

$$1. w = A^T y$$

$$\ell(y|w) = -\frac{1}{2} \log |A^T \Sigma A| - \frac{1}{2} w^T (A^T \Sigma A)^{-1} w + \text{const} \quad \text{Let } \Sigma = \sigma^2 \mathbb{I}$$

$$= -\frac{1}{2} \log (\sigma^2)^{\text{rank}(A)} - \frac{1}{2\sigma^2} w^T (A^T \oplus A)^{-1} w \quad \text{with } \sigma^2 = \frac{1}{\eta}$$

$$2. \ell(y|w) = -\text{rank}(A) \log \eta - \eta w^T (A^T \oplus A)^{-1} w$$

$$2 \frac{\partial \ell}{\partial \eta} = \frac{\text{rank}(A)}{\eta} - w^T (A^T \oplus A)^{-1} w = 0 \Rightarrow \sigma^2 = \frac{1}{\eta} = \frac{1}{\text{rank}(A)} w^T (A^T \oplus A)^{-1} w$$

$$(a) \quad \text{rank}(A) = 2 \quad A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -2 \end{pmatrix} \quad \mathbb{I} = I_3 \quad y = \frac{1}{2} n + \varepsilon$$

$$A(A^T A)^{-1} A^T = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{6} \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & -2 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

$$= I - P_{\mathcal{A}}$$

$$\hat{\sigma}^2 = \frac{1}{2} \left(\sum_{i=1}^3 (y_i - \bar{y})^2 \right)$$

$$(b) \quad A = \begin{pmatrix} 1 & 1 \\ -1 & -1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{rank}(A) = 2 \quad \mathbb{I} = \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 1 \end{pmatrix}$$

$$A^T \oplus A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \quad A(A^T \oplus A)^{-1} A^T = \begin{pmatrix} 1 & 1 \\ -1 & -1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix}$$

$$\therefore \hat{\sigma}^2 = \frac{1}{2} \left[2(y_1 - y_2)^2 + 2(y_3 - y_4)^2 \right]$$

$$= (y_1 - y_2)^2 + (y_3 - y_4)^2$$

$$4 \quad (a) \quad \ell(\theta) = \sum_i y_i \log \theta + (1 - y_i) \log (1 - \theta)$$

$$= n \bar{y} \cdot \log \frac{\theta}{1-\theta} + n \log (1-\theta)$$

$$(b) \quad \dot{\ell}(\theta) = \frac{n \bar{y}}{\theta} - \frac{n - n \bar{y}}{1-\theta} = 0$$

$$(c) \quad \bar{y} - \bar{y} \cdot \theta - \theta - \bar{y} \cdot \theta = 0 \Rightarrow \theta = \bar{y}$$

$$(d) \quad \ddot{\ell}(\theta) = -\frac{n \bar{y}}{\theta^2} - \frac{n - n \bar{y}}{(1-\theta)^2} < 0$$

Since $\ddot{\ell}(\theta) < 0$ for all $\theta \in [0, 1]$, $\ell(\theta)$ is maximized

when $\dot{\ell}(\theta) = 0$

$$(e) \quad -\mathbb{E}[\ddot{\ell}(\theta)] = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)} = I(\theta)$$

$$(f) \quad I^{-1}(\theta) = \frac{\theta(1-\theta)}{1}$$

$$(g) \quad \text{Var}(\bar{y}) = \frac{1}{n} \text{Var}(y_1) = \frac{1}{n} \theta(1-\theta) = I^{-1}(\theta)$$

$$(h) \quad \hat{\text{Var}}(\bar{y}) = \frac{\hat{\theta}(1-\hat{\theta})}{n} = \frac{\bar{y}(1-\bar{y})}{n}$$

hw10

Yonghyun Kwon

4/15/2020

Problem 2

(a)

```
dat <- read.csv("https://dnett.github.io/S510/hw10GenotypeYield.txt", sep = "\t")
dat$genotype <- factor(dat$genotype)
```

#(a)

```
(model1 <- lm(yield ~ 0 + genotype, data = dat))
```

##

Call:

lm(formula = yield ~ 0 + genotype, data = dat)

