

# Correlation and Mixed Effects

Code ▾

## Introduction

In this tutorial we want to evaluate how the parameters we measure with the MultispeQ relate to crop outputs. We will use the same experiment that we used in the previous tutorial: “ANOVA and multivariate analysis.”

However, in this tutorial, the data was transformed in a csv file so we cannot generate the dataframe directly from photosynq.org.

The first step will be to generate a correlation matrix of crop yield against the mean of each PhotosynQ parameter. Next, we will use a mixed effect model to account for factors that affect the PhotosynQ parameters: 1) light intensity, 2) time of measurement, 3) location within the field, etc.

## Import Libraries

```
library("broom", lib.loc="C:/Users/Dan/Documents/R/win-library/3.3")
library("dplyr", lib.loc="C:/Users/Dan/Documents/R/win-library/3.3")
library("data.table", lib.loc="C:/Users/Dan/Documents/R/win-library/3.3")
library("stringr", lib.loc="C:/Users/Dan/Documents/R/win-library/3.3")
library("Hmisc", lib.loc="C:/Users/Dan/Documents/R/win-library/3.3")
library("lme4", lib.loc="C:/Users/Dan/Documents/R/win-library/3.3")
```

Hide

## Simple Correlation

We will use the dataset ‘sun2.’ This dataset contains the mean values of each PhotosynQ parameter and the crop yield for each plot.

First, we need to open up our data file.

```
# Set the main folder location for your data
your_location = "C:/Users/Dan/Desktop/"
# The main data file, located in the main folder, is called...
file_name = "sun2.csv"
# Let's pull that data file into R and call it data_raw
my_data = read.csv(str_c(your_location, file_name))
View(my_data)
```

Hide

Now, lets generate a correlation matrix.

```
my_data <- my_data[c(3,4,5,6,7)]
head(my_data, 6)
```

Hide

	Yield <int>	Phi2 <dbl>	PhiNPQ <dbl>	PhiNO <dbl>	SPAD <dbl>
1	1556	0.4017500	0.2746250	0.3237500	45.56125
2	2667	0.2710000	0.4264444	0.3025556	42.81889

	Yield <int>	Phi2 <dbl>	PhiNPQ <dbl>	PhiNO <dbl>	SPAD <dbl>
3	1778	0.2536000	0.4234000	0.3232000	45.60200
4	2222	0.3087500	0.3706667	0.3136667	45.91750
5	2000	0.4288750	0.2431250	0.3280000	42.57625
6	1444	0.4229167	0.2138182	0.3446364	45.34083

6 rows

```
res2 <- rcorr(as.matrix(my_data))
res2
```

Hide

```

      Yield  Phi2 PhiNPQ PhiNO  SPAD
Yield  1.00 -0.28  0.48 -0.26  0.52
Phi2   -0.28  1.00 -0.92 -0.65  0.01
PhiNPQ 0.48 -0.92  1.00  0.31  0.23
PhiNO  -0.26 -0.65  0.31  1.00 -0.49
SPAD   0.52  0.01  0.23 -0.49  1.00

n= 16

P
      Yield  Phi2  PhiNPQ PhiNO  SPAD
Yield           0.2888 0.0603 0.3317 0.0380
Phi2  0.2888           0.0000 0.0067 0.9821
PhiNPQ 0.0603 0.0000           0.2423 0.3816
PhiNO  0.3317 0.0067 0.2423           0.0524
SPAD  0.0380 0.9821 0.3816 0.0524

```

The first correlation matrix presents the correlation coefficients and the second matrix contains the corresponding p values.

There was a significant correlation between crop yield and SPAD, but crop yield was not significantly correlated with any other PhotosynQ parameter.

## Transforming our data with a Mixed Effect Model

Now, we will use a mixed effect model to generate adjusted coefficients for each PhotosynQ parameter in each unique plot. The mixed effect model will take into account light intensity, the time of measurement, the location within the field (i.e. block or replicate), and each unique plot. Our unique plot, which we generated by using the 'concatenate' function in excel, combines the variety and block to give us the unique plot. For example, Variety A + Block 1 = plot A1.

