

ANOVA and Multivariate Analysis

Code ▾

Introduction

Many PhotosynQ users are interested in comparing the performance of different treatments, crop varieties, etc. A common approach to separate different groups is to use Analysis of Variance (ANOVA). However, many of the parameters measured by the MultispeQ device are affected by numerous factors including:

- light intensity
- time of measurement
- temperature (the temperature data on the MultispeQ beta device was not accurate and can be ignored. However, if you are using the new MultispeQ v1.0, you should take temperature into account)
- spatial location in the field, this is usually as a identified by `block` or `replicate`
- leaf age and position in the canopy
- growth stage when data was collected
- instrument/user variation. This could be caused user bias in selection of which leaf to measure or by device to device variation (especially in the MultispeQ Beta instruments, the new MultispeQ v1.0 instruments are much more consistent across instruments)
- this is not an exhaustive list, but rather a summary of the most common factors affecting photosynthesis measurements. In this tutorial, we will compare the performance of 4 varieties of sunflower. We will use the dataset `sun` if you want to follow along with the tutorial.

In this example, four varieties of sunflower were planted in a complete block design with 4 replicates (Blocks). Between flowering and seed fill six upper canopy leaves were measured in each plot. The data collector measured each block at multiple times so that we had a wide range of light intensities and times of day.

Lets get started!

Import required Libraries

First, lets load the libraries that we will need. If you haven't installed those packages already, please install them first.

```
library(PhotosynQ)
library(broom)
library(stringr)
library(dplyr)
```

Hide

```
Attaching package: dplyr
```

```
The following objects are masked from package:stats:
```

```
filter, lag
```

```
The following objects are masked from package:base:
```

```
intersect, setdiff, setequal, union
```

Get Data from PhotosynQ

Now, we need to get the Data from the PhotosynQ platform. More detailed instructions about how to import project data into R studio can be found in the "Import PhotosynQ Data into R" tutorial.

```
# Get data from PhotosynQ
ID <- 243
df <- PhotosynQ::getProject("teravesd@msu.edu", ID)
```

Hide

```
[1] "Successfully signed in as: Dan TerAvest"
[1] "Project: EVALUATION OF SUNFLOWER HYBRID GENOTYPES  "
[1] "Project data received, generating data frame."
```

```
# Create data-frames with only the submitted data and the Top leaves
chlorophyll <- subset(df$`Chlorophyll content (SPAD) I`, status == "submitted" & Leaf.location == "Top")
photosynthesis <- subset(df$`The One Protocol (Phi2, PSI, NPQ) II`, status == "submitted" & Leaf.location == "Top")
# View the data frames
View(chlorophyll)
View(photosynthesis)
# We are calculating the square root of the light intensity (PAR)
photosynthesis$sqrtPAR <- sqrt(photosynthesis$light_intensity)
# We translate the date into a number (0 to 24)
photosynthesis$TimeOfDay <- as.POSIXlt(photosynthesis$time)$hour + as.POSIXlt(photosynthesis$time)$min/60 + as.POSIXlt(photosynthesis$time)$sec/3600
```

Hide

One Way ANOVA

First, we will go thru and analyze the differences in each PhotosynQ parameter using simple One-Way ANOVA.

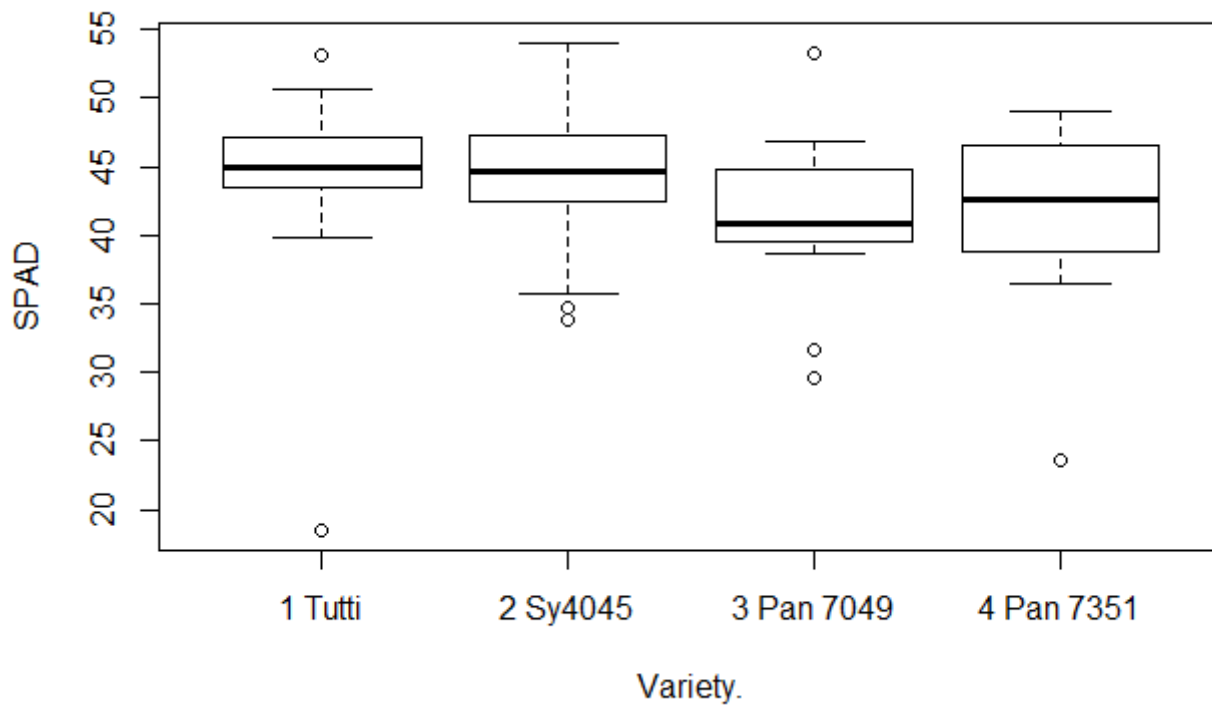
```
SPAD1 <- aov(SPAD ~ Variety., chlorophyll)
summary(SPAD1)
```

Hide

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety.	3	189.2	63.05	2.042	0.114
Residuals	92	2841.4	30.89		

