1. Suppose $X = [x_1, \ldots, x_r]$ is an $n \times r$ matrix of rank $r$. A technique known as *Gram-Schmidt Orthogonalization* can be used to obtain an $n \times r$ matrix $W = [w_1, \ldots, w_r]$ that has orthogonal columns and the same column space as $X$. This can be useful because $P_W$ is relatively easy to compute when the columns of $W$ are orthogonal, and $\mathcal{C}(X) = \mathcal{C}(W)$ implies $P_X = P_W$. An algorithm for carrying out Gram-Schmidt Orthogonalization is as follows:

   - Let $w_1 = x_1$.
   - Let $w_2 = (I - P_{[w_1]})x_2$.
   - Let $w_3 = (I - P_{[w_1, w_2]})x_3$.
   $\vdots$
   - Let $w_r = (I - P_{[w_1, \ldots, w_{r-1}]})x_r$.

   Now consider an experiment with two factors: $A$ and $B$. Suppose that the levels of factor $A$ are indexed by $i = 1, 2$. Suppose the levels of factor $B$ are indexed by $j = 1, 2$. For $i = 1, 2$ and $j = 1, 2$, let $n_{ij}$ be the number of observations for the treatment combination of level $i$ of factor $A$ and level $j$ of factor $B$. For $i = 1, 2$ and $j = 1, 2$ and $k = 1, \ldots, n_{ij}$, suppose

   $$y_{ijk} = \mu_{ij} + e_{ijk},$$

   where the $\mu_{ij}$ terms are unknown parameters and the $e_{ijk}$ terms are independent and identically distributed as $N(0, \sigma^2)$. The following table contains response averages and the number of observations for each treatment group.

   | Level of Factor $A$ | Level of Factor $B$ | Average Response ($\bar{y}_{ij\cdot}$) | Number of Observations ($n_{ij}$) |
   |---|---|---|---|
   | 1 | 1 | 3.0 | 2 |
   | 1 | 2 | 5.0 | 8 |
   | 2 | 1 | 7.0 | 6 |
   | 2 | 2 | 3.0 | 4 |

   (a) Provide a model matrix $W$ with orthogonal columns that corresponds to the additive model where $\mu_{ij} = \mu + \alpha_i + \beta_j$ for all $i = 1, 2$, $j = 1, 2$, and some unknown parameters $\mu$, $\alpha_1$, $\alpha_2$, $\beta_1$, and $\beta_2$.

   (b) Without the aid of a computer, find the value of $P_W y$. Note that it is usually easiest to compute this by finding $(W'W)^{-1}$, $W'y$, $(W'W)^{-1}W'y$, and then finally $W(W'W)^{-1}W'y$. Even though you don't have all the elements of $y$ it is still possible to compute $P_W y$ from the averages in the table above.

   (c) Without the aid of a computer, find the Type II sum of squares for factor $B$. (This was an exam question, but instead of following the solution given for the exam problem, answer the question by making use of part (b).)

2. An experiment was designed to compare the effect of three drugs (A, B, and C) on the heart rate of women. Five women were randomly assigned to each drug. The heart rate (in beats per minute) of each woman was measured at 0, 5, 10, and 15 minutes after the drug was administered. The data are provided in the file

Let $y_{ijk}$ denote the heart rate at the $k$th time point for the $j$th woman treated with the $i$th drug. Suppose

$$y_{ijk} = \mu_{ik} + w_{ij} + e_{ijk},$$

where $\mu_{ik}$ is an unknown constant for each combination of $i = 1, 2, 3$ and $j = 1, 2, 3, 4, 5$; $w_{ij} \sim N(0, \sigma_w^2)$ for all $i$ and $j$; $e_{ijk} \sim N(0, \sigma_e^2)$ for all $i$, $j$, and $k$; and all random effects are independent.

