Cats Example: Two-Factor Repeated Measures with Baseline as Covariate

A study was done to compare 2 Treatments (A and C) for diabetes in cats. 6 Cats were randomly assigned to each treatment. Hence cat is nested within treatment. Or equivalently, treatment is the between-subjects factor. Blood sugar (Y) measurements were taken at 4 equally spaced Times (0, 1, 2, 3). The first time point (Time=0) was pre-treatment and serves as a baseline.

We consider several approaches:

1. Basic Repeated measures
2. Repeated measures with Time0 (baseline) as covariate
3. Repeated measures on Difference values

Notes:

1. We start by creating unique IDs for each of the cats. Originally labelled as 1-6 for Trt A and 1-6 for Trt C.
2. Different covariance structures were NOT investigated, but certainly could be.

```
library(tidyverse)
library(lme4)
library(lmerTest)
library(pbkrtest)
library(emmeans)
CatsData <- read.csv("C:/hess/STAT512/RNotes/Random3/R3_Cats.csv")
str(CatsData)
```

```
## 'data.frame':    48 obs. of  4 variables:
##  $ Trt : Factor w/ 2 levels "A","C": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Cat : int  1 1 1 1 2 2 2 2 3 3 ...
##  $ Time: int  0 1 2 3 0 1 2 3 0 1 ...
##  $ Y   : num  70.9 64.9 61.7 56.6 70.4 ...
```

```
#Important: Need to define things as.factor!!!
CatsData$Cat <- as.factor(CatsData$Cat)
CatsData$Time <- as.factor(CatsData$Time)
#Original Cat IDs are NOT Unique
with(table(Cat, Trt), data = CatsData)
```

```
##    Trt
## Cat A C
##   1 4 4
##   2 4 4
##   3 4 4
##   4 4 4
##   5 4 4
##   6 4 4
```

```
#Create unique IDs
CatsData$newID <- paste(CatsData$Trt, CatsData$Cat, sep = "")
with(table(newID, Trt), data = CatsData)
```

