HW9

Giang Vu

4/22/2021

#Homework 9

1.

(a)

Use REML + GLS to fit the model you proposed in part (c)(i) of Problem 4. Test the null hypothesis that there is no patient × treatment interaction. Make sure to report your null and alternative models.

Null hypothesis model is the model without interaction term. $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$

Model proposed in part (c)(i) is also our alternative hypothesis model
$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$
where:
$i = 1, 2, ..., 10$
$j = 1, 2, 3, 4$
Y_{ij} is average hours of sleep for patient i under treatment j.
μ is population mean hours of sleep, fixed.
α_i is random effect of patient i.
β_j is fixed effect for treatment j.
$(\alpha\beta)_{ij}$ is interaction term for patient i and treatment j.
ϵ_{ij} is random error with patient i and treatment j.
Additional conditions:
$\sum(\alpha\beta)_{ij} = 0$, $(\alpha\beta)_{ij} \sim N(0, \frac{3}{4}\sigma^2_{\alpha\beta})$
$\epsilon_{ij} \sim iid.N(0, \sigma^2)$, $\sum \beta_j = 0$
$\alpha_i \sim independent\ N(0, \sigma^2_\alpha)$

H_0: No interaction term is needed.
H_1: There should be an interaction term.

After fitting both models with REML, and use ANOVA to carry out the test to see if an interaction term is needed, I got a very small p-value for the test (0.0001594).
Assuming the null hypothesis is true, the probability of getting results as extreme as we got from our sample is very small.
It gives me evidence to believe that there should be an interaction term in our model.

(b)

I used the coefficient estimates from model in previous part and replace the overall intercept with y0 like proposed in part (c)(ii) on Problem 4 last homework, and predicted the average hours of sleep for each individual (from 1 to 10), when they are given different drugs (B, C, D) given that they sleep 2 and 6 hours before treatment, respectively.
The results I got show that drug C and drug D have more potent impacts than drug B, and whether the patients sleep 2 or 6 hours before treatment does have an impact on the result after treatment.
However, I am not sure what drug to recommend to each patient.

Predicted avg hours of sleep for each patient (from 1 to 10, respectively) given that they sleep 2 hours before treatment and then received:
- drug B is: 9.693294 22.544908 43.387362 20.069915 28.715824 9.393797 34.231670 18.905062 17.246995 24.841478
- drug C is: 35.67511 48.52673 69.36918 46.05173 54.69764 35.37561 60.21349 44.88688 43.22881 50.82330
- drug D is: 35.82057 48.67218 69.51463 46.19719 54.84310 35.52107 60.35894 45.03233 43.37427 50.96875

Predicted avg hours of sleep for each patient (from 1 to 10, respectively) given that they sleep 6 hours before treatment and then received: - drug B is: 13.69329 26.54491 47.38736 24.06991 32.71582 13.39380 38.23167 22.90506 21.24699 28.84148
- drug C is: 39.67511 52.52673 73.36918 50.05173 58.69764 39.37561 64.21349 48.88688 47.22881 54.82330
- drug D is: 39.82057 52.67218 73.51463 50.19719 58.84310 39.52107 64.35894 49.03233 47.37427 54.96875

2) a) Suppose $A_x(A_x^T A_x)^{-1} A_x^T = Q$

$\Rightarrow A_x(A_x^T A_x)^{-1} A_x^T A_x = Q A_x$

$\Rightarrow A_x = Q A_x$

Also $Q = I_n - X(X^TX)^{-1}X^T$

and $X^T A_x = 0$ (A_x has columns forming basis for $\ker(X^T)$)

$\Rightarrow A_x = (I_n - X(X^TX)^{-1}X^T)A_x$

$\Rightarrow A_x = A_x - X(X^TX)^{-1}\underbrace{X^TA_x}_{=0}$

$\Rightarrow A_x = A_x$ (always true)

$\Rightarrow A_x(A_x^TA_x)^{-1}A_x^T = Q$

b) $\tilde{Y} = QY$

$\Rightarrow E(\tilde{Y}) = E(QY) = E((I_n - X(X^TX)^{-1}X^T)Y)$

$= E(Y - X(X^TX)^{-1}X^T(X\beta + \epsilon))$

$= E(Y) - E(X\beta) - E(X(X^TX)^{-1}X\epsilon)$

$= E(X\beta) + \cancel{E(\epsilon)}^{0} - E(X\beta) - X(X^TX)^{-1}X\cancel{E(\epsilon)}^{0}$

$= E(X\beta) - E(X\beta) = 0$

$Var(\tilde{Y}) = Var(QY) = Q\,Var(Y)Q^T = Q\,V(\theta)\,Q$

(Q is orthogonal projection matrix, Q is symmetric & idempotent)

$\tilde{Y} \sim N(0, QV(\theta)Q)$; $\tilde{Y} = QY \in \ker(X^T)$ (because A forms basis for $\ker(X^T)$)

$\ker(X^T) \subset \mathbb{R}^n \Rightarrow \tilde{Y}$ has a degenerate distribution.