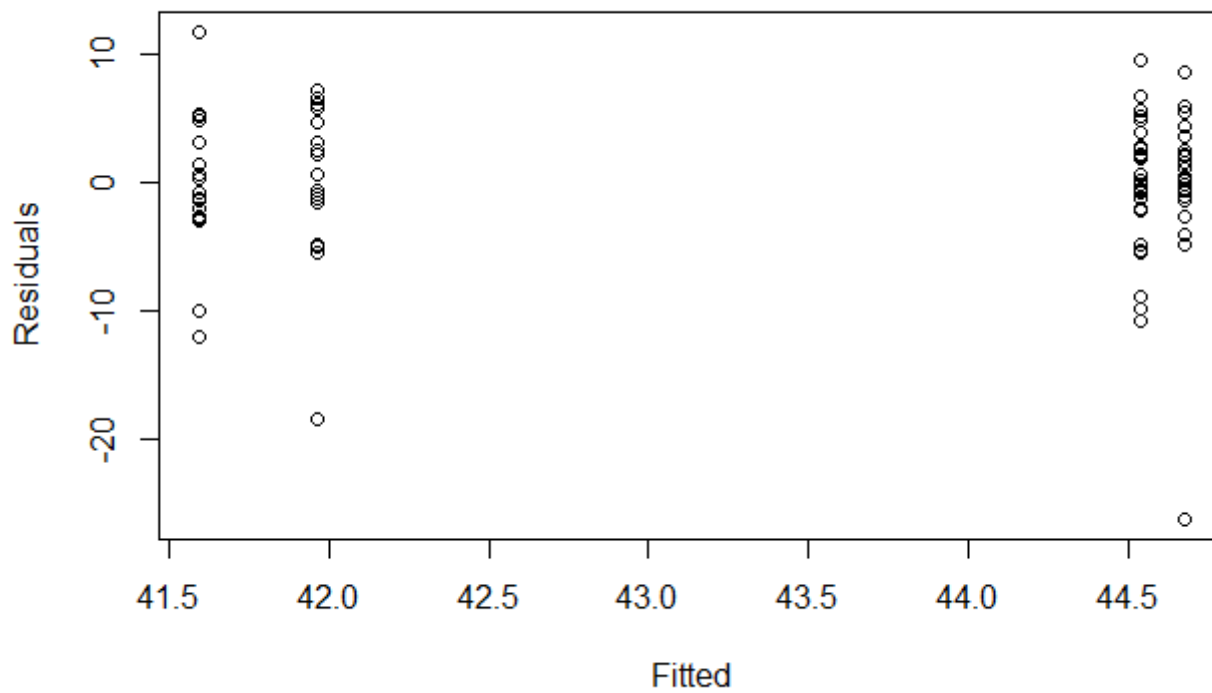
```
plot(SPAD ~ Variety., chlorophyll)
```

Hide



```
plot(SPAD1$fitted,SPAD1$res,xlab="Fitted",ylab="Residuals")
```

Hide



```
TukeyHSD(SPAD1, conf.level = 0.95)
```

Hide

Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = SPAD ~ Variety., data = chlorophyll)
```

```
$Variety.
```

		diff	lwr	upr	p adj
2	Sy4045-1 Tutti	-0.1362821	-4.032665	3.760101	0.9997233
3	Pan 7049-1 Tutti	-3.0831868	-7.349645	1.183272	0.2390753
4	Pan 7351-1 Tutti	-2.7146154	-7.103536	1.674305	0.3734381
3	Pan 7049-2 Sy4045	-2.9469048	-7.084326	1.190516	0.2508821
4	Pan 7351-2 Sy4045	-2.5783333	-6.841924	1.685258	0.3936892
4	Pan 7351-3 Pan 7049	0.3685714	-4.235674	4.972817	0.9967284

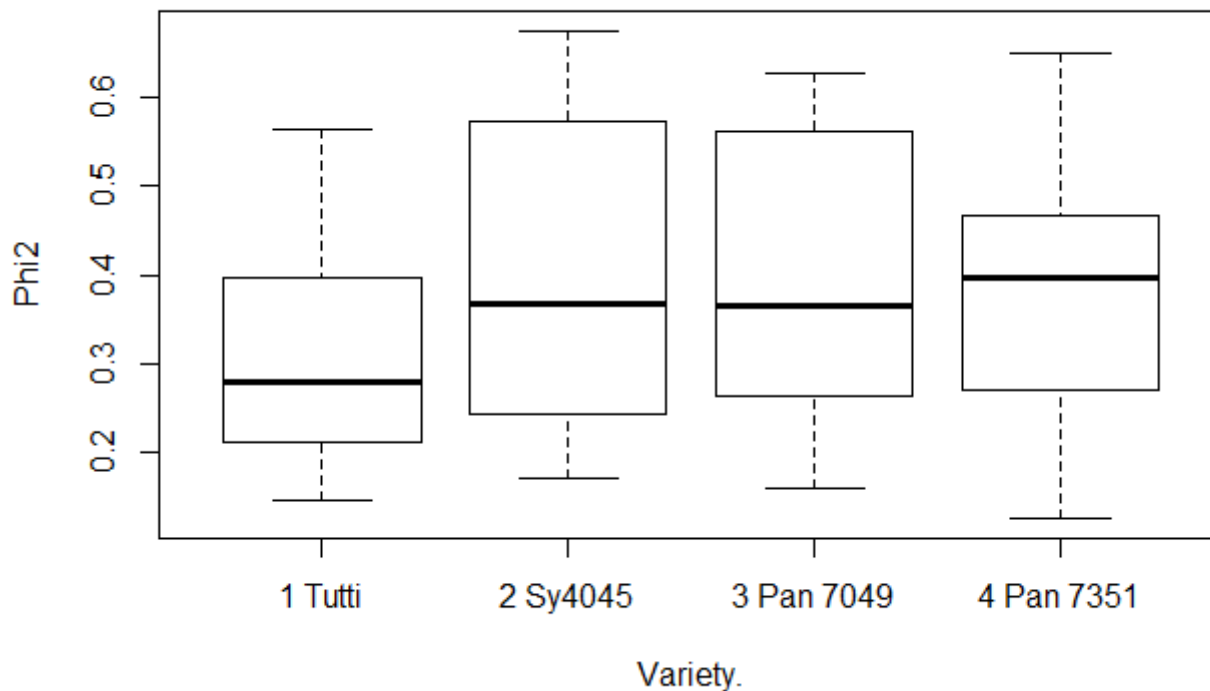
```
Phi21 <- aov(Phi2 ~ Variety., photosynthesis)  
summary(Phi21)
```

Hide

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety.	3	0.1017	0.03391	1.54	0.21
Residuals	92	2.0261	0.02202		

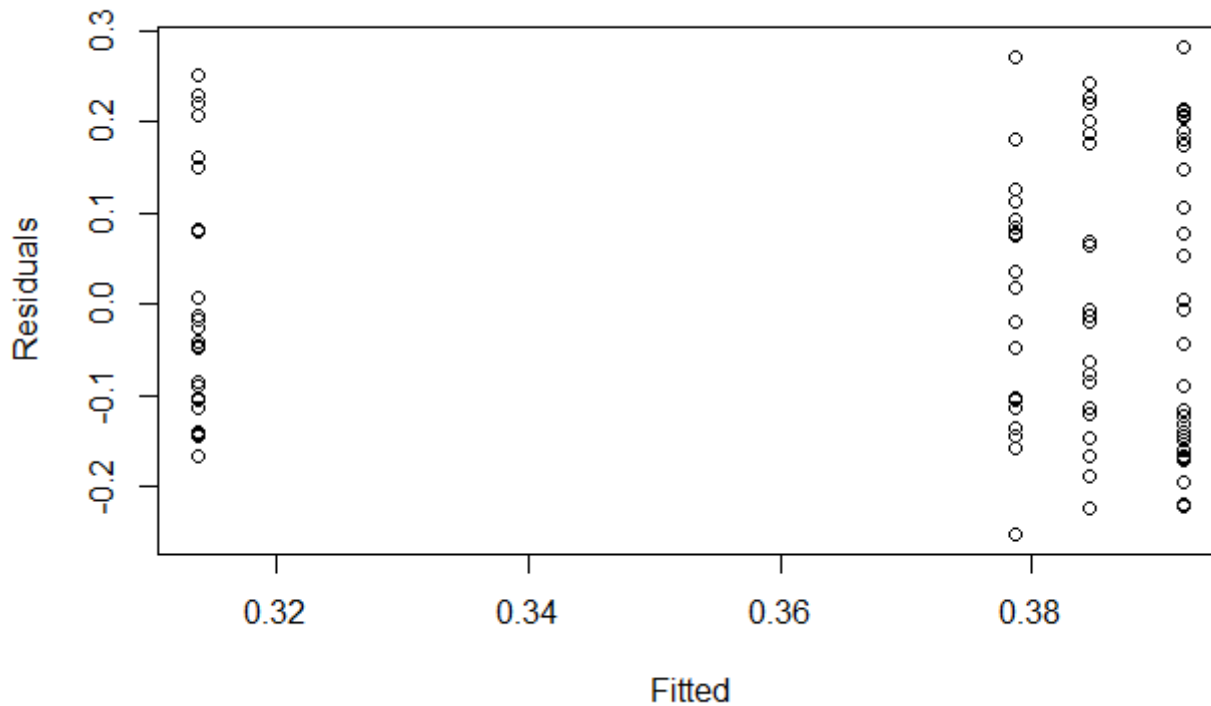
```
plot(Phi2 ~ Variety., photosynthesis)
```

Hide



