BIOS 6612 Final Exam - Spring 2021

You may refer to your course notes, homework assignments, labs, textbooks, etc. You may use R to check any work or answers, but you **must show all work.**

Answers without work will not receive partial credit. You will enter your responses via Canvas as a quiz that will be in the same order as the questions in the final. **To allow for time to transfer your answers, you will have until 4:30 pm to submit your answers.** Then you will have until 7:00 pm to upload your work (e.g., via images, Rmd files, etc.) to Canvas to be able to receive partial credit for grading.

Read the UCD Honor Code statement below. You must indicate your agreement with this statement when you submit your answers.

I understand that my participation in this examination and in all academic and professional activities as a UCD student is bound by the provisions of the UCD Honor Code. I understand that work on this exam is to be done independently and according to the guidelines provided.

On my honor, I have neither given nor received aid on this examination.

Problem 1 (20 points). Answer the following questions. Circle TRUE or FALSE: (4 points each).

1a. TRUE FALSE When outcome data is missing at random (MAR), GEEs will provide unbiased inference.

1b. TRUE FALSE Wald tests provide appropriate inference for GEE models.

1c. TRUE FALSE GLMMs can be used to make inferences about individuals in a population.

1d. TRUE FALSE In a generalized linear mixed effects model with a random slope and a random intercept, we assume that the random slope and the random intercept are independent.

1e. TRUE FALSE The exponentiated coefficients in Poisson regression can be interpreted as rate ratios.

Using the following data to answer Questions 2, 3 and 4.

A study was conducted to determine the effect of alcohol on blood glucose. For the study, blood glucose levels were measured at 10 time points for 7 volunteers given alcohol at time 0. The same experiment was repeated on another occasion with the same subjects but with a dietary additive administered to all subjects. There was no dropout or missing data for this study. The variables in the data set are:

- `Subject`: a factor with levels 1 to 7 identifying the subject whose glucose level is measured.
- `trt`: a factor with levels `control` and `additive` indicating whether the subject received treatment or control at that time point.
- `Time:` : a numeric vector giving the time since alcohol ingestion (in minutes/10).
- `glucose`: a numeric vector giving the blood glucose level (in mg/dl).

The data for Subject=1 is copied below:

```
##           trt    6    8   10   12   15   18   21   24   27   30
## 1   control   6.3  4.3  3.0  2.0  4.5  3.8  3.2  2.6  2.6  2.6
## 2  additive   5.8  4.5  3.6  3.3  3.2  3.3  3.0  3.0  3.2  3.1
```

