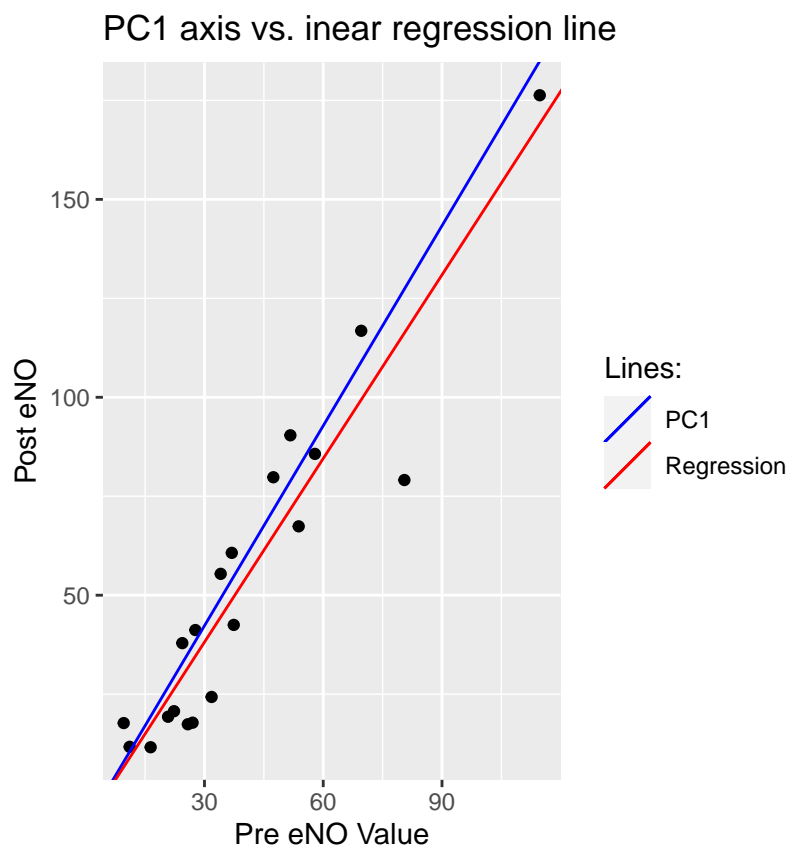# Homework 2

Joseph Froelicher

October 6, 2021

Question 1

The plot above indicates that the PC1 axis and linear regression provide similar context for the general trend of the data, or rather the association between pre and post. As we would expect, pre-eNO value is linearly associated with post-eNO value. We should note that the slope of these two lines is slightly difference, and this is a result of their differing methods of calculation.

Question 2

Part A

General linear models contain at a minimum an outcome, modeled linearly by its intercept. They allow for additional parameters to define the slope of the line, which we typically refer to as β coefficients. There can be any number of β coefficients. These are referred to as fixed effects. Linear mixed models allow for the addition of what we refer to as random effects. The random effects could include random intercepts and random slopes. Also importantly in general linear models, we assume that the errors are normally distributed, that is $\epsilon \sim N(0, \sigma^2)$. One of the many uses of linear mixed models is that we can relax this assumption and model the errors with different structures. General linear models are a special case of linear mixed model that contain no random effects and have a simple error covariance structure. Both methods use commonly use both maximum likelihood estimation and restricted maximum likelihood estimation to estimate parameters.

Part B

A profile likelihood is the standard likelihood function but where some subset of the parameters are estimated using maximum likelihood, and then the remaining parameters can be solved for analytically, sometimes referred to as profiling out. A restricted likelihood instead of estimating those parameters that were pofiled out before, they are eliminated by integrating the the log-likelihood function over the mean, which is effectively computing a marignal probability. This results in an unbiased estimate when compared to maximum likelihood estimation.

