

BIOS767__HW3

Cheyenna Crowley

2/6/2019

Question 1: Given

Y_1, \dots, Y_n ($n > 1$) are iid $N(\beta, \theta)$

$R_i = Y_i - Y_n$ $R = (R_1, \dots, R_m)^T$, $m = n-1$

R is MVN

Question 1A.i: Find $E(R_i)$

$$E[R_i] = E[Y_i - Y_n] = E[Y_i] - E[Y_n] = \beta - \beta = 0$$

Question 1A.ii: Find $Var(R_i)$

$$Var[R_i] = Var[Y_i - Y_n] = Var[Y_i] + Var[Y_n] - Cov(Y_i, Y_n) = \theta + \theta = 2\theta$$

Note: Y_i 's are iid so $Cov(Y_i, Y_j) = 0$ for all $i, j = 1, \dots, n$ where $i \neq j$

Question 1A.iii: Find $Cov(R_i, R_j)$ for $i \neq j$

$$Cov(R_i, R_j) = Cov(Y_i - Y_n, Y_j - Y_n) = Cov(Y_i, Y_j) - Cov(Y_n, Y_j) - Cov(Y_i, Y_n) + Cov(Y_n, Y_n) = Var(Y_n) = \theta$$

Note: Y_i 's are iid so $Cov(Y_i, Y_j) = 0$ for all $i, j = 1, \dots, n$ where $i \neq j$

$Cov(Y_n, Y_n) = Var(Y_n)$ by definition of Covariance

Question 1A.iv: Find $m \times m$ covariance matrix R written as $(a-b)I + bJ$

$$\Sigma = \begin{bmatrix} 2\theta & \theta & \dots & \theta \\ \theta & 2\theta & \dots & \theta \\ \theta & \theta & \ddots & \theta \\ \theta & \dots & \theta & 2\theta \end{bmatrix}_{m \times m} = (a-b)I_{m \times m} + bJ_{m \times m} = (2\theta - \theta)I_{m \times m} + \theta J_{m \times m} = \theta I_{m \times m} + \theta J_{m \times m}$$

$$b = \theta$$

$$(a-b) + b = 2\theta \Rightarrow a = 2\theta$$

With this format we see that $a-b = 2\theta - \theta = \theta \neq 0$ and $a + (m-1)b = 2\theta + (m-1)\theta = 2\theta + (n-2)\theta = 2\theta + n\theta - 2\theta = n\theta \neq 0$ since we know $n > 1$.

With this we can calculate the Σ^{-1} and $|\Sigma|$ where:

$$\Sigma^{-1} = [(a-b)I + bJ]^{-1} = \frac{1}{a-b} [I - \frac{b}{a+(m-1)b} J]$$

$$= \frac{1}{2\theta - \theta} [I - \frac{\theta}{2\theta + (m-1)\theta} J]$$

$$= \frac{1}{\theta} [I - \frac{\theta}{2\theta + m\theta - \theta} J]$$

$$= \frac{1}{\theta} [I - \frac{\theta}{\theta(1+m)} J]$$

$$\begin{aligned}
&= \frac{1}{\theta} [I - \frac{1}{(1+m)} J] \\
|\Sigma| &= \det[(a-b)I + bJ] = (a-b)^{m-1} [a + (m-1)b] \\
&= (2\theta - \theta)^{m-1} [2\theta + (m-1)\theta] \\
&= (\theta)^{m-1} [2\theta + m\theta - \theta] \\
&= (\theta)^{m-1} [\theta + m\theta] \\
&= (\theta)^{m-1} [\theta(1+m)] \\
&= (1+m)\theta^m
\end{aligned}$$

Question 1A.v: Develop an explicit expression for the log-likelihood for θ based on R by deriving the multivariate normal pdf of R .

$$P(R; \Sigma) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp -1/2 (R^T \Sigma^{-1} R)$$

Using the results from the previous problem, where:

$$\begin{aligned}
\Sigma^{-1} &= \frac{1}{\theta} [I - \frac{1}{(1+m)} J] \text{ and } |\Sigma| = (1+m)\theta^m \\
f(R; \theta) &= (2\pi)^{-m/2} ((1+m)\theta^m)^{-1/2} \exp [-1/2 (R^T (\frac{1}{\theta} [I - \frac{1}{(1+m)} J]) R)] \\
&= (2\pi)^{-m/2} ((1+m)\theta^m)^{-1/2} \exp [\frac{-1}{2\theta} R^T [I - \frac{1}{(1+m)} J] R] \\
l(\theta; R) &= \log[(2\pi)^{-m/2} ((1+m)\theta^m)^{-1/2} \exp [\frac{-1}{2\theta} R^T [I - \frac{1}{(1+m)} J] R]] \\
&= \frac{-m}{2} \log(2\pi) - \frac{1}{2} \log(1+m) - \frac{1}{2} \log(\theta^m) - \frac{1}{2\theta} [R^T (I - \frac{1}{1+m} J) R]
\end{aligned}$$

