Blunt-nosed Leopard Lizard CEWL Measurement Triplicate Tests

Savannah Weaver

April 2021

Contents

Packages

```
if (!require("tidyverse")) install.packages("tidyverse")
library("tidyverse")
if (!require("UsingR")) install.packages("UsingR")
library("UsingR")
if (!require("MASS")) install.packages("MASS")
library("MASS")
if (!require("lme4")) install.packages("lme4")
library("lme4")
```

Load in Data

This data was collected during fieldwork with Blunt-nosed Leopard Lizards (*Gambelia sila*) in the Elkhorn Plain of the Carrizo Plain National Monument in California, USA during the 2021 active season (April-August).

This is a preliminary analysis to decide whether CEWL measurements should be taken in triplicate or can be just one measurement.

Load CEWL Data

Only the first weekend is loaded for now, but we can check with other replicates later.

```r
# weekend 1
CEWL_April_23_25 <- read.csv("./data/BNLL CEWL 4-23-2021 to 4-25-2021.csv") %>%
  dplyr::mutate(weekend = "1") # add weekend index for throughout season

# weekend 2
# etc etc

# merge all CEWL datafiles & reformat
CEWL <- CEWL_April_23_25 %>%
  # join
  #rbind(., CEWL_May_xx, CEWL_May_xx, etc
  #        ) %>%
  # reformat data
  dplyr::select(Date, Time,
                ID = Comments,
                TEWL_g_m2h = TEWL..g..m2h.., # rename
                CV = CV...., # rename
                SSWL_g_m2 = SSWL..g..m2.., # rename
                ambient_temp_C = AmbT..C., # rename
                ambient_RH_percent = AmbRH...., # rename
                weekend
                ) %>%
  dplyr::mutate(date_time = as.POSIXct(paste(Date, Time), format = "%m/%d/%y %I:%M:%S %p"),
                # date only
                Date = as.Date(Date, format = "%m/%d/%y"),
                # time with arbitrary date, need to figure out how to remove
                Time = as.POSIXct(Time, format = "%I:%M:%S %p"),
                # get length of current ID for editing
                length_ID = nchar((ID)),
                # extract CEWL replicate number
                replicate = as.numeric(substring(ID, length_ID, length_ID)),
                # extract group (male, female, widespread)
                ID_group = substring(ID, 1, 1),
                # extract ID number within group
                ID_number = substring(ID, 2, (length_ID-2)),
                # merge IDs to match format in other data files
                ID = paste(ID_group, "-", ID_number, sep = ""),
                weekend = as.numeric(weekend)
                ) %>%
  dplyr::select(-length_ID, # remove these since no longer need
                -ID_group,
                -ID_number
                )
summary(CEWL)
```

```
##       Date                    Time                          ID
##  Min.   :2021-04-23   Min.   :2021-05-05 11:37:51   Length:198
##  1st Qu.:2021-04-23   1st Qu.:2021-05-05 13:53:22   Class :character
##  Median :2021-04-24   Median :2021-05-05 14:38:46   Mode  :character
##  Mean   :2021-04-23   Mean   :2021-05-05 15:01:37
##  3rd Qu.:2021-04-24   3rd Qu.:2021-05-05 15:58:40
```

```
##   Max.   :2021-04-25   Max.    :2021-05-05 19:56:00
##     TEWL_g_m2h           CV            SSWL_g_m2       ambient_temp_C
##   Min.   : 0.140   Min.   : 0.1000   Min.   :0.00000   Min.   :18.90
##   1st Qu.: 8.213   1st Qu.: 0.4125   1st Qu.:0.00900   1st Qu.:28.52
##   Median :10.615   Median : 0.5900   Median :0.02439   Median :30.50
##   Mean   :11.650   Mean   : 1.0383   Mean   :0.03701   Mean   :29.07
##   3rd Qu.:14.150   3rd Qu.: 0.8175   3rd Qu.:0.05147   3rd Qu.:31.60
##   Max.   :65.310   Max.   :42.8500   Max.   :0.18780   Max.   :33.70
##   ambient_RH_percent   weekend    date_time                    replicate
##   Min.   :11.60        Min.   :1   Min.   :2021-04-23 12:57:55   Min.   :1.000
##   1st Qu.:15.00        1st Qu.:1   1st Qu.:2021-04-23 14:55:45   1st Qu.:1.000
##   Median :15.90        Median :1   Median :2021-04-24 12:57:24   Median :2.000
##   Mean   :19.31        Mean   :1   Mean   :2021-04-24 07:23:26   Mean   :2.015
##   3rd Qu.:20.98        3rd Qu.:1   3rd Qu.:2021-04-24 15:36:36   3rd Qu.:3.000
##   Max.   :37.90        Max.   :1   Max.   :2021-04-25 14:41:17   Max.   :4.000
```

