\cdot},\ \hat{\boldsymbol{\beta}} = \hat{\mu} = Y_{\cdot\cdot}\ /\ N$. Hence,

$$D_0 = \frac{1}{\sigma^2}(\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}) = \frac{1}{\sigma^2}\left(\sum_{j=1}^{J}\sum_{k=1}^{K} Y_{jk}^2 - \frac{Y_{\cdot\cdot}^2}{N}\right)$$

which is $\chi^2(N-1)$. On the other hand, under the current model,

$$D_1 = \frac{1}{\sigma^2}(\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}) = \frac{1}{\sigma^2}\left(\sum_{j=1}^{J}\sum_{k=1}^{K} Y_{jk}^2 - \frac{1}{K}\sum_{j=1}^{J} Y_{j\cdot}^2\right)$$

which is $\chi^2(N-J)$. Therefore, the test statistic is

$$\Delta D = D_0 - D_1 = \frac{1}{\sigma^2}\left(\frac{1}{K}\sum_{j=1}^{J} Y_{j\cdot}^2 - \frac{1}{N}Y_{\cdot\cdot}^2\right)$$

which is $\chi^2(J-1)$. Hence, if σ^2 is known, then we reject H_0 if

$$\Delta D > \chi^2_\alpha(J-1)$$

If σ^2 is unknown, we use $F-$ statistic, i.e,

$$F = \frac{D_0 - D_1}{(N-1)-(N-J)} \Bigg/ \frac{D_1}{(N-J)}$$

which follows $F(J-1,\ N-J)$ distribution. Hence, we reject H_0 if $F > F_\alpha(J-1,\ N-J)$.

Ex.8.8 (p.321)

Heights of a plant for 3 types of fertilizers

A	4.17	5.58	5.18	6.11
B	4.81	4.17	4.41	3.59
C	6.31	5.12	5.54	5.50

In this data, $J = 3,\ K = 4,\ N = 12$, and the ANOVA model is

$$E(Y_{jk}) = \mu + \alpha_j, \quad j = 1,\ 2,\ 3\ ;\ k = 1,\ 2,\ 3,\ 4$$

and $\boldsymbol{X}$ and $\boldsymbol{\beta}$ are given by

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad , \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

and therefore,

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 12 & 4 & 4 & 4 \\ 4 & 4 & 0 & 0 \\ 4 & 0 & 4 & 0 \\ 4 & 0 & 0 & 4 \end{bmatrix} \quad , \quad \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} = \begin{bmatrix} 60.49 \\ 21.04 \\ 16.98 \\ 22.47 \end{bmatrix}$$

Since $\boldsymbol{X}'\boldsymbol{X}$ is singular, we assign the sum-to-zero constraint, $\alpha_1 + \alpha_2 + \alpha_3 =$

0, then

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \bar{Y}_{\cdot\cdot} \\ \bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot} \\ \bar{Y}_{2\cdot} - \bar{Y}_{\cdot\cdot} \\ \bar{Y}_{3\cdot} - \bar{Y}_{\cdot\cdot} \end{bmatrix} = \begin{bmatrix} 5.04 \\ \\ 0.22 \\ \\ -0.80 \\ \\ 0.58 \end{bmatrix}$$

Also, we must have $\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 = 0$. Now, $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = 308.95$ gives $D = (\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}) \ / \ \sigma^2 = (312.52 - 308.95) \ / \ \sigma^2 = 3.57 \ / \ \sigma^2$.

Next, under the corner-point constraint $\alpha_1 = 0$, $\boldsymbol{X}$ and $\boldsymbol{\beta}$ are given by

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} , \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

and therefore,

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 12 & 4 & 4 \\ 4 & 4 & 0 \\ 4 & 0 & 4 \end{bmatrix} , \quad \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} Y_{\cdot\cdot} \\ Y_{2\cdot} \\ Y_{3\cdot} \end{bmatrix} = \begin{bmatrix} 60.49 \\ 16.98 \\ 22.47 \end{bmatrix}$$

Hence,

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \bar{Y}_{1\cdot} \\ \bar{Y}_{2\cdot} - \bar{Y}_{1\cdot} \\ \bar{Y}_{3\cdot} - \bar{Y}_{1\cdot} \end{bmatrix} = \begin{bmatrix} 5.26 \\ -1.02 \\ 0.36 \end{bmatrix}$$

Note that $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = 308.95$ which is the same as the value under sum-to-zero constraint.

Now, consider $H_0: \ \alpha_1 = \alpha_2 = \alpha_3 = 0$, then the deviance under H_0 is

$$D_0 = \frac{1}{\sigma^2}\left(\sum_{j=1}^{3}\sum_{k=1}^{4} Y_{jk}^2 - \frac{Y_{\cdot\cdot}^2}{12}\right) = 7.60 \ / \ \sigma^2$$

so that $\Delta D = D_0 - D_1 = 4.03 \ / \ \sigma^2$. Since σ^2 is unknown, we use F-statistic, i.e.,

$$F = \frac{(4.03 \ / \ \sigma^2)}{(11-9)} \Bigg/ \frac{(3.57 \ / \ \sigma^2)}{9} = 5.08$$

which is larger than $F_{.05}(2, \ 9) = 4.26$, we reject H_0. On the other hand, if we replace σ^2 in ΔD by its estimator $3.57/9 = 0.40$, the we have $\Delta D = 10.08$ which is larger than $\chi^2_{.05}(2) = 5.99$. Therefore, we have the same result as $F-$statistic.

