1. Please read and run the R code available at

   http://www.public.iastate.edu/~dnett/S511/SoupSolution.R

2. Please read and run the R code available at

   http://www.public.iastate.edu/~dnett/S511/profile.R

3. (a) These represent the number of model parameters for each model. 49 is 21 parameters for the mean structure
   (3 programs $\times$ 7 times) plus 28 parameters for the variance matrix (7 variances plus $6+5+4+3+2+1 = 21$
   covariances). 23 is 21 parameters for the mean structure (3 programs $\times$ 7 times) plus 2 parameters for the
   variance matrix (1 variance and 1 covariance).

   (b)
```
> b=coef(d.timemle)
> b
  (Intercept)        Program2        Program3            Time      I(Time^2)
  79.17218621      1.22163780      0.28933129      0.29309380     -0.01084112
Program2:Time Program3:Time
   0.05088628     -0.13225739
>
> #program 1 intercept
>
> b[1]
(Intercept)
   79.17219
>
> #program 1 linear coefficient
>
> b[4]
      Time
0.2930938
>
> #program 1 quadratic coefficient
>
> b[5]
  I(Time^2)
-0.01084112
>
> #program 2 intercept
>
> b[1]+b[2]
(Intercept)
   80.39382
>
> #program 2 linear coefficient
>
```

```
> b[4]+b[6]
      Time
0.3439801
>
> #program 2 quadratic coefficient
>
> b[5]
  I(Time^2)
-0.01084112
>
> #program 3 intercept
>
> b[1]+b[3]
(Intercept)
   79.46152
>
> #program 3 linear coefficient
>
> b[4]+b[7]
      Time
0.1608364
>
> #program 3 quadratic coefficient
>
> b[5]
  I(Time^2)
-0.01084112
```

(c)
```
> #Examine first and last subjects.
>
> d[1,]
  Program Subj Time Strength Timef
1       3    1    2       85     2
>
> d[399,]
    Program Subj Time Strength Timef
399       2   57   14       83    14
>
> #First subject is in program 3.
>
> matrix(c(
+ 1,0,1,0,0,0,0,
+ 0,0,0,1,0,0,1,
+ 0,0,0,0,1,0,0),byrow=T,nrow=3)%*%
+ fixef(d.timer)+ranef(d.timer)[1,]
  (Intercept)      Time   I(Time^2)
1    83.83955 0.3087833 -0.01234835
>
> #Last subject is in program 2.
```

```
>
> matrix(c(
+ 1,1,0,0,0,0,0,
+ 0,0,0,1,0,1,0,
+ 0,0,0,0,1,0,0),byrow=T,nrow=3)%*%
+ fixef(d.timer)+ranef(d.timer)[57,]
    (Intercept)     Time I(Time^2)
57     79.66128 0.3175383 -0.008349
```

4.
```
#read in data
read.table("http://www.public.iastate.edu/~dnett/S511/HeartRate.txt",
+ header=T)->heart

#attach the data frame
attach(heart)

#change explanatory variables to factors
woman=as.factor(woman)
drug=as.factor(drug)
time=as.factor(time)

#load library
library(nlme)

#fit a model with unstructured correlation structures
> gls(y~drug*time,correlation=corSymm(form=~1|woman),
weight = varIdent(form = ~ 1|time),method="REML")->model.sy
#fit a model with AR(1) correlation structures
gls(y~drug*time,correlation=corAR1(form=~1|woman),
+ method="REML")->model.ar1
#fit a model with compound symmetric correlation structures
gls(y~drug*time,correlation=corCompSymm(form=~1|woman),
+ method="REML")->model.cs

#compare these three models
> anova(model.sy,model.cs)
         Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.sy     1 22 322.8481 364.0145 -139.4240
model.cs     2 14 317.9204 344.1172 -144.9602 1 vs 2 11.07227  0.1976

> anova(model.sy,model.ar1)
          Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.sy      1 22 322.8481 364.0145 -139.4240
model.ar1     2 14 313.9425 340.1394 -142.9713 1 vs 2 7.094446  0.5265
```
From the above result, we can see that based upon the likelihood ratio test both compound symmetric and AR(1) correlation structures are acceptable, but compound symmetric correlation structure is not as good as AR(1) correlation structure. What's more, both the AIC and BIC for AR(1) correlation structures are the smallest among these three models. So AR(1) correlation structures seems most appropriate for this data set.

