Giang Vu
STAT 2132
April 8, 2021

Homework 8

1) a) $Y = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix}_{n\times 1}$ (n_1 rows ... n_r rows)

$X = \left[x_{11}^T \cdots x_{1n_1}^T \cdots x_{r1}^T \cdots x_{rn_r}^T\right]^T$

$\delta_{r\times 1} = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_r \end{bmatrix}$; $\delta \sim N(0, \sigma_\delta^2 I_r)$
$\epsilon \sim N(0, \sigma^2 I_n)$

$\underset{n\times 1}{\epsilon} = \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \vdots \\ \epsilon_{r1} \\ \vdots \\ \epsilon_{rn_r} \end{bmatrix}$; $\beta \in \mathbb{R}^p$

$$Y_{ij} = x_{ij}^T \beta + \delta_i + \epsilon_{ij}$$

$$\Rightarrow Y = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix} = \begin{bmatrix} x_{11}^T \\ \vdots \\ x_{1n_1}^T \\ x_{r1}^T \\ \vdots \\ x_{rn_r}^T \end{bmatrix}_{n\times p} \beta_{p\times 1} + Z \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_r \end{bmatrix}_{r\times 1} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \vdots \\ \epsilon_{r1} \\ \vdots \\ \epsilon_{rn_r} \end{bmatrix}$$

There has to be an $n\times r$ matrix here to make the dimensions match

To keep $Y_{ij} = x_{ij}^T \beta + \delta_i + \epsilon_{ij}$ satisfied, Z has to be

$$\underset{n\times r}{Z} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_{r \text{ columns}} \begin{array}{l} \} \, n_1 \text{ rows} \\ \cdots \\ \} \, n_r \text{ rows} \end{array}$$

Therefore $Y = X\beta + Z\delta + \epsilon$ with Z defined as above

b) $\tilde{y} = Q^T y = Q^T(X\beta + Z\delta + \epsilon)$

$= Q^T X\beta + Q^T Z\delta + Q^T\epsilon \quad = Q^T X\beta + (Q^T Z)\delta + \tilde{\epsilon}$

$Q^T \in \mathbb{R}^{n\times(n-p)}$; $Q^T\epsilon = \tilde{\epsilon} \sim N(0, \sigma^2 I_{n-p})$

We need a matrix Q such that $Q^T X = 0$ to make $\tilde{y} = (Q^T Z)\delta + \tilde{\epsilon}$

Q has $(n-p)$ orthonormal columns , X has full column ranks so p columns vecters of X are linearly independent

$\Rightarrow$ We need Q to have columns that are the orthonormal basis of the orthogonal complement of the column space of X

(essentially we need the columns of Q to be orthogonal to the columns of X)
(or rows of Q^T)

In a subspace spanned by p linearly independent vectors ($C(X)$) in a n-dimension space ($\mathbb{R}^n$), we can always find at most $(n-p)$ vectors orthogonal to that subspace $C(X)$.

$\Rightarrow$ The orthonormal basis for $C(X)^\perp$ is not unique.

$\Rightarrow$ We can find different Q's that satisfy $\tilde{y} = (Q^T Z)\delta + \tilde{\epsilon}$;

Q is not unique.

c) (i) $\left.\begin{array}{l} \delta \sim N(0, \sigma_\delta^2 I_r) \\ \epsilon \sim N(0, \sigma^2 I_n) \end{array}\right\} \Rightarrow \tilde{y} = (Q^T Z)\delta + \tilde{\epsilon}$

$\sim N(0, Q^T Z \sigma_\delta^2 I_r Z^T Q + \sigma^2 I_{n-p})$

$\sim N(0, \tilde{Z}\sigma_\delta^2 I_r \tilde{Z}^T + \sigma^2 I_{n-p})$

$\tilde{H}$ is orthogonal projection for $\tilde{Z}$ $\Rightarrow$ $\tilde{H}$ is idempotent and symmetric

$\Rightarrow \quad (I_{n-p} - \tilde{H})\tilde{H} = \tilde{H} - \tilde{H}\tilde{H} = \tilde{H} - \tilde{H} = 0$

$\Rightarrow$ $(I_{n-p} - \tilde{H})$ and $\tilde{H}$ are orthogonal

$\Rightarrow$ The two quadratic forms

$SSE = \tilde{y}^T(I_{n-p} - \tilde{H})\tilde{y}$

$SSTR = \tilde{y}^T\tilde{H}\tilde{y}$

are independent when $\tilde{y} \sim N$ and $(I_{n-p} - \tilde{H})$ and $\tilde{H}$ are orthogonal.

