DNA damage-Mixed effects model

Code

Or Ben-Zvi

load packages

Hide

```
library(car)
library(ggplot2)
library(lme4)
library(lmerTest)
library(dplyr)
library(MASS)
library(cowplot)
library(doBy)
```

read data

Hide

```
data_dna<- read.csv(file.choose())
```

Hide

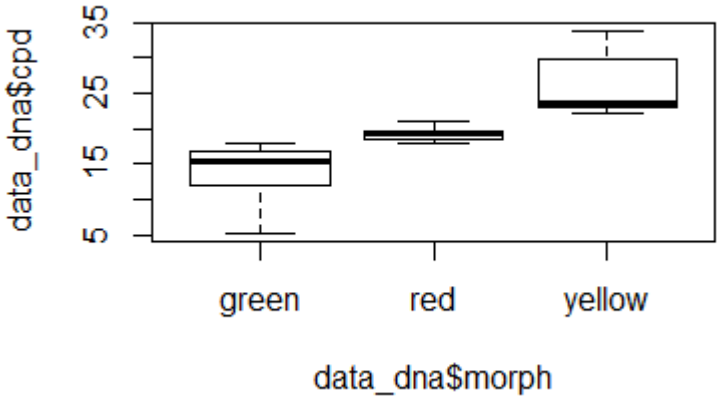
```
str(data_dna)
```

```
'data.frame':   26 obs. of  5 variables:
 $ treatment: Factor w/ 2 levels "par","paruv": 2 2 2 2 2 2 2 2 2 2 ...
 $ morph    : Factor w/ 3 levels "green","red",..: 1 1 1 1 1 1 1 2 2 2 ...
 $ colony   : Factor w/ 13 levels "a","c","d","e",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ cpd      : num  16.71 7.43 33.86 21 12.07 ...
 $ pp       : num  3.49 4.74 5.52 3.23 4.64 ...
```

Plot the data

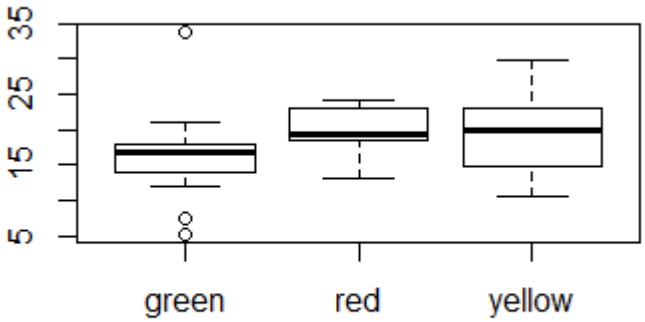
Hide

```
qqplot(data_dna$morph,data_dna$cpd)
```



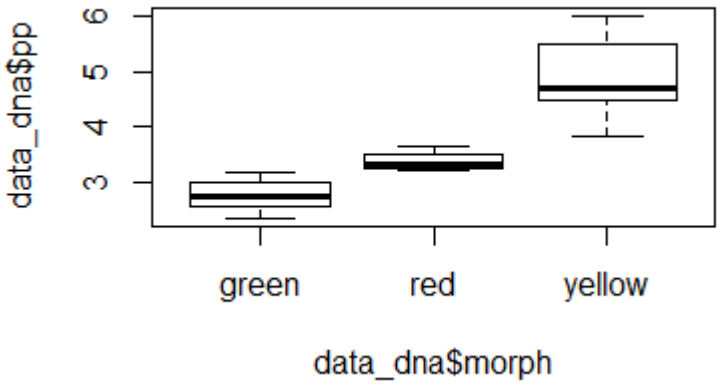
Hide

```
plot(data_dna$morph,data_dna$cpd)
```



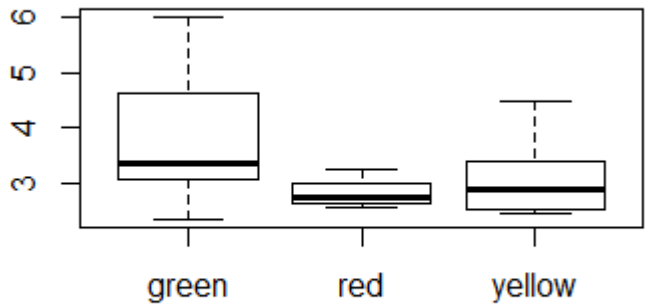
Hide

```
qqplot(data_dna$morph,data_dna$pp)
```