Scanned with CamScanner

2c) (i) ⊛ $X^{n\times p}$ and $A^{n\times(n-p)}$ are full rank matrices

$$\left.\begin{aligned} \Rightarrow \quad A^+ &= (A^TA)^{-1}A^T \\ X^+ &= (X^TX)^{-1}X^T \end{aligned}\right\} \Rightarrow Q = I_n - XX^+ = AA^+$$

We also have $A^TX = 0$

($V^{\frac{1}{2}}$ exists because V is symmetric, positive definite matrix)

$$\Rightarrow \quad A^T(V^{\frac{1}{2}})^T V^{-\frac{1}{2}} X = 0$$

$$\Rightarrow \quad (V^{\frac{1}{2}}A)^T V^{-\frac{1}{2}}X = 0$$

Therefore, this result below for A and X holds for $V^{\frac{1}{2}}A$ and $V^{-\frac{1}{2}}X$

From $I - X(X^TX)^{-1}X = A(A^TA)^{-1}A^T$

$$\Rightarrow \quad I - V^{-\frac{1}{2}}X(X^TV^{-1}X)^{-1}X^TV^{-\frac{1}{2}} = V^{\frac{1}{2}}A(A^TVA)^{-1}A^TV^{\frac{1}{2}}$$

$$\Rightarrow \quad V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1} = A(A^TVA)^{-1}A^T$$

Denote left hand side $= P$

$$\Rightarrow \quad P = A(A^TVA)^{-1}A^T$$

⊛ In next page, using

$$Q = I - XX^+ = AA^+$$

we will prove

$$P = Q\cdot(QVQ)^+Q$$
$$= (QVQ)^+$$

and therefore show

$$(QVQ)^+ = A(A^TVA)^{-1}A^T \quad (=P)$$

Scanned with CamScanner

2c) (i) continued

We have:

$$A^T Q V Q A [A^+ (QVQ)^- (A^+)^T] A^T Q V Q A$$

$$= A^T Q V Q (QVQ)^- Q V Q A$$

$$= A^T Q V Q A$$

$$\Rightarrow (A^T Q V Q A)^- = A^+ (QVQ)^- (A^+)^T$$

From earlier:

$$P = V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1} = A (A^T V A)^{-1} A^T$$

$$\Rightarrow P = QA (A^T V A)^{-1} A^T Q$$

(Because $QA = A$ proven in 2a)

$$= QA (A^T Q V Q A)^{-1} A^T Q$$

$$= QA (A^+ (QVQ)^+ (A^+)^T) A^T Q$$

$$= Q (QVQ)^+ Q$$

$$\Rightarrow Q(QVQ)^+ Q = A (A^T V A)^{-1} A^T$$

$$\Rightarrow Q(QVQ)^+ Q = QA (A^T V A)^{-1} A^T Q$$

$$\Rightarrow (QVQ)^+ = A (A^T V A)^{-1} A^T$$

$$\Rightarrow (Q V(\theta) Q)^+ = A_x (A_x^T V(\theta) A_x)^{-1} A_x^T$$

(Throughout this problem, I called A_x as "A" and $V(\theta)$ as "V" to make things short)

2 c) (ii) Suppose B has $\lambda_1, \dots, \lambda_n$ as its eigenvalues

B is positive semi-definite $\Rightarrow$ these are n non-negative eigenvalues

Let $\text{rank}(B) = r$ = number of non-zero eigenvalues

Let's say $\lambda_1, \dots, \lambda_r$ are non zero, the remaining $(n-r)$ are zero

$\Rightarrow$ $B + \epsilon I_n$ will also have n eigenvalues, but now

- r of them will be $\lambda_1 + \epsilon, \dots, \lambda_r + \epsilon$
- $(n-r)$ of these will be ϵ

$$\Rightarrow \lim_{\epsilon \to 0} \frac{\det(B + \epsilon I_n)}{\epsilon^{n-r}} = \lim_{\epsilon \to 0} \frac{\left[\prod_{i=1}^{r} (\lambda_i + \epsilon)\right] \epsilon^{n-r}}{\epsilon^{n-r}} = \lim_{\epsilon \to 0} \prod_{i=1}^{r} (\lambda_i + \epsilon) = \prod_{i=1}^{r} \lambda_i$$

