

# BIOS 6643 HW 2

Charlie Carpenter

9/12/2018

3.

$$L = a^t X \tilde{\beta} = (X^t X)^{-1} X^t Y$$

$$L \tilde{\beta} = a^t X (X^t X)^{-1} X^t Y \sim N[a^t X (X^t X)^{-1} X^t X \beta, (a^t X (X^t X)^{-1} X^t) I \sigma^2 (a^t X (X^t X)^{-1} X^t)^t] \quad (1)$$

$$\sim N[a^t P_x X \beta, (a^t X (X^t X)^{-1} X^t) (X (X^t X)^{-1} X^t a) \sigma^2] \quad (2)$$

$$\sim N[a^t X \beta, (a^t X (X^t X)^{-1} X^t a) \sigma^2] \quad (3)$$

$$\sim N[L \beta, L (X^t X)^{-1} L^t \sigma^2] \quad (4)$$

8.

One-way effects model:  $Y_{ij} = \mu + \kappa_i + \epsilon_{ij} \quad i = 1, 2 \quad j = 1, 2, 3$

Myostatin\*Time (i):  $(Yes, No) \times (24hr, 48hr, 72hr)$

Means model:  $Y_{ijk} = \mu_{ijk} + \epsilon_{ijk} \quad i = 1, 2 \quad j = 1, 2, 3 \quad k = 1, 2$

Two-way effects model:  $Y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + \epsilon_{ijk} \quad i = 1, 2 \quad j = 1, 2, 3 \quad k = 1, 2$

Myostatin (i): 1 = No, 2 = Yes ; Time (j): 1 = 24hr, 2 = 48hr, 3 = 72hr ; Rep (k)

1.  $\mu_{2,2}$
2.  $\mu + \alpha_2 + \tau_2 + \gamma_{2,2}$
3.  $\kappa_5 - \kappa_6$
4.  $\tau_2 + \gamma_{2,2} - (\tau_3 + \gamma_{2,3})$

9

The normal equations are  $X^t X \beta = X^t Y$ . Substituting  $\tilde{\beta} = (X^t X)^{-1} X^t Y$  in for  $\beta$  gives us

$$X^t X \tilde{\beta} = X^t X (X^t X)^{-1} X^t Y \quad \text{This is already true with Full Rank X} \quad (5)$$

$$= X^t P_x Y \quad (6)$$

$$= (P_x^t X)^t Y \quad (7)$$

$$= [(X X^\dagger)^t X]^t Y \quad (8)$$

$$= (X X^\dagger X)^t Y \quad (9)$$

$$= X^t Y \quad (10)$$

11.

a) The full rank model would be

$$Y_{ij} = \beta_0 + \beta_1 Grp1_i + \beta_2 Grp2_i + \beta_3 Grp3_i + \beta_4 Time_i + \quad (11)$$

$$\beta_5 Grp1_i \times Time_j + \beta_6 Grp2_i \times Time_j + \beta_7 Grp3_i \times Time_j + \epsilon_{ij} \quad (12)$$

There would be 8 columns in  $\mathbf{X}$ .

$$\begin{bmatrix} 1 & 0 & 0 & 0 & t_1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & t_2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & t_i & t_i & 0 & 0 \\ 1 & 1 & 0 & 0 & t_{i+i} & t_{i+1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 & t_{i+j} & 0 & t_{i+j} & 0 \\ 1 & 0 & 1 & 0 & t_{i+j+1} & 0 & t_{i+j+1} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 & t_{i+j+k} & 0 & 0 & t_{i+j+k} \\ 1 & 0 & 0 & 1 & t_{i+j+k+1} & 0 & 0 & t_{i+j+k+1} \end{bmatrix}$$

The LTFR model would be

$$Y_{ij} = \beta_0 + \beta_1 Grp1_i + \beta_2 Grp2_i + \beta_3 Grp3_i + \beta_4 Grp4_i + \beta_5 Time_i + \quad (13)$$

$$\beta_6 Grp1_i \times Time_j + \beta_7 Grp2_i \times Time_j + \beta_7 Grp3_i \times Time_j + \beta_8 Grp3_i \times Time_j + \epsilon_{ij} \quad (14)$$

There would be 10 columns in  $\mathbf{X}$ .

