1. (a) For the jth woman treated with the ith drug,

$$\begin{aligned} \boldsymbol{W} &= \text{Var}(y_{ij1}, y_{ij2}, y_{ij3}, y_{ij4})' \\ &= \begin{bmatrix} \sigma_w^2 + \sigma_e^2 & \sigma_w^2 & \sigma_w^2 & \sigma_w^2 \\ \sigma_w^2 & \sigma_w^2 + \sigma_e^2 & \sigma_w^2 & \sigma_w^2 \\ \sigma_w^2 & \sigma_w^2 & \sigma_w^2 + \sigma_e^2 & \sigma_w^2 \\ \sigma_w^2 & \sigma_w^2 & \sigma_w^2 & \sigma_w^2 + \sigma_e^2 \end{bmatrix} \\ &= \sigma_w^2 \mathbf{1}\mathbf{1}'_{4\times 4} + \sigma_e^2 \boldsymbol{I}_{4\times 4}. \end{aligned}$$

We know that $\text{Var}(\boldsymbol{y})$ is block diagonal with blocks $\boldsymbol{W}$. There are a total of $3 \cdot 5 = 15$ blocks, so that

$$\text{Var}(\boldsymbol{y}) = \boldsymbol{I}_{15\times 15} \otimes \boldsymbol{W} = \boldsymbol{I}_{15\times 15} \otimes (\sigma_w^2 \mathbf{1}\mathbf{1}'_{4\times 4} + \sigma_e^2 \boldsymbol{I}_{4\times 4}).$$

(b) The null hypothesis of no drug-by-time interactions is $H_0 : \mu_{ij} - \mu_{ij*} = \mu_{i*j} - \mu_{i*j*}$ for all $i \neq i*$ and $j \neq j*$. The test statistic $F = 7.12$ on $(6, 36)$ degrees of freedom with $p < 0.001$. We reject the null hypothesis and conclude that there is significant evidence for drug-by-time interactions on heart rate.

(c) The null hypothesis for testing the same mean heart rate 10 minutes after treatment for all three drugs is $H_0 : \mu_{13} = \mu_{23} = \mu_{33}$. The test statistic $F = 1.85$ on $(2, 17.1)$ degrees of freedom with $p = 0.188 > 0.05$. We fail to reject the null hypothesis and conclude that there is no significant evidence for the same mean heart rate 10 minutes after treatment for all three drugs.

(d) An approximate 95% confidence interval for $\mu_{13} - \mu_{23}$ is $(-3.77, 12.57)$ with df $= 17.1$ by the SAS code below.

Note: for part (c-d), df $= 17.1$ was computed by Cochran-Satterthwaite since it is for the difference between simple effects with different whole-plot factors. The easiest way to get this is to use SAS with ddfm=satterthwaite option.

SAS code:

```
proc import datafile="./HeartRate.txt"
 dbms=TAB replace out=d;
run;
proc print data=d (obs=14);
run;
proc mixed;
class woman drug time;
model y=drug time drug*time /ddfm = satterthwaite;
random woman(drug);
contrast "same mean for drug A B C at 10 min"
drug 1 -1 0 drug*time 0 0 1 0 0 0 -1 0 0 0 0 0 ,
drug 1 0 -1 drug*time 0 0 1 0 0 0 0 0 0 0 -1 0 ;
estimate "drug A - drug B at 10 min"
 drug 1 -1 0 drug*time 0 0 1 0 0 0 -1 0 0 0 0 0 /cl;
run;
```

