HW4

Jinhong Du - 12243476

2020/01/03

Contents

8.5. Sometimes sample proportions are continuous rather than of the binomial form $\frac{\textbf{number of successes}}{\textbf{number of trials}}$. Each observation is any real number between 0 and 1, such as the proportion of a tooth surface that is covered with plaque. For independent responses $\{y_i\}$, Bartlett (1937) modeled logit$(y_i) \sim \mathcal{N}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2)$. Then $\boldsymbol{y_i}$ itself has a *logit-normal distribution*.

a. Expressing a $\mathcal{N}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2)$ variate as $\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z$, where z is a standard normal variate, show that $y_i = \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}+\sigma z)}{1+\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}+\sigma z)}$ and for small σ,

$$y_i = \frac{e^{\boldsymbol{x}_i^\top \beta}}{1+e^{\boldsymbol{x}_i^\top \beta}} + \frac{e^{\boldsymbol{x}_i^\top \beta}}{1+e^{\boldsymbol{x}_i^\top \beta}}\frac{1}{1+e^{\boldsymbol{x}_i^\top \beta}}\sigma z + \frac{e^{\boldsymbol{x}_i^\top \beta}\left(1-e^{\boldsymbol{x}_i^\top \beta}\right)}{2\left(1+e^{\boldsymbol{x}_i^\top \beta}\right)^3}\sigma^2 z^2 + \cdots .$$

Proof. Since $\text{logit}(y_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z$ where $\boldsymbol{x}_i, \boldsymbol{\beta} \in \mathbb{R}^p$, we have

$$\log\left(\frac{y_i}{1-y_i}\right) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z$$

$$\implies \qquad y_i = \frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+\sigma z}}{1+e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+\sigma z}}.$$

Let $f_i(t) = \frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+t}}{1+e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+t}} = 1 - \frac{1}{1+e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+t}}$. Since

$$f_i'(t) = \frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+t}}{(1+e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+t})^2}$$

$$f_i''(t) = \frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+t}(1-e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+t})}{(1+e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}+t})^3},$$

the Taylor expansion of $f(t)$ near $t=0$ is given by

$$f(z) = \frac{e^{\boldsymbol{x}_i^\top \beta}}{1+e^{\boldsymbol{x}_i^\top \beta}} + \frac{e^{\boldsymbol{x}_i^\top \beta}}{1+e^{\boldsymbol{x}_i^\top \beta}}\frac{1}{1+e^{\boldsymbol{x}_i^\top \beta}}t + \frac{e^{\boldsymbol{x}_i^\top \beta}\left(1-e^{\boldsymbol{x}_i^\top \beta}\right)}{2\left(1+e^{\boldsymbol{x}_i^\top \beta}\right)^3}t^2 + \cdots .$$

Therefore,

$$y_i = f(\sigma z) = \frac{e^{\boldsymbol{x}_i^\top \beta}}{1+e^{\boldsymbol{x}_i^\top \beta}} + \frac{e^{\boldsymbol{x}_i^\top \beta}}{1+e^{\boldsymbol{x}_i^\top \beta}}\frac{1}{1+e^{\boldsymbol{x}_i^\top \beta}}\sigma z + \frac{e^{\boldsymbol{x}_i^\top \beta}\left(1-e^{\boldsymbol{x}_i^\top \beta}\right)}{2\left(1+e^{\boldsymbol{x}_i^\top \beta}\right)^3}\sigma^2 z^2 + \cdots .$$

□

b. Letting $\mu_i = \frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}}$, when σ is close to 0 show that

$$\mathbb{E}(y_i) \approx \mu_i, \qquad \mathbf{Var}(y_i) \approx [\mu_i(1-\mu_i)]^2\sigma^2.$$