$\Rightarrow$ This is product of r non-zero eigenvalues of B

Scanned with CamScanner

2 c) (ii) continued

$$\det_+ (QVQ) = \det_+ (A_x (A_x^T A_x)^{-1} A_x^T V(\theta) A_x (A_x^T A_x)^{-1} A_x^T)$$

$$= \det (A_x^T A_x)^{-1} \det (A_x^T V(\theta) A_x)$$

Scanned with CamScanner

2d) $\tilde{\ell}(\theta) = -\frac{1}{2}\log[\det_+(QV(\theta)Q)] - \frac{1}{2}\tilde{y}^T(QV(\theta)Q)^+\tilde{y}$

$$\ell_{REML}(\theta) = -\frac{1}{2}\log[\det\{A_x^T V(\theta) A_x\}] - \frac{1}{2}(A_x^T Y)^T\{A_x^T V(\theta)A_x\}^{-1}(A_x^T Y)$$

$$= -\frac{1}{2}\log\left(\frac{\det_+(QV(\theta)Q)}{\det(A_x^T A_x)^{-1}}\right) - \frac{1}{2}\underbrace{Y^T}_{\tilde{y}^T}\underbrace{(\overbrace{QA}^{=A(\text{from } 2a)})\{A_x^T V(\theta)A_x\}^{-1}A_x^T}_{=(QV(\theta)Q)^+}\underbrace{QY}_{\tilde{y}}$$

$$= -\frac{1}{2}\log[\det_+(QV(\theta)Q)] + \frac{1}{2}\log(\det(A_x^T A_x)^{-1}) - \frac{1}{2}\tilde{y}^T(QV(\theta)Q)^+\tilde{y}$$

$$= \tilde{\ell}(\theta) + \frac{1}{2}\log(\det(A_x^T A_x)^{-1})$$

Here $c = \frac{1}{2}\log(\det(A_x^T A_x)^{-1})$ doesn't depend on θ

So when we maximize $\ell_{REML}(\theta)$, it's equivalent to maximizing $\tilde{\ell}(\theta)$. The value $\hat{\theta}_{REML}$ is the value that maximizes $\tilde{\ell}(\theta)$ as well. Therefore, no matter what A_x we choose, in the end we're still ending up maximizing the original loglikelihood function $\tilde{\ell}(\theta)$.

2e) $\underset{\theta\in\Theta}{\arg\max}\, \ell_{REML}(\theta) = \underset{\theta\in\Theta}{\arg\max}\, \tilde{\ell}(\theta)$

$$= \underset{\theta\in\Theta}{\arg\max}\left[-\frac{1}{2}\log[\det_+(Q\sigma^2 I_n Q)] - \frac{1}{2}\tilde{y}^T(Q\sigma^2 I_n Q)^+\tilde{y}\right]$$

$$\nabla_{\sigma^2}(\tilde{\ell}(\sigma^2)) = -\frac{1}{2}\operatorname{tr}((Q\sigma^2 I_n Q)^+ Q I_n Q) - \frac{1}{2\sigma^4}\tilde{y}^T(QI_nQ)^+\tilde{y}$$

$$= -\frac{1}{2\sigma^2}(QI_nQ)^+(n-p) - \frac{1}{2\sigma^4}\tilde{y}^T(QI_nQ)^+\tilde{y} = 0$$

$$\Rightarrow \frac{n-p}{\hat{\sigma}^2_{REML}} = \frac{\tilde{y}^T\tilde{y}}{\hat{\sigma}^4_{REML}}$$

$$\Rightarrow \hat{\sigma}^2_{REML} = \frac{\tilde{y}^T\tilde{y}}{n-p} = MSE$$

Scanned with CamScanner

3.

(a)

For age 1,
The estimation of the expected difference in log-concentration between patients with $Wheeze = 2$ and those with $Wheeze = 0$ is the coefficient estimate for Wheeze2 in the model I ran, which is -0.314. This estimate has a 95% confidence interval of (-0.867, 0.239).

