Case Study 6

Golden Girls

Nitrogen transport from agricultural fields in the upper Midwest down the Mississippi River into the Gulf of Mexico can lead to severe environmental effects if there is an excess of nitrogen. This experiment examined whether streams simply serve as pipes transporting nitrogen downstream or whether biological processes in the streams retain or even all of the nitrogen instead of moving it down stream. This study involved 8 locations that were scattered throughout the US and Puerto Rico to be representative of similar locations. At each location nine streams were studied: three were undisturbed, 3 were agricultural, and 3 were urban. The ammonium uptake distance was measured on each stream for a total of 72 observations. Large uptake distance means that the stream is acting like a pipe and moving nitrogen downstream while a small value means the stream is retaining nitrogen.

```
#read in dataset
streams<-read_csv("streams.csv")
#convert to factor
streams <- streams %>%
  mutate(location = factor(location),
         type = factor(type))
```

1. [4pts] The researchers are interested in comparing the three types of streams, accounting for the differences among the eight locations, and the possibility of inconsistent treatment effects (type of stream) across the locations. Based on this information, 1) identify if a factor is fixed or random and 2) identify if the factors are nested or crossed. If nested, describe the nesting structure. If crossed identify if the interaction is fixed or random.

Answer: Location is a random effect and type of stream is fixed. Each location and type of stream combination is used so location and type of stream is crossed. The interaction is random.

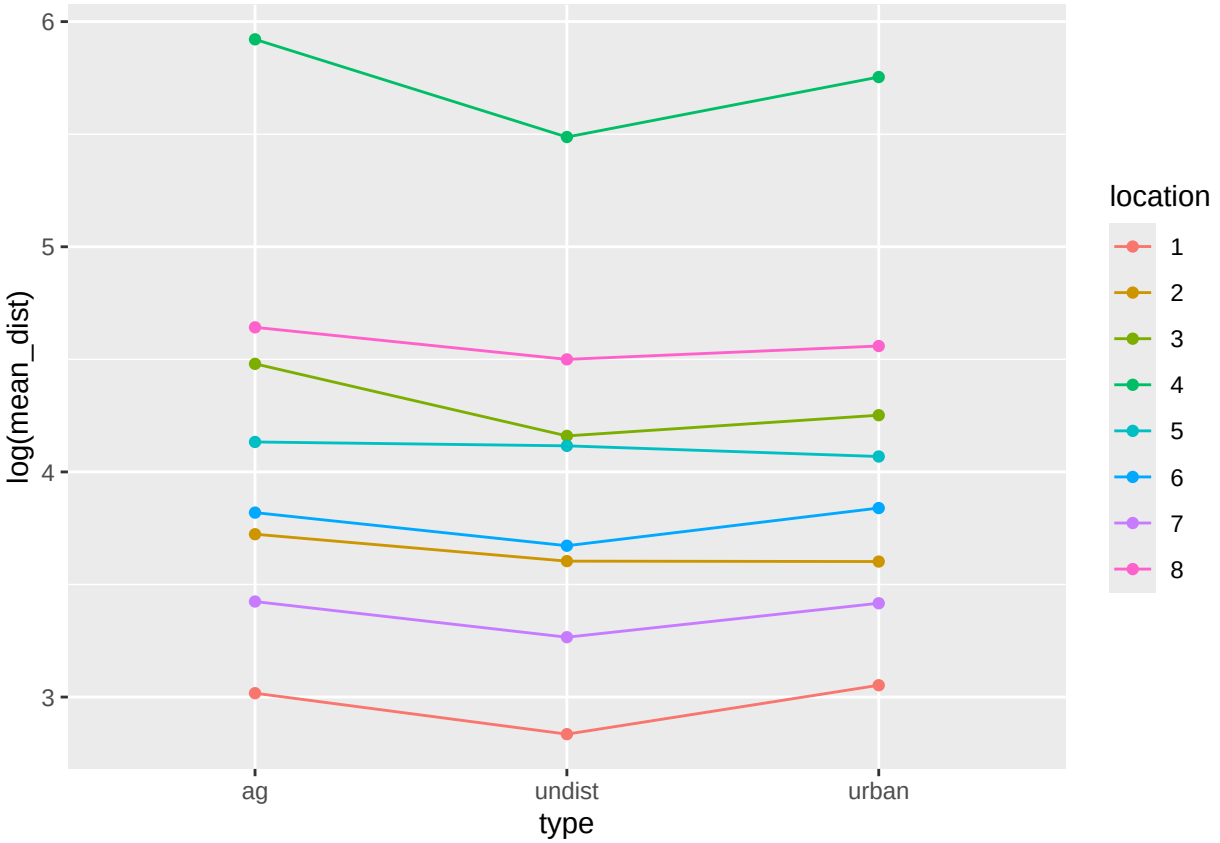
For this analysis use log(dist) as the response!