Proof. From a. we have

$$\begin{aligned}
\mathbb{E}(y_i) &= \mu_i + \mu_i(1-\mu_i)\sigma\mathbb{E}(z) + \frac{e^{\boldsymbol{x}_i^\top \beta}\left(1-e^{\boldsymbol{x}_i^\top \beta}\right)}{2\left(1+e^{\boldsymbol{x}_i^\top \beta}\right)^3}\sigma^2\mathbb{E}(z^2) + \cdots \\
&= \lim_{\sigma\to 0+}\left[\mu_i + \mu_i(1-\mu_i)\sigma\mathbb{E}(z) + \frac{e^{\boldsymbol{x}_i^\top \beta}\left(1-e^{\boldsymbol{x}_i^\top \beta}\right)}{2\left(1+e^{\boldsymbol{x}_i^\top \beta}\right)^3}\sigma^2\mathbb{E}(z^2) + \cdots\right] \\
&= \mu_i
\end{aligned}$$

as the standard normal variate z has finite momemnts. Thus, $\mathbb{E}(y_i) \approx \mu_i$ when σ is close to 0.

Also,

$$\begin{aligned}\mathrm{Var}(y_i) &= [\mu_i(1-\mu_i)]^2\sigma^2\mathrm{Var}(z) + \frac{e^{\boldsymbol{x}_i^\top\beta}\left(1-e^{\boldsymbol{x}_i^\top\beta}\right)}{2\left(1+e^{\boldsymbol{x}_i^\top\beta}\right)^3}\sigma^2\mathbb{E}(z^2) + \cdots \\ &= [\mu_i(1-\mu_i)]^2\sigma^2\mathrm{Var}(z) + \frac{e^{2\boldsymbol{x}_i^\top\beta}\left(1-e^{\boldsymbol{x}_i^\top\beta}\right)^2}{4\left(1+e^{\boldsymbol{x}_i^\top\beta}\right)^6}\sigma^4\mathbb{E}(z^2) + \cdots \\ &\approx [\mu_i(1-\mu_i)]^2\sigma^2\end{aligned}$$

since higher terms of σ tends to 0. Thus, $\mathrm{Var}(y_i) \approx [\mu_i(1-\mu_i)]^2\sigma^2$ when σ is close to 0. □

c. The approximate moments for the logit-normal motivate a QL approach with $\nu(\mu_i) = \phi[\mu_i(1-\mu_i)]^2$ for unknown ϕ. Explain why this approach provides similar results as fitting an ordinary linear model to the sample logits, assuming constant variance. (The QL approach has the advantage of not requiring adjustment of 0 or 1 observations, for which sample logits do not exist. Papke and Wooldridge (1996) proposed an alternative QL approach using a sandwich covariance adjustment.)

Since in a. we assume that $\mathrm{logit}(y_i) = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z \sim \mathcal{N}(\boldsymbol{x}_i^\top\boldsymbol{\beta}, \sigma^2)$, this is an ordinary linear model to the sample logits, assuming constant variance σ^2. In this setting, the estimated $\boldsymbol{\beta}$ is the MLE. From b. we know that $\mathbb{E}(y_i) \approx \mu_i$ and $\mathrm{Var}(y_i) \approx [\mu_i(1-\mu_i)]^2\sigma^2$. Therefore, if we use quasi-likelihood method with $\nu(\mu_i) = \phi[\mu_i(1-\mu_i)]^2$, which assume the approximate true mean-variance relationship, we will get similar estimates of $\boldsymbol{\beta}$ with the linear model. By assuming the same link function, the estimated $\boldsymbol{\mu}$ are also similar.

d. Wedderburn (1974) used QL to model the proportion of a leaf showing a type of blotch. Envision an approximation of binomial form based on cutting each leaf into a very large number of tiny regions of the same size and observing for each region whether it is covered with blotch. Explain why this suggests using $\nu(\mu_i) = \phi\mu_i(1-\mu_i)$. What violation of the binomial assumptions might make this questionable? (Recall that the parametric family of beta distributions has variance function of this form.)