Columns indicate time of measurement (in units of minutes/10); the two rows correspond to the two treatments. A spaghetti plot of the data, faceted by treatment, is given below.

```
glucose %>%
  ggplot(aes(Time, glucose)) +
  geom_line(alpha = 0.8, aes(group = Subject)) +
  geom_smooth(method = "loess", se = FALSE) +
  facet_wrap(~trt)
```

```
## `geom_smooth()` using formula 'y ~ x'
```
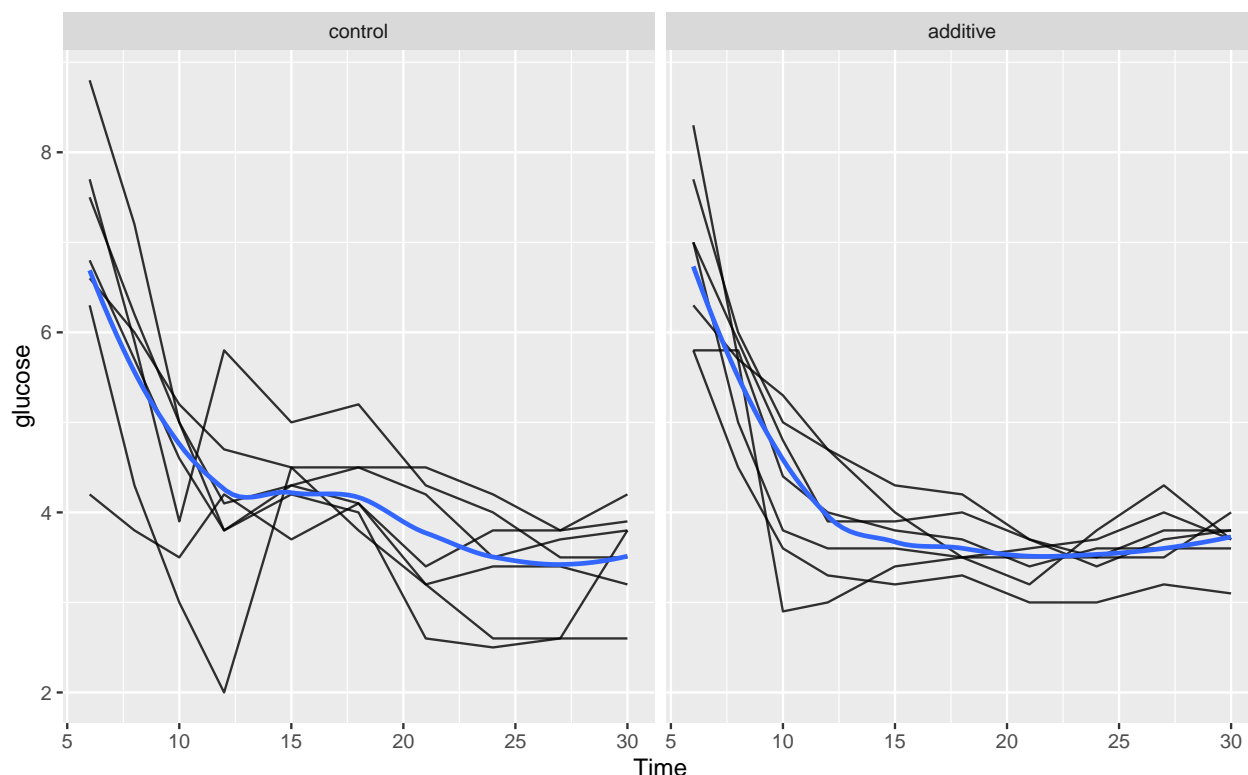
Problem 2 (30 points).

A linear mixed effects model was fitted to this data with a random intercept for `Subject` and fixed effects for `Time`, `trt`, and the interaction between `Time` and `trt`. This is Model 1; output appears in attached document.

```
model1 = lmer(glucose ~ (1 | Subject) + Time*trt,
              data = glucose)

summary(model1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: glucose ~ (1 | Subject) + Time * trt
##    Data: glucose
##
## REML criterion at convergence: 391.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6140 -0.6184 -0.0345  0.5053  3.5801
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 0.1049   0.3239
##  Residual             0.8167   0.9037
## Number of obs: 140, groups:  Subject, 7
##
## Fixed effects:
##                   Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)        6.19087    0.28660   66.68610  21.601  < 2e-16 ***
## Time              -0.10799    0.01377 130.00000  -7.839 1.43e-12 ***
## trtadditive       -0.32766    0.36647 130.00000  -0.894    0.373
## Time:trtadditive   0.01223    0.01948 130.00000   0.628    0.531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Time   trtddt
## Time        -0.822
## trtadditive -0.639  0.643
## Tim:trtddtv  0.581 -0.707 -0.909
```

```
anova(model1)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##          Sum Sq Mean Sq NumDF DenDF  F value Pr(>F)
## Time     89.342  89.342     1   130 109.3917 <2e-16 ***
## trt       0.653   0.653     1   130   0.7994 0.3729
## Time:trt  0.322   0.322     1   130   0.3940 0.5313
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2a. Calculate the estimated ICC for this data.

2b. Interpret your ICC estimate.

2c. Interpret the interaction between the `Time` and `trt` variables. Is this interaction significant? Provide a test statistic and p-value to support your conclusion.

2d Based on trends observed in the spaghetti plot, what other term(s) might we consider adding to our model?