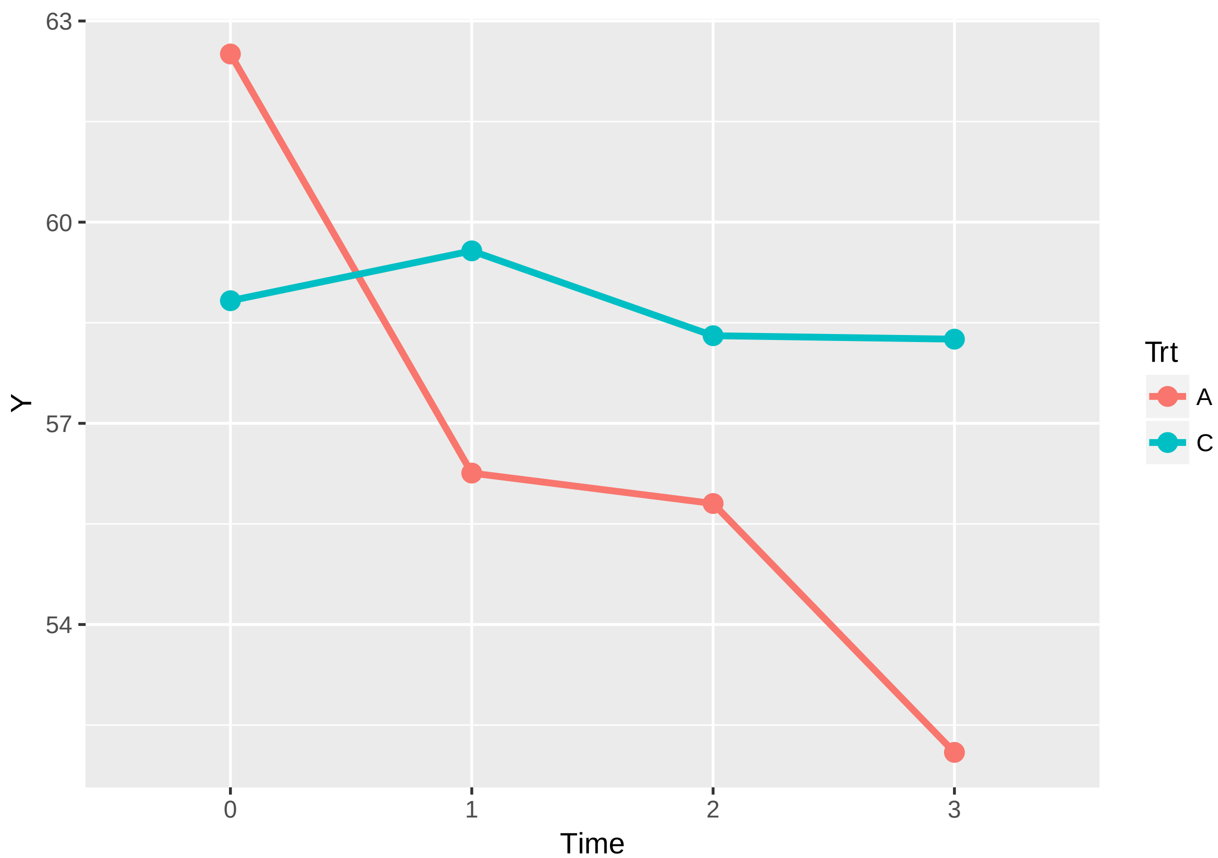
```
##      Trt
## newID A C
##    A1 4 0
```

```
##     A2 4 0
##     A3 4 0
##     A4 4 0
##     A5 4 0
##     A6 4 0
##     C1 0 4
##     C2 0 4
##     C3 0 4
##     C4 0 4
##     C5 0 4
##     C6 0 4
```

```r
head(CatsData)
```

```
##   Trt Cat Time        Y newID
## 1   A   1    0 70.89283    A1
## 2   A   1    1 64.92599    A1
## 3   A   1    2 61.68156    A1
## 4   A   1    3 56.56081    A1
## 5   A   2    0 70.41208    A2
## 6   A   2    1 67.64846    A2
```

```r
SumStats <- aggregate(Y ~ Trt + Time, FUN = mean, data = CatsData)
qplot(x = Time, y = Y, colour = Trt, group = Trt, data = SumStats) +  geom_line(size=1.2) +
  geom_point(size=3)
```



Model1A: Basic Model (non-unique IDS)

Since (1) cats are nested within trt and (2) cats were not given unique IDS, we need to explicitly specify the nesting in the model.

```
Model1A <- lmer(Y ~ Trt*Time + (1|Trt:Cat), data = CatsData)
Model1A
```

```
## Linear mixed model fit by REML ['merModLmerTest']
## Formula: Y ~ Trt * Time + (1 | Trt:Cat)
##    Data: CatsData
## REML criterion at convergence: 249.0432
## Random effects:
##  Groups   Name        Std.Dev.
##  Trt:Cat  (Intercept) 7.084
##  Residual             3.091
## Number of obs: 48, groups:  Trt:Cat, 12
## Fixed Effects:
## (Intercept)         TrtC        Time1        Time2        Time3
##      62.508       -3.679       -6.250       -6.704      -10.415
##  TrtC:Time1   TrtC:Time2   TrtC:Time3
##       6.993        6.182        9.842
```

```
anova(Model1A, ddf = "Kenward-Roger")
```

```
## Analysis of Variance Table of type III  with  Kenward-Roger
## approximation for degrees of freedom
##          Sum Sq Mean Sq NumDF DenDF F.value   Pr(>F)
## Trt       2.347   2.347     1    10  0.2457 0.630853
## Time    187.831  62.610     3    30  6.5532 0.001535 **
## Trt:Time 154.615  51.538     3    30  5.3943 0.004328 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model1B: Basic Model (unique IDS)

Now we rerun the basic model. Since we are now working with unique IDS, the nesting does not need to be explicitly specified in the model. The results are the same as Model1A.