The binomial approximation would imply that for a single region $\nu(\mu_i) = \phi\mu_i(1-\mu_i)$. This approach is inappropriate when $n_i = 1$ since in that case $\phi = 1$ since $\mathrm{Var}(y_i) = \mu_i(1-\mu_i)$ for Bernoulli random variable y_i. Regardless of n_i, the binomial distribution assumes the small regions are independent, but contiguous regions would likely have dependent results.

8.8. Ordinary linear models assume that $\nu(\mu_i) = \sigma^2$ is constant. Suppose instead that actually $\mathrm{Var}(y_i) = \mu_i$. Using the QL approach for the null model $\mu_i = \beta$, $i = 1, \ldots, n$, show that $u(\beta) = \frac{1}{\sigma^2}\sum_i(y_i - \beta)$, so $\hat{\beta} = \overline{y}$ and $V = \frac{\sigma^2}{n}$. Find the model-based estimate of Var($\hat{\beta}$), the actual variance, and the robust estimate of that variance that adjusts for misspecification of the variance.

For the null model, $\mu_i = \beta$ and $\nu(\mu_i) = \sigma^2$,

$$u(\beta) = \sum_{i=1}^n \frac{\partial\mu_i}{\partial\beta}\frac{y_i-\mu_i}{\nu(\mu_i)} = \sum_{i=1}^n \frac{y_i-\beta}{\sigma^2}.$$

Therefore, $\hat{\beta} = \overline{y}$ and $V = \mathrm{Var}(\hat{\beta}) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(y_i) = \frac{\sigma^2}{n}$ under the null model. A model-based estimate of V is given by $\widehat{\mathrm{Var}}(\hat{\beta}) = \frac{1}{n^2}\sum_{i=1}^n \widehat{\mathrm{Var}}(y_i) = \frac{1}{n^2}\sum_{i=1}^n (y_i - \overline{y})^2$. The actual variance is given by

$$\mathrm{Var}(\hat{\beta}) \approx V\left(\sum_{i=1}^n \frac{\partial\mu_i}{\partial\beta}\frac{\mathrm{Var}(y_i)}{\nu(\mu_i)^2}\frac{\partial\mu_i}{\partial\beta}\right)V = \frac{\sigma^2}{n}\left(\sum_{i=1}^n\frac{\beta}{\sigma^4}\right)\frac{\sigma^2}{n} = \frac{\beta}{n}.$$

The robust estimate of the variance that adjusts for misspecification of the variance is given by

$$\mathrm{Var}(\hat{\beta}) \approx V\left(\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} \frac{\widehat{\mathrm{Var}}(y_i)}{\nu(\mu_i)^2} \frac{\partial \mu_i}{\partial \beta}\right) V = \frac{\sigma^2}{n}\left(\sum_{i=1}^{n} \frac{(y_i - \overline{y})^2}{\sigma^4}\right)\frac{\sigma^2}{n} = \frac{1}{n^2}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

8.9. Suppose we assume $\nu(\mu_i) = \mu_i$ but actually $\mathrm{Var}(y_i) = \sigma^2$. For the null model $\mu_i = \beta$, find the model-based Var(β), the actual Var(β), and the robust estimate of that variance.

For the null model, $\nu(\mu_i) = \mu_i = \beta$,

$$u(\beta) = \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} \frac{y_i - \mu_i}{\nu(\mu_i)} = \sum_{i=1}^{n} \frac{y_i - \beta}{\beta}.$$

Therefore, $\hat{\beta} = \overline{y}$ and $V = \mathrm{Var}(\hat{\beta}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(y_i) = \frac{\beta}{n}$ under the null model. A model-based estimate of V is given by $\widehat{\mathrm{Var}}(\hat{\beta}) = \frac{1}{n^2}\sum_{i=1}^{n}\widehat{\mathrm{Var}}(y_i) = \frac{1}{n^2}\sum_{i=1}^{n}(y_i - \overline{y})^2$. The actual variance is given by

$$\mathrm{Var}(\hat{\beta}) \approx V\left(\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} \frac{\mathrm{Var}(y_i)}{\nu(\mu_i)^2} \frac{\partial \mu_i}{\partial \beta}\right) V = \frac{\beta}{n}\left(\sum_{i=1}^{n} \frac{\sigma^2}{\beta^2}\right)\frac{\beta}{n} = \frac{\sigma^2}{n}.$$