R code:

```
library(MASS)
library(nlme)
library(dplyr)

> d <- read.table("http://dnett.github.io/S510/HeartRate.txt", header = T)
> d$t <- factor((d$time + 5) / 5)   # Levels of time.
> fit <- lme(y ~ drug * t, random = ~ 1 | woman, data = d)
> anova(fit)
            numDF denDF   F-value p-value
(Intercept)     1    36 2900.7782  <.0001
drug            2    12    1.3517  0.2955
t               3    36   10.2159  0.0001
drug:t          6    36    7.1153  <.0001
>
> # Function from Dr. Nettleton's Notes
> test=function(lmout,C,d=0,df){
+    b=fixed.effects(lmout)
+    V=vcov(lmout)
+    dfn=nrow(C)
+    Cb.d=C %*% b - d
+    Fstat=drop(t(Cb.d)%*%solve(C%*%V%*%t(C))%*%Cb.d/dfn)
+    pvalue=1-pf(Fstat,dfn,df)
+    cbind(Fstat=Fstat,pvalue=pvalue)
+ }
> C1 <- matrix(c(0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
+                0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0), byrow=T,nrow = 2)
> test(fit,C1,df=17.1) # df computed by Cochran-Satterthwaite via SAS.
        Fstat    pvalue
[1,] 1.847521 0.1877415
>
> # Function from Dr. Nettleton's Notes
> ci <- function(lmeout, C, df, a = 0.05) {
+   b = fixed.effects(lmeout)
+   V = vcov(lmeout)
+   Cb = C %*% b
+   se = sqrt(diag(C %*% V %*% t(C)))
+   tval = qt(1 - a / 2, df)
+   low = Cb - tval * se
+   up = Cb + tval * se
+   m = cbind(C, Cb, se, low, up)
+   dimnames(m)[[2]] = c(paste("c", 1 : ncol(C), sep = ""),
+                        "estimate", "se", paste(100 * (1 - a), "% Conf.", sep = ""), "limits")
+   return(m)
+ }
> C2 <- matrix(c(0, -1, 0, 0, 0, 0, 0, 0, -1, 0, 0, 0), nrow = 1)
> ci(fit, C2, 17.1) # df computed by Cochran-Satterthwaite via SAS.
     c1 c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12 estimate       se
[1,]  0 -1  0  0  0  0  0  0  -1   0   0   0      4.4 3.872564
     95% Conf.   limits
[1,] -3.76676 12.566758
```

2. (a) Under a compound symmetry assumption,

$$\boldsymbol{W} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix},$$

where the REML estimates for the heart rate data are $\hat{\sigma} = 6.12$ and $\hat{\rho} = 0.777$. $(\hat{\sigma}_s^2 = 29.13, \hat{\sigma}_e^2 = 8.36)$

(b) Using R, AIC =-2(-144.9602)+2(14)=317.92 and BIC =-2(-144.9602)+(14)log(48)= 344.12. Using SAS, AIC =-2(-144.9602)+2(2)= 293.9 and BIC = -2(-144.9602)+(2)log(15)=295.3.

(c) Under an AR(1) assumption,

$$\boldsymbol{W} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix},$$

where the REML estimates for the heart rate data are $\hat{\sigma} = 6.00$ and $\hat{\rho} = 0.828$.

(d) Using R, AIC =-2(-142.9713)+2(14)=313.94 and BIC=-2(-142.9713)+(14)log(48)= 340.14. Using SAS, AIC =-2(-142.9713)+2(2)= 289.9 and BIC =-2(-142.9713)+(2)log(15)= 291.4.

(e) Under a general symmetry assumption,

$$\boldsymbol{W} = \sigma^2 \begin{bmatrix} 1 & \rho_{12}\delta_2 & \rho_{13}\delta_3 & \rho_{14}\delta_4 \\ \rho_{12}\delta_2 & \delta_2^2 & \rho_{23}\delta_2\delta_3 & \rho_{24}\delta_2\delta_4 \\ \rho_{13}\delta_3 & \rho_{23}\delta_2\delta_3 & \delta_3^2 & \rho_{34}\delta_3\delta_4 \\ \rho_{14}\delta_4 & \rho_{24}\delta_2\delta_4 & \rho_{34}\delta_3\delta_4 & \delta_4^2 \end{bmatrix},$$

where the REML estimates for the heart rate data are

$$\begin{aligned} &\hat{\sigma} = 6.10, \\ &\hat{\delta}_2 = 1.08, \quad \hat{\delta}_3 = 0.995, \quad \hat{\delta}_4 = 0.928, \\ &\hat{\rho}_{12} = 0.850, \quad \hat{\rho}_{13} = 0.889, \quad \hat{\rho}_{14} = 0.625, \\ &\hat{\rho}_{23} = 0.870, \quad \hat{\rho}_{24} = 0.631, \quad \hat{\rho}_{34} = 0.794. \end{aligned}$$

(f) Using R, AIC =-2(-139.424)+2(22)=322.85 and BIC =-2(-139.424)+(22)log(48)= 364.01. Using SAS, AIC =-2(-139.424)+2(10)= 298.8 and BIC =-2(-139.424)+(10)log(15)= 305.9.

(g) The model with an AR(1) correlation structure has the smallest AIC and BIC of the three (regardless of whether you used R or SAS). Consequently, the AR(1) correlation structure is preferred for this dataset.