2. [3pts] Researchers are concerned that the effects of stream type on (log) nitrogen uptake distance may differ across locations. Create the appropriate plot to examine this and comment on whether this seems to be the case and why or why not.

```
streams_sum <- streams %>%
  group_by(location, type) %>%
  summarize(mean_dist = mean(dist))
```

```
## `summarise()` has grouped output by 'location'. You can override using the
## `.groups` argument.
```

```
ggplot(data=streams_sum, aes(x=type, y=log(mean_dist)))+
  geom_line(aes(group=location, color=location))+
  geom_point(aes(color=location))
```



Answer: There is variation in the (log) nitrogen levels between locations and differences in (log) nitrogen levels for the different types of stream. Effects of stream type on (log) nitrogen uptake distance do not differ across locations.

3. [3pts] Fit the model of interest (note that the assumptions for this analysis are reasonably satisfied when using log nitrogen uptake distance). Using the appropriate output, explain if there is evidence that the effect of stream type differs by location.

```
streams_int <- lmer(log(dist)~type+(1|location)+(1|location:type), data=streams, REML=T)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
ranova(streams_int)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## log(dist) ~ type + (1 | location) + (1 | location:type)
##                      npar   logLik    AIC    LRT Df Pr(>Chisq)
## <none>                  6   3.1815  5.637
## (1 | location)          5 -25.1389 60.278 56.641  1  5.231e-14 ***
## (1 | location:type)     5   3.1815  3.637  0.000  1          1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: With a p-value of 1 ($\chi^2 = 0$) there is no evidence that location and type interact.

4. [2pts] For better or worse you decide to not include an interaction term and use the additive model. Fit this model. What is the correlation between two streams at the same location?

```
streams_reml <- lmer(log(dist)~type+(1|location), data=streams, REML=T)
summary(streams_reml)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: log(dist) ~ type + (1 | location)
##    Data: streams
##
## REML criterion at convergence: -6.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.53933 -0.73348  0.06301  0.53603  2.38606
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  location (Intercept) 0.70765  0.8412
##  Residual             0.02667  0.1633
## Number of obs: 72, groups:  location, 8
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  4.14077    0.29928  7.11720  13.836  2.1e-06 ***
## typeundist  -0.19331    0.04714 62.00000  -4.100 0.000122 ***
## typeurban   -0.08990    0.04714 62.00000  -1.907 0.061169 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) typnds
## typeundist -0.079
## typeurban  -0.079  0.500
```

Answer: The correlation between two streams at the same location is 0.7076/(0.7076+0.0267) = 0.963.

5. [4pts] Somewhat surprisingly there is a treatment effect (F=8.4204, df=2 and 62, p-value=0.0006). Using the appropriate output, summarize the effects of stream type on log nitrogen uptake distance.

```
stream_means <- emmeans(streams_reml, ~type)
confint(pairs(stream_means, adjust="tukey"))
```

```
##  contrast       estimate     SE df lower.CL upper.CL
##  ag - undist      0.1933 0.0471 62   0.0801  0.30651
##  ag - urban       0.0899 0.0471 62  -0.0233  0.20310
##  undist - urban  -0.1034 0.0471 62  -0.2166  0.00979
##
```