We will use the dataset 'sun,' which is the same dataset used in the "ANOVA and multivariate analysis tutorial."

First, we need to read our data file

Hide

```
# Set the main folder location for your data
your_location = "C:/Users/Dan/Desktop/"
# The main data file, located in the main folder, is called...
file_name = "sun.csv"
# Let's pull that data file into R and call it data_raw
data = read.csv(str_c(your_location, file_name))
View(data)
```

We need to set block to a categorical variable

```
data$Block <- factor(data$Block, ordered = TRUE)
```

Hide

Now we will run a mixed effect model for SPAD. We will set block as a random effect. Plot is the fixed effect whose coefficient we are interested in.

```
SPAD <- lmer(SPAD ~ Plot + (1|Block), data)
```

Hide

```
unable to evaluate scaled gradient Hessian is numerically singular: parameters are not uniquely determined
```

```
summary(SPAD)
```

Hide

```
Linear mixed model fit by REML ['lmerMod']
```

```
Formula: SPAD ~ Plot + (1 | Block)
```

```
Data: data
```

```
REML criterion at convergence: 526.7
```

```
Scaled residuals:
```

Min	1Q	Median	3Q	Max
-4.4318	-0.4194	0.0907	0.5720	1.6051

```
Random effects:
```

Groups	Name	Variance	Std.Dev.
Block	(Intercept)	1.373	1.172
	Residual	30.137	5.490

```
Number of obs: 96, groups: Block, 4
```

```
Fixed effects:
```

	Estimate	Std. Error	t value
(Intercept)	45.56125	2.26713	20.096
PlotA2	-2.74236	3.14025	-0.873
PlotA3	0.04075	3.54118	0.012
PlotA4	0.35625	3.74791	0.095
PlotB1	-2.98500	2.74484	-1.087
PlotB2	-0.22042	3.00400	-0.073
PlotB3	1.12589	3.28906	0.342
PlotB4	-4.01458	4.06917	-0.987
PlotC1	-6.17000	2.74484	-2.248
PlotC2	-1.04292	3.39638	-0.307
PlotC3	-6.63125	4.06917	-1.630
PlotC4	-1.96375	3.74791	-0.524
PlotD1	-4.97125	2.84118	-1.750
PlotD2	-0.44625	3.39638	-0.131
PlotD3	-1.73792	4.06917	-0.427
PlotD4	-8.57792	4.06917	-2.108

```
Correlation matrix not shown by default, as p = 16 > 12.
```

```
Use print(x, correlation=TRUE) or
```

```
vcov(x) if you need it
```

```
convergence code: 0
```

```
unable to evaluate scaled gradient
```

```
Hessian is numerically singular: parameters are not uniquely determined
```

```
SPAD = tidy(SPAD)
setnames(SPAD, "estimate", "SPAD")
write.csv(SPAD, file = "C:/Users/Dan/Desktop/SPAD.csv")
```

Hide

For the photosynthesis parameters (Phi2, PhiNPQ, PhiNO, qL, NPQt, and LEF) we will include sqrtPAR and TimeofDay has fixed effects. Again, block is set as a random effect.

```
Phi2 <- lmer(Phi2 ~ sqrtPAR + TimeofDay + Plot + (1|Block), data)
```

[Hide](#)

```
unable to evaluate scaled gradient Hessian is numerically singular: parameters are not uniquely determined
```

```
summary(Phi2)
```

[Hide](#)

```
Linear mixed model fit by REML ['lmerMod']  
Formula: Phi2 ~ sqrtPAR + TimeofDay + Plot + (1 | Block)  
Data: data
```

```
REML criterion at convergence: -187.7
```

```
Scaled residuals:
```

Min	1Q	Median	3Q	Max
-2.95051	-0.47831	-0.05387	0.53481	2.31369

```
Random effects:
```

Groups	Name	Variance	Std.Dev.
Block	(Intercept)	0.001034	0.03216
Residual		0.003005	0.05482

```
Number of obs: 96, groups: Block, 4
```

```
Fixed effects:
```