This doesn't depend on the choice of Q. As long as Q has requirements specified in part (b), we will always have $\tilde{y} \sim N$ and $(I_{n-p} - \tilde{H})$ and $\tilde{H}$ being orthogonal

1 c) (ii)

$$\frac{\tilde{Y}^T \tilde{H} \tilde{Y}}{\frac{1}{d} tr(\tilde{Z}^T \tilde{Z}) \sigma_g^2 + \sigma^2} \sim \chi^2_d$$

$$\left(rank \left(\frac{\tilde{H}}{\frac{1}{d} tr(\tilde{Z}^T \tilde{Z}) \sigma_g^2 + \sigma^2} \right) = rank(\tilde{H}) = d \right)$$

$$\Rightarrow \frac{MSTR\ (d)}{\frac{1}{d} tr(\tilde{Z}^T \tilde{Z}) \sigma_g^2 + \sigma^2} \sim \chi^2_d$$

$$\Rightarrow E\left(\frac{MSTR(d)}{\frac{1}{d} tr(\tilde{Z}^T \tilde{Z}) \sigma_g^2 + \sigma^2} \right) = d$$

$$\Rightarrow d \frac{E(MSTR)}{\frac{1}{d} tr(\tilde{Z}^T \tilde{Z}) \sigma_g^2 + \sigma^2} = d$$

$$\Rightarrow E(MSTR) = \frac{1}{d} tr(\tilde{Z}^T \tilde{Z}) \sigma_g^2 + \sigma^2$$

$\Rightarrow$ MSTR is unbiased est. for $d^{-1} tr(\tilde{Z}^T \tilde{Z}) \sigma_g^2 + \sigma^2$

Similarly $\frac{\tilde{Y}^T (I_{n-p} - \tilde{H}) \tilde{Y}}{\sigma^2} \sim \chi^2_{n-p-d}$ $\left(rank \left(\frac{1}{\sigma^2} (I_{n-p} - \tilde{H}) \right) = n-p-d \right)$

$$\Rightarrow \frac{MSE(n-p-d)}{\sigma^2} \sim \chi^2_{n-p-d}$$

$$\Rightarrow E(MSE) \cdot \frac{n-p-d}{\sigma^2} = n-p-d$$

$$\Rightarrow E(MSE) = \sigma^2$$

$\Rightarrow$ MSE is unbiased est. for σ^2

1 c) (iii)

when $\sigma_g^2 = 0$: $\begin{cases} MSTR \sim \frac{\sigma^2}{d} \chi^2_d \\ MSE \sim \frac{\sigma^2}{n-p-d} \chi^2_{n-p-d} \end{cases}$ $\Rightarrow \frac{MSTR}{MSE} \sim \frac{\chi^2_d / d}{\chi^2_{n-p-d} / (n-p-d)}$

$$\Rightarrow \frac{MSTR}{MSE} \sim F_{d,\, n-p-d}$$

2a) If we stack all the Y's into a vector Y

$$Var(Y) = \sigma_p^2 B + \sigma^2 I_n$$

where B is a partition matrix to account for the fact that $Y_1, \dots, Y_k$ do not necessarily have same variance

Then $Corr(Y_r, Y_s) \geq \frac{-1}{\lambda_{max} - 1}$

where λ_{max} is the largest eigenvalue of B, or number of samples per individual, which in our case is k

$$\Rightarrow Corr(Y_r, Y_s) = \rho \geq \frac{-1}{\lambda_{max} - 1} = \frac{-1}{k-1}$$

2b) Because λ_{max} is k, if we keep increasing k, the lower bound $\frac{-1}{k-1}$ of ρ will also be increasing

This makes sense because we are taking samples from the same individual, as we take more and more sample, we expect the value/measurements each time to be close to each other $\Rightarrow$ It's reasonable to expect corr between observations from the same individual to be non-negative