Hide

```
plot(data_dna$morph,data_dna$pp)
```



Test with mixed effect model

Hide

```
model_cpd <- lmer(cpd ~ treatment*morph + (1|morph/colony),
                  data=data_dna, REML=FALSE)
summary(model_cpd)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: cpd ~ treatment * morph + (1 | morph/colony)
   Data: data_dna

     AIC      BIC   logLik deviance df.resid
   182.4    193.8    -82.2    164.4       17

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8034 -0.4268 -0.0516  0.3517  2.8354

Random effects:
 Groups       Name        Variance Std.Dev.
 colony:morph (Intercept)  7.143   2.673
 morph        (Intercept)  0.000   0.000
 Residual                 26.315   5.130
Number of obs: 26, groups:  colony:morph, 13; morph, 3

Fixed effects:
                           Estimate Std. Error      df t value Pr(>|t|)
(Intercept)                 16.8163     2.1863 24.8667   7.692 4.95e-08 ***
treatmentparuv              -0.8163     2.7420 13.0000  -0.298    0.771
morphred                     2.2789     3.9915 24.8667   0.571    0.573
morphyellow                  2.6361     3.9915 24.8667   0.660    0.515
treatmentparuv:morphred      1.8878     5.0062 13.0000   0.377    0.712
treatmentparuv:morphyellow   1.4116     5.0062 13.0000   0.282    0.782
---
Signif. codes:  0 □***□ 0.001 □**□ 0.01 □*□ 0.05 □.□ 0.1 □ □ 1

Correlation of Fixed Effects:
               (Intr) trtmnt mrphrd mrphyl trtmntprv:mrphr
treatmntprv    -0.627
morphred       -0.548  0.343
morphyellow    -0.548  0.343  0.300
trtmntprv:mrphr  0.343 -0.548 -0.627 -0.188
trtmntprv:mrphy  0.343 -0.548 -0.188 -0.627  0.300
```

Hide

```
anova(model_cpd)
```

```
Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
treatment       0.447  0.4466     1    13  0.0170 0.8983
morph          45.147 22.5737     2    13  0.8578 0.4468
treatment:morph  4.566  2.2831     2    13  0.0868 0.9174
```

Hide

```
model_cpd1 <- lmer(cpd ~ treatment + (1|morph/colony),
                   data=data_dna, REML=FALSE)
model_cpd2 <- lmer(cpd ~ morph + (1|morph/colony),
                   data=data_dna, REML=FALSE)
anova(model_cpd, model_cpd1)
```

```
Data: data_dna
Models:
model_cpd1: cpd ~ treatment + (1 | morph/colony)
model_cpd: cpd ~ treatment * morph + (1 | morph/colony)
           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model_cpd1  5 176.23 182.52 -83.115   166.23
model_cpd   9 182.45 193.77 -82.223   164.45 1.7839      4     0.7754
```

Hide

```
anova(model_cpd, model_cpd2)
```

```
Data: data_dna
Models:
model_cpd2: cpd ~ morph + (1 | morph/colony)
model_cpd: cpd ~ treatment * morph + (1 | morph/colony)
           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model_cpd2  6 176.62 184.17 -82.310   164.62
model_cpd   9 182.45 193.77 -82.223   164.45 0.1731      3     0.9818
```

Hide

```
model_pp <-  lmer(pp ~ treatment*morph + (1|morph/colony),
                   data=data_dna, REML=FALSE)
summary(model_pp)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: pp ~ treatment * morph + (1 | morph/colony)
   Data: data_dna

     AIC      BIC   logLik deviance df.resid
    72.1     83.4    -27.1     54.1       17

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.94930 -0.39648 -0.02814  0.48214  2.38478

Random effects:
 Groups       Name        Variance Std.Dev.
 colony:morph (Intercept) 0.0000   0.000
 morph        (Intercept) 0.0000   0.000
 Residual                 0.4692   0.685
Number of obs: 26, groups:  colony:morph, 13; morph, 3

Fixed effects:
                           Estimate Std. Error      df t value Pr(>|t|)
(Intercept)                  3.1153     0.2589 26.0000  12.033 3.95e-12 ***
treatmentparuv               1.2683     0.3661 26.0000   3.464  0.00186 **
morphred                    -0.4931     0.4727 26.0000  -1.043  0.30646
morphyellow                 -0.3179     0.4727 26.0000  -0.673  0.50715
treatmentparuv:morphred     -0.8989     0.6685 26.0000  -1.345  0.19031
treatmentparuv:morphyellow  -0.6717     0.6685 26.0000  -1.005  0.32426
---
Signif. codes:  0 □***□ 0.001 □**□ 0.01 □*□ 0.05 □.□ 0.1 □ □ 1

Correlation of Fixed Effects:
               (Intr) trtmnt mrphrd mrphyl trtmntprv:mrphr
treatmntprv    -0.707
morphred       -0.548  0.387
morphyellow    -0.548  0.387  0.300
trtmntprv:mrphr  0.387 -0.548 -0.707 -0.212
trtmntprv:mrphy  0.387 -0.548 -0.212 -0.707  0.300
```