##

Coefficients:

## genotype1	genotype2	genotype3	genotype4	genotype5	genotype6
## 194.9	184.2	191.4	198.6	194.2	197.7
## genotype7	genotype8	genotype9	genotype10	genotype11	genotype12
## 193.6	196.6	202.3	196.1	182.1	186.1
## genotype13	genotype14	genotype15	genotype16	genotype17	genotype18
## 184.0	170.2	189.9	187.4	192.6	189.3
## genotype19	genotype20	genotype21	genotype22	genotype23	genotype24
## 185.7	197.8	199.3	192.8	183.1	200.8
## genotype25	genotype26	genotype27	genotype28	genotype29	genotype30
## 190.6	182.7	192.8	187.2	195.5	194.0
## genotype31	genotype32	genotype33	genotype34	genotype35	genotype36
## 178.7	203.7	185.5	189.6	196.0	187.1
## genotype37	genotype38	genotype39	genotype40	genotype41	genotype42
## 188.2	190.3	185.4	191.5	186.2	187.2
## genotype43	genotype44	genotype45	genotype46	genotype47	genotype48
## 189.7	179.6	189.1	190.5	185.7	206.2
## genotype49	genotype50	genotype51	genotype52	genotype53	genotype54
## 192.2	194.3	197.7	184.5	193.2	182.2
## genotype55	genotype56	genotype57	genotype58	genotype59	genotype60
## 192.1	188.9	185.4	193.0	198.3	192.4
## genotype61	genotype62	genotype63	genotype64	genotype65	genotype66
## 189.2	181.5	192.2	189.3	196.1	201.5
## genotype67	genotype68	genotype69	genotype70	genotype71	genotype72
## 194.0	194.4	181.1	201.3	185.4	183.7
## genotype73	genotype74	genotype75	genotype76	genotype77	genotype78
## 194.6	196.1	196.8	179.6	191.5	196.1

```
## genotype79 genotype80 genotype81 genotype82 genotype83 genotype84
## 190.1 186.5 194.1 189.4 184.5 191.5
## genotype85 genotype86 genotype87 genotype88 genotype89 genotype90
## 177.4 193.0 194.7 196.2 197.4 190.5
## genotype91 genotype92 genotype93 genotype94 genotype95 genotype96
## 189.9 193.9 177.6 183.8 189.3 178.2
## genotype97 genotype98 genotype99 genotype100
## 200.3 188.8 192.9 191.8
```

```
BLUE <- model1$coefficients
```

(b)

```
(model2 <- lmer(yield ~ 0 + 1|genotype, data = dat))
```

```
## Linear mixed model fit by REML ['lmerModLmerTest']
## Formula: yield ~ 0 + 1 | genotype
## Data: dat
## REML criterion at convergence: 2261.146
## Random effects:
## Groups Name Std.Dev.
## genotype (Intercept) 2.687
## Residual 9.669
## Number of obs: 304, groups: genotype, 100
## Fixed Effects:
## (Intercept)
## 190.7
```

```
#sigma_g^2 = 2.687^2 = 7.22
#sigma_e^2 = 9.669^2 = 93.49
```

$\sigma_g^2 = 2.687^2 = 7.22$ while $\sigma_e^2 = 9.669^2 = 93.49$

(c)