```
plot(Phi21$fitted,Phi21$res,xlab="Fitted",ylab="Residuals")
```

Hide



```
TukeyHSD(Phi2l, conf.level = 0.95)
```

Hide

Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = Phi2 ~ Variety., data = photosynthesis)
```

```
$Variety.
```

	diff	lwr	upr	p adj
2 Sy4045-1 Tutti	0.078307692	-0.02573828	0.1823537	0.2071488
3 Pan 7049-1 Tutti	0.070879121	-0.04304906	0.1848073	0.3681997
4 Pan 7351-1 Tutti	0.065044534	-0.05215377	0.1822428	0.4704903
3 Pan 7049-2 Sy4045	-0.007428571	-0.11791104	0.1030539	0.9980516
4 Pan 7351-2 Sy4045	-0.013263158	-0.12711476	0.1005884	0.9901053
4 Pan 7351-3 Pan 7049	-0.005834586	-0.12878276	0.1171136	0.9993109

```
PhiNPQ1 <- aov(PhiNPQ ~ Variety., photosynthesis)
summary(PhiNPQ1)
```

Hide

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety.	3	0.2013	0.06710	3.764	0.0137 *
Residuals	85	1.5153	0.01783		

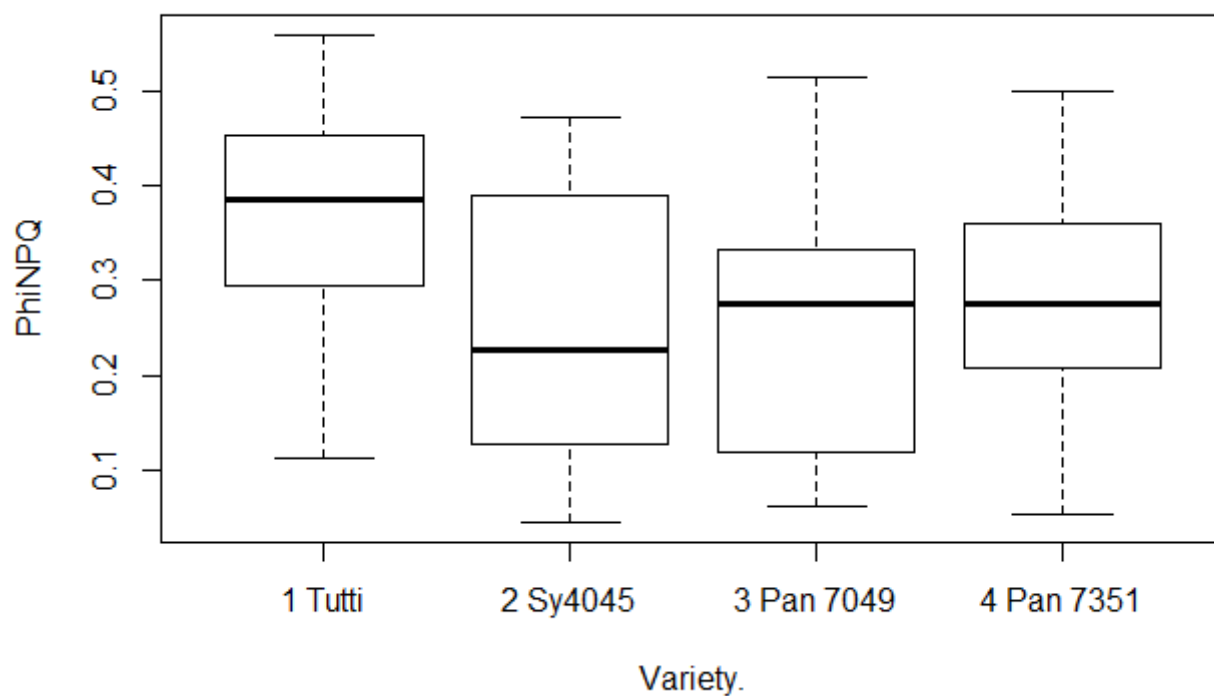
```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
7 observations deleted due to missingness
```