Hide

```
anova(model_pp)
```

```
Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
treatment      3.0830 3.08300     1    26  6.5708 0.01650 *
morph          4.3666 2.18332     2    26  4.6533 0.01873 *
treatment:morph 1.0350 0.51748     2    26  1.1029 0.34694
---
Signif. codes:  0 □***□ 0.001 □**□ 0.01 □*□ 0.05 □.□ 0.1 □ □ 1
```

Hide

```
model_pp1 <-  lmer(pp ~ treatment + (1|morph/colony),
                   data=data_dna, REML=FALSE)
model_pp2 <-  lmer(pp ~ morph + (1|morph/colony),
                   data=data_dna, REML=FALSE)
anova(model_pp, model_pp1)
```

```
Data: data_dna
Models:
model_pp1: pp ~ treatment + (1 | morph/colony)
model_pp: pp ~ treatment * morph + (1 | morph/colony)
          Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model_pp1  5 71.953 78.243 -30.976   61.953
model_pp   9 72.110 83.433 -27.055   54.110 7.8428      4    0.09751 .
---
Signif. codes:  0 □***□ 0.001 □**□ 0.01 □*□ 0.05 □.□ 0.1 □ □ 1
```

Hide

```
anova(model_pp, model_pp2)
```

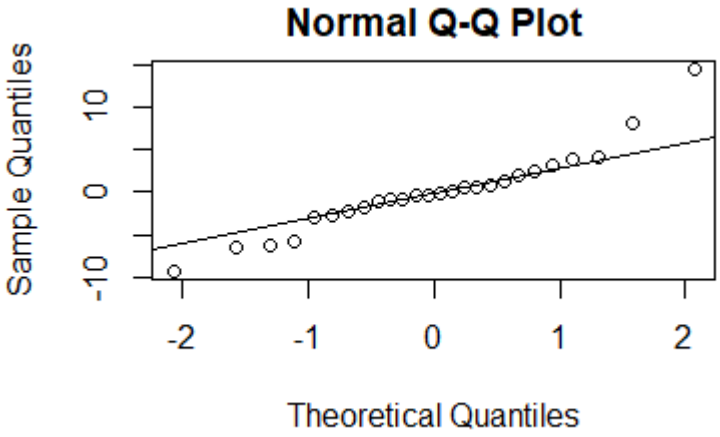
```
Data: data_dna
Models:
model_pp2: pp ~ morph + (1 | morph/colony)
model_pp: pp ~ treatment * morph + (1 | morph/colony)
          Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model_pp2  6 77.031 84.579 -32.515   65.031
model_pp   9 72.110 83.433 -27.055   54.110 10.921      3    0.01216 *
---
Signif. codes:  0 □***□ 0.001 □**□ 0.01 □*□ 0.05 □.□ 0.1 □ □ 1
```

Check the LMM assumptions

Check the normality of model residuals

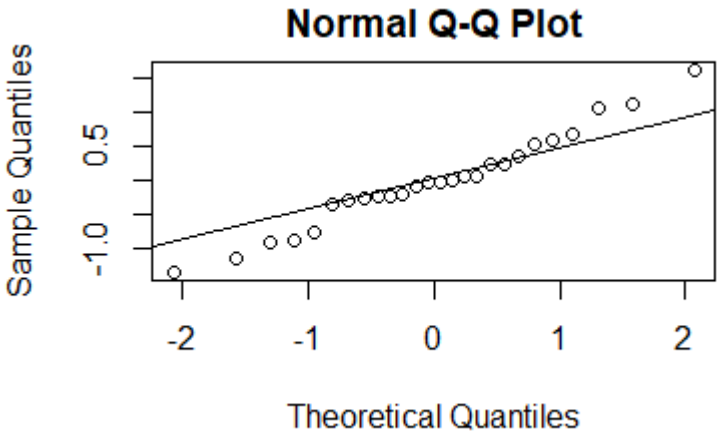
Hide

```
qqnorm(resid(model_cpd))
qqline(resid(model_cpd))
```