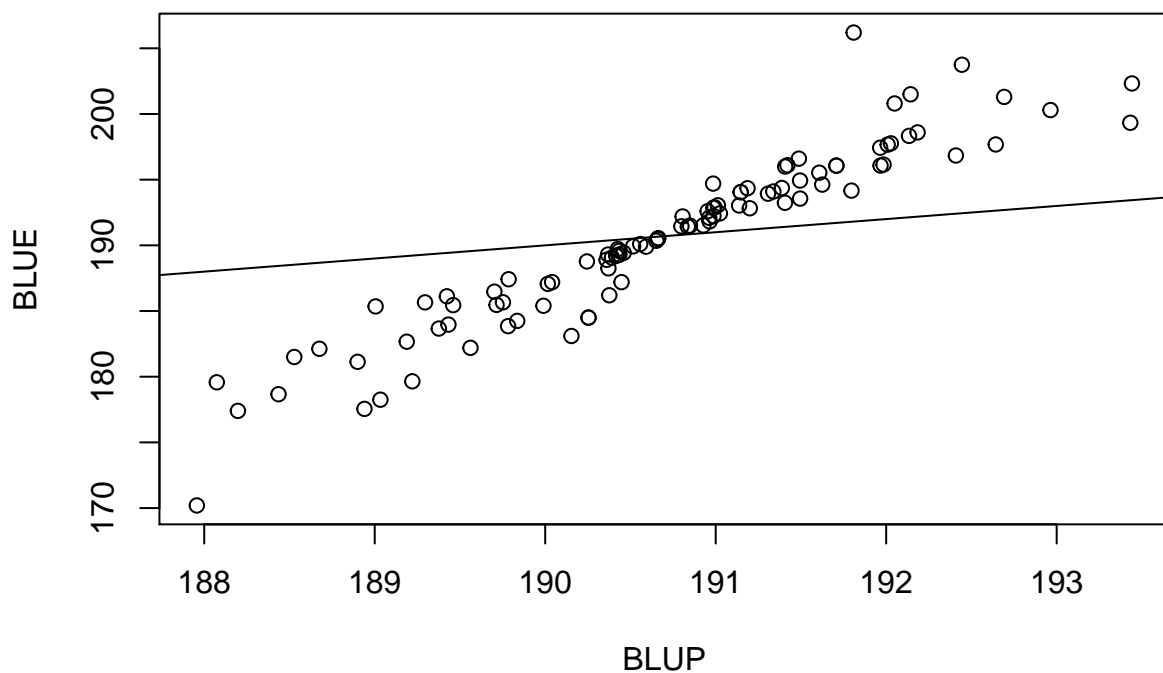
```
BLUP <- unlist(ranef(model2)) + fixef(model2)
names(BLUP) <- names(BLUE)
BLUP
```

```
## genotype1 genotype2 genotype3 genotype4 genotype5 genotype6
## 191.4947 189.8359 190.8365 192.1842 191.7963 192.6427
## genotype7 genotype8 genotype9 genotype10 genotype11 genotype12
## 191.4953 191.4877 193.4415 191.4208 188.6755 189.4232
## genotype13 genotype14 genotype15 genotype16 genotype17 genotype18
## 189.4324 187.9567 190.5159 189.7853 190.9527 190.4291
## genotype19 genotype20 genotype21 genotype22 genotype23 genotype24
## 189.2951 192.0275 193.4319 190.9861 190.1538 192.0494
## genotype25 genotype26 genotype27 genotype28 genotype29 genotype30
## 190.6633 189.1880 191.2001 190.0405 191.6075 191.1466
## genotype31 genotype32 genotype33 genotype34 genotype35 genotype36
## 188.4358 192.4440 189.7145 190.4333 191.4074 190.0154
## genotype37 genotype38 genotype39 genotype40 genotype41 genotype42
## 190.3709 190.6517 189.9897 190.7989 190.3759 190.4476
```

##	genotype43	genotype44	genotype45	genotype46	genotype47	genotype48
##	190.4259	188.0738	190.3915	190.6610	189.7521	191.8093
##	genotype49	genotype50	genotype51	genotype52	genotype53	genotype54
##	190.8059	191.1867	192.0087	190.2541	191.4062	189.5617
##	genotype55	genotype56	genotype57	genotype58	genotype59	genotype60
##	190.9619	190.3601	189.0052	191.0129	192.1341	191.0246
##	genotype61	genotype62	genotype63	genotype64	genotype65	genotype66
##	190.4166	188.5280	190.9870	190.3684	191.7078	192.1430
##	genotype67	genotype68	genotype69	genotype70	genotype71	genotype72
##	191.1466	191.3881	188.8996	192.6920	189.4600	189.3760
##	genotype73	genotype74	genotype75	genotype76	genotype77	genotype78
##	191.6248	191.9669	192.4088	189.2206	190.9272	191.7078
##	genotype79	genotype80	genotype81	genotype82	genotype83	genotype84
##	190.5571	189.7019	191.3380	190.4604	190.2541	190.8491
##	genotype85	genotype86	genotype87	genotype88	genotype89	genotype90
##	188.1976	191.1374	190.9851	191.9846	191.9648	190.6610
##	genotype91	genotype92	genotype93	genotype94	genotype95	genotype96
##	190.5915	191.3067	188.9397	189.7824	190.4354	189.0334
##	genotype97	genotype98	genotype99	genotype100		
##	192.9638	190.2445	190.9928	190.9641		