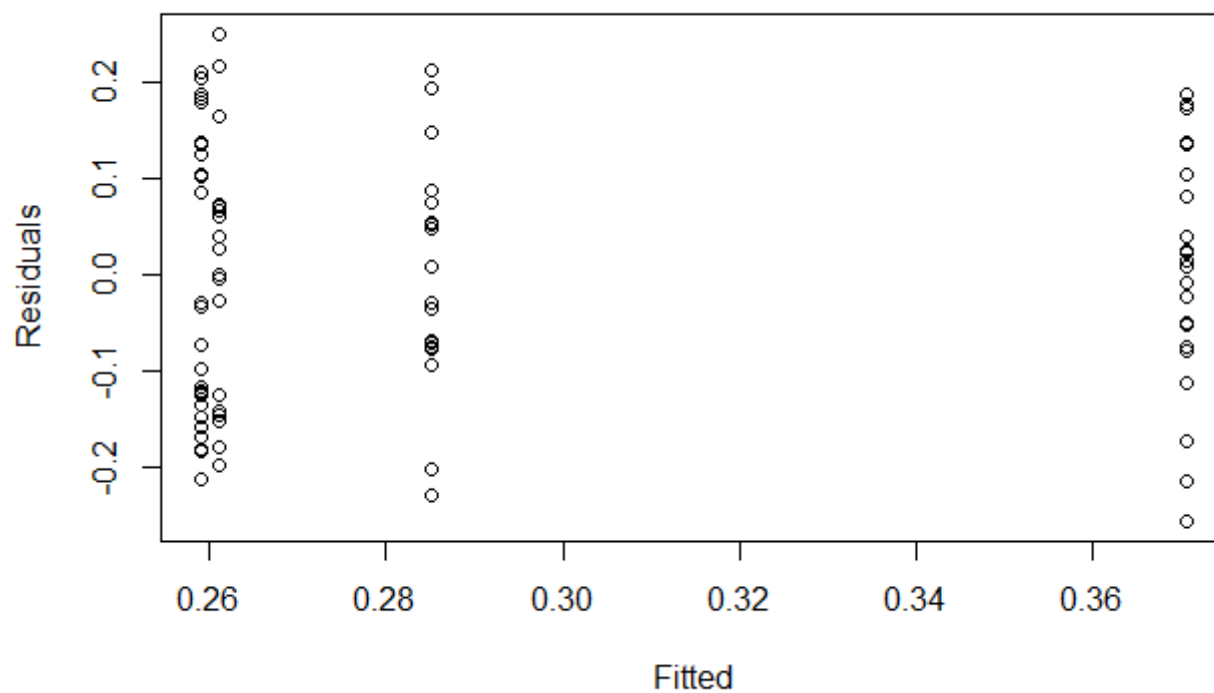
```
plot(PhiNPQ ~ Variety., photosynthesis)
```

Hide



```
plot(PhiNPQ1$fitted,PhiNPQ1$res,xlab="Fitted",ylab="Residuals")
```

Hide



```
TukeyHSD(PhiNPQ1, conf.level = 0.95)
```

Hide

Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = PhiNPQ ~ Variety., data = photosynthesis)
```

```
$Variety.
```

		diff	lwr	upr	p adj
2	Sy4045-1 Tutti	-0.111560000	-0.20783848	-0.015281521	0.0164618
3	Pan 7049-1 Tutti	-0.109393333	-0.21755373	-0.001232936	0.0463926
4	Pan 7351-1 Tutti	-0.085448889	-0.19360929	0.022711509	0.1712169
3	Pan 7049-2 Sy4045	0.002166667	-0.10354038	0.107873716	0.9999439
4	Pan 7351-2 Sy4045	0.026111111	-0.07959594	0.131818161	0.9162081
4	Pan 7351-3 Pan 7049	0.023944444	-0.09268791	0.140576801	0.9494966