Hide

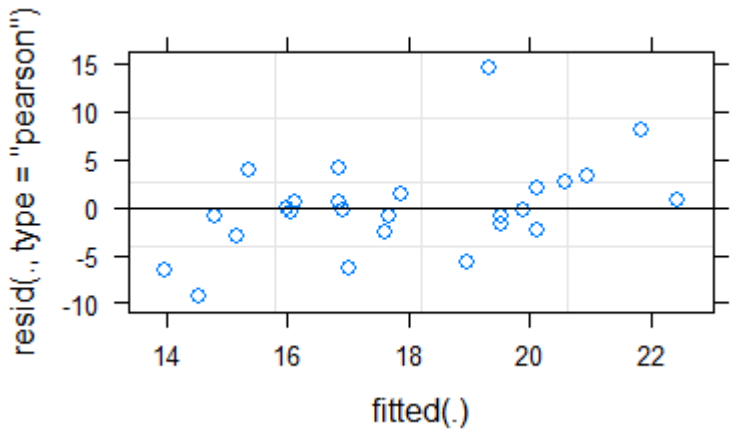
```
qqnorm(resid(model_pp))
qqline(resid(model_pp))
```



Check for homogenity of variance

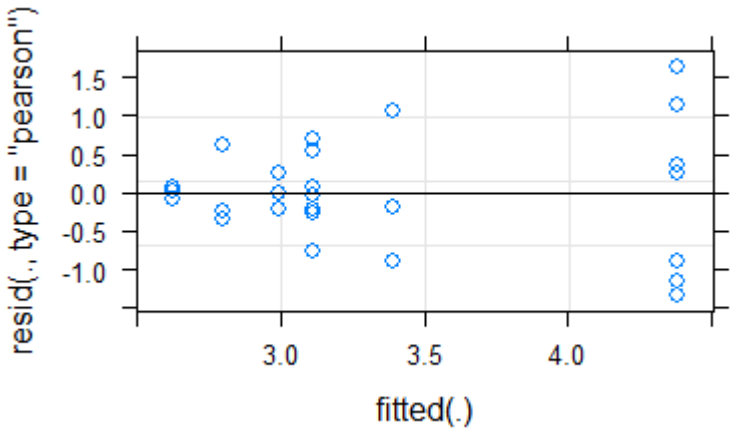
Hide

```
plot(model_cpd)
```



Hide

```
plot(model_pp)
```



Obtain mean and SD

Hide

```
sumfun <- function(x, ...){
  c(mean=mean(x, na.rm=TRUE, ...), sd=sd(x, na.rm=TRUE, ...), l=length(x))
}
mean(data_dna$cpd, na.rm = TRUE)
```

```
[1] 17.92308
```

Hide

```
sd(data_dna$cpd, na.rm = TRUE)
```

```
[1] 6.145412
```

Hide

```
summaryBy(cpd ~ morph * treatment, data=data_dna, FUN=sumfun)
```

	morph <fctr>	treatment <fctr>	cpd.mean <dbl>	cpd.sd <dbl>	cpd.l <dbl>
1	green	par	16.81633	1.6587519	7
2	green	paruv	16.00000	9.5587158	7
3	red	par	19.09524	0.5455447	3
4	red	paruv	20.16667	6.1063422	3
5	yellow	par	19.45238	9.7502834	3
6	yellow	paruv	20.04762	4.4654760	3

6 rows

Hide

```
summaryBy(cpd ~ morph, data=data_dna, FUN=sumfun)
```

	morph <fctr>	cpd.mean <dbl>	cpd.sd <dbl>	cpd.l <dbl>
1	green	16.40816	6.604517	14
2	red	19.63095	3.921530	6
3	yellow	19.75000	6.790412	6

3 rows

Hide

```
summaryBy(cpd ~ treatment, data=data_dna, FUN=sumfun)
```

	treatment <fctr>	cpd.mean <dbl>	cpd.sd <dbl>	cpd.l <dbl>
1	par	17.95055	4.348780	13
2	paruv	17.89560	7.730836	13

2 rows

Hide

```
mean(data_dna$pp, na.rm = TRUE)
```

```
[1] 3.38101
```

Hide

```
sd(data_dna$pp, na.rm = TRUE)
```

```
[1] 0.9577882
```

Hide

```
summaryBy(pp ~morph * treatment, data=data_dna, FUN=sumfun)
```

	morph <fctr>	treatment <fctr>	pp.mean <dbl>	pp.sd <dbl>	pp.l <dbl>
1	green	par	3.115260	0.49754546	7
2	green	paruv	4.383523	1.15892611	7
3	red	par	2.622159	0.07908756	3
4	red	paruv	2.991477	0.23469761	3
5	yellow	par	2.797348	0.52793030	3
6	yellow	paruv	3.393939	0.99374307	3