```
Model1B <- lmer(Y ~ Trt*Time + (1|newID), data = CatsData)
summary(Model1B)
```

```
## Linear mixed model fit by REML t-tests use Satterthwaite approximations
##   to degrees of freedom [lmerMod]
## Formula: Y ~ Trt * Time + (1 | newID)
##    Data: CatsData
##
## REML criterion at convergence: 249
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9770 -0.6500  0.0400  0.6279  1.6428
##
## Random effects:
```

```
## Groups   Name        Variance Std.Dev.
## newID    (Intercept) 50.182   7.084
## Residual              9.554   3.091
## Number of obs: 48, groups:  newID, 12
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   62.508      3.155  12.832  19.810 5.30e-11 ***
## TrtC          -3.679      4.462  12.832  -0.825 0.424685
## Time1         -6.250      1.785  30.000  -3.502 0.001468 **
## Time2         -6.704      1.785  30.000  -3.756 0.000742 ***
## Time3        -10.415      1.785  30.000  -5.836 2.21e-06 ***
## TrtC:Time1     6.993      2.524  30.000   2.771 0.009503 **
## TrtC:Time2     6.182      2.524  30.000   2.449 0.020364 *
## TrtC:Time3     9.842      2.524  30.000   3.900 0.000503 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) TrtC   Time1  Time2  Time3  TrC:T1 TrC:T2
## TrtC       -0.707
## Time1      -0.283  0.200
## Time2      -0.283  0.200  0.500
## Time3      -0.283  0.200  0.500  0.500
## TrtC:Time1  0.200 -0.283 -0.707 -0.354 -0.354
## TrtC:Time2  0.200 -0.283 -0.354 -0.707 -0.354  0.500
## TrtC:Time3  0.200 -0.283 -0.354 -0.354 -0.707  0.500  0.500
```

```r
anova(Model1B, ddf = "Kenward-Roger")
```

```
## Analysis of Variance Table of type III  with  Kenward-Roger
## approximation for degrees of freedom
##           Sum Sq Mean Sq NumDF DenDF F.value   Pr(>F)
## Trt        2.347   2.347     1    10  0.2457 0.630853
## Time     187.831  62.610     3    30  6.5532 0.001535 **
## Trt:Time 154.615  51.538     3    30  5.3943 0.004328 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
emmeans(Model1B, pairwise ~ Trt|Time)
```

```
## $emmeans
## Time = 0:
##  Trt   emmean       SE    df lower.CL upper.CL
##  A   62.50801 3.155318 12.83 55.68230 69.33372
##  C   58.82859 3.155318 12.83 52.00288 65.65430
##
## Time = 1:
##  Trt   emmean       SE    df lower.CL upper.CL
##  A   56.25795 3.155318 12.83 49.43224 63.08366
##  C   59.57181 3.155318 12.83 52.74610 66.39752
##
## Time = 2:
##  Trt   emmean       SE    df lower.CL upper.CL
##  A   55.80448 3.155318 12.83 48.97877 62.63019
```

```
##  C   58.30690 3.155318 12.83 51.48119 65.13261
##
## Time = 3:
##  Trt   emmean       SE    df lower.CL upper.CL
##  A   52.09279 3.155318 12.83 45.26708 58.91849
##  C   58.25529 3.155318 12.83 51.42958 65.08100
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
##
## $contrasts
## Time = 0:
##  contrast  estimate       SE    df t.ratio p.value
##  A - C     3.679417 4.462293 12.83   0.825  0.4247
##
## Time = 1:
##  contrast  estimate       SE    df t.ratio p.value
##  A - C    -3.313859 4.462293 12.83  -0.743  0.4711
##
## Time = 2:
##  contrast  estimate       SE    df t.ratio p.value
##  A - C    -2.502422 4.462293 12.83  -0.561  0.5846
##
## Time = 3:
##  contrast  estimate       SE    df t.ratio p.value
##  A - C    -6.162502 4.462293 12.83  -1.381  0.1909
```

Reformatting

In order to use Time0 (baseline) as a covariate or use Difference as the response, we need to do some reformatting. Specifically, we need to add a column with the Time0 values. This column can be used to calculate a difference (downstream time points versus Time 0). Since we are using Time0 as a covariate (or calculating difference versus Time0), we "drop" that period from the analysis.
Note: The rename() function from dplyr is very handy!