```
PhiNO1 <- aov(PhiNO ~ Variety., photosynthesis)  
summary(PhiNO1)
```

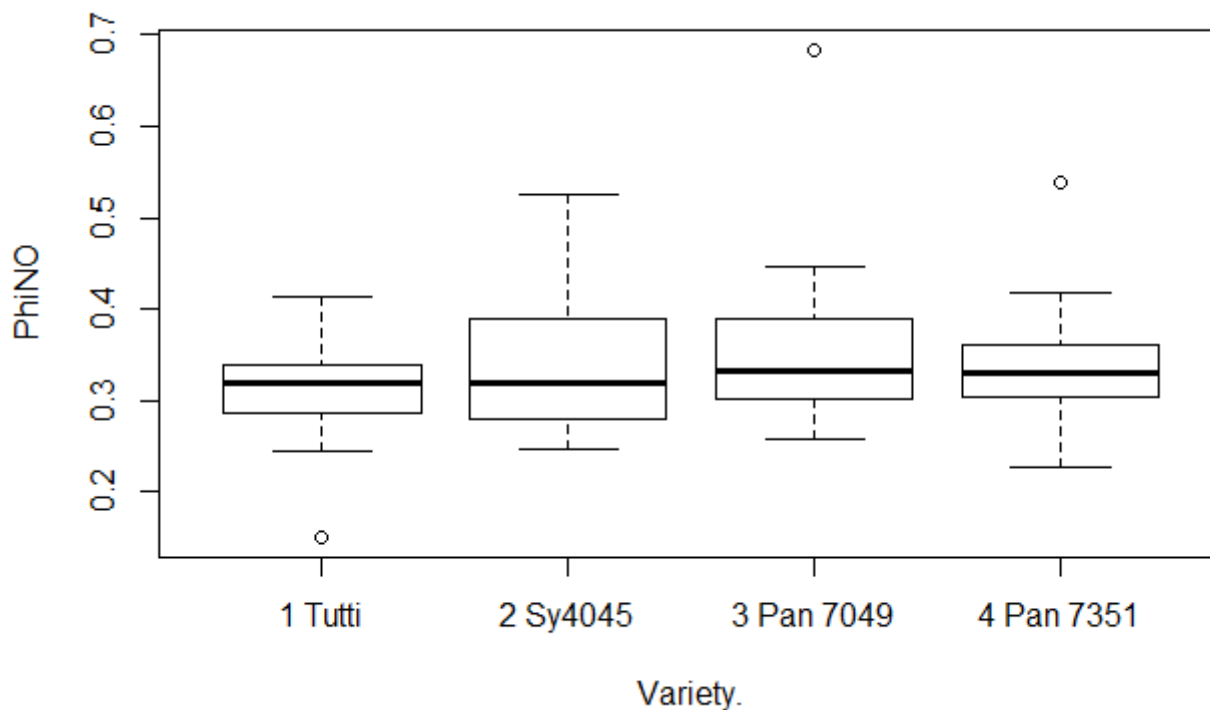
Hide

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety.	3	0.0234	0.007816	1.54	0.21
Residuals	85	0.4314	0.005075		

7 observations deleted due to missingness

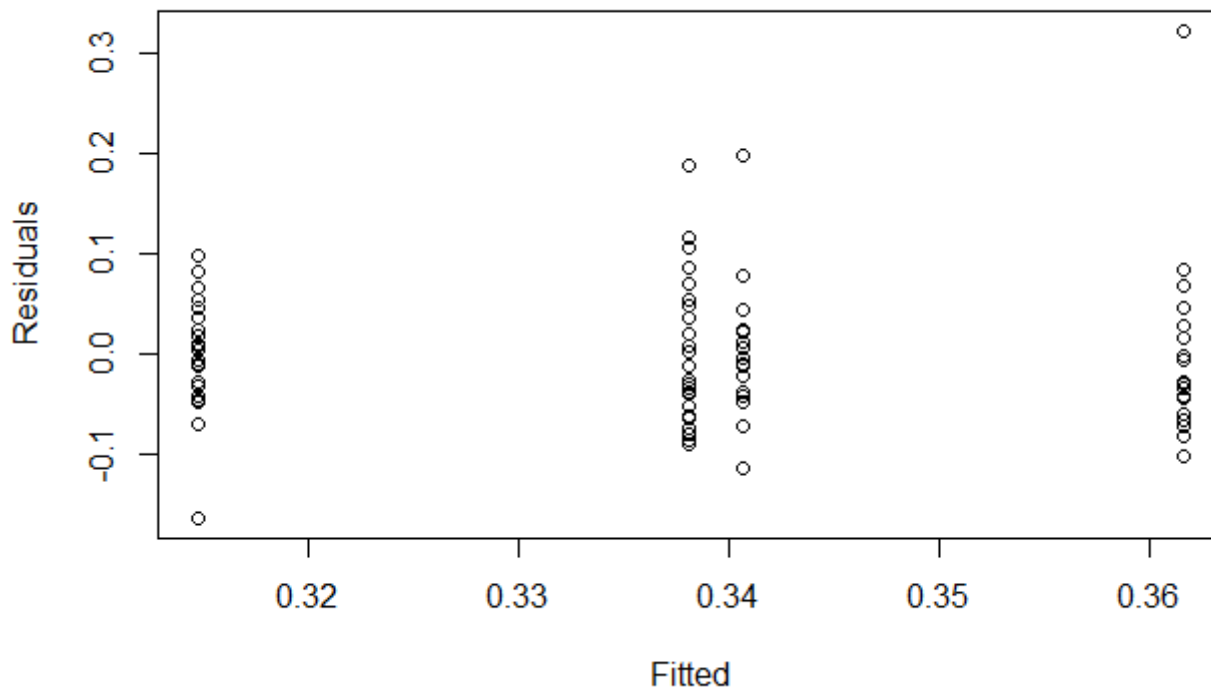
```
plot(PhiNO ~ Variety., photosynthesis)
```

Hide



```
plot(PhiNO1$fitted, PhiNO1$res, xlab="Fitted", ylab="Residuals")
```

Hide



TukeyHSD(PhiNO1, conf.level = 0.95)

Hide

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = PhiNO ~ Variety., data = photosynthesis)

\$Variety.

	diff	lwr	upr	p adj
2 Sy4045-1 Tutti	0.023307143	-0.02806318	0.07467746	0.6354469
3 Pan 7049-1 Tutti	0.046755556	-0.01095448	0.10446559	0.1540925
4 Pan 7351-1 Tutti	0.025866667	-0.03184337	0.08357670	0.6444464
3 Pan 7049-2 Sy4045	0.023448413	-0.03295262	0.07984944	0.6968851
4 Pan 7351-2 Sy4045	0.002559524	-0.05384150	0.05896055	0.9993939
4 Pan 7351-3 Pan 7049	-0.020888889	-0.08311922	0.04134144	0.8153844

Using One-Way ANOVA, we found significant differences in PhiNPQ between varieties. PhiNPQ was significantly greater in Tutti than in Sy4045 and Pan7049. We did not identify any significant differences in SPAD, Phi2 or PhiNO.

Multivariate Analysis

In the next set of analyses below, we will use a multiple linear regression model to account for the factors that effect photosynthesis measurements. By controlling for these covariates we hope to identify additional varietal differences that were not visible above the noise of differing light intensities and measurement times from the above analysis.

In our data set, Blocks are labelled 1 to 4. We need to treat Block as a categorical variable and not a numerical variable

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

Hide

We do not expect light intensity or time of day to affect SPAD measurements, so in this model we add Block as a covariate

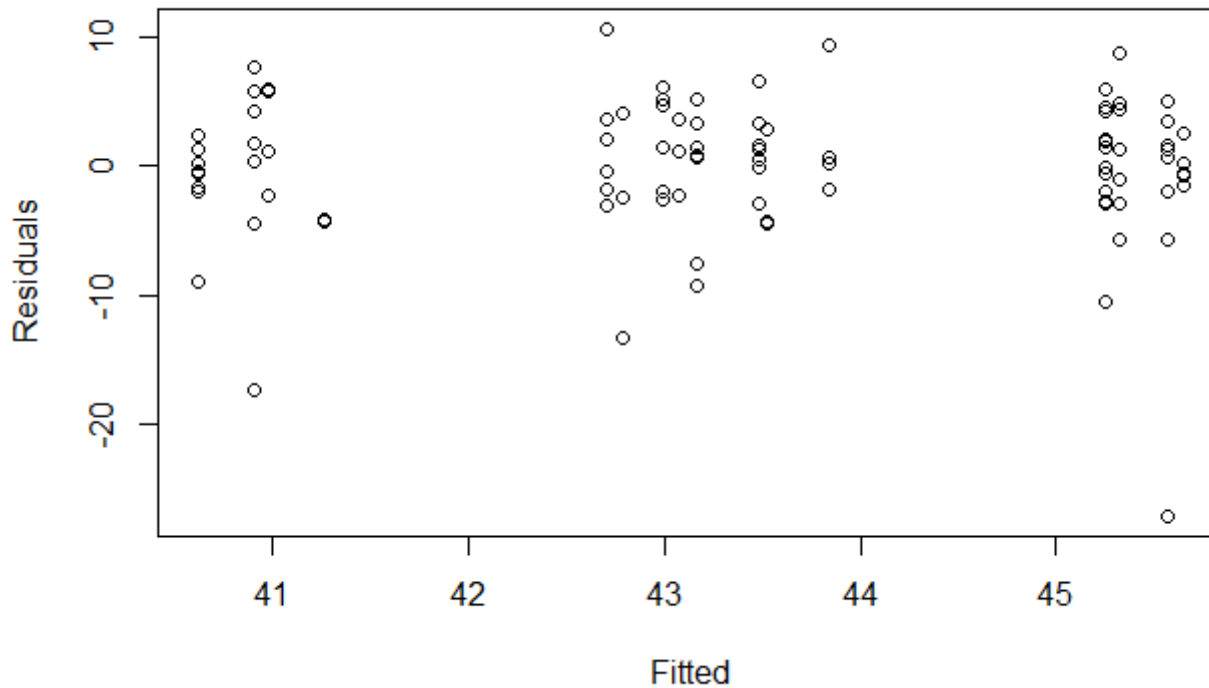
```
SPAD2 <- aov(SPAD ~ Block + Variety., chlorophyll)  
summary(SPAD2)
```

Hide

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	132.0	44.01	1.426	0.240
Variety.	3	152.1	50.68	1.642	0.185
Residuals	89	2746.5	30.86		