	Estimate	Std. Error	t value
(Intercept)	0.8346472	0.0443797	18.807
sqrtPAR	-0.0130663	0.0007015	-18.627
TimeofDay	0.0010404	0.0010073	1.033
PlotA2	-0.0773681	0.0538900	-1.436
PlotA3	-0.0287904	0.0555531	-0.518
PlotA4	0.0099678	0.0568005	0.175
PlotB1	0.0275708	0.0274103	1.006
PlotB2	0.0573877	0.0524106	1.095
PlotB3	-0.0308226	0.0554752	-0.556
PlotB4	0.0454166	0.0590560	0.769
PlotC1	0.0573289	0.0274110	2.091
PlotC2	-0.0164455	0.0560657	-0.293
PlotC3	-0.0378036	0.0594238	-0.636
PlotC4	0.0444594	0.0566751	0.784
PlotD1	0.0014681	0.0283776	0.052
PlotD2	0.0276461	0.0543171	0.509
PlotD3	-0.0304939	0.0588505	-0.518
PlotD4	0.0014337	0.0592383	0.024

```
Correlation matrix not shown by default, as p = 18 > 12.
```

```
Use print(x, correlation=TRUE) or  
vcov(x) if you need it
```

```
convergence code: 0
unable to evaluate scaled gradient
Hessian is numerically singular: parameters are not uniquely determined
```

Hide

```
Phi2 = tidy(Phi2)
setnames(Phi2, "estimate", "Phi2")
write.csv(Phi2, file = "C:/Users/Dan/Desktop/Phi2.csv")
PhiNPQ <- lmer(PhiNPQ ~ sqrtPAR + TimeofDay + Plot + (1|Block), data)
summary(PhiNPQ)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: PhiNPQ ~ sqrtPAR + TimeofDay + Plot + (1 | Block)
Data: data
```

REML criterion at convergence: -117.6

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-3.2005	-0.4525	0.0229	0.4807	2.7235

Random effects:

Groups	Name	Variance	Std.Dev.
Block	(Intercept)	0.0009506	0.03083
Residual		0.0061612	0.07849

Number of obs: 89, groups: Block, 4

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-0.044948	0.053966	-0.833
sqrtPAR	0.010065	0.001029	9.782
TimeofDay	-0.004823	0.001555	-3.101
PlotA2	0.153504	0.060392	2.542
PlotA3	0.057124	0.063182	0.904
PlotA4	0.019630	0.069208	0.284
PlotB1	-0.030705	0.039249	-0.782
PlotB2	-0.048570	0.058052	-0.837
PlotB3	0.085652	0.065148	1.315
PlotB4	-0.025410	0.069402	-0.366
PlotC1	-0.074502	0.042393	-1.757
PlotC2	0.061841	0.067318	0.919
PlotC3	-0.036901	0.070065	-0.527
PlotC4	-0.041492	0.065179	-0.637
PlotD1	-0.001545	0.040634	-0.038
PlotD2	-0.006358	0.060909	-0.104
PlotD3	0.024222	0.076133	0.318
PlotD4	-0.078141	0.069719	-1.121

Correlation matrix not shown by default, as  $p = 18 > 12$ .  
Use `print(x, correlation=TRUE)` or  
`vcov(x)` if you need it

[Hide](#)

```
PhiNPQ = tidy(PhiNPQ)
setnames(PhiNPQ, "estimate", "PhiNPQ")
write.csv(PhiNPQ, file = "C:/Users/Dan/Desktop/PhiNPQ.csv")
PhiNO <- lmer(PhiNO ~ sqrtPAR + TimeofDay + Plot + (1|Block), data)
summary(PhiNO)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: PhiNO ~ sqrtPAR + TimeofDay + Plot + (1 | Block)
Data: data
```

REML criterion at convergence: -151.6

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-2.1356	-0.4583	-0.1241	0.4630	3.9859

Random effects:

Groups	Name	Variance	Std.Dev.
Block	(Intercept)	0.0000675	0.008216
Residual		0.0038155	0.061770

