BIOS 6612 final exam

A study was conducted to determine the effect of an antagonist MDL 72222 on the change in blood pressure experienced by a sample of rabbits with increasing dosage of phenylbiguanide (PBG). For the study, each of five rabbits was exposed to increasing doses of phenylbiguanide after having either a placebo or the HD5-antagonist MDL 72222 administered. The variables in the data set are:

- `Rabbit`: a factor with levels 1 to 5 identifying the subject.
- `Treatment`: a factor with levels `'Placebo'` and `'MDL 72222'` indicating the treatment administered.
- `dose`: a numeric vector giving the dose of PBG administered.
- `deltaBP`: a numeric vector giving the outcome, change in blood pressure.

The data for `Rabbit=1` is printed below:

```
## Grouped Data: deltaBP ~ dose | Rabbit
##    deltaBP   dose Treatment Rabbit
## 1     0.50   6.25   Placebo      1
## 2     4.50  12.50   Placebo      1
## 3    10.00  25.00   Placebo      1
## 4    26.00  50.00   Placebo      1
## 5    37.00 100.00   Placebo      1
## 6    32.00 200.00   Placebo      1
## 31    1.25   6.25 MDL 72222      1
## 32    0.75  12.50 MDL 72222      1
## 33    4.00  25.00 MDL 72222      1
## 34    9.00  50.00 MDL 72222      1
## 35   25.00 100.00 MDL 72222      1
## 36   37.00 200.00 MDL 72222      1
```

Answer the following questions based on this data set. Output from R is given, but you are free to fit any models necessary yourself, in SAS or R, to help you answer the questions.

1. A linear mixed effects model was fitted to this data with a random intercept for `Rabbit`. This is Model 1; output appears below.

```
# Model 1
mod1 <- lme(deltaBP ~ dose*Treatment,
            random=~1|Rabbit,
            data=PBG)

summary(mod1)

## Linear mixed-effects model fit by REML
##  Data: PBG
```

```
##        AIC      BIC    logLik
##   400.9387 413.0908 -194.4694
##
## Random effects:
##  Formula: ~1 | Rabbit
##         (Intercept) Residual
## StdDev:     2.14037 5.735298
##
## Fixed effects: deltaBP ~ dose * Treatment
##                               Value Std.Error DF   t-value p-value
## (Intercept)                4.180348 1.7446163 52  2.396142  0.0202
## dose                       0.142903 0.0154724 52  9.235945  0.0000
## TreatmentMDL 72222        -4.288955 2.0627406 52 -2.079251  0.0425
## dose:TreatmentMDL 72222   -0.005959 0.0218813 52 -0.272322  0.7865
##  Correlation:
##                           (Intr) dose   TMDL72
## dose                      -0.582
## TreatmentMDL 72222        -0.591  0.492
## dose:TreatmentMDL 72222    0.412 -0.707 -0.696
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.2444284 -0.6080561 -0.2439825  0.3231946  2.7545985
##
## Number of Observations: 60
## Number of Groups: 5
```

```
anova(mod1)
```

```
##                numDF denDF   F-value p-value
## (Intercept)        1    52  85.93657  <.0001
## dose               1    52 163.56558  <.0001
## Treatment          1    52   9.98783  0.0026
## dose:Treatment     1    52   0.07416  0.7865
```

(a) Interpret the interaction between the `dose` and `Treatment` variables. Is this interaction significant? Provide a test statistic and p-value to support your conclusion. (**10 points**)

```
# solution:
# interaction represents the difference in slopes of the mean of deltaBP with
respect to dose between treatment and control conditions
# interaction not significant, F stat of 0.07416 (p=0.7865)
```

(b) Calculate the estimated ICC for this data. Interpret the estimate. (**10 points**)

```
# solution:
2.14037^2/(2.14037^2+ 5.735298^2)
```

```
## [1] 0.122247
```

```
# ICC measures the proportion of total variability due to between-subject
variance,
```

```
# so small value means this portion is relatively small;
# this also means that there is relatively little correlation within subjects
```

(c) Now consider a linear regression model for this data with the same fixed effects but no random effect.

```
mod0 <- lm(deltaBP ~ dose*Treatment,data=PBG)
anova(mod0)
```

```
## Analysis of Variance Table
##
## Response: deltaBP
##                Df Sum Sq Mean Sq  F value    Pr(>F)
## dose            1 5380.3  5380.3 146.1233 < 2.2e-16 ***
## Treatment       1  328.5   328.5   8.9227  0.004174 **
## dose:Treatment  1    2.4     2.4   0.0663  0.797819
## Residuals      56 2061.9    36.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compare this ANOVA table with the ANOVA table for Model 1 above; *ignore any differences in output format due to this model using `lm()` with no random effect and Model 1 using `lme()` including a random effect*. Why are they similar? (**4 points**)

```
# solution:
# similar results because ICC is low, so there isn't much gain in power using
the within-subject design here
```

2. Model 2 is fitted to the data, exchanging the linear effect for `dose` with a categorical one.