(a) Using Kronecker product notation, provide an expression for the variance of

$$\boldsymbol{y} = [y_{111}, y_{112}, y_{113}, y_{114}, y_{121}, y_{122}, y_{123}, y_{124}, \ldots, y_{351}, y_{352}, y_{353}, y_{354}]'.$$

(b) Test the null hypothesis of no drug-by-time interactions. Compute a test statistic, state its degrees of freedom, find a $p$-value, and provide a brief conclusion.

(c) Is the mean heart rate 15 minutes after treatment the same for all three drugs? State a formal null hypothesis corresponding to this question. Compute a test statistic, state its degrees of freedom, find a $p$-value, and provide a brief conclusion.

(d) Compute a 95% confidence interval for the mean heart rate 15 minutes after treatment with drug A minus the mean heart rate 15 minutes after treatment with drug B.

3. Again consider the data on heart rate from problem 2. Now suppose

$$y_{ijk} = \mu_{ik} + \epsilon_{ijk},$$

where $\mu_{ik}$ is an unknown constant for each combination of $i = 1, 2, 3$ and $k = 1, 2, 3, 4, 5$ and $\epsilon_{ijk}$ is a normally distributed error term with mean 0 for all $i = 1, 2, 3$, $j = 1, 2, 3, 4, 5$, and $k = 1, 2, 3, 4$. For all $i = 1, 2, 3$ and $j = 1, 2, 3, 4, 5$, let

$$\epsilon_{ij} = (\epsilon_{ij1}, \epsilon_{ij2}, \epsilon_{ij3}, \epsilon_{ij4})'.$$

Suppose all the $\epsilon_{ij}$ vectors are mutually independent, and let $\boldsymbol{W}$ be the variance-covariance matrix of $\epsilon_{ij}$, which is assumed to be the same for all $i = 1, 2, 3$ and $j = 1, 2, 3, 4, 5$.

(a) Find the REML estimate of $\boldsymbol{W}$ under the assumption that $\boldsymbol{W}$ is a positive definite, compound symmetric matrix.

(b) Find AIC and BIC for the case where $\boldsymbol{W}$ is a positive definite, compound symmetric matrix.

(c) Find the REML estimate of $\boldsymbol{W}$ under the assumption that $\boldsymbol{W}$ is a positive definite matrix with constant variance and an AR(1) correlation structure.

(d) Find AIC and BIC for the case where $\boldsymbol{W}$ is a a positive definite matrix with constant variance and an AR(1) correlation structure.

(e) Find the REML estimate of $\boldsymbol{W}$ under the assumption that $\boldsymbol{W}$ is a positive definite, symmetric matrix.

(f) Find AIC and BIC for the case where $\boldsymbol{W}$ is a positive definite, symmetric matrix.

(g) Which of the three structures for $\boldsymbol{W}$ is preferred for this dataset? Explain briefly.

(h) Using the preferred structure for $\boldsymbol{W}$, compute a 95% confidence interval for the mean heart rate 15 minutes after treatment with drug A minus the mean heart rate 15 minutes after treatment with drug B.

4. Let $y_{i1}$ denote the weight (in kg) gained by a calf $i$ from birth to 1 week of age. Let $y_{i2}$ denote the weight (in kg) gained by calf $i$ from 1 week of age to 12 weeks of age. Suppose it is known that

$$\mathrm{Var}(y_{i1}) = \mathrm{Var}(y_{i2}) = 4 \text{ for all } i$$

and that the correlation between $y_{i1}$ and $y_{i2}$ is 0.5 for all $i$. Suppose weight gained by any one calf is independent of the weight gained by any other calf. Suppose the following information is available for three randomly selected calves. (Note that the periods in the table denote data that are missing completely at random.)

| Calf Number ($i$) | $y_{i1}$ | $y_{i2}$ |
|:---:|:---:|:---:|
| 1 | 51 | 54 |
| 2 | 48 | . |
| 3 | 52 | . |

(a) Determine the best linear unbiased estimator of the expected total weight gained by a calf from birth to 12 weeks of age.

(b) Predict the weight gains that are missing from the table above; i.e., predict $y_{22}$ and $y_{32}$.