Part C

$$MLE[\beta] = \hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}X)^{-1}\mathbf{X}^T\mathbf{V}^{-1}y$$

$$Var\left[\hat{\beta}\right] = Var\left[(\mathbf{X}^T\mathbf{V}^{-1}X)^{-1}\mathbf{X}^T\mathbf{V}^{-1}y\right]$$

$$Var\left[\hat{\beta}\right] = (\mathbf{X}^T\mathbf{V}^{-1}X)^{-1}\mathbf{X}^T\mathbf{V}^{-1}Var\left[y\right]\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$

$$Var\left[\hat{\beta}\right] = (\mathbf{X}^T\mathbf{V}^{-1}X)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$

$$Var\left[\hat{\beta}\right] = (\mathbf{X}^T\mathbf{V}^{-1}X)^{-1}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$

$$Var\left[\hat{\beta}\right] = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$

Question 3

Parts A & B

```
proc mixed data = data;
  class prepar time;
    model y = prepar time prepar*time / solution;
    repeated / subject=ID(prepar) type=UN;
  contrast "30 mg vs. 60 mg"
    prepar*time 0 0 0 0 0 0 0 0 0 0 1 -1 0 0 0 -1 1 0 0 0,
    prepar*time 0 0 0 0 0 0 0 0 0 0 1 0 -1 0 0 -1 0 1 0 0,
    prepar*time 0 0 0 0 0 0 0 0 0 0 1 0 0 -1 0 -1 0 0 1 0,
    prepar*time 0 0 0 0 0 0 0 0 0 0 1 0 0 0 -1 -1 0 0 0 1;
  contrast "12-week vs. baseline"
    prepar*time 1 0 0 0 -1 -1 0 0 0 1 0 0 0 0 0 0 0 0 0 0,
    prepar*time 1 0 0 0 -1 0 0 0 0 0 -1 0 0 0 1 0 0 0 0 0,
    prepar*time 1 0 0 0 -1 0 0 0 0 0 0 0 0 0 -1 0 0 0 0 1,
    prepar*time 0 0 0 0 0 1 0 0 0 -1 -1 0 0 0 1 0 0 0 0 0,
    prepar*time 0 0 0 0 0 1 0 0 0 -1 0 0 0 0 0 -1 0 0 0 1,
    prepar*time 0 0 0 0 0 0 0 0 0 0 1 0 0 0 -1 -1 0 0 0 1;
run;
```

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
30 mg vs. 60 mg	4	19	3.34	0.0313
12-week vs. baseline	3	19	2.40	0.0997

Figure 1: Resutls of two contrasts for questions A and B.

Part C

```
data poly;
 set data;
 time_sq = time*time;
 time_cu = time*time*time;
run;

proc mixed data=poly;
  class prepar;
  model y = prepar time time_sq time_cu prepar*time prepar*time_sq prepar*time_cu / solution;
  repeated / subject=ID(prepar) type=UN;
run;
```

Time as a class variable offers the most flexibility and imposes no parametric constraints imposed across levels of time. However, it uses more degrees of freedom in the model. Time as a continuous variable is recommended when there are many times of observation, possibly unequally spaced times, different times of measurement for subjects, and when interpolating estimates and predicted values may be of interest. We know that the quartic polynomial is equivalent to treating time as a class variable. Howwever seeing as we have been restricted to no higher than cubic, this are not effectively equivalent. The cubic model with continuous time may be a good choice because of unequally space time points. The model including time as categorical and all of its interactions has a significantly lower AIC (1054) than the cubic model (1174).

Part D

```
proc import datafile = 'C:\Users\jofro\dev\longi\data\beta_carotene_univar_CLEAN.csv'
 out = clean
 dbms = CSV;
run;

proc mixed data=clean;
  class prepar time;
  model y = prepar baseline time / solution;
  repeated / subject=ID(prepar) type=UN;
  estimate 'linear' time -3 -1 1 3;
  estimate 'quadratic' time 1 -1 -1 1;
  estimate 'cubic' time -1 3 -3 1;
run;
```

One significant advantage to modeling the data in this way, is that by removing the baseline timepoint from the repeated measures outcome, the outcome now has equally spaced time points. It now makes much more sense to model time a categorical variable, because we have equally space time points, and we now can consider structuring the covariance matrix, particularly as AR(1). However, by removing the baseline measurement we can no longer specify a "start value" for each participant, we now just have the mean estimates (fixed effects).