Number of obs: 89, groups: Block, 4

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	0.2171820	0.0358098	6.065
sqrtPAR	0.0028837	0.0008097	3.562
TimeofDay	0.0029660	0.0012239	2.423
PlotA2	-0.0669853	0.0348748	-1.921
PlotA3	-0.0271227	0.0378120	-0.717
PlotA4	-0.0340018	0.0438607	-0.775
PlotB1	0.0032473	0.0308868	0.105
PlotB2	-0.0081475	0.0323204	-0.252
PlotB3	-0.0283234	0.0398242	-0.711
PlotB4	-0.0204652	0.0440500	-0.465
PlotC1	0.0210644	0.0333608	0.631
PlotC2	-0.0123546	0.0419996	-0.294
PlotC3	0.0760843	0.0446960	1.702
PlotC4	-0.0033037	0.0398558	-0.083
PlotD1	0.0001316	0.0319768	0.004
PlotD2	-0.0205350	0.0354265	-0.580
PlotD3	0.0177115	0.0504687	0.351
PlotD4	0.0781634	0.0443591	1.762

Correlation matrix not shown by default, as  $p = 18 > 12$ .

Use `print(x, correlation=TRUE)` or

`vcov(x)` if you need it

[Hide](#)

```
PhiNO = tidy(PhiNO)
setnames(PhiNO, "estimate", "PhiNO")
write.csv(PhiNO, file = "C:/Users/Dan/Desktop/PhiNO.csv")
```

We wrote all of our results into a csv and saved them to our computer. Using these results we will create a new csv file that includes the coefficients of all of our PhotosynQ parameters and crop yield data (you can set the coefficient for plot A1 to 0). We will use this file, taking into account light intensity, time of day, and block, to re-run the correlation matrix we did in the beginning of the tutorial. Hopefully, we will identify more significant relationships between PhotosynQ parameters and crop yield than we did using plot averages!

## Correlations using our new values

We need to read our new dataset, 'sun3.'

```
# Set the main folder location for your data
your_location = "C:/Users/Dan/Desktop/"
# The main data file, located in the main folder, is called...
file_name = "sun3.csv"
# Let's pull that data file into R and call it data_raw
data = read.csv(str_c(your_location, file_name))
View(data)
```

Hide

Now, we will generate a correlation matrix using crop yield and the coefficients from the mixed effects model above.

```
data <- data[c(4,5,6,7,8)]
head(data, 6)
```

Hide

	<b>Yield</b> <int>	<b>Phi2</b> <dbl>	<b>PhiNPQ</b> <dbl>	<b>PhiNO</b> <dbl>	<b>SPAD</b> <dbl>
1	1556	0.00000000	0.00000000	0.00000000	0.00000000
2	2667	-0.07736807	0.15350358	-0.066985307	-2.7423611
3	1778	-0.02879036	0.05712399	-0.027122671	0.0407500
4	2222	0.00996781	0.01963012	-0.034001825	0.3562500
5	2000	0.02757076	-0.03070520	0.003247293	-2.9850000
6	1444	0.05738775	-0.04856991	-0.008147453	-0.2204167

6 rows

```
res2 <- rcorr(as.matrix(data))
res2
```

Hide



	Yield	Phi2	PhiNPQ	PhiNO	SPAD
Yield	1.00	-0.46	0.67	-0.57	0.52
Phi2	-0.46	1.00	-0.77	0.12	-0.07
PhiNPQ	0.67	-0.77	1.00	-0.71	0.53
PhiNO	-0.57	0.12	-0.71	1.00	-0.72
SPAD	0.52	-0.07	0.53	-0.72	1.00

n= 16

P

	Yield	Phi2	PhiNPQ	PhiNO	SPAD
Yield		0.0739	0.0043	0.0206	0.0380
Phi2	0.0739		0.0005	0.6504	0.7864
PhiNPQ	0.0043	0.0005		0.0021	0.0366
PhiNO	0.0206	0.6504	0.0021		0.0015
SPAD	0.0380	0.7864	0.0366	0.0015	

After transforming our data using a mixed effects model, the correlations between PhotosynQ parameters and crop yield are much stronger than when using averages. While the SPAD correlation is unchanged, we now see a significantly positive correlation between PhiNPQ and crop yield and a significantly negative correlation between PhiNO and crop yield.