Plot Replicates

```r
CEWL %>%
  ggplot(data = .) +
  geom_point(aes(x = ID,
                 y = TEWL_g_m2h,
                 color = ID
                 ),
             size = 1,
             alpha = 0.6) +
  theme_classic() +
  xlab("BNLL Individual ID") +
  ylab("CEWL replicate values (g/m2h)") +
  theme(text = element_text(color = "black",
                            family = "sans",
                            size = 12),
        axis.text = element_text(color = "black",
                                 family = "sans",
                                 size = 10),
        legend.text.align = 0,
        legend.position = "none"
)
```
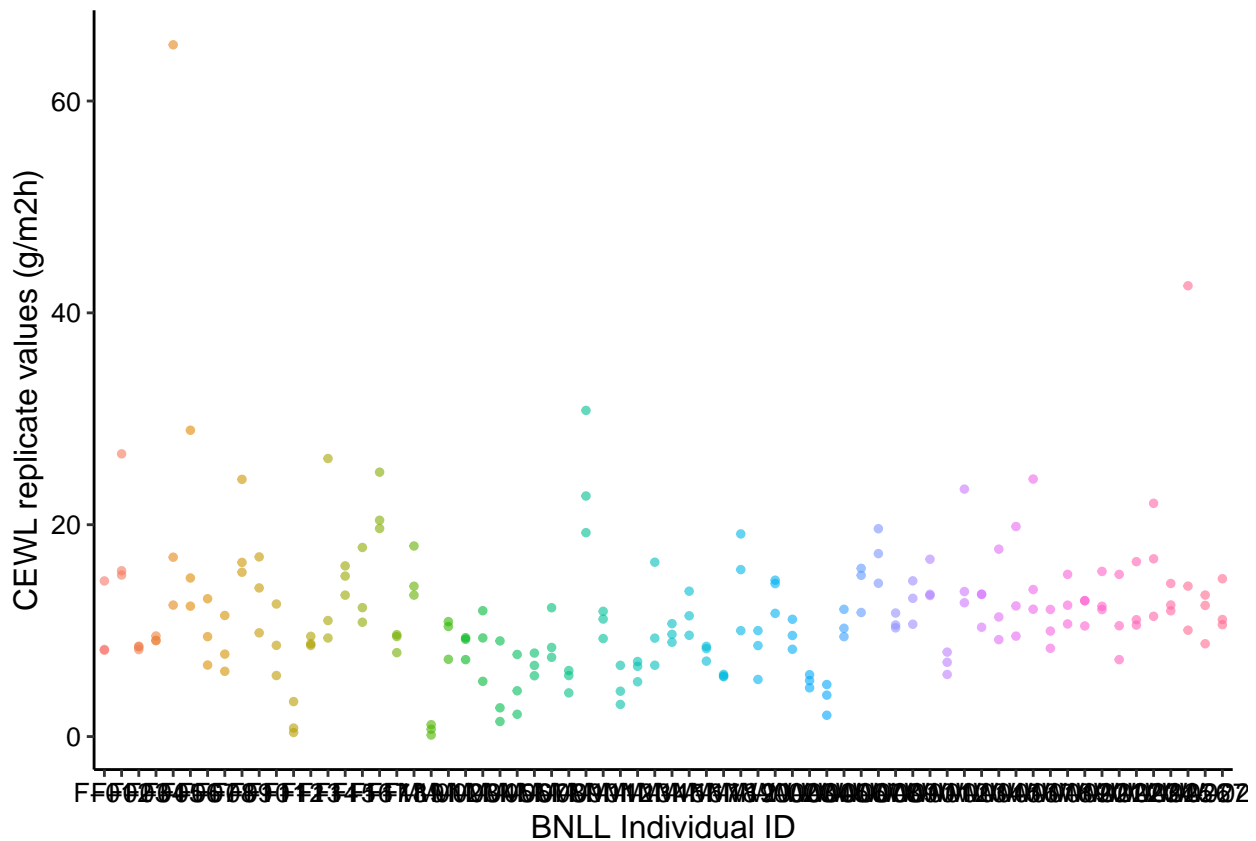
It looks like there are two unusually high measurements - one replicate for the fifth lizard listed and one replicate for the third-to-last lizard listed. I think they're both above 40, so I could see how filtering them out affects replicability.