```
plot(SPAD2$fitted,SPAD2$res,xlab="Fitted",ylab="Residuals")
```

Hide



```
coef(summary.lm(SPAD2))
```

Hide

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	44.6337448	1.106817	40.32621010	5.729894e-59
Block.L	0.2610971	1.254572	0.20811645	8.356136e-01
Block.Q	-1.9460359	1.225135	-1.58842514	1.157366e-01
Block.C	0.0283575	1.163424	0.02437418	9.806087e-01
Variety.2 Sy4045	-0.3180746	1.493178	-0.21301856	8.317997e-01
Variety.3 Pan 7049	-2.8631212	1.634854	-1.75130089	8.333926e-02
Variety.4 Pan 7351	-2.5781101	1.678641	-1.53583196	1.281272e-01

```
TukeyHSD(SPAD2, which = "Variety.")
```

Hide

Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = SPAD ~ Block + Variety., data = chlorophyll)
```

```
$Variety.
```

	diff	lwr	upr	p adj
2 Sy4045-1 Tutti	-0.3513250	-4.248518	3.545869	0.9953390
3 Pan 7049-1 Tutti	-2.8253784	-7.092724	1.441968	0.3126440
4 Pan 7351-1 Tutti	-2.5560935	-6.945927	1.833740	0.4272960
3 Pan 7049-2 Sy4045	-2.4740534	-6.612335	1.664228	0.4035943
4 Pan 7351-2 Sy4045	-2.2047686	-6.469246	2.059709	0.5316033
4 Pan 7351-3 Pan 7049	0.2692848	-4.335918	4.874488	0.9987120

When analyzing photosynthetic parameters (Phi2, PhiNPQ, PhiNO, qL, NPQt, LEF), we should include both light intensity (PAR) and time of day. Because the effect of light intensity is not linear, we use the square root of light intensity (sqrtPAR) rather than the raw PAR.

```
Phi22 <- aov(Phi2 ~ Block + sqrtPAR + TimeOfDay + Variety., photosynthesis)
summary(Phi22)
```

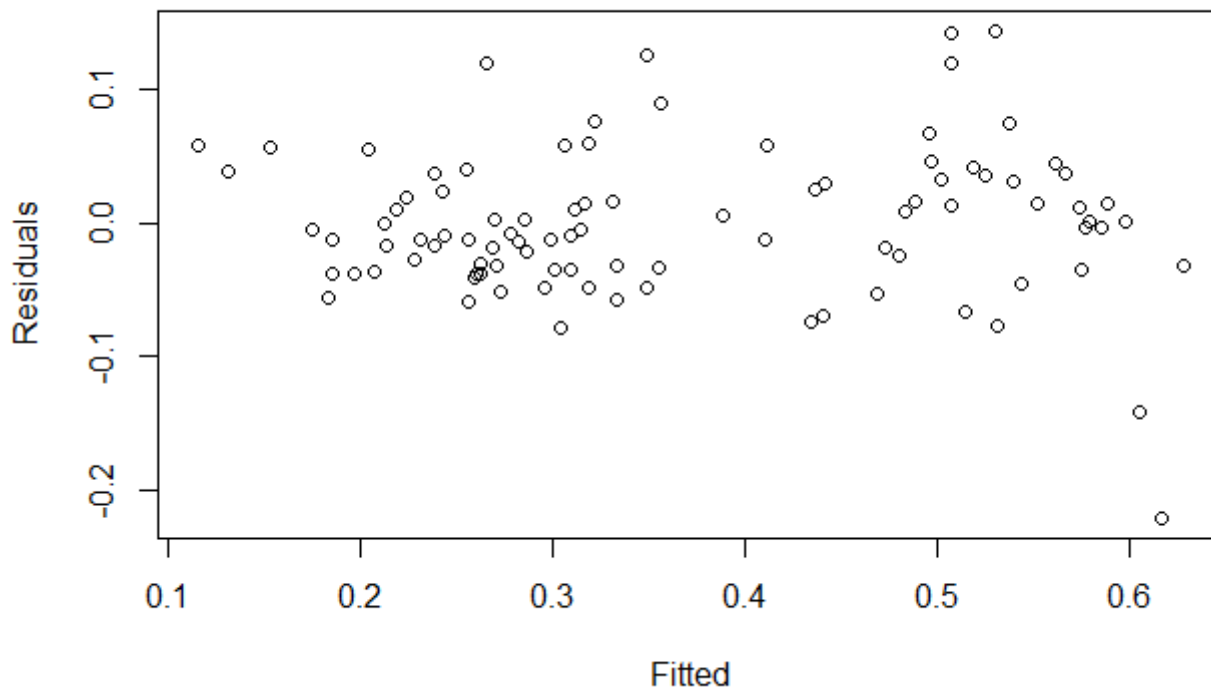
Hide

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	0.3178	0.1059	31.652	6.3e-14 ***
sqrtPAR	1	1.4596	1.4596	436.091	< 2e-16 ***
TimeOfDay	1	0.0000	0.0000	0.002	0.96413
Variety.	3	0.0592	0.0197	5.893	0.00104 **
Residuals	87	0.2912	0.0033		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
plot(Phi22$fitted, Phi22$res, xlab="Fitted", ylab="Residuals")
```

Hide