For age 3,
Similarly, I obtained the point estimate -0.0844, with a 95% confidence interval of (-0.410, 0.241).

```
##
## Call:
## lm(formula = Bilirubin ~ ., data = age1)
##
## Coefficients:
##  (Intercept)  IndividualID       Wheeze1       Wheeze2         Diet1
##   24.7452997    -0.0008183     0.3017068    -0.3142903     0.1987372
##     Daycare1          Sex1
##   -0.1624301     0.0964383
```

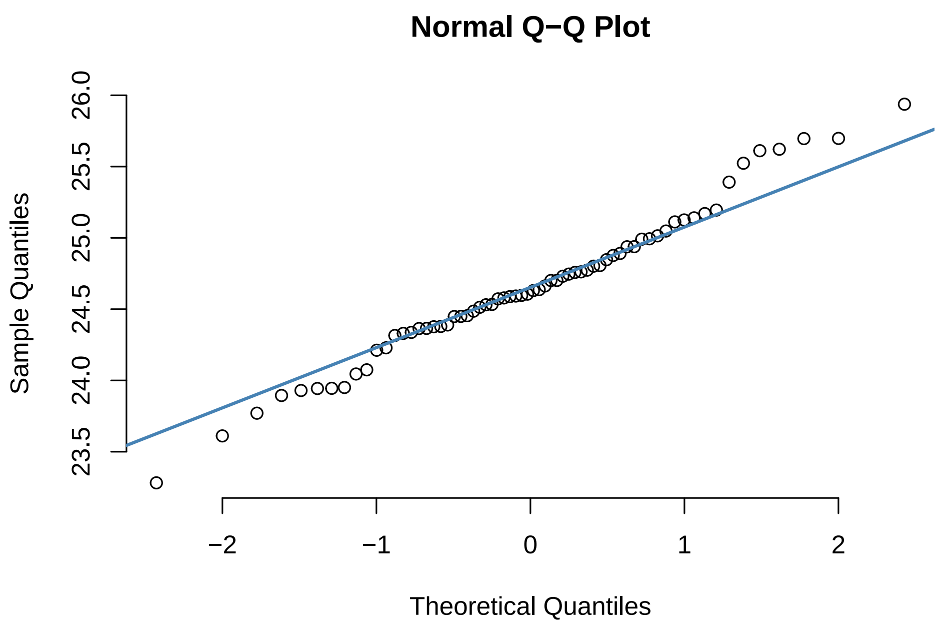
```
##
## Call:
## lm(formula = Bilirubin ~ ., data = age3)
##
## Coefficients:
##  (Intercept)  IndividualID       Wheeze1       Wheeze2         Diet1
##    2.502e+01    -9.186e-05    -8.523e-02    -8.436e-02    -4.809e-02
##     Daycare1          Sex1
##   -5.870e-02    -2.041e-02
```

```
##                     2.5 %       97.5 %
## (Intercept)  24.445371217 25.045228184
## IndividualID -0.002777774  0.001141246
## Wheeze1      -0.067341194  0.670754848
## Wheeze2      -0.867090741  0.238510069
## Diet1        -0.252035908  0.649510340
## Daycare1     -0.433021071  0.108160892
## Sex1         -0.177595520  0.370472152
```

```
##                     2.5 %       97.5 %
## (Intercept)  24.823992254 2.521607e+01
## IndividualID -0.001100446 9.167235e-04
## Wheeze1      -0.322610953 1.521423e-01
## Wheeze2      -0.409619234 2.408944e-01
## Diet1        -0.431460785 3.352716e-01
## Daycare1     -0.222677470 1.052830e-01
## Sex1         -0.187134273 1.463185e-01
```

(b)

Below is the QQ plot to compare the sample distribution of the log concentration bilirubin data versus the theoretical normal distribution.
We can see that it's quite close so this assumption of normal distribution or t-distribution might be trustworthy.



(c)

I combined the data of both ages together and refit the model, and the point estimate for β is -0.158 and the 95% CI is (-0.445, 0.130).

```
##                     2.5 %       97.5 %
## (Intercept)  24.761335185 25.099255474
## IndividualID -0.001172882  0.000662565
## Wheeze1      -0.180408409  0.223957891
## Wheeze2      -0.445476327  0.129634161
## Diet1        -0.287557534  0.303768217
## Daycare1     -0.230742140  0.057849172
## Sex1         -0.142417330  0.147632368
```

(d)

The CI in part (c) compared to CIs in part (a) is actually narrower.
I think it is appropriate to assume normality like we did in (c) because as shown in (b) with the QQ plot,

normality assumption is reliable.

(e)

(i) Because ϕ is the covariance from same person at different age, so we would expect it to be positive.

(ii)

The confidence intervals calculated in part (a) for age 1 and age 3 data don't overlap, i.e., no CI completely lies inside the other CI, so that give us no evidence to say that their variances σ_1^2 and σ_3^2 should be different.
It is reasonable to assume $\sigma_1^2 = \sigma_3^2$.