```
## Degrees-of-freedom method: kenward-roger
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
```

Answer: On average, log nitrogen uptake distance for agricultural streams is significantly higher than undisturbed streams. Agricultural and undisturbed streams are similar to urban streams.

Phlebitis is the inflammation of a blood vein and is a potential complication during the intravenous administration of some drugs. This study was conducted to compare three affects of three different intravenous treatments (1=drug in a solution designed to carry the drug, 2=carrier solution only, and 3=saline solution). In the study 15 test animals were randomly assigned to one of the three treatments with five animals per treatment.

The treatments were administered in one ear of the animal and the difference in temperature between the treated and untreated ear was measured at the beginning of the study, and 30, 60, and 90 minutes after the treatment was administered. It is believed that increased temperature is an early sign of phlebitis.

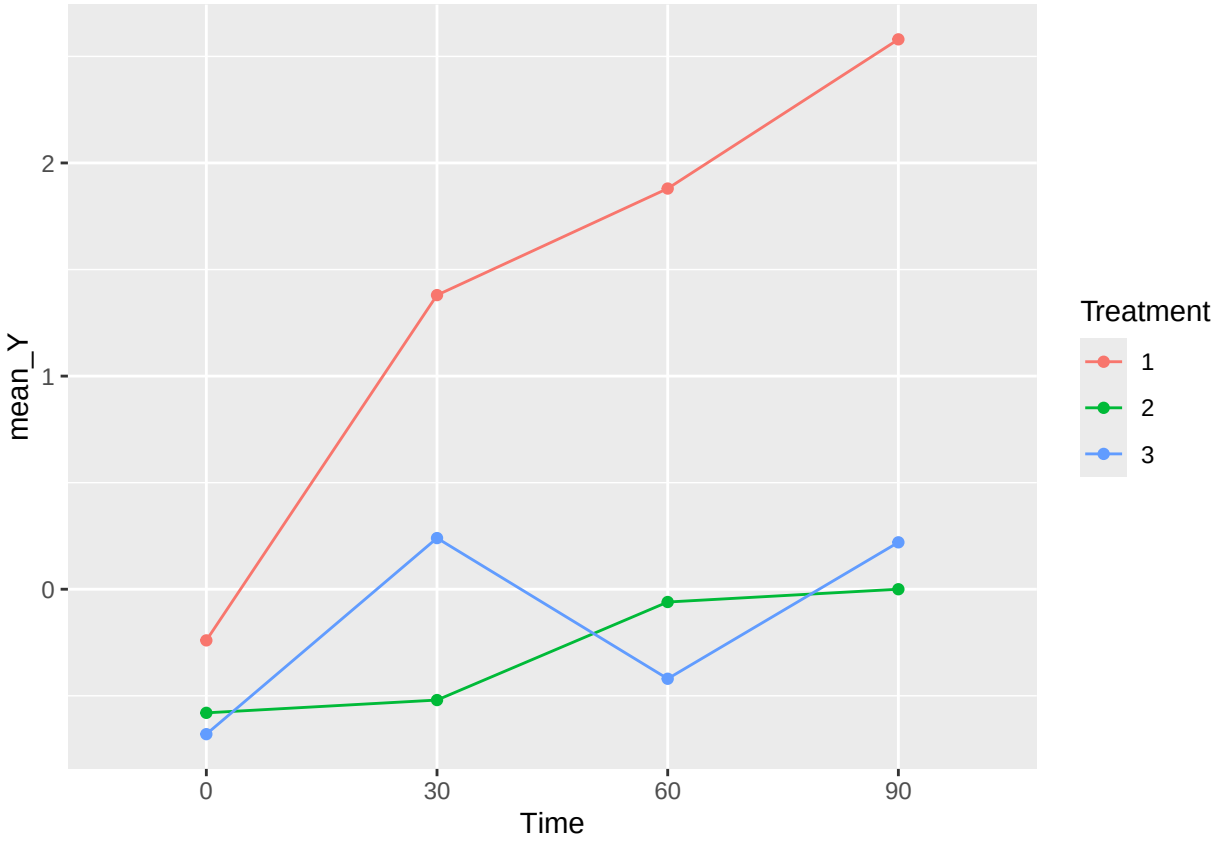
```
#load data
phleb <- read_csv("phlebitis.csv")
#convert to factor as needed
phleb <- phleb %>%
  mutate(Animal = factor(Animal),
         Treatment = factor(Treatment),
         Time = factor(Time))
```

6. [3pts] Obtain a plot of average difference in temperatures for each treatment over time. Explain if the effect of treatment over time differs by treatment.

```
means_phleb<- phleb %>%
  group_by(Time, Treatment) %>%
  summarize(mean_Y = mean(Y))
```

```
## `summarise()` has grouped output by 'Time'. You can override using the
## `.groups` argument.
```

```
ggplot(data=means_phleb, aes(x=Time, y=mean_Y))+
  geom_line(aes(group=Treatment,color=Treatment))+
  geom_point(aes(color=Treatment))
```



Answer: The interaction plot suggest that effect of treatment over time differs by treatment as we see the lines are not roughly parallel for the treatments.

7. [2pts] Let's proceed assuming that all assumptions are reasonably met (feel free to check them–they are!). Using the parameter estimates, explain if there is more variability between or within animals.

```
phleb_rm <- lmer(Y~Treatment*Time+(1|Animal), data=phleb)
summary(phleb_rm)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Y ~ Treatment * Time + (1 | Animal)
##    Data: phleb
##
## REML criterion at convergence: 134.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0120 -0.4932  0.1116  0.5573  1.6701
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Animal   (Intercept) 0.08336  0.2887
##  Residual             0.57831  0.7605
## Number of obs: 60, groups:  Animal, 15
```