```
mod2 <- lme(deltaBP ~ as.factor(dose)*Treatment,
            random=~1|Rabbit,
            data=PBG)

summary(mod2)
```

```
## Linear mixed-effects model fit by REML
##  Data: PBG
##        AIC      BIC    logLik
##   320.0336 346.2304 -146.0168
##
## Random effects:
##  Formula: ~1 | Rabbit
##         (Intercept) Residual
## StdDev:     2.46726 3.848636
##
## Fixed effects: deltaBP ~ as.factor(dose) * Treatment
##                                         Value Std.Error DF   t-value
## (Intercept)                              1.00  2.044474 44  0.489123
## as.factor(dose)12.5                      1.35  2.434091 44  0.554622
## as.factor(dose)25                        4.60  2.434091 44  1.889822
```

```
## as.factor(dose)50                          16.40  2.434091 44  6.737627
## as.factor(dose)100                         26.80  2.434091 44 11.010269
## as.factor(dose)200                         26.20  2.434091 44 10.763770
## TreatmentMDL 72222                          0.68  2.434091 44  0.279365
## as.factor(dose)12.5:TreatmentMDL 72222     -1.34  3.442325 44 -0.389272
## as.factor(dose)25:TreatmentMDL 72222       -3.98  3.442325 44 -1.156195
## as.factor(dose)50:TreatmentMDL 72222      -13.88  3.442325 44 -4.032158
## as.factor(dose)100:TreatmentMDL 72222     -11.28  3.442325 44 -3.276855
## as.factor(dose)200:TreatmentMDL 72222      -1.68  3.442325 44 -0.488042
##                                           p-value
## (Intercept)                                0.6272
## as.factor(dose)12.5                        0.5820
## as.factor(dose)25                          0.0654
## as.factor(dose)50                          0.0000
## as.factor(dose)100                         0.0000
## as.factor(dose)200                         0.0000
## TreatmentMDL 72222                         0.7813
## as.factor(dose)12.5:TreatmentMDL 72222     0.6990
## as.factor(dose)25:TreatmentMDL 72222       0.2538
## as.factor(dose)50:TreatmentMDL 72222       0.0002
## as.factor(dose)100:TreatmentMDL 72222      0.0021
## as.factor(dose)200:TreatmentMDL 72222      0.6279
##  Correlation:
##                                           (Intr) a.()12 a.()25 a.()50 a.()10
## as.factor(dose)12.5                       -0.595
## as.factor(dose)25                         -0.595  0.500
## as.factor(dose)50                         -0.595  0.500  0.500
## as.factor(dose)100                        -0.595  0.500  0.500  0.500
## as.factor(dose)200                        -0.595  0.500  0.500  0.500  0.500
## TreatmentMDL 72222                        -0.595  0.500  0.500  0.500  0.500
## as.factor(dose)12.5:TreatmentMDL 72222     0.421 -0.707 -0.354 -0.354 -0.354
## as.factor(dose)25:TreatmentMDL 72222       0.421 -0.354 -0.707 -0.354 -0.354
## as.factor(dose)50:TreatmentMDL 72222       0.421 -0.354 -0.354 -0.707 -0.354
## as.factor(dose)100:TreatmentMDL 72222      0.421 -0.354 -0.354 -0.354 -0.707
## as.factor(dose)200:TreatmentMDL 72222      0.421 -0.354 -0.354 -0.354 -0.354
##                                           a.()20 TMDL72 a.()127 a.()257
## as.factor(dose)12.5
## as.factor(dose)25
## as.factor(dose)50
## as.factor(dose)100
## as.factor(dose)200
## TreatmentMDL 72222                         0.500
## as.factor(dose)12.5:TreatmentMDL 72222    -0.354 -0.707
## as.factor(dose)25:TreatmentMDL 72222      -0.354 -0.707  0.500
## as.factor(dose)50:TreatmentMDL 72222      -0.354 -0.707  0.500   0.500
## as.factor(dose)100:TreatmentMDL 72222     -0.354 -0.707  0.500   0.500
## as.factor(dose)200:TreatmentMDL 72222     -0.707 -0.707  0.500   0.500
##                                           a.()57 a.()107
## as.factor(dose)12.5
## as.factor(dose)25
```