There are many ways to answer this question. The following is one example. The profile plot of the heart rate means is shown is Figure 1. From the plot we can see that

(a) The heart rate mean of drug C change gently compared to other two drugs.

- The polylines corresponding to drug B and drug C are somewhat parallel, which means that the effect of drug B and drug C with time have similar patterns. The pattern for drug B is more exaggerated than the pattern for drug C, but the slopes of each line segment have the same sign for drugs B and C. The effect of drug A with time is much different from drug B and drug C though the significance of this difference needs to be evaluated.

The plot suggests that it might be difficult to make concise statements about how heart rate might be affected by the factors drug and time. Let's begin by taking a look at the ANOVA table of the model with AR(1) correlation structure:

```
> anova(model.ar1)
Denom. DF: 48
            numDF   F-value  p-value
(Intercept)     1 3120.9699  <.0001
drug            2    2.1437  0.1283
time            3   14.5263  <.0001
drug:time       6    8.5720  <.0001
```

From the above ANOVA table we can see that, both drug and time have effect on heart rate on the whole. The test for drug main effects is not significant, but both time and drug by time interaction tests are highly significant.

The mean response (under the default constraint of R) of each treatment group is presented in Table 1, where D and T stand for factor drug and time respectively and DT stands for the interaction of drug and time. We can answer various questions that might be of interest by testing the significance of various linear combinations of the estimated parameters. For example, to determine if the the mean heart rate for subjects assigned to drug A changes significantly from baseline (time=0) at each subsequent time point, we can test whether T_2, T_3 and T_4 each differ from zero.

The function $test$ used to carry out these inferences is as follows:

```
test=function(lmout,C,dfd=lmout$df,d=0,b=coef(lmout),V=vcov(lmout))
{
  if(is.vector(C))
    C=t(C)
  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,dfd)
  list(Fstat=Fstat,pvalue=pvalue)
}
```

One way to do inference in this case is to use a Wald approach. The calculations below use this approach by setting the denominator degrees of freedom equal to infinity. This approach assumes that Σ is known or that our sample size is so large that the error in our estimator of Σ is negligible. A more conservative approach that is usually preferred for data analyses like these would be to use t-tests or F-tests with the denominator degrees of freedom equal to $n - \text{rank}(X) = 60 - 12 = 48$. Even this approach is too liberal for some tests. This approach is equivalent to assuming that $\Sigma = \sigma^2 V$, where V is known (or extremely well estimated) but σ^2 is unknown. However, for this data set, most of the p-values are either quite large or very small, so most of the conclusions below are not really in doubt even though the inference is only approximate.

The following are the code and results for testing drug A:

```
zeros=rep(0,12)
names(zeros)=names(coef(model.ar1))

#Mean for drug A Time=5m different from mean for drug A Time=0?
#Test whether T2=0.

CA5=zeros
CA5[4]=1
CA5
test(model.ar1,CA5,Inf)
$Fstat
[1] 47.0193
$pvalue
[1] 7.029155e-12


#Mean for drug A Time=10m different from mean for drug A Time=0?
#Test whether T3=0.
CA10=zeros
CA10[5]=1
CA10
test(model.ar1,CA10,Inf)
$Fstat
[1] 22.94588
$pvalue
[1] 1.666269e-06


#Mean for drug A Time=15m different from mean for drug A Time=0?
#Test whether T4=0.
CA15=zeros
CA15[6]=1
CA15
test(model.ar1,CA15,Inf)
$Fstat
[1] 0.5197333
$pvalue
[1] 0.4709555
```

From the above results we can see significant differences from baseline at time=5 and 10(minutes) for drug A.