The robust estimate of the variance that adjusts for misspecification of the variance is given by

$$\mathrm{Var}(\hat{\beta}) \approx V\left(\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} \frac{\widehat{\mathrm{Var}}(y_i)}{\nu(\mu_i)^2} \frac{\partial \mu_i}{\partial \beta}\right) V = \frac{\beta}{n}\left(\sum_{i=1}^{n} \frac{(y_i - \overline{y})^2}{\beta^2}\right)\frac{\beta}{n} = \frac{1}{n^2}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

8.12. Let y_{ij} denote the response to a question about belief in life after death (1 = yes, 0 = no) for person j in household i, $j = 1, \ldots, n_i$, $i = 1, \ldots, n$. In modeling $\mathbb{P}(y_{ij} = 1)$ with explanatory variables, describe a scenario in which you would expect binomial overdispersion. Specify your preferred method for dealing with it, presenting your reasoning for that choice.

For grouped binary data, the real $\{y_i\}$ may exhibit more variability than the binomial allows. This can happen in two common ways.

1. Heterogeneity of π_i. Observations at a particular setting of explanatory variables have success probabilities that vary according to values of unobserved variables. For example, in some families, parents are superstitious while the children believe no based on their scientific knowledge. In other families, parents and children may have the same opinions.

2. Positively correlated Bernoulli trials. The Bernoulli trials at each i are positively correlated. Usually, the belief of the children will be affacted by their parents.

I would prefer to use beta-binomial distribution as it is easy to use. We don't need to specify any mean-variance relationship. If the outcomes are not satisfied, then I will try to choose a more suitable ν and use quasi-likelihood methods.

8.13. Use QL methods to construct a model for the horseshoe crab satellite counts, using `weight`, `color`, and `spine` condition as explanatory variables. Compare results with those obtained with zero-inflated GLMs in Section 7.5.

As we can see from the following results, in the quasi-likelihood model, among the three covariates, only $\beta_{\texttt{weight}}$ is significant. However, in the zero-inflated Poisson model, both $\beta_{\texttt{weight}}$ and $\beta_{\texttt{color}}$ are significant in the zero-inflation model, which may account for the zero-inflation phenomenon of y.

```r
Crabs <- read.table("Crabs.dat", header=T)
fit_p <- glm(y ~ weight + color + spine, data=Crabs, family=quasi(link="log",variance="mu"))
summary(fit_p)
```

```
##
## Call:
## glm(formula = y ~ weight + color + spine, family = quasi(link = "log",
##     variance = "mu"), data = Crabs)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0359  -1.8986  -0.5127   0.9335   4.9750
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.009716   0.513775  -0.019    0.985
## weight       0.558179   0.125358   4.453 1.54e-05 ***
## color       -0.190301   0.118156  -1.611    0.109
## spine        0.043999   0.101337   0.434    0.665
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasi family taken to be 3.223873)
##
##     Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 552.18  on 169  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```r
library(pscl)
fit_zip <- zeroinfl(y ~ weight + color + spine | weight + color + spine,
                 dist = "poisson", data = Crabs)
summary(fit_zip)
```

```
##
## Call:
## zeroinfl(formula = y ~ weight + color + spine | weight + color + spine,
##     data = Crabs, dist = "poisson")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.7959 -0.8296 -0.3343  0.7190  4.5094
##
## Count model coefficients (poisson with log link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.11812    0.30165   3.707  0.00021 ***
## weight       0.17131    0.07748   2.211  0.02705 *
## color        0.05319    0.07120   0.747  0.45508
## spine       -0.07735    0.05729  -1.350  0.17695
##
## Zero-inflation model coefficients (binomial with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.5120     1.2255   2.050  0.04038 *
```

```
## weight       -1.6836     0.3962  -4.250 2.14e-05 ***
## color         0.6571     0.2495   2.633  0.00845 **
## spine        -0.3261     0.2468  -1.322  0.18629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 15
## Log-likelihood: -359.1 on 8 Df
```