Get mean values for each lizard:

```
CEWL_rep_avg <- CEWL %>%
  group_by(Date, ID, weekend) %>% # put all keeps here
  # summaries
  summarise(mean_TEWL_g_m2h = mean(TEWL_g_m2h),
            sd_TEWL_g_m2h = sd(TEWL_g_m2h),
            mean_CV = mean(CV),
            sd_CV = sd(CV),
            mean_SSWL_g_m2 = mean(SSWL_g_m2),
            sd_SSWL_g_m2 = sd(SSWL_g_m2),
            mean_temp_C = mean(ambient_temp_C),
            sd_temp_C = sd(ambient_temp_C),
            mean_RH_percent = mean(ambient_RH_percent),
            sd_RH_percent = sd(ambient_RH_percent)
            )
```

`summarise()` regrouping output by 'Date', 'ID' (override with `.groups` argument)

```
avg_TEWL_sd <- mean(CEWL_rep_avg$sd_TEWL_g_m2h)
avg_TEWL_sd
```

[1] 3.322823

Plot means:

```
CEWL_rep_avg %>%
  ggplot(data = .) +
  geom_point(aes(x = ID,
                 y = mean_TEWL_g_m2h,
                 color = ID
                 ),
             size = 1,
             alpha = 0.6) +
  theme_classic() +
  xlab("BNLL Individual ID") +
  ylab("Mean CEWL (g/m2h)") +
  theme(text = element_text(color = "black",
                            family = "sans",
                            size = 12),
        axis.text = element_text(color = "black",
                                 family = "sans",
                                 size = 10),
        legend.text.align = 0,
        legend.position = "none"
)
```


Model Replicates

make a GLMM to figure out whether replicates have significantly different CEWL values:

```
mod <- lme4::lmer(data = CEWL, TEWL_g_m2h ~ replicate + (1|ID))
mod2 <- nlme::lme(data = CEWL, TEWL_g_m2h ~ replicate, random = ~ 1|ID) # for p-values
summary(mod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: TEWL_g_m2h ~ replicate + (1 | ID)
##    Data: CEWL
##
## REML criterion at convergence: 1270.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3402 -0.3939 -0.0861  0.2671  7.7263
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 22.56    4.750
##  Residual             23.09    4.805
## Number of obs: 198, groups:  ID, 66
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   16.217      1.079  15.028
## replicate     -2.267      0.417  -5.436
##
## Correlation of Fixed Effects:
##           (Intr)
## replicate -0.779
```

```
summary(mod2)
```

```
## Linear mixed-effects model fit by REML
##   Data: CEWL
##        AIC      BIC    logLik
##   1278.673 1291.785 -635.3363
##
## Random effects:
##  Formula: ~1 | ID
##         (Intercept) Residual
## StdDev:    4.749645 4.804734
##
## Fixed effects:  TEWL_g_m2h ~ replicate
##                 Value Std.Error  DF   t-value p-value
## (Intercept) 16.21740 1.0791558 131 15.027857       0
## replicate   -2.26668 0.4170131 131 -5.435512       0
##  Correlation:
##           (Intr)
## replicate -0.779
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.34017065 -0.39388613 -0.08605866  0.26705911  7.72628368
##
## Number of Observations: 198
```

```
## Number of Groups: 66
```

Remove Unusual Points

```r
CEWL_sub <- CEWL %>%
  dplyr::filter(TEWL_g_m2h <= 40)
```