(f)

I decided to fit this model on the stacked data of age1 and age3 datasets, with an added factor age that specifies if the measurement is taken at age 1 or age 3.
Treating the IndividualID as random effect, and Age and the rest as fixed effect, I got the estimate for σ_j^2 and ϕ as 0.097 and 0.121, respectively.

```
## [1] 0.09658999
```

```
## [1] 0.1209939
```

3) g) $\hat{\beta}_{OLS}$ seeks to mimizes e^Te where

$$e = y - Z\hat{\beta} - \tilde{X}\gamma$$

$$\Rightarrow e^Te = (y - Z\hat{\beta} - \tilde{X}\gamma)^T(y - Z\hat{\beta} - \tilde{X}\gamma)$$

$$= y^Ty - 2\hat{\beta}^TZ^Ty - y^T\tilde{X}\gamma + \hat{\beta}^TZ^TZ\hat{\beta} + 2\hat{\beta}Z^T\tilde{X}\gamma$$

$$-(\tilde{X}\gamma)^Ty + (\tilde{X}\gamma)^T\tilde{X}\gamma$$

$$\Rightarrow \frac{\partial e^Te}{\partial \hat{\beta}} = -2Z^Ty + 2Z^TZ\hat{\beta} + 2Z^T\tilde{X}\gamma = 0$$

$$\Rightarrow Z^TZ\hat{\beta} = Z^Ty - Z^T\tilde{X}\gamma$$

$$\Rightarrow Z^T\tilde{Q}Z\hat{\beta} = Z^T\tilde{Q}y - Z^T\tilde{Q}\tilde{X}\gamma$$

$$\Rightarrow \hat{\beta} = (Z^T\tilde{Q}Z)^{-1}Z^T\tilde{Q}y - (Z^T\tilde{Q}Z)^{-1}Z^T\underbrace{\tilde{Q}\tilde{X}}_{=0}\gamma$$

$$\tilde{Q}\tilde{X} = (I - \tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T)\tilde{X} = \tilde{X} - \tilde{X} = 0$$

$$\Rightarrow \hat{\beta}_j = (Z_j^T\tilde{Q}_jZ_j)^{-1}Z_j^T\tilde{Q}_jY_j$$

Scanned with CamScanner

(h)

I created a new variable Z as defined by the question in each dataset for age 1 and age 3, and fit the models again using Z instead of Wheeze, and obtained the OLS estimates for β_1 and β_3 as -0.386 and -0.068, respectively.
The variances for these estimates are 0.076 and 0.026, respectively.

```
##
## Call:
## lm(formula = Bilirubin ~ ., data = age1[, -3])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.01338 -0.27371 -0.00111  0.31214  1.31863
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.7497241  0.1519440 162.887   <2e-16 ***
## IndividualID -0.0004314  0.0009635  -0.448    0.656
## Diet1         0.1970744  0.2283969   0.863    0.392
## Daycare1     -0.1171240  0.1341977  -0.873    0.386
## Sex1          0.0525867  0.1361625   0.386    0.701
## Z1           -0.3856973  0.2765763  -1.395    0.168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5235 on 60 degrees of freedom
## Multiple R-squared:  0.06028,    Adjusted R-squared:  -0.01803
## F-statistic: 0.7698 on 5 and 60 DF,  p-value: 0.5753
```

```
##
## Call:
## lm(formula = Bilirubin ~ ., data = age3[, -3])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34436 -0.28390 -0.00569  0.26648  1.48963
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.501e+01  9.819e-02 254.722   <2e-16 ***
## IndividualID -6.925e-05  5.074e-04  -0.136    0.892
## Diet1        -5.168e-02  1.932e-01  -0.268    0.790
## Daycare1     -6.479e-02  8.221e-02  -0.788    0.432
## Sex1         -2.884e-02  8.320e-02  -0.347    0.729
## Z1           -6.821e-02  1.624e-01  -0.420    0.675
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4593 on 121 degrees of freedom
## Multiple R-squared:  0.009198,   Adjusted R-squared:  -0.03174
## F-statistic: 0.2247 on 5 and 121 DF,  p-value: 0.9512
```

(i)

I rerun (c) but this time with new variable Z instead of Wheeze.
The estimate for β is -0.162 with a 95% CI of (-0.446, 0.123), which is quite close the result in part (c).

```
##                     2.5 %       97.5 %
## (Intercept)  24.763682959 2.509972e+01
## IndividualID -0.001166193 6.632307e-04
## Diet1        -0.285962611 3.036683e-01
## Daycare1     -0.226288010 5.815699e-02
## Sex1         -0.141434725 1.476999e-01
## Z1           -0.446383950 1.232778e-01
```