9.2. How does positive correlation affect the SE for between-cluster effects with binary data? Let $y_{11}, \dots, y_{1d}$ be Bernoulli trials with $\mathbb{E}(y_{1j}) = \pi_1$ and let $y_{21}, \dots, y_{2d}$ be Bernoulli trials with $\mathbb{E}(y_{2j}) = \pi_2$. Suppose Corr$(y_{ij}, y_{ik}) = \rho$ for $i = 1, 2$ and Corr$(y_{1j}, y_{2k}) = 0$ for all j and k. Find the SE of $\hat{\pi}_1 - \hat{\pi}_2$. Show it is larger when $\rho > 0$ than when $\rho = 0$.

Since $y_{ij} \sim \text{Bernoulli}(\pi_i)$ for $i = 1, 2$ and $j = 1, \dots, d$, we have $\text{Var}(y_{ij}) = \pi_i(1 - \pi_i)$. Since $\text{Corr}(y_{ij}, y_{ik}) = \rho$ for $j \neq k$, we have $\text{Cov}(y_{ij}, y_{ik}) = \rho\pi_i(1 - \pi_i)$. Since $\text{Corr}(y_{1j}, y_{2k}) = 0$, we have $\text{Cov}(y_{1j}, y_{2k}) = 0$. Therefore,

$$
\begin{aligned}
\text{Var}(\hat{\pi}_1 - \hat{\pi}_2) &= \text{Var}\left(\frac{1}{d}\sum_{j=1}^{d} y_{1j} - \frac{1}{d}\sum_{k=1}^{d} y_{2k}\right) \\
&= \frac{1}{d^2}\left(\sum_{1 \le j_1, j_2 \le d} \text{Cov}(y_{1j_1}, y_{1j_2}) + \sum_{1 \le k_1, k_2 \le d} \text{Cov}(y_{2k_1}, y_{2k_2})\right) \\
&= \frac{1}{d^2}\sum_{i=1}^{2}\left[d\pi_i(1 - \pi_i) + d(d-1)\rho\pi_i(1 - \pi_i)\right] \\
&= \frac{\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)}{d}\left[1 + (d-1)\rho\right] \\
&> \frac{\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)}{d},
\end{aligned}
$$

i.e. it is larger when $\rho > 0$ than when $\rho = 0$.

9.6. A crossover study comparing $d = 2$ drugs observes a continuous response (y_{i1}, y_{i2}) for each subject for each drug. Let $\mu_1 = \mathbb{E}(y_{i1})$ and $\mu_2 = \mathbb{E}(y_{i2})$ and consider $H_0 : \mu_1 = \mu_2$.

a. Construct the normal linear mixed model that generates a paired-difference t test (with test statistic $t = \frac{\sqrt{n}\bar{d}}{s}$, using mean and standard deviation of the differences $\{d_i = y_{i2} - y_{i1}\}$) and the corresponding confidence interval for $\mu_1 - \mu_2$.