```
coef(summary.lm(Phi22))
```

Hide

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.8347394762	0.0268861224	31.0472244	7.583721e-49
Block2	-0.0103615074	0.0162937142	-0.6359205	5.264976e-01
Block3	-0.0479020335	0.0183769505	-2.6066367	1.075697e-02
Block4	0.0059893833	0.0186514275	0.3211220	7.488881e-01
sqrtPAR	-0.0132949714	0.0006519689	-20.3920341	6.733577e-35
TimeOfDay	-0.0003849237	0.0009369921	-0.4108078	6.822235e-01
Variety.2 Sy4045	0.0635114141	0.0156515819	4.0578272	1.077998e-04
Variety.3 Pan 7049	0.0475527338	0.0170508739	2.7888737	6.496381e-03
Variety.4 Pan 7351	0.0281567310	0.0177368825	1.5874679	1.160347e-01

```
TukeyHSD(Phi22, which = "Variety.")
```

Hide

```
non-factors ignored: sqrtPARnon-factors ignored: TimeOfDay
```

Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = Phi2 ~ Block + sqrtPAR + TimeOfDay + Variety., data = photosynthesis)
```

```
$Variety.
```

		diff	lwr	upr	p adj
2	Sy4045-1 Tutti	0.06229844	0.021693388	0.10290349	0.0007055
3	Pan 7049-1 Tutti	0.04733887	0.002877175	0.09180056	0.0323343
4	Pan 7351-1 Tutti	0.02910457	-0.016633325	0.07484246	0.3474135
3	Pan 7049-2 Sy4045	-0.01495958	-0.058076540	0.02815739	0.8001958
4	Pan 7351-2 Sy4045	-0.03319387	-0.077625681	0.01123793	0.2124338
4	Pan 7351-3 Pan 7049	-0.01823430	-0.066216140	0.02974754	0.7523991

We see from the output above that sqrtPAR, Block, and Variety had a significant affect on Phi2. This is different from the One-Way ANOVA we conducted above, where there were no significant differences between the Phi2 of different varieties.

Now we have generated new unique coefficients for Sy4045, Pan 7049, and Pan 7351. To calculate adjusted Phi2 values for each Variety we do the following:

1. Determine the mean of Tutti. This mean has not changed from the One-Way ANOVA we ran before. It appears that the mean is about 0.28
2. To calculate the adjusted Phi2 for Sy4045, we add the mean of Tutti to the coefficient for Variety B (0.06).
 $0.28 + 0.06 = 0.34$

Based on the results of this analysis, we see that both Sy4045 and Pan 7049 have significantly greater Phi2 than Tutti.

```
PhiNPQ2 <- aov(PhiNPQ ~ Block + sqrtPAR + TimeOfDay + Variety., photosynthesis)
summary(PhiNPQ2)
```

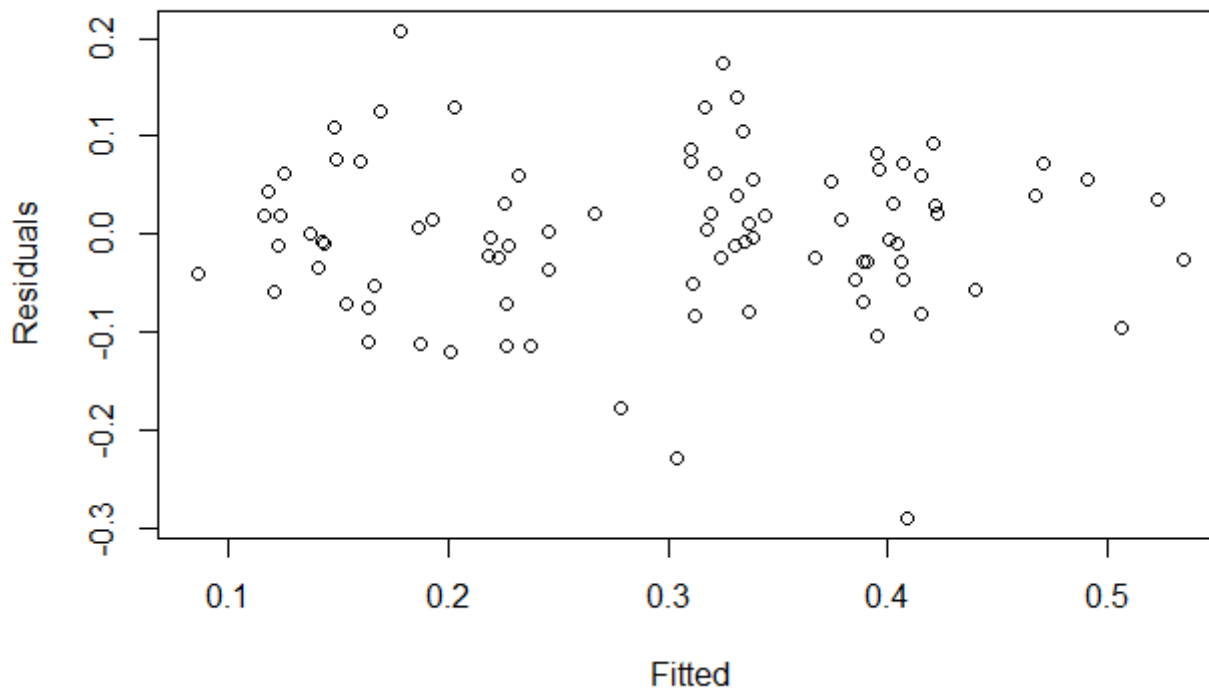
Hide

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	0.2294	0.0765	10.645	5.79e-06 ***
sqrtPAR	1	0.7612	0.7612	105.951	2.60e-16 ***
TimeOfDay	1	0.0235	0.0235	3.275	0.07409 .
Variety.	3	0.1276	0.0425	5.919	0.00106 **
Residuals	80	0.5748	0.0072		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
7 observations deleted due to missingness

```
plot(PhiNPQ2$fitted,PhiNPQ2$res,xlab="Fitted",ylab="Residuals")
```

Hide



```
TukeyHSD(PhiNPQ2, which = "Variety.")
```

Hide

```
non-factors ignored: sqrtPARnon-factors ignored: TimeOfDay
```

Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = PhiNPQ ~ Block + sqrtPAR + TimeOfDay + Variety., data = photosynthesis)
```

```
$Variety.
```

		diff	lwr	upr	p adj
2	Sy4045-1 Tutti	-0.0878396771	-0.14903805	-0.026641307	0.0017612
3	Pan 7049-1 Tutti	-0.0869801976	-0.15573118	-0.018229215	0.0072907
4	Pan 7351-1 Tutti	-0.0669804795	-0.13573146	0.001770503	0.0590187
3	Pan 7049-2 Sy4045	0.0008594795	-0.06633206	0.068051018	0.9999863
4	Pan 7351-2 Sy4045	0.0208591977	-0.04633234	0.088050736	0.8474205
4	Pan 7351-3 Pan 7049	0.0199997181	-0.05413637	0.094135809	0.8937113

In the analysis of PhiNPQ, we actually end up with the same basic results that we got from the One-Way ANOVA above, with Tutti having a significantly higher PhiNPQ than Sy4045 and Pan 7049. However, we see that by including key covariates, the F statistic increased from 3.76 to 5.95 and the p-value decreased from 0.014 to 0.001.