(h) There are several ways to find a 95% confidence interval for $\mu_{13} - \mu_{23}$ using the model with an AR(1) correlation structure. In question 1 (d), we used a split-plot design to get a confidence interval, for which it is clear that we should compute the degrees of freedom using Cochran-Satterthwaite. However, it is less clear for the model using AR(1). Regardless of which degrees of freedom you use, you should have $\widehat{\mu_{13} - \mu_{23}} = 4.4$ with $\sqrt{\widehat{\text{Var}}(\widehat{\mu_{13} - \mu_{23}})} = 3.795$.

We can use the Cochran-Satterthwaite method in SAS by specifying the "ddfm = satterthwaite" option, which gives the interval $(-3.54, 12.34)$ based on df $= 19.2$. The default "ddfm" method in SAS uses df $= 36$, which gives the interval $(-3.30, 12.10)$.

In R, gls computes df $= n - \text{rank}(X) = 48$, which leads to the interval $(-3.23, 12.03)$.

Of these intervals, I would prefer the one where the degrees of freedom are computed by Cochran-Satterthwaite because it is the widest and hence the most conservative in terms of inference about the value of $\mu_{13} - \mu_{23}$.

(i) An approximate 95% confidence interval for $\mu_{13} - \mu_{12}$ is $(-3.7946,\ 2.5946)$ which is obtained by using the Cochran-Satterthwaite method (df=35.9) and $(-3.7942,\ 2.5943)$ which is obtained by the default "ddfm" method (df=36) in SAS.

In R, the approximated interval is $(-3.7667,\ 2.5668)$ with df=48.

SAS code:

```
proc mixed;
     class woman drug time;
     model y = drug time drug*time;
     repeated time / subject = woman type = cs r rcorr;
run;
proc mixed;
     class woman drug time;
     model y = drug time drug*time / ddfm = satterthwaite;
     repeated time / subject = woman type = ar(1) r rcorr;
     estimate 'drug A - drug B at 10 minutes'
          drug 1 -1 0 drug * time 0 0 1 0   0 0 -1 0   0 0 0 0 / cl;
     estimate '10 minutes - 5 minutes within drug A'
          time 0 -1 1 0 drug * time 0 -1 1 0   0 0 0 0   0 0 0 0 / cl;

run;
proc mixed;
     class woman drug time;
     model y = drug time drug*time;
     repeated time / subject = woman type = un  r rcorr;
run;
```

R code:

```
attach(d)
woman <- as.factor(woman)
drug <- as.factor(drug)
time <- as.factor(time)
model.cs <- gls(y ~ drug * time,
                correlation = corCompSymm(form = ~1 | woman),
                method = "REML")
model.ar <- gls(y ~ drug * time,
                correlation = corAR1(form = ~1 | woman),
                method = "REML")
model.sy <- gls(y ~ drug * time,
                correlation = corSymm(form = ~1 | woman),
                weight = varIdent(form = ~1 | time),
                method = "REML")
summary(model.cs)
getVarCov(model.cs)
summary(model.ar)
getVarCov(model.ar)
summary(model.sy)
getVarCov(model.sy)
ci.gls <- function(lmeout, C, df, a = 0.05) {
  b = coef(lmeout)
  V = vcov(lmeout)
  Cb = C %*% b
  se = sqrt(diag(C %*% V %*% t(C)))
  tval = qt(1 - a / 2, df)
  low = Cb - tval * se
  up = Cb + tval * se
  m = cbind( Cb, se, low, up)
  dimnames(m)[[2]] = c("estimate", "se", paste(100 * (1 - a), "% Conf.", sep = ""), "limits")
  return(m)
}
ci.gls(model.ar, C2, 19.2) # Cheated and took Cochran-Satterthwaite df value from SAS.
ci.gls(model.ar, C2, 36) # Default df method in SAS.
ci.gls(model.ar, C2, 48) # Default df method in R.

C3 <- matrix(c(0, 0, 0, -1, 1, 0, 0, 0, 0, 0, 0, 0), nrow = 1) # problem(i)
ci.gls(model.ar, C3, 35.9) # Cheated and took Cochran-Satterthwaite df value from SAS.
ci.gls(model.ar, C3, 36) # Default df method in SAS.
ci.gls(model.ar, C3, 48) # Default df method in R.
```

3. The model we used in this problem is $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zs} + \boldsymbol{e}$ where $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2', ..., \boldsymbol{y}_{50}')', \boldsymbol{e} = (\boldsymbol{e}_1', \boldsymbol{e}_2', ..., \boldsymbol{e}_{50}')', \boldsymbol{s} = (s_1, s_2, ..., s_{50})'$, $\boldsymbol{y}_1 = (y_{11}, y_{12})'$, $\boldsymbol{e}_1 = (e_{11}, e_{12})'$, $\boldsymbol{y}_i = (y_{i1}, y_{i2}, y_{i3})'$ and $\boldsymbol{e}_i = (e_{i1}, e_{i2}, e_{i3})'$ for $i = 2, ..., 50$.