ANOVA table of the heights of a plant for 3 fertilizers

s.v.	d.f.	SS	MS	F
treatment	2	4.03	2.02	5.08
error	9	3.57	0.40	
total	11	7.60		

Ex.8.9 (p.324)

Now, we extend to twoway classification defined as

$Y_{jkl} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{jkl}, \ j = 1, \cdots, J; \ k = 1, \cdots, K; \ l = 1, \cdots, L,$

where μ is an overall effect, α_j is the main effect of factor A, β_k is the main effect of factor B, and $(\alpha\beta)_{jk}$ is the *interaction effect* of factors A and B. At each level, L replications are done. The data in Table 8.8 represents the effect of temperature and pressure on the viscosity of tire. Here, factor A is the temperature with $J = 3$, factor B is the pressure with $K =)$, and $L = 2$ replications are done at each level. Therefore, the total number of observations is $N = JKL = 12$.

Table 8.8 Temperature, pressure, and viscosity of tire

temperature pressure	B_1		B_2		total
A_1	6.8	6.6	5.3	6.1	24.8
A_2	7.5	7.4	7.2	6.5	28.6
A_3	7.8	9.1	8.8	9.1	34.8
합	45.2		43.0		88.2

Now, we are interested in 3 questions;

A : Is there the effect of temperature?

B : Is there the effect of pressure?

I : Is there the interaction effect of temperature and pressure?

which can be modelled as follows;

$$\begin{aligned}
A &: Y_{jkl} = \mu + \alpha_j + \varepsilon_{jkl} \\
B &: Y_{jkl} = \mu + \beta_k + \varepsilon_{jkl} \\
I &: Y_{jkl} = \mu + \alpha_j + \beta_k + \varepsilon_{jkl}
\end{aligned}$$

Now, under the full model, we have

$$\boldsymbol{X} = \begin{bmatrix} 110010100000 \\ 110010100000 \\ 110001010000 \\ 110001010000 \\ 101010001000 \\ 101010001000 \\ 101001000100 \\ 101001000100 \\ 100110000010 \\ 100110000010 \\ 100101000001 \\ 100101000001 \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \\ (\alpha\beta)_{31} \\ (\alpha\beta)_{32} \end{bmatrix}, \quad \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} Y_{...} \\ Y_{1..} \\ Y_{2..} \\ Y_{3..} \\ Y_{.1.} \\ Y_{.2.} \\ Y_{11.} \\ Y_{12.} \\ Y_{21.} \\ Y_{22.} \\ Y_{31.} \\ Y_{32.} \end{bmatrix} = \begin{bmatrix} 88.2 \\ 24.8 \\ 28.6 \\ 34.8 \\ 45.2 \\ 43.0 \\ 13.4 \\ 11.4 \\ 14.9 \\ 13.7 \\ 16.9 \\ 17.9 \end{bmatrix}$$

Note that $\boldsymbol{X}$ is 12×12 matrix, however, only 6 column vectors are linearly independent. Therefore, we need $6 = 12 - 6$ constraints. If we use the sum-to-zero constraints, then

$$\begin{array}{ll} \alpha_1 + \alpha_2 + \alpha_3 = 0 & \beta_1 + \beta_2 = 0 \\ (\alpha\beta)_{11} + (\alpha\beta)_{12} = 0 & (\alpha\beta)_{21} + (\alpha\beta)_{22} = 0 \\ (\alpha\beta)_{31} + (\alpha\beta)_{32} = 0 & (\alpha\beta)_{11} + (\alpha\beta)_{21} + (\alpha\beta)_{31} = 0 \end{array}$$

Under this constraints, we have

$$\hat{\boldsymbol{\beta}} = (7.35, \ -1.15, \ -0.2, \ 1.35, \ 0.18, \ -0.18, \ 0.32, \ -0.32, \ 0.12, \ -0.12, \ -0.43, \ 0.43)'$$

and $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = 662.62$.

On the other hand, the corner point constraints can be written as

$$\alpha_1 = \beta_1 = (\alpha\beta)_{11} = (\alpha\beta)_{12} = (\alpha\beta)_{21} = (\alpha\beta)_{31} = 0$$

and under this constraints, we have

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ (\alpha\beta)_{22} \\ (\alpha\beta)_{32} \end{bmatrix}$$

$$
\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} Y_{\dots} \\ Y_{2..} \\ Y_{3..} \\ Y_{12.} \\ Y_{22.} \\ Y_{32.} \end{bmatrix} = \begin{bmatrix} 88.2 \\ 28.6 \\ 34.8 \\ 43.0 \\ 13.7 \\ 17.9 \end{bmatrix}
$$

Therefore, we have $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = (6.7,\ 0.75,\ 1.75,\ -1.0,\ 0.4,\ 1.5)'$ and $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = 662.62$ which is the same as the value obtained under the sum-to-zero constraints.

Finally, we can test 3 current models $A,\ B,\ I$, and these can be done by the following ANOVA table.

Table 8.9 ANOVA table for the viscosity of tire data

s.v.	d.f.	S.S.	MS	F
A(temperature)	2	12.74	6.37	25.82
B(pressure)	1	0.40	0.40	1.63
$A \times B$	2	1.21	0.60	2.45
error	6	1.48	0.25	
total	11	15.83		

We see that the main effect of temperature is significant, however, the main effect of pressure and the interaction effect of temperature and pressure are not significant.