```
## as.factor(dose)50
## as.factor(dose)100
## as.factor(dose)200
## TreatmentMDL 72222
## as.factor(dose)12.5:TreatmentMDL 72222
## as.factor(dose)25:TreatmentMDL 72222
## as.factor(dose)50:TreatmentMDL 72222
## as.factor(dose)100:TreatmentMDL 72222   0.500
## as.factor(dose)200:TreatmentMDL 72222   0.500  0.500
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -1.73841154 -0.41355540  0.06431518  0.40703912  2.39129789
##
## Number of Observations: 60
## Number of Groups: 5
```

```
anova(mod2)
```

```
##                            numDF denDF  F-value p-value
## (Intercept)                    1    44 85.94376  <.0001
## as.factor(dose)                5    44 81.30602  <.0001
## Treatment                      1    44 22.18039  <.0001
## as.factor(dose):Treatment      5    44  5.67019   4e-04
```

(a) Should `dose` be treated as categorical or continuous in this model? Justify your answer: include the results of a formal statistical test. The following additional output may be helpful: (**6 points**)

```
## refitting model(s) with ML (instead of REML)

## Data: PBG
## Models:
## mod1: deltaBP ~ dose * Treatment + (1 | Rabbit)
## mod2: deltaBP ~ as.factor(dose) * Treatment + (1 | Rabbit)
##      Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## mod1  6 392.58 405.15 -190.29   380.58
## mod2 14 355.51 384.83 -163.76   327.51 53.07      8  1.046e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# solution:
# models are nested (linear within 5th degree polynomial), so LRT works:
# test stat is 53.07 on 8 degrees of freedom, p<0.001
```

(b) Using Model 2, what is the average `deltaBP` value estimated for subjects at `Treatment='MDL 72222'` and `dose=50`? (**10 points**)

```
# solution:
# intercept
1.000 +
  # dose effect at 50
```

```
  16.400 +
  # treatment effect
  0.680 +
  # interaction
  -13.880
```

```
## [1] 4.2
```

```
summary(lmer(deltaBP ~ 0+as.factor(dose):Treatment+(1|Rabbit),data=PBG))$coef
```

```
##                                        Estimate Std. Error    t value
## as.factor(dose)6.25:TreatmentPlacebo       1.00   2.044474  0.4891233
## as.factor(dose)12.5:TreatmentPlacebo       2.35   2.044474  1.1494398
## as.factor(dose)25:TreatmentPlacebo         5.60   2.044474  2.7390905
## as.factor(dose)50:TreatmentPlacebo        17.40   2.044474  8.5107454
## as.factor(dose)100:TreatmentPlacebo       27.80   2.044474 13.5976278
## as.factor(dose)200:TreatmentPlacebo       27.20   2.044474 13.3041538
## as.factor(dose)6.25:TreatmentMDL 72222     1.68   2.044474  0.8217271
## as.factor(dose)12.5:TreatmentMDL 72222     1.69   2.044474  0.8266184
## as.factor(dose)25:TreatmentMDL 72222       2.30   2.044474  1.1249836
## as.factor(dose)50:TreatmentMDL 72222       4.20   2.044474  2.0543179
## as.factor(dose)100:TreatmentMDL 72222     17.20   2.044474  8.4129208
## as.factor(dose)200:TreatmentMDL 72222     26.20   2.044474 12.8150305
```

(c) Using results from Model 2, provide an interpretation of the effect of `dose`, `Treatment`, and their interaction on mean glucose levels; give p-values to support your conclusions. (**14 points**)

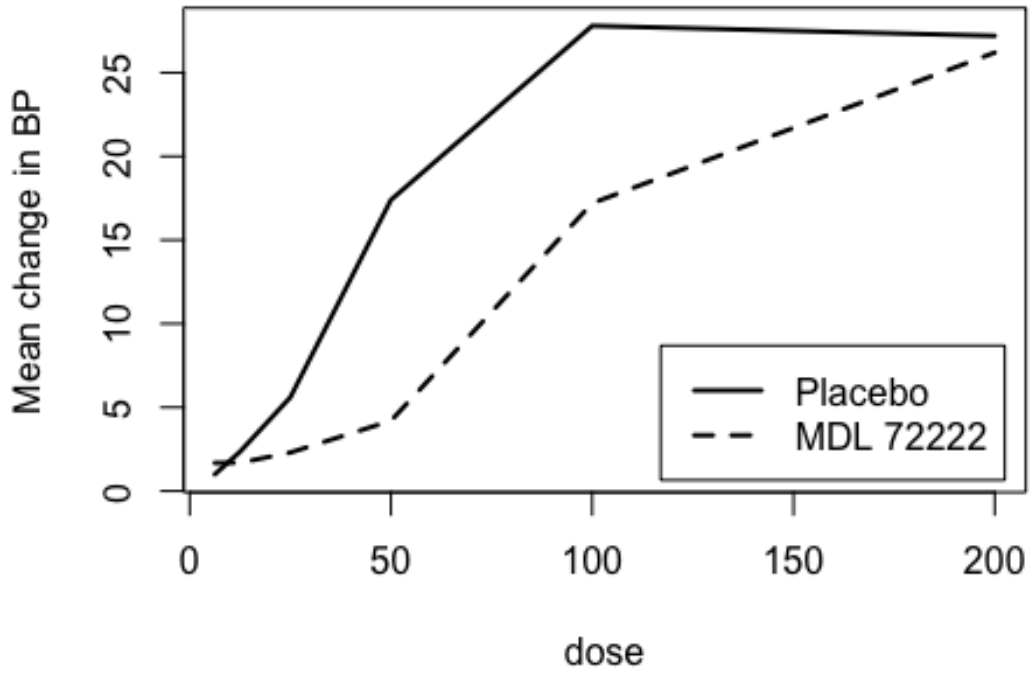
```
mod2 <- lmer(deltaBP ~ as.factor(dose)*Treatment+(1|Rabbit),
             data=PBG)
# solution:
# interaction is significant (p<0.001), so we can't interpret the effects of
dose and treatment separately
# the effect of the treatment MDL 72222 is to attenuate the increase in
deltaBP caused by administration of PBG
# in the control condition, deltaBP increases with increasing dose
nonlinearly, leveling off around dose=100
# in the treatment condition, deltaBP increases with dose but by a lesser
amount:
L1 <- rbind(c(0,1,0,0,0,0,0,1,0,0,0,0),
            c(0,0,1,0,0,0,0,0,1,0,0,0),
            c(0,0,0,1,0,0,0,0,0,1,0,0),
            c(0,0,0,0,1,0,0,0,0,0,1,0),
            c(0,0,0,0,0,1,0,0,0,0,0,1))
L1 %*% summary(mod2)$coef[,1]
```