```r
head(CatsData)
```

```
##   Trt Cat Time        Y newID
## 1   A   1    0 70.89283    A1
## 2   A   1    1 64.92599    A1
## 3   A   1    2 61.68156    A1
## 4   A   1    3 56.56081    A1
## 5   A   2    0 70.41208    A2
## 6   A   2    1 67.64846    A2
```

```r
Time0 <- CatsData %>%
        filter(Time == 0) %>%
        rename(Time0 = Y) %>%
        select(-Trt, -Cat, -Time)
head(Time0)
```

```
##      Time0 newID
## 1 70.89283    A1
## 2 70.41208    A2
```

```
## 3 69.79313    A3
## 4 54.96961    A4
## 5 55.04248    A5
## 6 53.93793    A6
```

```
nrow(Time0)
```

```
## [1] 12
```

```
WithTime0 <- full_join(CatsData, Time0) %>%
            mutate(Diff = Y - Time0) %>%
            filter(Time != 0)
```

```
## Joining, by = "newID"
```

```
head(WithTime0)
```

```
##   Trt Cat Time        Y newID    Time0       Diff
## 1   A   1    1 64.92599    A1 70.89283  -5.966840
## 2   A   1    2 61.68156    A1 70.89283  -9.211274
## 3   A   1    3 56.56081    A1 70.89283 -14.332024
## 4   A   2    1 67.64846    A2 70.41208  -2.763622
## 5   A   2    2 68.98425    A2 70.41208  -1.427830
## 6   A   2    3 62.08641    A2 70.41208  -8.325670
```

```
nrow(WithTime0)
```

```
## [1] 36
```

Model2: Using baseline (Time0) as a covariate

After reformatting, it is easy to include Time0 as a covariate!

```
Model2 <- lmer(Y ~ Time0 + Trt*Time + (1|newID), data = WithTime0)
summary(Model2)
```

```
## Linear mixed model fit by REML t-tests use Satterthwaite approximations
##   to degrees of freedom [lmerMod]
## Formula: Y ~ Time0 + Trt * Time + (1 | newID)
##    Data: WithTime0
##
## REML criterion at convergence: 175.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.03044 -0.60779 -0.01715  0.68440  1.46115
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  newID    (Intercept)  4.588   2.142
##  Residual             10.269   3.205
## Number of obs: 36, groups:  newID, 12
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   7.9260     6.8494  9.4520   1.157   0.2756
```

```
## Time0          0.7732     0.1066  9.0000   7.250 4.81e-05 ***
## TrtC           6.1588     2.2597 22.7390   2.725   0.0121 *
## Time2         -0.4535     1.8501 20.0000  -0.245   0.8089
## Time3         -4.1652     1.8501 20.0000  -2.251   0.0358 *
## TrtC:Time2    -0.8114     2.6165 20.0000  -0.310   0.7597
## TrtC:Time3     2.8486     2.6165 20.0000   1.089   0.2892
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) Time0  TrtC   Time2  Time3  TrC:T2
## Time0      -0.973
## TrtC       -0.329  0.174
## Time2      -0.135  0.000  0.409
## Time3      -0.135  0.000  0.409  0.500
## TrtC:Time2  0.095  0.000 -0.579 -0.707 -0.354
## TrtC:Time3  0.095  0.000 -0.579 -0.354 -0.707  0.500
```

```
anova(Model2, ddf = "Kenward-Roger")
```

```
## Analysis of Variance Table of type III  with  Kenward-Roger
## approximation for degrees of freedom
##          Sum Sq Mean Sq NumDF DenDF F.value    Pr(>F)
## Time0    539.80  539.80     1     9  52.566 4.815e-05 ***
## Trt      170.00  170.00     1     9  16.555  0.002805 **
## Time      47.16   23.58     2    20   2.296  0.126533
## Trt:Time  22.17   11.08     2    20   1.079  0.358776
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
emmeans(Model2, pairwise ~ Trt|Time)
```