Re-Plot

```r
CEWL_sub %>%
  ggplot(data = .) +
  geom_point(aes(x = ID,
                 y = TEWL_g_m2h,
                 color = ID
                 ),
             size = 1,
             alpha = 0.6) +
  theme_classic() +
  xlab("BNLL Individual ID") +
  ylab("CEWL replicate values (g/m2h)") +
  theme(text = element_text(color = "black",
                            family = "sans",
                            size = 12),
        axis.text = element_text(color = "black",
                                 family = "sans",
                                 size = 10),
        legend.text.align = 0,
        legend.position = "none"
)
```
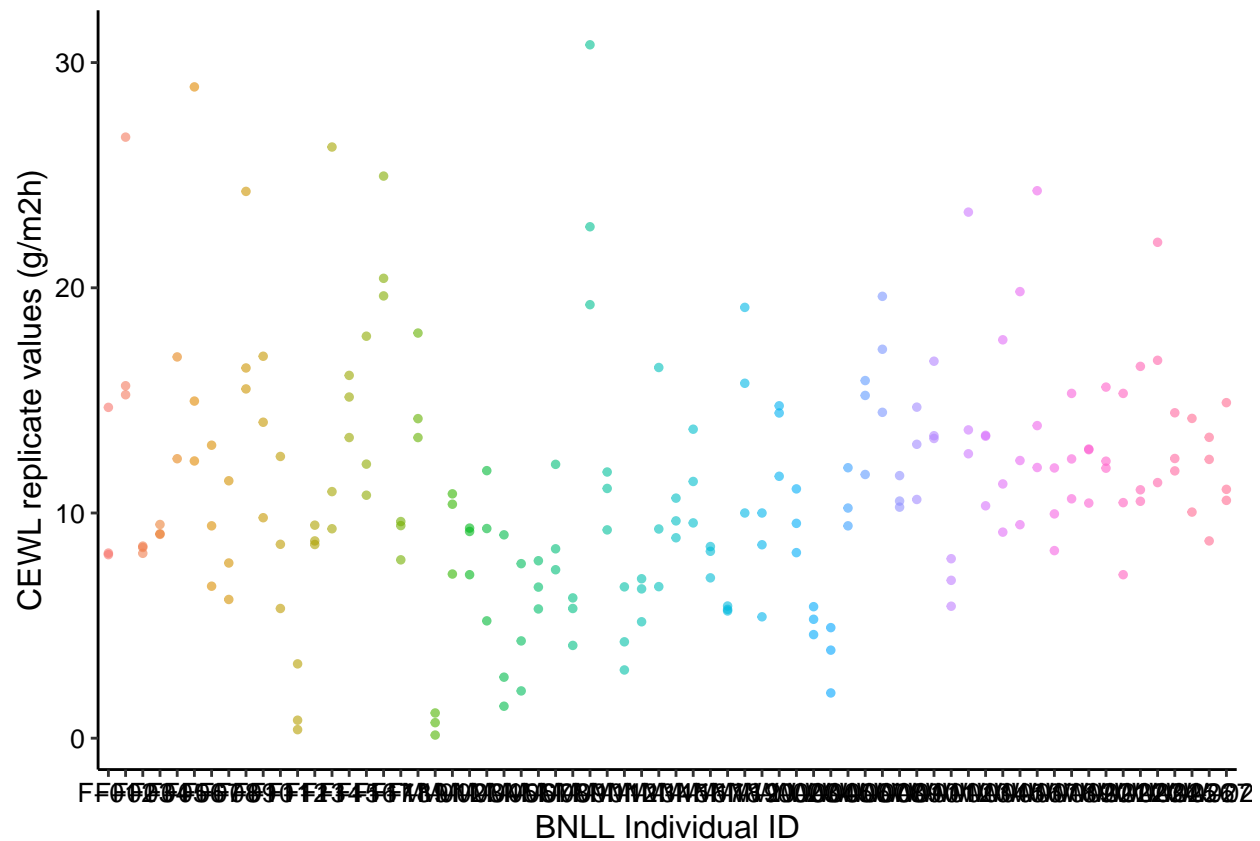
This looks a lot better....

Re-Model

```
mod3 <- lme4::lmer(data = CEWL_sub, TEWL_g_m2h ~ replicate + (1|ID))
mod4 <- nlme::lme(data = CEWL_sub, TEWL_g_m2h ~ replicate, random = ~ 1|ID) # for p-values
summary(mod3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: TEWL_g_m2h ~ replicate + (1 | ID)
##    Data: CEWL_sub
##
## REML criterion at convergence: 1097.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0156 -0.5504 -0.0698  0.4158  3.3646
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 18.687   4.323
##  Residual              7.935   2.817
## Number of obs: 196, groups:  ID, 66
##
## Fixed effects:
##             Estimate Std. Error t value
```

```
## (Intercept)   14.8771     0.7584    19.61
## replicate    -1.7947     0.2472    -7.26
##
## Correlation of Fixed Effects:
##          (Intr)
## replicate -0.661
```

`summary(mod4)`

```
## Linear mixed-effects model fit by REML
##   Data: CEWL_sub
##       AIC      BIC    logLik
##   1105.523 1118.594 -548.7614
##
## Random effects:
##  Formula: ~1 | ID
##         (Intercept) Residual
## StdDev:    4.322866 2.816841
##
## Fixed effects:  TEWL_g_m2h ~ replicate
##                 Value Std.Error  DF   t-value p-value
## (Intercept) 14.877077 0.7584422 129 19.615307       0
## replicate   -1.794696 0.2471906 129 -7.260373       0
##  Correlation:
##          (Intr)
## replicate -0.661
##
## Standardized Within-Group Residuals:
##        Min          Q1         Med          Q3         Max
## -2.01563852 -0.55038247 -0.06984615  0.41576743  3.36460933
##
## Number of Observations: 196
## Number of Groups: 66
```

The p-values are still 0 even after removing the two unusually high points...

Conclusion

There's definitely a lot of variation among triplicate measurements for a given lizard, and the models show that variation is significant. I would conclude that it is important to take these measurements in triplicate to even out the variation in measurements. Taking measurements in triplicate will also allow us to exclude unusual points like the two in the first weekend's dataset, while still having the other replicates for that lizard for us to analyse.