Let $\boldsymbol{z}_{ij} = (1, \delta_{j1})^\top$ where $\delta_{j1} = \mathbb{1}_{\{j=1\}}$ Since $y_{ij} = \boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \boldsymbol{z}_{ij}^\top\boldsymbol{u}_i + \epsilon_{ij}$, we have $d_i = y_{i1} - y_{i2} = (\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2})^\top\boldsymbol{\beta} + (\boldsymbol{z}_{i1} - \boldsymbol{z}_{i2})^\top\boldsymbol{u}_i + (\epsilon_{i1} - \epsilon_{i2})$. Then, d_i follows a normal distribution with $\mathbb{E}(d_i) = (\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2})^\top\boldsymbol{\beta} = \mu_1 - \mu_2$ and $\text{Var}(d_i) = (\boldsymbol{z}_{i1} - \boldsymbol{z}_{i2})^\top\sigma_u^2(\boldsymbol{z}_{i1} - \boldsymbol{z}_{i2}) + 2\sigma_\epsilon^2$. The paired-difference t test statistic is given by $t = \frac{\overline{d}}{\sqrt{\frac{s^2}{n}}}$ where $\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d}_i)^2 \overset{H_0}{\sim} t_{n-1}$. Reject H_0 if $|t| > t_{n-1,1-\frac{\alpha}{2}}$. The corresponding $(1-\alpha)\times 100\%$ confidence interval for $\mu_1 - \mu_2$ is $(\overline{d} - t_{n-1,1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \overline{d} + t_{n-1,1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}})$.

b. Show the effect of the relative sizes of the variances of the random error and random effect on Corr(y_{i1}, y_{i2}). Based on this, to compare two means, explain why it can be more efficient to use a design with dependent samples than with independent samples.

$$\mathrm{Cov}(y_{i1}, y_{i2}) = \mathrm{Cov}(\boldsymbol{z}_{i1}^\top \boldsymbol{u}_i + \epsilon_{i1}, \boldsymbol{z}_{i2}^\top \boldsymbol{u}_i + \epsilon_{i2}) = \boldsymbol{z}_{i1}^\top \sigma_u^2 \boldsymbol{z}_{i2} + \mathbb{1}_{\{j_1=j_2\}}\sigma_\epsilon^2$$
$$\mathrm{Corr}(y_{i1}, y_{i2}) = \frac{\boldsymbol{z}_{i1}^\top \sigma_u^2 \boldsymbol{z}_{i2}}{\sqrt{(\boldsymbol{z}_{i1}^\top \sigma_u^2 \boldsymbol{z}_{i1} + \sigma_\epsilon^2)(\boldsymbol{z}_{i2}^\top \sigma_u^2 \boldsymbol{z}_{i2} + \sigma_\epsilon^2)}}$$

If the variances of the random error are much larger than σ_ϵ^2, then $\mathrm{Corr}(y_{i1}, y_{i2})$ goes to ± 1. If the variances of the random error are much smaller than σ_ϵ^2, then $\mathrm{Corr}(y_{i1}, y_{i2})$ goes to 0. With dependent samples, the variances of the random error will be larger than σ_ϵ^2 in general, then we can ignore the random error and thus it will be more efficient.

9.7. For the normal linear mixed model (9.6), derive expression (9.7) for Var(y_i).

The normal linear mixed model is given by $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{u}_i + \boldsymbol{\epsilon}_i$ where $\boldsymbol{y}_i, \boldsymbol{\beta}, \boldsymbol{u}_i, \boldsymbol{\epsilon}_i \in \mathbb{R}^d$, $\boldsymbol{X}_i, \boldsymbol{Z}_i \in \mathbb{R}^{d\times p}$. Here $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_u)$ and $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 I_d)$ are independent. Then

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{y}_i) &= \mathrm{Var}(\boldsymbol{Z}_i\boldsymbol{u}_i) + \mathrm{Var}(\boldsymbol{\epsilon}_i) \\
&= \boldsymbol{Z}_i \mathrm{Var}(\boldsymbol{u}_i)\boldsymbol{Z}_i^\top + \mathrm{Var}(\boldsymbol{\epsilon}_i) \\
&= \boldsymbol{Z}_i\boldsymbol{\Sigma}_u\boldsymbol{Z}_i^\top + \sigma_\epsilon^2 I_d.
\end{aligned}$$