By dropping additive constants (as done per the notes) the log likelihood for θ based on R is:

$$l(\theta; R) = -\frac{m}{2} \log(\theta) - \frac{1}{2\theta} [R^T (I - \frac{1}{1+m} J) R]$$

Question 1A.vi: Find the MLE of θ based on the REML likelihood

$$\begin{aligned}
l(\theta; R) &= -\frac{m}{2} \log(\theta) - \frac{1}{2\theta} [R^T (I - \frac{1}{1+m} J) R] \\
\frac{dl(\theta; R)}{d\theta} &= -\frac{m}{2\theta} + \frac{1}{2\theta^2} [R^T (I - \frac{1}{1+m} J) R] \\
\text{set } &= 0 \text{ and solve for } \theta \\
-\frac{m}{2\theta} + \frac{1}{2\theta^2} [R^T (I - \frac{1}{1+m} J) R] &= 0 \\
\frac{1}{2\theta^2} [R^T (I - \frac{1}{n} J) R] &= \frac{m}{2\theta} \\
\hat{\theta} &= \frac{1}{m} [R^T (I - \frac{1}{n} J) R]
\end{aligned}$$

Question 1Bi: Derive the REML log-likelihood using the REML formula.

From the equation in the REML.pdf the REML formula is:

$$l_{REML}(\theta; R) = -\frac{1}{2} \log(|\Sigma|) - \frac{1}{2} \log(|X^T \Sigma^{-1} X|) - \frac{1}{2} Q(R, \Sigma, X)$$

$$\text{Where: } Q(R, \Sigma, X) = R^T [\Sigma^{-1} - \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}] R$$

$$\text{From previous parts, } \Sigma^{-1} = \frac{1}{\theta} [I - \frac{1}{(1+m)} J] \text{ and } |\Sigma| = (1+m)\theta^m \text{ and } E[Y_i] = X\beta = 0$$

First, solving for $Q(R, \Sigma, X)$:

$$Q(R, \Sigma, X) = R^T [\Sigma^{-1} - \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}] R$$

$$Q(R, \Sigma, X) = R^T [\Sigma^{-1}] R = R^T \Sigma^{-1} R = R^T \frac{1}{\theta} [I - \frac{1}{(1+m)} J] R$$

Using this result,

$$l_{REML}(\theta; R) = -\frac{1}{2} \log(|\Sigma|) - \frac{1}{2} \log(|X^T \Sigma^{-1} X|) - \frac{1}{2} R^T \frac{1}{\theta} [I - \frac{1}{(1+m)} J] R$$

By dropping additive constants (as done per the notes) the log likelihood then

$$= -\frac{1}{2} \log((1+m)\theta^m) - \frac{1}{2\theta} R^T [I - \frac{1}{(1+m)} J] R$$

$$= -\frac{1}{2} \log((1+m)) - \frac{1}{2} \log(\theta^m) - \frac{1}{2\theta} R^T [I - \frac{1}{(1+m)} J] R$$

Again by dropping additive constants (as done per the notes) the log likelihood then:

$$l_{REML}(\theta; R) = -\frac{m}{2} \log(\theta) - \frac{1}{2\theta} R^T [I - \frac{1}{(1+m)} J] R$$

Question 1Bii: Compare to the direct derivation above.

The derivation from 1B is the same as the derivation as 1A.