(d)

```
plot(BLUE ~ BLUP)
abline(0, 1)
```



We can see that the scatters has higher slope than 1. This is because BLUEs have higher variability than BLUPs.

(e)

```
sort(BLUE, decreasing = TRUE)[1:5]

## genotype48 genotype32 genotype9 genotype66 genotype70
##      206.200      203.750      202.325      201.500      201.300
```

(f)

```
sort(BLUP, decreasing = TRUE)[1:5]

## genotype9 genotype21 genotype97 genotype70 genotype6
##      193.4415      193.4319      192.9638      192.6920      192.6427
```

Problem 3

```
y = c(14, 9, 10, 5, 18, 9, 9, 17, 10, 17, 18, 13, 17, 16)
G = rep(c(1,2), each = 7)
G = as.factor(G)

model1 = glm(y ~ 0 + G, family = "poisson")
s1 = summary(model1)

model2 = glm(y ~ 1, family = "poisson")
s2 = summary(model2)
```

##(a)

```
AIC(model1)
```

```
## [1] 78.21563
```

##(b)

```
BIC(model1)
```

```
## [1] 79.49374
```

##(c)

```
AIC(model2)
```

```
## [1] 82.60474
```

##(d)

```
BIC(model2)
```

```
## [1] 83.2438
```

##(e) According to AIC, the first model is preferred.

##(f) According to BIC, the first model is preferred.

```
##(g)
- 2 * logLik(model2) + 2 * logLik(model1)

## 'log Lik.' 6.389118 (df=1)

##(h)
aov <- anova(model1, model2, test = "LRT")
aov$Pr(>Chi)[2]

## [1] 0.0114822

##(i)
V <- s1$cov.unscaled
thetahat <- s1$coefficients[,1]
(thetahat[2] - thetahat[1]) / sqrt(V[1,1] + V[2,2])

##          G2
## 2.505302

##(j)
aov <- anova(model1, model2, test = "Rao")
aov$Pr(>Chi)[2]

## [1] 0.01172719

##(g) While the BLUPs are the weighted average of all the sample means and so BLUPs have smaller
variabilities in this case, the BLUE's are simply the average of i-th genotype.
```

Problem 5

(a)

```
phat = 0.17
n = 100
margin = 1.96 * sqrt(phat * (1-phat) / n)
phat - margin; phat + margin

## [1] 0.09637597
## [1] 0.243624
```

(b)

Find the solution of non-linear equation using bisection method in “NLRRoot” package.

```
ftn1 <- function(x){
  17 * log(x) + 83 * log(1 - x)
}

lphat = ftn1(phat)
lik2 = lphat - 1/2*qchisq(1- 0.05, 1)

ftn2 <- function(x){
```

```
17 * log(x) + 83 * log(1 - x) - lik2
}
```

```
BFfzero(ftn2, 0, phat)
```

```
## [1] 1
## [1] 0.1051833
## [1] 0.0001715579
## [1] "finding root is successful"
```

```
#BFfzero(ftn2, phat, 1) #0.2515018
```

We can see that the confidence interval is (0.1051833, 0.251507)

Problem 6

We can construct a confidence interval using the following asymptotic distribution

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\text{Var}(\hat{p}_1 - \hat{p}_2)}} \rightarrow N(0, 1)$$

```
p1 = 172 / 350
p2 = 137 / 350
n = 350
margin = 1.96 * sqrt(p1*(1-p1) / n + p2*(1-p2) / n)
p1 - p2 - margin
```

```
## [1] 0.02680293
```

```
p1 - p2 + margin
```

```
## [1] 0.1731971
```

So the 95 % confidence interval for $p_1 - p_2$ is(0.02680293, 0.1731971)