As k increases, $\rho \geq 0 \Rightarrow \rho > -1$

3a) $\nabla_{\sigma^2} \left[-\frac{1}{2} \log \{ \det(\sigma^2 I_n) \} - \frac{1}{2\sigma^2} (Y - X\hat{\beta})^T (Y - X\hat{\beta}) \right]$

$$= -\frac{1}{2} tr[(\hat{\sigma}^2 I_n)^{-1} I_n] + \frac{1}{2\hat{\sigma}^4} (Y - X\hat{\beta})^T (Y - X\hat{\beta}) = 0$$

$$\Rightarrow \frac{-n}{\hat{\sigma}^2} + \frac{1}{\hat{\sigma}^4} (Y - X\hat{\beta})^T (Y - X\hat{\beta}) = 0$$

$$\Rightarrow \hat{\sigma}^2_{ML} = \frac{(Y - X\hat{\beta})^T (Y - X\hat{\beta})}{n}$$

$$E(\hat{\sigma}^2_{ML}) = \frac{1}{n} E(\hat{\epsilon}^T \hat{\epsilon}) = \frac{1}{n} E(\epsilon^T (I_n - H) \epsilon)$$ (H is hat matrix)

$$= \frac{1}{n} Tr((I - H) Var(\epsilon)) + E(\cancelto{0}{\epsilon^T})(I - H) E(\cancelto{0}{\epsilon})$$

$$= \frac{1}{n}(n-p)\sigma^2 = \frac{n-p}{n} \sigma^2 < \sigma^2$$

HW8

Giang Vu

4/3/2021

4.

(a)

The first model's estimate for σ is 11.6, and the second model's estimate is 4.3.

The second model with only treatments C and D has smaller estimate, probably because C and D are of the same drug type so the variance among them is smaller, unlike when we include A and B which are completely different treatments.

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Individual   9   9070  1007.8   7.504 2.02e-05 ***
## Treatment    3   4108  1369.4  10.197 0.000116 ***
## Residuals   27   3626   134.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] 11.58879
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Individual   9   5595   621.7  34.024 6.85e-06 ***
## Treatment    1      0     0.1   0.003    0.959
## Residuals    9    164    18.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] 4.27785
```

(b)

Because the mean squared error for the model 1 (full data) is much larger than model 2 (only treatments B and C), and also the estimate for the treatment effect in model 2 is not statistically significant, it would be reasonable to expect that model 1 fails to account for some of the treatment effect difference between A, B, and C and D. Therefore, we have some evidence to believe that adding an interaction term in our model would be useful here to eliminate this issue with problem 1.

4 c) (i) We concluded that we need an interaction term so here is the model assuming treatment is a fixed effect.

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij} \qquad \begin{matrix} i = 1, \dots, 10 \\ j = 1, \dots, 4 \end{matrix}$$

Y_{ij} - avg. hours of sleep for patient i under treatment (trt) j
μ - population mean hrs of sleep (w/o trt, or trt A)
α_i - random effect of i^{th} patient (each patient is randomly sampled from the population)
β_j - fixed effect of j^{th} treatment
$(\alpha\beta)_{ij}$ - interaction term for patient i and treatment j
ϵ_{ij} - random error with patient i and treatment j

Additional conditions: $\sum_j (\alpha\beta)_{ij} = 0$, $(\alpha\beta)_{ij} \sim N(0, \frac{3}{4}\sigma^2_{\alpha\beta})$
$\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, $\sum_j \beta_j = 0$, $\alpha_i \underset{independent}{\sim} N(0, \sigma_\alpha^2)$
μ is a constant; $\alpha_i, (\alpha\beta)_{ij}, \epsilon_{ij}$ pairwise independent

(ii) Given that avg hrs of sleep before trt is y_0, the new model is

$$Y_{ij} = y_0 + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

Y_{ij} - avg hrs of sleep for patient i under trt j.
y_0 - avg hrs of sleep before trt, or with trt A.
α_i - random effect of patient i
β_j - fixed effect of trt j compared to trt A
$(\alpha\beta)_{ij}$ - interaction term for patient i and trt j
ϵ_{ij} - random error

Conditions. $\sum_j (\alpha\beta)_{ij} = 0$, $(\alpha\beta)_{ij} \sim N(0, \frac{3}{4}\sigma^2_{\alpha\beta})$, $\alpha_i \sim N(0, \sigma_\alpha^2)$
$\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, y_0 is a constant, $\alpha_i, (\alpha\beta)_{ij}, \epsilon_{ij}$ pairwise $\perp\!\!\!\perp$
$\beta_A = y_0$

(iii) The mean effect of each drug based on model in part (ii) depends on the avg hrs of sleep before treatment. Therefore, for patients with different hrs of sleep before trt, we cannot give them the same drug. This is why this mean effect is not a quantity of clinical importance