```
## $emmeans
## Time = 1:
##  Trt   emmean       SE    df lower.CL upper.CL
##  A   54.83547 1.585782 23.01 51.55514 58.11579
##  C   60.99430 1.585782 23.01 57.71397 64.27462
##
## Time = 2:
##  Trt   emmean       SE    df lower.CL upper.CL
##  A   54.38199 1.585782 23.01 51.10167 57.66232
##  C   59.72939 1.585782 23.01 56.44906 63.00971
##
## Time = 3:
##  Trt   emmean       SE    df lower.CL upper.CL
##  A   50.67030 1.585782 23.01 47.38998 53.95062
##  C   59.67777 1.585782 23.01 56.39745 62.95810
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
##
## $contrasts
## Time = 1:
##  contrast  estimate       SE    df t.ratio p.value
##  A - C    -6.158829 2.259733 22.74  -2.725  0.0121
```

```
##
## Time = 2:
##  contrast  estimate       SE    df t.ratio p.value
##  A - C    -5.347393 2.259733 22.74  -2.366  0.0269
##
## Time = 3:
##  contrast  estimate       SE    df t.ratio p.value
##  A - C    -9.007472 2.259733 22.74  -3.986  0.0006
```

Model3: Using Diff (from Baseline) as the Response

```r
Model3 <- lmer(Diff ~ Trt*Time + (1|newID), data = WithTime0)
summary(Model3)
```

```
## Linear mixed model fit by REML t-tests use Satterthwaite approximations
##   to degrees of freedom [lmerMod]
## Formula: Diff ~ Trt * Time + (1 | newID)
##    Data: WithTime0
##
## REML criterion at convergence: 177.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.78233 -0.60578 -0.00503  0.50279  1.70926
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  newID    (Intercept)  7.41    2.722
##  Residual             10.27    3.205
## Number of obs: 36, groups:  newID, 12
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  -6.2501     1.7165 22.2000  -3.641  0.00143 **
## TrtC          6.9933     2.4275 22.2000   2.881  0.00863 **
## Time2        -0.4535     1.8501 20.0000  -0.245  0.80887
## Time3        -4.1652     1.8501 20.0000  -2.251  0.03577 *
## TrtC:Time2   -0.8114     2.6165 20.0000  -0.310  0.75967
## TrtC:Time3    2.8486     2.6165 20.0000   1.089  0.28922
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) TrtC   Time2  Time3  TrC:T2
## TrtC       -0.707
## Time2      -0.539  0.381
## Time3      -0.539  0.381  0.500
## TrtC:Time2  0.381 -0.539 -0.707 -0.354
## TrtC:Time3  0.381 -0.539 -0.354 -0.707  0.500
```

```r
anova(Model3, ddf = "Kenward-Roger")
```

```
## Analysis of Variance Table of type III  with  Kenward-Roger
```

```
## approximation for degrees of freedom
##           Sum Sq Mean Sq NumDF DenDF F.value   Pr(>F)
## Trt      167.400 167.400     1    10 16.3015 0.002371 **
## Time      47.164  23.582     2    20  2.2964 0.126533
## Trt:Time  22.169  11.085     2    20  1.0794 0.358776
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
emmeans(Model3, pairwise ~ Trt|Time)
```

```
## $emmeans
## Time = 1:
##  Trt      emmean       SE   df   lower.CL  upper.CL
##  A    -6.2500588 1.716537 22.2  -9.808082 -2.692036
##  C     0.7432166 1.716537 22.2  -2.814806  4.301239
##
## Time = 2:
##  Trt      emmean       SE   df   lower.CL  upper.CL
##  A    -6.7035326 1.716537 22.2 -10.261555 -3.145510
##  C    -0.5216935 1.716537 22.2  -4.079716  3.036329
##
## Time = 3:
##  Trt      emmean       SE   df   lower.CL  upper.CL
##  A   -10.4152255 1.716537 22.2 -13.973248 -6.857203
##  C    -0.5733070 1.716537 22.2  -4.131330  2.984716
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
##
## $contrasts
## Time = 1:
##  contrast  estimate      SE   df t.ratio p.value
##  A - C    -6.993275 2.42755 22.2  -2.881  0.0086
##
## Time = 2:
##  contrast  estimate      SE   df t.ratio p.value
##  A - C    -6.181839 2.42755 22.2  -2.547  0.0183
##
## Time = 3:
##  contrast  estimate      SE   df t.ratio p.value
##  A - C    -9.841918 2.42755 22.2  -4.054  0.0005
```