```
PhiNO2 <- aov(PhiNO ~ Block + sqrtPAR + TimeOfDay + Variety., photosynthesis)
summary(PhiNO2)
```

Hide

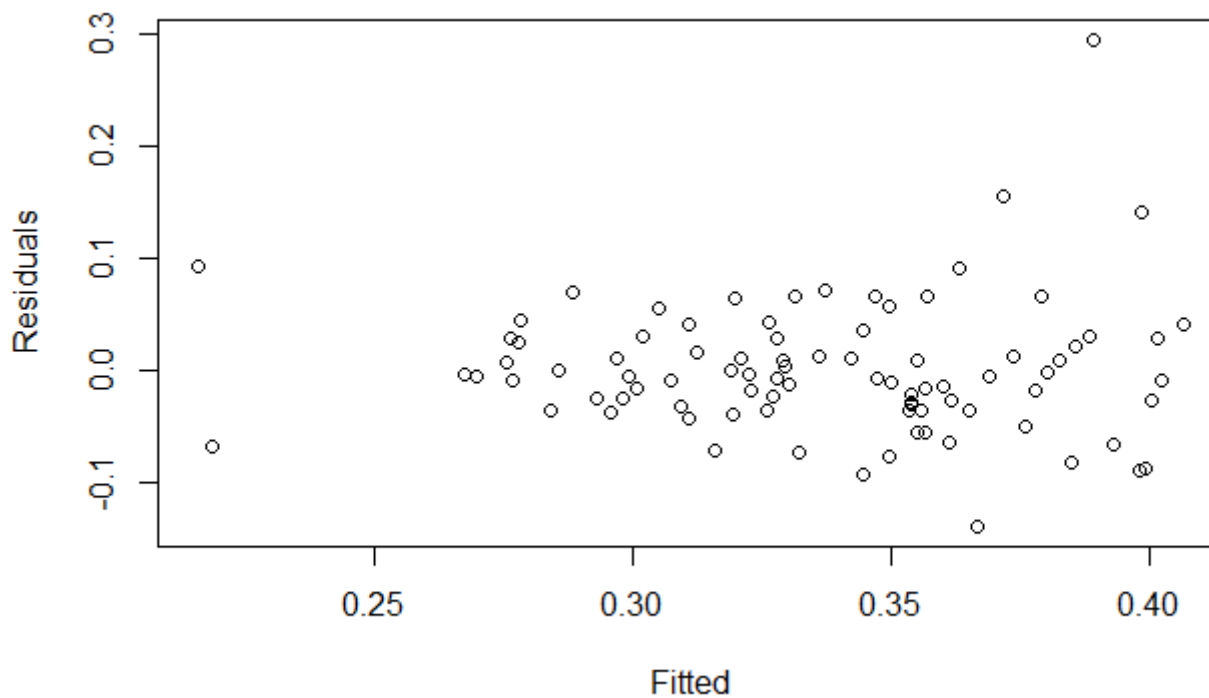
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	0.01223	0.00408	1.051	0.3747
sqrtPAR	1	0.08537	0.08537	22.003	1.11e-05 ***
TimeOfDay	1	0.01840	0.01840	4.742	0.0324 *
Variety.	3	0.02843	0.00948	2.443	0.0702 .
Residuals	80	0.31040	0.00388		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

7 observations deleted due to missingness

```
plot(PhiNO2$fitted,PhiNO2$res,xlab="Fitted",ylab="Residuals")
```

Hide



```
coef(summary.lm(PhiNO2))
```

Hide

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.184939885	0.0296982299	6.2273033	2.068951e-08
Block2	-0.026893389	0.0182344643	-1.4748658	1.441734e-01
Block3	-0.004880776	0.0208608706	-0.2339680	8.156075e-01
Block4	-0.003823646	0.0208636611	-0.1832682	8.550513e-01
sqrtPAR	0.003288345	0.0007302639	4.5029541	2.253612e-05
TimeOfDay	0.002202323	0.0010433043	2.1109118	3.790274e-02
Variety.2 Sy4045	0.026373384	0.0173479693	1.5202577	1.323889e-01
Variety.3 Pan 7049	0.048205101	0.0194344324	2.4803966	1.522222e-02
Variety.4 Pan 7351	0.039018296	0.0195731703	1.9934582	4.961971e-02

```
TukeyHSD(PhiNO2, which = "Variety.")
```

Hide

```
non-factors ignored: sqrtPARnon-factors ignored: TimeOfDay
```

```
Tukey multiple comparisons of means  
95% family-wise confidence level
```

```
Fit: aov(formula = PhiNO ~ Block + sqrtPAR + TimeOfDay + Variety., data = photosynthesi  
s)
```

```
$Variety.
```

		diff	lwr	upr	p adj
2	Sy4045-1 Tutti	0.026347396	-0.01862487	0.07131967	0.4203897
3	Pan 7049-1 Tutti	0.047556017	-0.00296637	0.09807840	0.0725421
4	Pan 7351-1 Tutti	0.038434141	-0.01208825	0.08895653	0.1980946
3	Pan 7049-2 Sy4045	0.021208621	-0.02816779	0.07058503	0.6740263
4	Pan 7351-2 Sy4045	0.012086745	-0.03728967	0.06146316	0.9179295
4	Pan 7351-3 Pan 7049	-0.009121876	-0.06360157	0.04535781	0.9714635

Similar to SPAD, there were no significant differences when using multivariate analysis to analyze PhiNO.

Conclusions

When we used a multivariate analysis to account for the factors that affect PhotosynQ parameters, we were able to identify differences that were not apparent using One-Way ANOVA. This can be critical when interpreting your results. In this example, using One-Way ANOVA we know that PhiNPQ was greater in Tutti than Sy4045 and Pan 7049. However, we cannot determine the cause of that increase. When we used the more sophisticated multivariate analysis, we now see that, Phi2 was also greater in Sy4045 and Pan 7049 than in Tutti.

Therefore, we now know that Tutti was diverting more light energy away from photosystem II, which is what accounted for the greater PhiNPQ. In the next tutorial we will use correlation matrices and mixed effect models to determine how PhotosynQ parameters are related to crop productivity.