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ & \vdots & \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ \boldsymbol{Z} = diag(\boldsymbol{1}_2, \boldsymbol{I}_{49} \otimes \boldsymbol{1}_3), \ \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \text{ and } \boldsymbol{\epsilon} = \boldsymbol{Zs} + \boldsymbol{e} \sim N\left(\boldsymbol{0}_{149\times 1}, \boldsymbol{\Sigma}\right).$$

$Var(\boldsymbol{y}) = \boldsymbol{\Sigma} = diag(\boldsymbol{W}_1, \boldsymbol{I}_{49} \otimes \boldsymbol{W})$ is a block diagonal matrix with, for $i = 2, ..., 50$,

$$Var\left(\boldsymbol{y}_1\right) = \boldsymbol{W_1} = \sigma_s^2 \boldsymbol{1}_2 \boldsymbol{1}_2' + \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & \delta_2^2 \end{bmatrix},$$

$$VAR\left(\boldsymbol{y}_i\right) = \boldsymbol{W} = \sigma_s^2 \boldsymbol{1}_3 \boldsymbol{1}_3' + \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & \delta_2^2 & 0 \\ 0 & 0 & \delta_3^2 \end{bmatrix}.$$

(a) Under a model above, the REML estimates for the exam score data are

$$\hat{\sigma}_s = 13.43525, \quad \hat{\sigma} = 7.933829 \quad \hat{\delta}_2 = 0.978757, \quad \hat{\delta}_3 = 0.522279$$
$$\hat{\sigma}_s^2 = 180.5059, \quad \hat{\sigma}_1^2 = \hat{\sigma}^2(1) = 62.9456, \quad \hat{\sigma}_2^2 = \hat{\sigma}^2\hat{\delta}_2^2 = 60.2914, \quad \hat{\sigma}_3^2 = \hat{\sigma}^2\hat{\delta}_3^2 = 17.1676.$$

```
> library(nlme)
> d=read.table("http://dnett.github.io/S510/ExamScores.txt",
+               header = T, colClasses = c("factor","factor","numeric"))
> o = lme(score ~ 0 + exam, random = ~ 1 | student,
+        weights = varIdent(form = ~ 1 | exam), data = d)
> o
Linear mixed-effects model fit by REML
  Data: d
  Log-restricted-likelihood: -552.6604
  Fixed: score ~ 0 + exam
   exam1    exam2    exam3
46.84000 58.16000 67.25471

Random effects:
 Formula: ~1 | student
        (Intercept) Residual
StdDev:    13.43525 7.933829
```

```
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | exam
 Parameter estimates:
        1         2         3
1.0000000 0.9787570 0.5222791
Number of Observations: 149
Number of Groups: 50
```

(b) The eBLUP for student 1's exam 3 score is 83.7345 by the R code below.

```
> fixef(o)[3] + ranef(o)[1, 1]
   exam3
83.73545
```

(c) Under the model,

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \hline y_{13} \end{pmatrix} \sim N\left(\left(\begin{array}{c} \mu_1 \\ \mu_2 \\ \hline \mu_3 \end{array} \right), \left[\begin{array}{cc|c} \sigma_1^2 + \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_2^2 + \sigma_s^2 & \sigma_s^2 \\ \hline \sigma_s^2 & \sigma_s^2 & \sigma_3^2 + \sigma_s^2 \end{array} \right] \right).$$

By using the conditional expectation formula of multivariate normal distribution, we can find the expression for $E(y_{13}|y_{11}, y_{12})$ in terms of model (1) paramters.

$$\begin{aligned} E(y_{13}|y_{11}, y_{12}) &= \mu_3 + \begin{pmatrix} \sigma_s^2 & \sigma_s^2 \end{pmatrix} \begin{bmatrix} \sigma_1^2 + \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_2^2 + \sigma_s^2 \end{bmatrix}^{-1} \begin{pmatrix} y_{11} - \mu_1 \\ y_{12} - \mu_2 \end{pmatrix} \\ &= \mu_3 + \frac{\sigma_s^2}{(\sigma_1^2 + \sigma_s^2)(\sigma_2^2 + \sigma_s^2) - \sigma_s^2 \sigma_s^2} \left\{ \sigma_2^2 \left(y_{11} - \mu_1 \right) + \sigma_1^2 \left(y_{12} - \mu_2 \right) \right\} \end{aligned}$$