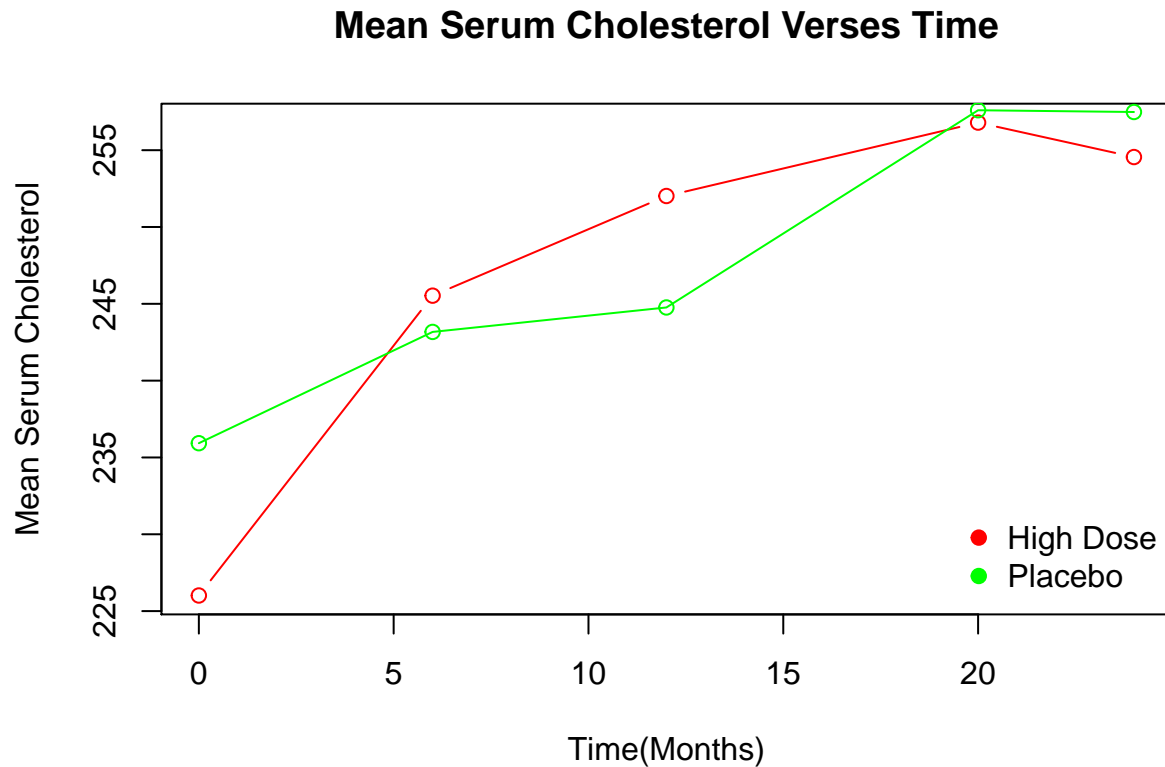
Question 1Biii: Which derivation is easier

Log-likelihood REML is easier.

Question 5.1.2: Calculate the sample means, standard deviations, and variances of the serum cholesterol levels at each occasion for each treatment group

Treatment	Time (Months)	Mean	Standard Deviation	Variance
High Dose	0	225.016	39.664	1573.26
	6	245.532	39.452	1556.48
	12	252.018	38.329	1469.13
	20	256.795	34.489	1189.52
	24	254.553	49.9620	2496.20
Placebo	0	235.927	55.875	3121.97
	6	243.171	49.240	2424.55
	12	244.763	46.111	2126.19
	20	257.600	51.142	2615.48
	24	257.484	49.388	2439.19

Question 5.1.3: On a single graph, construct a time plot that displays the mean serum cholesterol verses time (in months) for the two treatment group. Describe the general characteristics of the time trends for the two groups.



The graph shows that the mean serum cholesterol for high dose at baseline is lower than the mean serum cholesterol for placebo. It also shows that at the end of the study (t=24) patients that received the high dose had a lower mean serum cholesterol than patients that received placebo.

Question 5.1.5: Assuming an unstructured covariance matrix, conduct an analysis of response profiles. Determine whether the patterns of change over time differ in two treatment groups.

Parameter	Estimate
β_0	235.930
β_1	-9.911
β_2	7.244
β_3	8.848
β_4	23.103
β_5	21.124
β_6	12.272
β_7	16.4175
β_8	4.977
β_9	6.903

β_0 =Intercept where reference is time=baseline and treatment=placebo

β_1 =Effect for Treatment=High Dose

β_2 =Effect for Time=Week 6

β_3 =Effect for Time=Week 12

β_4 =Effect for Time=Week 20

β_5 =Effect for Time=Week 24

β_6 =Effect for Time=Week 6 and Treatment=High Dose

β_7 =Effect for Time=Week 12 and Treatment=High Dose

β_8 =Effect for Time=Week 20 and Treatment=High Dose

β_9 =Effect for Time=Week 24 and Treatment=High Dose

By conducting a Wald Test, we can test if the patterns of change over time differ in two treatment groups.

H_0 : the pattern of change over time does not differ between the two treatment groups

H_A : Otherwise

$F_{stat}=1.97$

P-value=0.1055

Conclusion: We fail to reject the null hypothesis and conclude that the pattern of change over time does not differ between the two treatment groups.