6 rows

Hide

```
summaryBy(pp ~morph, data=data_dna, FUN=sumfun)
```

	morph <fctr>	pp.mean <dbl>	pp.sd <dbl>	pp.l <dbl>
1	green	3.749391	1.0803734	14
2	red	2.806818	0.2558396	6
3	yellow	3.095644	0.7831159	6

3 rows

Hide

```
summaryBy(pp ~treatment, data=data_dna, FUN=sumfun)
```

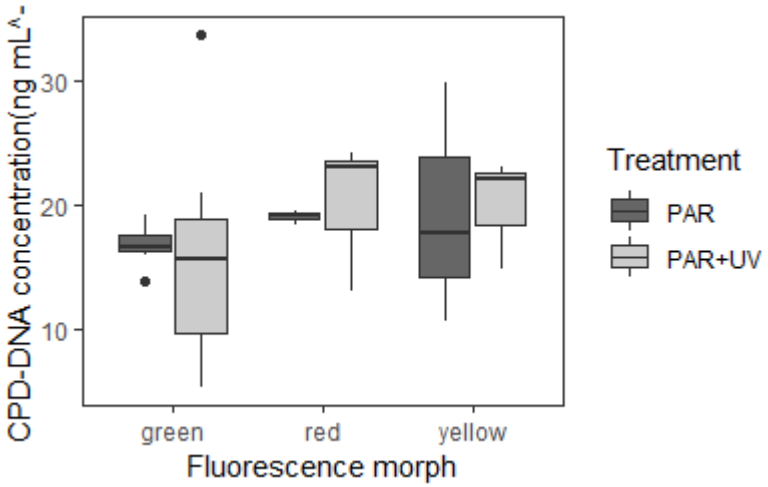
	treatment <fctr>	pp.mean <dbl>	pp.sd <dbl>	pp.l <dbl>
1	par	2.928103	0.4683774	13
2	paruv	3.833916	1.1168481	13

2 rows

Visualize the data as a box plot

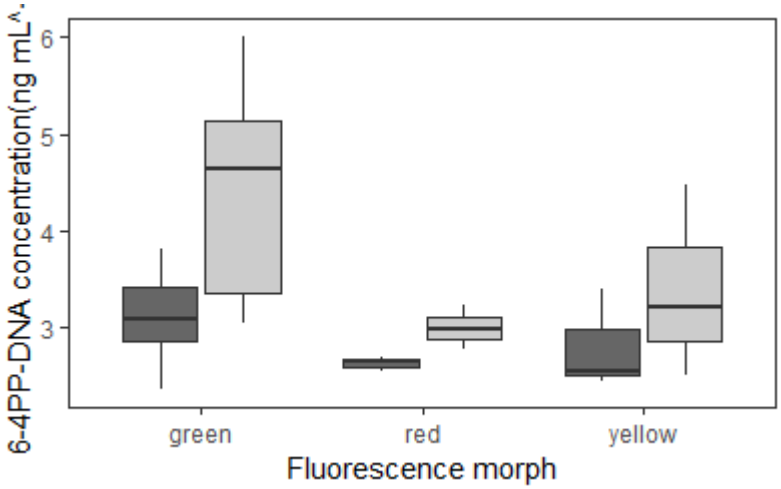
Hide

```
cpd_plot <- ggplot(data = data_dna, aes(x=morph, y=cpd)) +
  geom_boxplot(aes(fill=treatment)) +
  labs(x="Fluorescence morph", y="CPD-DNA concentration(ng mL^-1)") +
  scale_fill_grey(start = 0.4, end = 0.8, name="Treatment",breaks=c("par", "paruv"), labels=c("P
AR","PAR+UV")) +
  theme_bw() +
  theme(legend.position="right")+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
cpd_plot
```



Hide

```
pp_plot <- ggplot(data = data_dna, aes(x=morph, y=pp)) +
  geom_boxplot(aes(fill=treatment)) +
  labs(x="Fluorescence morph", y="6-4PP-DNA concentration(ng mL^-1)") +
  scale_fill_grey(start = 0.4, end = 0.8, name="Treatment",breaks=c("par", "paruv"), labels=c("P
AR","PAR+UV")) +
  theme_bw() +
  theme(legend.position="none")+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
pp_plot
```


6-4PP-DNA concentration(ng mL^-1
6
5
4
3
green
red
yellow
Fluorescence morph