

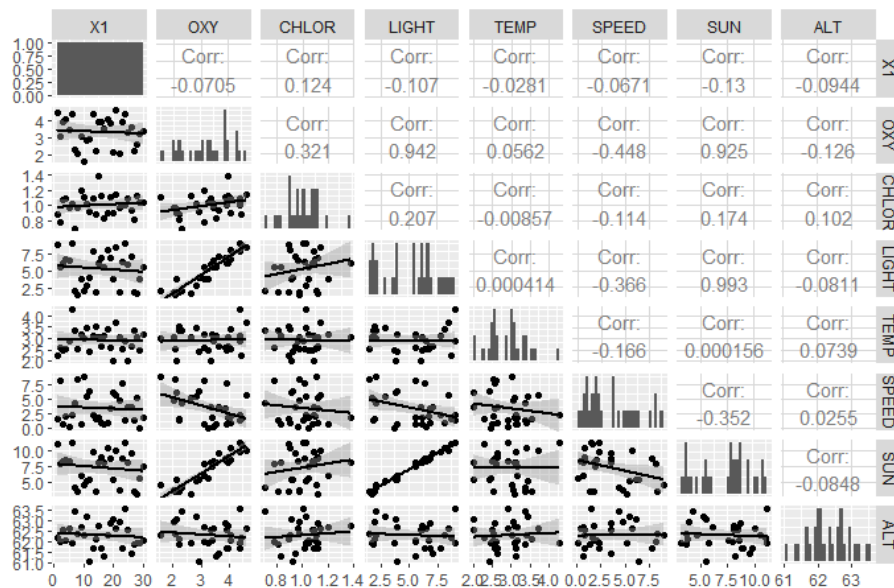
# ADSSM\_2

author : Ziliang Zhou Student Number: 690052545

## a. exploratory data analysis

Explore the relationship in explanatory variables and production of oxygen.

Using the scatter plot explore the relationship.



This plot shows the **Sun**, **chlorophyll** and **Light** has significant positive influence of oxygen production. The **Speed** have little negative influence of oxygen production. There are not obverse influence in other variable.

And it also shows the **Sun** and **Light** have highly positive relationship.

## b. build a model

Based on problem **a** , I build a liner module about oxygen production. Because of variable **Sun** and **light** have liner relationship, I only choose variable **Sun** in my model.

$$OXY = \beta_0 + \beta_1 * Chlor + \beta_2 * Sun$$

Getting a result by Using R to calculate the module. The model is :

$$OXY = 1.05Chlor + 0.31Sun$$

## c. report

In my model, I choose three variable to explain the oxygen production. Using liner regression to estimate how the oxygen production been influenced by this variable. Using the summary we can see the p value and R-square of each variation. Both **Chlor Speed** and **Sun** have p-value < 0.05. And the  $R^2$  of 88.2% and adjusted  $R^2$  of 87.3% tells us, this model explains more than 80% of oxygen production.

	ESTIMATE	STD. ERROR	T VALUE	PR(>T)
(Intercept)	-0.004289288	0.44036776	-0.00974024	9.923001e-01
CHLOR	1.055572804	0.42986188	2.45560924	2.078765e-02
SUN	0.309412243	0.02319611	13.33897072	2.125494e-13

I did not choose the Light as a explanatory variables, because it has significant relationship with Sun, if I add it to my model, the model will become incorrect. Using chisq.test their p-value is 0.2372, smaller than 0.8.

Other two factor **Temp** and **ALT** was not choose by me. We can build another model to exam the relationship with **Oxy**. R shows the adjust  $R^2$  of **Temp** is -0.03 and of **Alt** is -0.02. That means neither of this two variables can explain the oxygen production

## Appendix

a.

```
library(tidyverse)
library(GGally)
oxygenfull <- read_csv("oxygenfull.csv")
ggpairs(oxygenfull,
        lower = list(continuous = "smooth"),
        diag=list(continuous="barDiag"),
        axisLabels='show')
```

b.

```
mod_1 <- lm(formula = OXY ~ 1 + CHLOR + SUN + SPEED,data =
oxygenfull)
result <- summary(mod_1)
```

c.

```
#exam whether the light and sun are independent or not.
```

```
chisq.test(oxygenfull$LIGHT,oxygenfull$SUN)
```

```
mod_2 <-lm(formula = OXY ~ TEMP,data = oxygenfull)
```

```
mod_3 <-lm(formula = OXY ~ ALT, data = oxygenfull)
```

```
mod_4 <-lm(formula = OXY ~SPEED,data = oxygenfull)
```

```
summary(mod_2)$adj.r.squared
```

```
summary(mod_3)$adj.r.squared
```

```
summary(mod_4)$adj.r.squared
```