Question 5.1.6: Display the estimated 5x5 Covariance and correlation matrices for the five repeated measurements of serum cholesterol

Covariance Matrix:

$$\begin{bmatrix} 2186.61 & 1570.05 & 1407.50 & 1444.53 & 1326.20 \\ 1570.05 & 1900.27 & 1390.52 & 1454.97 & 1406.03 \\ 1407.50 & 1390.52 & 1692.74 & 1250.13 & 1347.66 \\ 1444.53 & 1454.97 & 1250.13 & 1756.27 & 1264.18 \\ 1326.20 & 1406.03 & 1347.66 & 1264.18 & 2321.34 \end{bmatrix}$$

Correlation Matrix:

$$\begin{bmatrix} 1.0000 & 0.7702 & 0.7316 & 0.7371 & 0.5886 \\ 0.7702 & 1.0000 & 0.7753 & 0.7964 & 0.6695 \\ 0.7316 & 0.7753 & 1.0000 & 0.7250 & 0.6799 \\ 0.7371 & 0.7964 & 0.7250 & 1.0000 & 0.6261 \\ 0.5886 & 0.6695 & 0.6799 & 0.6261 & 1.0000 \end{bmatrix}$$

Question 5.1.7: With baseline (month 0) and the placebo group (group 2) as the reference group, write out the regression model for the mean serum cholesterol that corresponds to the analysis of response profiles in Problem 5.1.5

$$E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9$$

β_0 =Intercept where reference is time=baseline and treatment=placebo

β_1 =Effect for Treatment=High Dose

β_2 =Effect for Time=Week 6

β_3 =Effect for Time=Week 12

β_4 =Effect for Time=Week 20

β_5 =Effect for Time=Week 24
 β_6 =Effect for Time=Week 6 and Treatment=High Dose
 β_7 =Effect for Time=Week 12 and Treatment=High Dose
 β_8 =Effect for Time=Week 20 and Treatment=High Dose
 β_9 =Effect for Time=Week 24 and Treatment=High Dose
 X_1 =I(Treatment=High Dose)
 X_2 =I(Time=Week 6)
 X_3 =I(Time=Week 12)
 X_4 =I(Time=Week 20)
 X_5 =I(Time=Week 24)
 X_6 =I(Time=Week 6)*I(Treatment=High Dose)
 X_7 =I(Time=Week 12)*I(Treatment=High Dose)
 X_8 =I(Time=Week 20)*I(Treatment=High Dose)
 X_9 =I(Time=Week 24)*I(Treatment=High Dose)

Question 5.1.8: Let L denote a matrix of known weights and β vector of linear regression parameters from the model assumed in Problem 5.1.7. The null hypothesis that the patterns change over time and did not differ in the two treatment groups can be expressed as $H_0 : L\beta = 0$. Describe an appropriate weight matrix L for this null hypothesis .

$$H_0 : L\beta = 0$$

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9)^T$$

$$L = \begin{bmatrix} 00000001000 \\ 00000000100 \\ 00000000010 \\ 00000000001 \end{bmatrix}$$

Question 5.1.10: With baseline (month 0) and the placebo group (group 2) as the reference group, provide an interpretation for each of the estimated regression coefficients in terms of the effect of the treatments on the patterns of change in mean serum cholesterol.

Parameter	Estimate
β_0	235.930
β_1	-9.911
β_2	7.244
β_3	8.848
β_4	23.103
β_5	21.124
β_6	12.272
β_7	16.418
β_8	4.977
β_9	6.903

$\beta_0 = 235.930$ = Reference Mean Serum Cholesterol (time=baseline and treatment=placebo)
 β_1 = Mean serum cholesterol decreases by 9.911 when treatment=High Dose
 β_2 = Mean serum cholesterol increases by 7.244 when Time=Week 6
 β_3 = Mean serum cholesterol increases by 8.848 when time=Week 12
 β_4 = Mean serum cholesterol increases by 23.103 when Time=Week 20
 β_5 = Mean serum cholesterol increases by 21.124 when Time=Week 24
 β_6 = Mean serum cholesterol increases by 12.272 when Time=Week 6 and Treatment=High Dose
 β_7 = Mean serum cholesterol increases by 16.4175 when Time=Week 12 and Treatment=High Dose
 β_8 = Mean serum cholesterol increases by 4.977 when Time=Week 20 and Treatment=High Dose
 β_9 = Mean serum cholesterol increases by 6.903 when Time=Week 24 and Treatment=High Dose
 β_1 through β_9 assumes the other effects are held constant