```
##
## Fixed effects:
##                    Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)         -0.2400     0.3638 45.8182  -0.660  0.51272
## Treatment2          -0.3400     0.5145 45.8182  -0.661  0.51199
## Treatment3          -0.4400     0.5145 45.8182  -0.855  0.39685
## Time30               1.6200     0.4810 36.0000   3.368  0.00181 **
## Time60               2.1200     0.4810 36.0000   4.408 9.02e-05 ***
## Time90               2.8200     0.4810 36.0000   5.863 1.06e-06 ***
## Treatment2:Time30   -1.5600     0.6802 36.0000  -2.294  0.02776 *
## Treatment3:Time30   -0.7000     0.6802 36.0000  -1.029  0.31028
## Treatment2:Time60   -1.6000     0.6802 36.0000  -2.352  0.02424 *
## Treatment3:Time60   -1.8600     0.6802 36.0000  -2.735  0.00963 **
## Treatment2:Time90   -2.2400     0.6802 36.0000  -3.293  0.00223 **
## Treatment3:Time90   -1.9200     0.6802 36.0000  -2.823  0.00771 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Trtmn2 Trtmn3 Time30 Time60 Time90 T2:T30 T3:T30 T2:T60
## Treatment2  -0.707
## Treatment3  -0.707  0.500
## Time30      -0.661  0.467  0.467
## Time60      -0.661  0.467  0.467  0.500
## Time90      -0.661  0.467  0.467  0.500  0.500
## Trtmnt2:T30  0.467 -0.661 -0.331 -0.707 -0.354 -0.354
## Trtmnt3:T30  0.467 -0.331 -0.661 -0.707 -0.354 -0.354  0.500
## Trtmnt2:T60  0.467 -0.661 -0.331 -0.354 -0.707 -0.354  0.500  0.250
## Trtmnt3:T60  0.467 -0.331 -0.661 -0.354 -0.707 -0.354  0.250  0.500  0.500
## Trtmnt2:T90  0.467 -0.661 -0.331 -0.354 -0.354 -0.707  0.500  0.250  0.500
## Trtmnt3:T90  0.467 -0.331 -0.661 -0.354 -0.354 -0.707  0.250  0.500  0.250
##             T3:T60 T2:T90
## Treatment2
## Treatment3
## Time30
## Time60
## Time90
## Trtmnt2:T30
## Trtmnt3:T30
## Trtmnt2:T60
## Trtmnt3:T60
## Trtmnt2:T90  0.250
## Trtmnt3:T90  0.500  0.500
```

Answer: There is more variability within animals compared to between animal variability (σ^2 = 0.578 vs σ_A^2 = 0.083)

8. [3pts] Is there evidence to suggest that the effects of the three treatments over time differ? Explain.

```
anova(phleb_rm)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##                Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
```

```
## Treatment      22.439 11.2196     2    12 19.4008 0.0001737 ***
## Time           16.083  5.3611     3    36  9.2704 0.0001117 ***
## Treatment:Time 10.063  1.6771     6    36  2.9000 0.0207002 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer:With a p-value of 0.0210 (F=2.9, df = 6 and 36) there is moderate evidence that the effects of the three treatments differ over time.

9. [4pts] Specifically of interest is identifying the treatment(s) that minimize the potential of phlebitis. Using the appropriate tools, identify the treatment(s) you would recommend and explain why.

```
phleb_means_time <- emmeans(phleb_rm, ~Treatment|Time)
confint(pairs(phleb_means_time, adjust="tukey"))
```

```
## Time = 0:
##  contrast                estimate    SE   df lower.CL upper.CL
##  Treatment1 - Treatment2     0.34 0.514 45.8   -0.906    1.586
##  Treatment1 - Treatment3     0.44 0.514 45.8   -0.806    1.686
##  Treatment2 - Treatment3     0.10 0.514 45.8   -1.146    1.346
##
## Time = 30:
##  contrast                estimate    SE   df lower.CL upper.CL
##  Treatment1 - Treatment2     1.90 0.514 45.8    0.654    3.146
##  Treatment1 - Treatment3     1.14 0.514 45.8   -0.106    2.386
##  Treatment2 - Treatment3    -0.76 0.514 45.8   -2.006    0.486
##
## Time = 60:
##  contrast                estimate    SE   df lower.CL upper.CL
##  Treatment1 - Treatment2     1.94 0.514 45.8    0.694    3.186
##  Treatment1 - Treatment3     2.30 0.514 45.8    1.054    3.546
##  Treatment2 - Treatment3     0.36 0.514 45.8   -0.886    1.606
##
## Time = 90:
##  contrast                estimate    SE   df lower.CL upper.CL
##  Treatment1 - Treatment2     2.58 0.514 45.8    1.334    3.826
##  Treatment1 - Treatment3     2.36 0.514 45.8    1.114    3.606
##  Treatment2 - Treatment3    -0.22 0.514 45.8   -1.466    1.026
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
```

Answer: Overall treatments 2 and 3 would be recommended based on this analysis. On average for time zero there is no significant difference in effects for the three treatments. For time 30 the effects of treatment 1 is higher than treatment 2 and treatment 3 is similar to treatments 1 and 2. For time 60 treatment 1 is higher than both treatment 2 and 3 with treatments 2 and 3 being similar. For time 90 similarly to time 60 treatment 1 is higher than treatment 2 and 3 with treatment 2 and 3 being similar.