The following are the code and results for testing drug B:

```
CB5=CA5
CB5[7]=1
CB5
test(model.ar1,CB5,Inf)
$Fstat
[1] 4.127894
```

```
$pvalue
[1] 0.0421818

CB10=CA10
CB10[9]=1
CB10
test(model.ar1,CB10,Inf)
$Fstat
[1] 3.184722
$pvalue
[1] 0.07432963

CB15=CA15
CB15[11]=1
CB15
test(model.ar1,CB15,Inf)
$Fstat
[1] 0.7763918
$pvalue
[1] 0.3782469
```

From the above results we can see that for drug B, there is a modestly significant difference between the mean at time=5(minutes) and baseline. Because that p-value is probably a little too liberal, the significance here is in doubt. There's no significant evidence of the difference for other time points.

The following are the code and results for testing drug C:

```
CC5=CA5
CC5[8]=1
CC5
test(model.ar1,CC5,Inf)
$Fstat
[1] 0.7901048
$pvalue
[1] 0.3740684

CC10=CA10
CC10[10]=1
CC10
test(model.ar1,CC10,Inf)
$Fstat
[1] 0.03528778
$pvalue
[1] 0.8509939

CC15=CA15
CC15[12]=1
CC15
test(model.ar1,CC15,Inf)
```

```
$Fstat
[1] 0.1604115
$pvalue
[1] 0.6887779
```

None of the differences are significant for drug C.

We can also conduct other inference that might be of interest. For instance, we can ask whether the change in mean heart rate from Time = 0 to Time=5 minutes is the same for both drugs A and B.

```
CIntB5=zeros
CIntB5[7]=1
CIntB5[7]
test(model.ar1,CIntB5,Inf)
$Fstat
[1] 11.64195
$pvalue
[1] 0.0006448089
```

We can conclude that the change in mean heart rate from Time = 0 to Time=5 minutes was different for drugs A and B. Since the coefficient of DT_{22} is negative, we know that drug A had a greater change than drug B.

In summary, it seems that for drug A, the mean heart rates were significantly elevated at 5 minutes and 10 minutes relative to baseline. After 15 minutes, the mean heart rate returned to a level not significantly different from that at baseline. The other drugs did not have strong effects on mean heart rate.

Figure 1: Observed Heart Rate Means
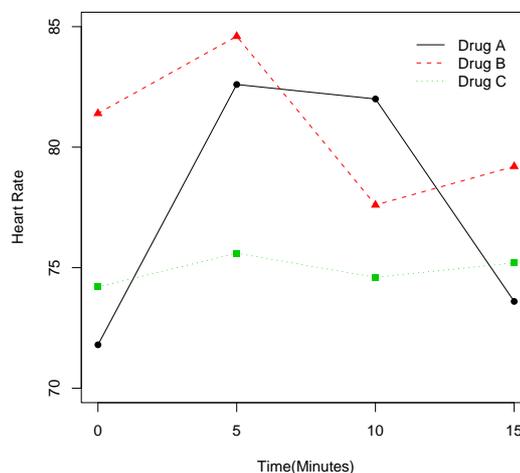
5. (a) If we invert both sides of the equation, we get

$$1/E(y) = (b_2 + x)/(b_1\,x) = 1/b_1 + (b_2/b_1)(1/x).$$

Thus, to get starting values, we might use a linear regression of $1/y$ on $1/x$. The estimate of the intercept could be inverted to get a starting value for b_1. Then, the estimate of the slope could be divided by the estimate