(d) By plugging the estimated parameters in (a) and the observed values of y_{11} and y_{12}, 76 and 68, respectively,

$$\begin{aligned} \hat{E}(y_{13}|y_{11}, y_{12}) &= \hat{\mu}_3 + \frac{\hat{\sigma}_s^2}{(\hat{\sigma}_1^2 + \hat{\sigma}_s^2)(\hat{\sigma}_2^2 + \hat{\sigma}_s^2) - (\hat{\sigma}_s^2)^2} \left\{ \hat{\sigma}_2^2 \left(y_{11} - \hat{\mu}_1 \right) + \hat{\sigma}_1^2 \left(y_{12} - \hat{\mu}_2 \right) \right\} \\ &= 83.73545 \end{aligned}$$

```
> m = fixef(o)
> W = getVarCov(o, individuals = 2, type = "marginal")
> W = W[[1]]
> W
         1        2        3
1 243.4515 180.5058 180.5058
2 180.5058 240.8056 180.5058
3 180.5058 180.5058 197.6759
> m[3] +  W[3, 1:2] %*% solve(W[1:2, 1:2]) %*% (c(d[1:2, 3]) - m[1:2])
         [,1]
[1,] 83.73545
```

(e) In the same manners in (c), we can find the expression of $E(y_{13}|y_{11}, y_{12})$ in terms of the model in (e).

$$
\begin{aligned}
E(y_{13}|y_{11}, y_{12}) &= \mu_3 + \begin{pmatrix} w_{13} & w_{23} \end{pmatrix} \begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix}^{-1} \begin{pmatrix} y_{11} - \mu_1 \\ y_{12} - \mu_2 \end{pmatrix} \\
&= \mu_3 + \frac{1}{w_{11}w_{22} - (w_{12})^2} \Big\{ (w_{13}w_{22} - w_{23}w_{12})(y_{11} - \mu_1) \\
&\qquad\qquad + (w_{11}w_{23} - w_{12}w_{13})(y_{12} - \mu_2) \Big\}
\end{aligned}
$$

Under a general symmetry assumption of $\boldsymbol{W}$, the estimated parameters are

$$
\hat{\boldsymbol{\mu}} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{pmatrix} = \begin{pmatrix} 46.84 \\ 58.16 \\ 67.18 \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{W}} = \begin{bmatrix} 230.99 & 205.27 & 164.28 \\ 205.27 & 289.85 & 199.54 \\ 164.28 & 199.54 & 184.78 \end{bmatrix},
$$

where $\widehat{\boldsymbol{W}}$ is the REML estimate of $\boldsymbol{W}$.
By plugging the above estimated parameters and observed values of y_{11} and y_{12}, 76 and 68, respectively,

$$\hat{E}(y_{13}|y_{11}, y_{12}) = 79.9037$$

```
> o.un = gls(score ~ 0 + exam,
+           correlation = corSymm(form = ~ 1 | student),
+           weights = varIdent(form = ~ 1 | exam), data = d)
> W = getVarCov(o.un, individual = 2)
> m = coef(o.un)
> m[3] + W[3, 1:2] %*% solve(W[1:2, 1:2]) %*% (c(d[1:2, 3]) - m[1:2])
        [,1]
[1,] 79.9037
```

(f) For i=2,...50, suppose y_i is the exam 3 score for student i. Then, model in (f) can be represented as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

where x_{ij} is the score for student i on exam j (j=1, 2) and $\epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$.
From the R code below, the estimated regression equation by using the data from student 2 through 50 is as following:

$$
\begin{aligned}
\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} \\
&= 25.6232 + 0.2682 x_{i1} + 0.4985 x_{i2}
\end{aligned}
$$

For given (x_{11}, x_{12}) = (76, 68), the predicted value for the exam 3 score of student 1, say $\hat{y}_1$, can be obtained as

$$\hat{y}_1 = 25.6232 + 0.2682 \times 76 + 0.4985 \times 68 = 79.9037$$

```
> d.multreg = matrix(d[-(1:2), 3], byrow = T, ncol = 3)
> d.multreg = data.frame(d.multreg)
> names(d.multreg) = c("exam1", "exam2", "final")
> o.lm = lm(final ~ exam1 + exam2, data = d.multreg)
> o.lm$coefficients
(Intercept)       exam1       exam2
 25.6231947   0.2681885   0.4985030
> newdata = data.frame(exam1 = d[1, 3], exam2 = d[2, 3])
> predict(o.lm, newdata)
       1
79.90372
```