Part E

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > |t|
linear	26.2609	39.6780	18	0.66	0.5165
quadratic	15.6522	13.0080	18	1.20	0.2445
cubic	4.6957	32.3936	18	0.14	0.8864

Figure 2: Results of continuous time trend estimates, when time is a class variable

Question 4

Part A

Section i, LTFR Model

$$\mathbf{Y}_{ijk} = \mu + \tau_j + \kappa_k + \gamma_{jk} + b_i + \epsilon_{ijk}$$

$$\mu : \text{joint mean}$$
$$\tau : \text{time catergory}$$
$$\kappa : \text{group catergory}$$
$$\gamma : \text{group time interaction}$$
$$i = 1,2 \ (subject)$$
$$j = 1,2,3 \ (time)$$
$$k = 1,2,3 \ (group)$$
$$b_i \sim \mathcal{N}(0, \mathbf{G}_i)$$
$$\epsilon_{ijk} \sim \mathcal{N}(0, \mathbf{R}_i)$$
$$\mathbf{G}_i = \sigma_b^2 \mathbf{I}$$
$$\mathbf{R}_i = \sigma_\epsilon^2 \mathbf{I}$$

$\mathbf{X}_{(18,16)} =$

	μ	τ_1	τ_2	τ_3	κ_1	κ_2	κ_3	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}	γ_{31}	γ_{32}	γ_{33}
subject 1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0
subject 1	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
subject 1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
subject 2	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0
subject 2	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
subject 2	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
subject 3	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0
subject 3	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0
subject 3	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0
subject 4	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0
subject 4	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0
subject 4	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0
subject 5	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
subject 5	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0
subject 5	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1
subject 6	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
subject 6	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0
subject 6	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1

Section ii, Set-to-0 Model

$$\mathbf{Y}_{ijk} = \mu + \tau_j + \kappa_k + \gamma_{jk} + b_i + \epsilon_{ijk}$$

$$\mu : \text{joint mean}$$
$$\tau : \text{time catergory}$$
$$\kappa : \text{group catergory}$$
$$\gamma : \text{group time interaction}$$
$$i = 1, 2 \, (subject)$$
$$j = 1, 2 \, (time, \text{ref=3})$$
$$k = 1, 2 \, (group, \text{ref=3})$$
$$b_i \sim \mathcal{N}(0, \mathbf{G}_i)$$
$$\epsilon_{ijk} \sim \mathcal{N}(0, \mathbf{R}_i)$$
$$\mathbf{G}_i = \sigma_b^2 \mathbf{I}$$
$$\mathbf{R}_i = \sigma_\epsilon^2 \mathbf{I}$$

$$
\mathbf{X}_{(8,8)} =
\begin{bmatrix}
& \mu & \tau_1 & \tau_2 & \kappa_1 & \kappa_2 & \gamma_{11} & \gamma_{12} & \gamma_{21} & \gamma_{22} \\
subject\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
subject\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
& - & - & - & - & - & - & - & - & - \\
subject\ 2 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
subject\ 2 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
& - & - & - & - & - & - & - & - & - \\
subject\ 3 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
subject\ 3 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
& - & - & - & - & - & - & - & - & - \\
subject\ 4 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
subject\ 4 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

Part B

```
X = matrix(data = NA, nrow = 18, ncol = 16)
X[1,] = c(1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)
X[2,] = c(1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0)
X[3,] = c(1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0)
X[4,] = c(1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)
X[5,] = c(1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0)
X[6,] = c(1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0)
X[7,] = c(1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0)
X[8,] = c(1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0)
X[9,] = c(1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0)
X[10,] = c(1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0)
X[11,] = c(1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0)
X[12,] = c(1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0)
X[13,] = c(1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0)
X[14,] = c(1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0)
X[15,] = c(1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1)
X[16,] = c(1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0)
X[17,] = c(1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0)
X[18,] = c(1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1)

H = ginv(t(X) %*% X, tol = .Machine$double.eps) %*% t(X) %*% X
L_t = c(0, 0, 0, 0, 1, -1, 0, 1, -1, 0, 0, 0, 0, 0, 0, 0)
L = t(L_t)

round((L %*% H), 1)-L
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
## [1,]    0    0    0    0    0    0    0    0    0     0     0     0     0     0
##      [,15] [,16]
## [1,]     0     0
```

Above, I took the design matrix $\mathbf{X}$ used it to calculate $\mathbf{H}$ the "hat" matrix using the `ginv()` function in R. To demonstrate that $\mathbf{L} = \mathbf{LH}$ we can just simply output $\mathbf{LH} - \mathbf{L}$, which should be a row vector of zeroes.