of the intercept to get a starting value for b_2. It's also not hard to come up with reasonable starting values by inspection. The starting values used below are 22 and 8.5 for b_1 and b_2, respectively. The R code to estimate the two parameters is as follows:

```
#read in data
read.table("http://www.public.iastate.edu/~dnett/S511/
+ EnzymeKinetics.txt",head=T)->enzyme
#attach data set
attach(enzyme)
> nls(velocity~b1*concentration/(b2+concentration),
+ start=c(b1=22,b2=8.5))->nlsout
> nlsout
Nonlinear regression model
  model:  velocity ~ b1 * concentration/(b2 + concentration)
   data:  parent.frame()
   b1     b2
28.14 12.57
 residual sum-of-squares: 4.302

Number of iterations to convergence: 5
Achieved convergence tolerance: 5.864e-07
```

So the estimates for b_1 and b_2 are $\hat{b}_1 = 28.14$ and $\hat{b}_2 = 12.57$ respectively. The scatter plot and the curve of expect velocity versus the concentration is shown in Figure 2. From Figure 2 we can see that the curve fits the data very well.

Table 1: Mean of Treatments

	Time 1	Time 2	Time 3	Time 4
Drug A	μ	$\mu + T_2$	$\mu + T_3$	$\mu + T_4$
Drug B	$\mu + D_2$	$\mu + D_2 + T_2 + (DT)_{22}$	$\mu + D_2 + T_3 + (DT)_{23}$	$\mu + D_2 + T_4 + (DT)_{24}$
Drug C	$\mu + D_3$	$\mu + D_3 + T_2 + (DT)_{32}$	$\mu + D_3 + T_3 + (DT)_{33}$	$\mu + D_3 + T_4 + (DT)_{34}$

Figure 2: Mean Velocity of Reaction
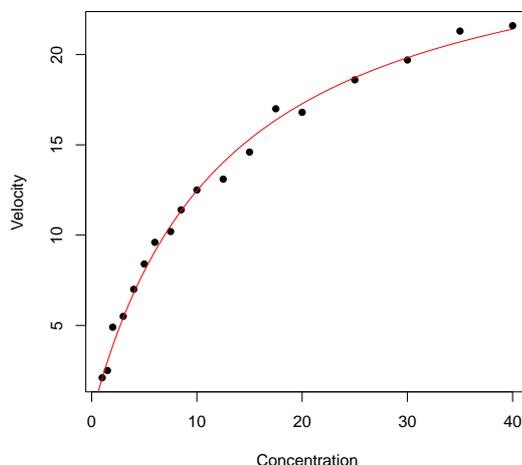
(b) Let $mean\ velocity = 15$ and solve the equation we get

$$x_{15} = \frac{15b_2}{b_1 - 15}.$$

Plugging in the estimate of b_1 and b_2 from part (a), we obtain the estimated concentration for x_{15} is $\hat{x}_{15} = 14.36$. The R code is as follows:

```
coef(nlsout)->bhat
> b1hat=bhat[1]
> b2hat=bhat[2]
> x15=deriv(y~15*b2/(b1-15),c("b1","b2"),function(b1,b2){})
> as.numeric(x15(b1hat,b2hat))->x15hat
> x15hat
[1] 14.35762
```

(c) An approximate 95% confidence interval is

$$\hat{x}_{15} \pm t_{df,\alpha/2}\sqrt{\left(\frac{\partial x_{15}}{\partial \boldsymbol{b}} \mid \boldsymbol{b} = \hat{\boldsymbol{b}}\right)' \left(\hat{\Sigma}_{\boldsymbol{b}} \mid \boldsymbol{b} = \hat{\boldsymbol{b}}\right)\left(\frac{\partial x_{15}}{\partial \boldsymbol{b}} \mid \boldsymbol{b} = \hat{\boldsymbol{b}}\right)}$$

```
> der=attr(x15(b1hat,b2hat),"gradient")
> x15hat+c(-1,1)*sqrt(der%*%vcov(nlsout)%*%t(der))*qt(1-0.05/2,16)
[1] 13.70911 15.00613
```

So the approximate 95% confidence interval is (13.71,15.01).