```
##       [,1]
## [1,]  0.01
## [2,]  0.62
## [3,]  2.52
```

```
## [4,] 15.52
## [5,] 24.52

# shows the increase in deltaBP with increasing dose with the treatment
condition
# the effect of MDL 72222 seems to be most pronounced at doses 50 and 100
(p<0.01 for both)
# (compare with main effect estimates for dose in interaction model)
```

3. A plot of mean change in BP with dose for each `Treatment` appears below.



(a) If we think about each point on this plot as a fitted value from some regression model for the data, then what variables would be included in this model, and what form would each take (e.g., categorical, continuous)? (**8 points**)

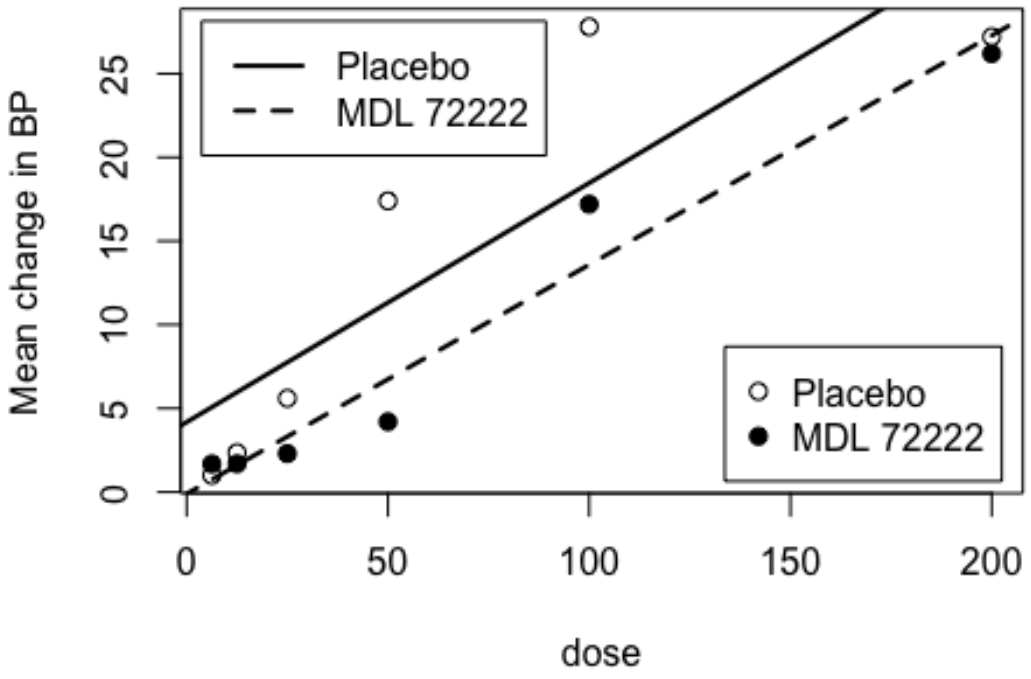
```
# solution:
# this plot corresponds to treating dose as categorical (because there are
separate line segments between each dose)
# and including main effects and interaction effects for dose and treatment
(because there is a line for each treatment)
```

(b) Estimated fixed effects parameters from Model 1 appear in the table below.