Part C

In part A, our error distribution is $\epsilon_{ijk} \sim \mathcal{N}(0, \mathbf{R}_i)$ where $\mathbf{R}_i = \sigma_\epsilon^2 \mathbf{I}$. When we incorporate AR(1) covariance structure, the off diagonals of the covariance matrix are no longer zeroes, they are some function of the variance times the correlation parameter ρ.

Part D

Less Than Full Rank

$$\mathbf{Y}_{ij} = \mu + \beta_{time} x_{1i} + \kappa_j + \gamma_j + b_i + \epsilon_{ij}$$

$$\mu : \text{joint mean}$$
$$\beta_{time} : \text{continuous time regression coefficient}$$
$$\kappa : \text{group catergory}$$
$$\gamma : \text{group time interaction}$$
$$i = 1, 2 \, (subject)$$
$$j = 1, 2, 3 \, (group)$$
$$b_i \sim \mathcal{N}(0, \mathbf{G}_i)$$
$$\epsilon_{ij} \sim \mathcal{N}(0, \mathbf{R}_i)$$
$$\mathbf{G}_i = \sigma_b^2 \mathbf{I}$$
$$\mathbf{R}_i = \sigma_\epsilon^2 \mathbf{I}$$

$$
\mathbf{X}_{(6,8)} =
\begin{bmatrix}
 & \mu & \beta_{time} & \kappa_1 & \kappa_2 & \kappa_3 & \gamma_{time \text{ x } group=1} & \gamma_{time \text{ x } group=2} & \gamma_{time \text{ x } group=3} \\
subject\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
subject\ 2 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
subject\ 3 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
subject\ 4 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
subject\ 5 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
subject\ 6 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1
\end{bmatrix}
$$

Part E

i. There are times when it is either more or less appropriate to use time as a class variable. One of those times is when the times are equally spaced. One suggestion is when observations were not collected over equally spaced times, to treat time as continuous. So in this case, we should not be treating time as a class variable.

ii. The particular covariance we have discussed about possibly using for when time is unequally spaced times is compound symmetry, this is a naive solution. There are some more sophisticated structures that account for points closer in time to be more correlated than those farther apart in time. One particular structure that adjusts for distance between measurements is the spatial power structure (See below).

$$
\mathbf{R}_i =
\begin{bmatrix}
\sigma^2 & \sigma^2 \rho^{d_{12}} & \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{14}} & \sigma^2 \rho^{d_{15}} & \sigma^2 \rho^{d_{16}} \\
\sigma^2 \rho^{d_{21}} & \sigma^2 & \sigma^2 \rho^{d_{23}} & \sigma^2 \rho^{d_{24}} & \sigma^2 \rho^{d_{25}} & \sigma^2 \rho^{d_{26}} \\
\sigma^2 \rho^{d_{31}} & \sigma^2 \rho^{d_{32}} & \sigma^2 & \sigma^2 \rho^{d_{34}} & \sigma^2 \rho^{d_{35}} & \sigma^2 \rho^{d_{36}} \\
\sigma^2 \rho^{d_{41}} & \sigma^2 \rho^{d_{42}} & \sigma^2 \rho^{d_{43}} & \sigma^2 & \sigma^2 \rho^{d_{45}} & \sigma^2 \rho^{d_{46}} \\
\sigma^2 \rho^{d_{51}} & \sigma^2 \rho^{d_{52}} & \sigma^2 \rho^{d_{53}} & \sigma^2 \rho^{d_{54}} & \sigma^2 & \sigma^2 \rho^{d_{56}} \\
\sigma^2 \rho^{d_{61}} & \sigma^2 \rho^{d_{62}} & \sigma^2 \rho^{d_{63}} & \sigma^2 \rho^{d_{64}} & \sigma^2 \rho^{d_{65}} & \sigma^2
\end{bmatrix}
$$

where d_{ij} is the estimated Euclidean distance between the ith and j^{th} measurements.