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & t_1 & t_1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & t_2 & t_2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 & 0 & t_i & 0 & t_i & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & t_{i+i} & 0 & t_{i+1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 & 0 & t_{i+j} & 0 & 0 & t_i & 0 \\ 1 & 0 & 0 & 1 & 0 & t_{i+j+1} & 0 & 0 & t_{i+1} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 1 & t_{i+j+k} & 0 & 0 & 0 & t_{i+j+k} \\ 1 & 0 & 0 & 0 & 1 & t_{i+j+k+1} & 0 & 0 & 0 & t_{i+j+k+1} \end{bmatrix}$$

b)

Treating time as a class variable could be more appropriate for unequally spaced time points. It is very possible (and common in biological processes) for trends to start out linear and then level off at some point. If we treat time as a class variable we will have a much more flexible model to work with that could capture this type of curve. We would have to take into consideration the extra degrees of freedom we use and the loss of our ability to interpolate as the tradeoffs for this flexibility.



$$\begin{bmatrix}
\mu & \alpha_1 & \alpha_2 & \tau_1 & \tau_2 & \gamma_{1,1} & \gamma_{1,2} & \gamma_{2,1} & \gamma_{2,2} \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0 \\
1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & -1 & -1 & 0 & 0 & -1 & -1 \\
1 & 0 & 1 & -1 & -1 & 0 & 0 & -1 & -1 \\
1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 \\
1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 \\
1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\
1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\
1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1
\end{bmatrix}$$

b)

	<i>Int</i>	<i>Group</i>			<i>Time</i>			<i>G × T</i>						
<i>A linear</i> :	1	1	0	0	-1	0	1	-1	0	1	0	0	0	0
<i>B linear</i> :	1	0	1	0	-1	0	1	0	0	0	-1	0	1	0
<i>L</i> :	0	1	-1	0	0	0	0	-1	0	1	1	0	-1	0

```

## X / Design Matrix
X <- matrix(c(1, 1,0,0, 1,0,0, 1,0,0,0,0,0,0,0,0,0,0,0,0,0,
              1, 1,0,0, 0,1,0, 0,1,0,0,0,0,0,0,0,0,0,0,0,0,
              1, 1,0,0, 0,0,1, 0,0,1,0,0,0,0,0,0,0,0,0,0,0,

              1, 0,1,0, 1,0,0, 0,0,0,1,0,0,0,0,0,0,0,0,0,0,
              1, 0,1,0, 0,1,0, 0,0,0,0,1,0,0,0,0,0,0,0,0,0,
              1, 0,1,0, 0,0,1, 0,0,0,0,0,1,0,0,0,0,0,0,0,0,

              1, 0,0,1, 1,0,0, 0,0,0,0,0,0,1,0,0,0,0,0,0,0,
              1, 0,0,1, 0,1,0, 0,0,0,0,0,0,0,1,0,0,0,0,0,0,
              1, 0,0,1, 0,0,1, 0,0,0,0,0,0,0,0,0,1,0,0,0,0,

              1, 1,0,0, 1,0,0, 1,0,0,0,0,0,0,0,0,0,0,0,0,0,
              1, 1,0,0, 0,1,0, 0,1,0,0,0,0,0,0,0,0,0,0,0,0,
              1, 1,0,0, 0,0,1, 0,0,1,0,0,0,0,0,0,0,0,0,0,0,

              1, 0,1,0, 1,0,0, 0,0,0,1,0,0,0,0,0,0,0,0,0,0,
              1, 0,1,0, 0,1,0, 0,0,0,0,1,0,0,0,0,0,0,0,0,0,
              1, 0,1,0, 0,0,1, 0,0,0,0,0,1,0,0,0,0,0,0,0,0,

              1, 0,0,1, 1,0,0, 0,0,0,0,0,0,1,0,0,0,0,0,0,0,
              1, 0,0,1, 0,1,0, 0,0,0,0,0,0,0,1,0,0,0,0,0,0,
              1, 0,0,1, 0,0,1, 0,0,0,0,0,0,0,0,0,1,0,0,0,0,
              nrow = 18, byrow = T)

## Moore-Penrose inverse
H <- ginv(t(X) %*% X) %*% t(X) %*% X

```

```
L <- c(0, 1,-1,0, 0,0,0, -1,0,1, 1,0,-1, 0,0,0)
```

```
L%*%H %>% as.vector %>% round
```

```
## [1] 0 1 -1 0 0 0 0 -1 0 1 1 0 -1 0 0 0
```

Which is our L, so comparing the linear trends of two groups is estimable.

14

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr >  t
CL vs CO time1	-0.8517	3.9722	60	-0.21	0.8310
CH vs CO time1	3.0333	3.9722	60	0.76	0.4481

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
trt F-test	2	15	1.20	0.3275
time F-test	4	60	0.65	0.6309

The two *Estimate* statements show that there is no evidence of pairwise differences between the control group and the two treatment groups at baseline in the Dog data. There is also no evidence of a difference in the overall means for the three treatments or the 5 time points.

We know that there is much more going on in these data, but I just wanted to confirm the full rank reparameterization theorem with contrasts.