```
##                              Value Std.Error DF t-value p-value
## (Intercept)                 4.1803    1.7446 52  2.3961  0.0202
## dose                        0.1429    0.0155 52  9.2359  0.0000
```

```
## TreatmentMDL 72222       -4.2890    2.0627 52 -2.0793  0.0425
## dose:TreatmentMDL 72222 -0.0060    0.0219 52 -0.2723  0.7865
```

Draw the fitted mean curve for `Treatment='MDL 72222'` corresponding to Model 1 on the plot below, which shows the mean curve for `Treatment='Placebo'` for this model, along with the mean change in BP at each time for each treatment group. (This doesn't need to be exact, just show approximately where the curve should be.) (**12 points**)



(c) Another variable in the data set, `Run`, indexes repeated measurements within each rabbit at each treatment condition. A researcher has erroneously used this as the subject variable and obtained the following table from a RMANOVA analysis.

```
##
## Error: as.factor(Run)
##           Df Sum Sq Mean Sq F value Pr(>F)
## Treatment  1  328.5   328.5    5.87 0.0417 *
## Residuals  8  447.7    56.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##                            Df Sum Sq Mean Sq F value   Pr(>F)
## as.factor(dose)             5   6022  1204.3  86.727  < 2e-16 ***
## as.factor(dose):Treatment  5    420    84.0   6.048 0.000294 ***
```

```
## Residuals                 40    555    13.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Use this to calculate the estimated random intercept variance for this model. **Do not** fit this model to answer the question. (**10 points**)

```
# solution:
# summary()
# there are
length(unique(PBG$dose))
```

```
## [1] 6
```

```
# measurements per run, so
MS_S <- 56.0
MS_R <- 13.9
sigma.e <- MS_R^.5
sigma.b <- sqrt((MS_S-MS_R)/length(unique(PBG$dose)))
getVarCov(lme(deltaBP ~ as.factor(dose)*Treatment,random=~1|Run,
              data=PBG))
```

```
## Random effects variance covariance matrix
##             (Intercept)
## (Intercept)      7.0132
##   Standard Deviations: 2.6482
```

```
sigma.b^2
```

```
## [1] 7.016667
```

4. Now consider a polynomial trend model for the effect of `dose`. The table below gives values of AIC and BIC for orthogonal polynomial models for the effect of `dose` on mean glucose levels; the models also include `Treatment` and the interaction between `dose` and `Treatment`.

```
##    Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## .L     1  6 392.5815 405.1475 -190.2907
## .Q     2  8 353.7388 370.4936 -168.8694 1 vs 2 42.84264  <.0001
## .C     3 10 350.4396 371.3831 -165.2198 2 vs 3  7.29917  0.0260
## ^4     4 12 351.6234 376.7555 -163.8117 3 vs 4  2.81623  0.2446
## ^5     5 14 355.5118 384.8327 -163.7559 4 vs 5  0.11157  0.9457
```

(a) Which model is best according to AIC? What about BIC? Why are these answers different? Note that R bases the BIC penalty on total number of observations rather than number of independent subjects. (**8 points**)

```
# best for AIC
rownames(poly.tab)[which.min(poly.tab$AIC)]
```

```
## [1] ".C"
```

```r
# best for BIC
rownames(poly.tab)[which.min(poly.tab$BIC)]
```

```
## [1] ".Q"
```

```r
# different because BIC penalizes more heavily than AIC for excessive
parameters
```

(b) Assuming the fitting methods used are correct, were these models fitted using ML or REML? Justify your answer. (**8 points**)

```r
# solution:
# we need to use ML to compare models with different fixed effects, so these
should be fitted using ML
```