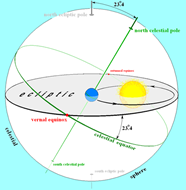
**Final Report-Group6**

*[Part1. Introduction, Background Theory, and Question Refinement]*

The question of whether the earth is round or flat has been much debated throughout history. These days, the majority of people believe that the earth is round while we wonder what made people have no doubts in believing that idea. One possible reason is that people can easily observe the earth's round-shaped shadow on the moon during the lunar eclipse. Also, by measuring shadow casting from two different cities, the spherical shape of the earth can be proved. Furthermore, the earth images taken by satellite directly gives us the fact that the earth is round. Last but not least, the sunrise time can actually prove the theory that the earth is round. In this project, we will use this evidence to prove that the earth is round.

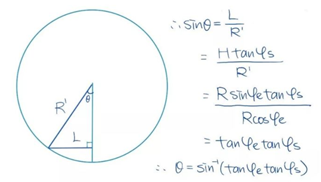
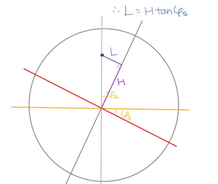
How can we connect the sunrise time with the round earth theory? To begin this test, the first thing we have to find is what features about the sunrise will be different between the flat earth theory and the round earth theory. Before we start, we have to know the relationship between the sunrise time, longitude, and latitude.

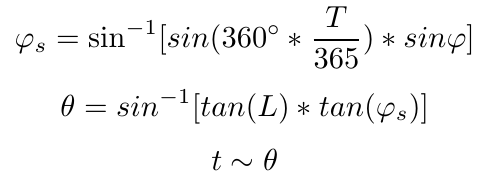


To begin with, we need some basic theories. According to the heliocentric theory, the earth auto-rotates around the sun one round per day. The equator has a tilt with the earth’s orbit, which makes four seasons. The angle of the tilt is called the Obliquity of the Ecliptic, which equals 23.44˚, written as φ.

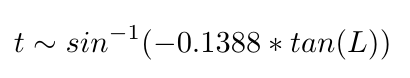
During the rotation around the sun, the angle between the equator plane and the sun-earth line changes. It equals 0 at the vernal equinox (about March 20th) and the autumnal equinox. The angle equals the positive/negative Obliquity of the Ecliptic when it is the summer solstice and winter solstice. This angle is called the Sun Declination angle, written as , it has: In this equation, T means the number of days from the vernal equinox. 

No matter what date it is, the equator is always equally divided by day and night, which means that the sunrise time is always at 6 am on the equator. But on the other latitude, the sunrise time depends on the sun's declination angle.



In this equation, θ means a given location that passes the terminator at the same time as the equator which has a longitude difference of θ. We then can calculate the sunrise time and correct the time zone difference by:

where: T: number of days from the vernal equinox, L: latitude of the location, φ: Obliquity of the Ecliptic, 23.44˚ : Sun Declination angle t: sunrise time, θ: a given location passes the terminator at the same time with someplace on the equator which has a longitude difference of θ

For Oct 10th, where T=203, we can get: 

(notice that L > 82 degrees is out of the domain of this equation because a place north than 82N will have a polar night on Oct 10th which means that there’s no sunrise time)

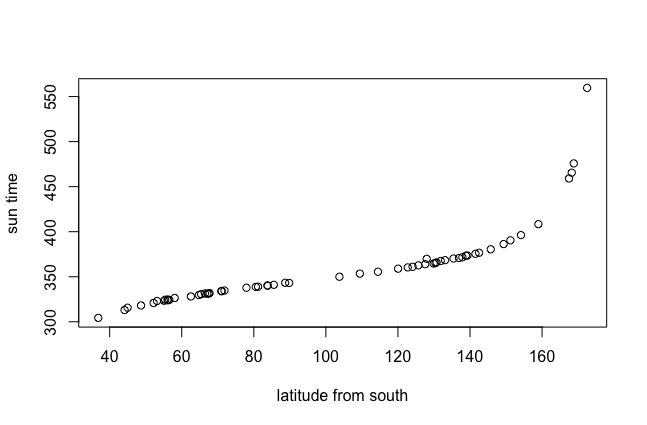
Based on the theory, the sunrise time has a clear relationship with the latitude. If the earth is flat and the sun moves in a circular motion, locations with the same longitude will have the same sunrise time. Thus, the latitude will not have any influence when they have the same longitude. However, if we only use one explanatory factor, we cannot build an accurate model since there are many other factors that influence the sunrise time. So we refined the question into, **If the earth is round, are we able to predict the sunrise time of different cities?** Then we are going to design an experiment, find other explanatory variables, build the model, and finish the prediction.

*[Part2. The Design Of Experiment]*

**Step 1. Model Design**

Our group assumed that the earth is round and the latitude affects the sunrise time. Then we found proof to explain our assumption by adding two variables: altitude and light refraction rate. First, the altitude influences the time of sunrise. This is easy to understand with an example of one person standing on the top of a cliff near the sea, and another person standing at the bottom of the cliff. Based on the triangle this produces, when the sun rises from the horizon, the person who is on the top of the cliff will observe the sun first. In another word, it means that the place with a higher altitude will have an early sunrise time. Second, the refraction rate influences the sunrise time because there is water vapor, air, and other substances in the atmosphere. Therefore, the refraction of light will cause us to see the sun before the sun has risen from the horizon. In other words, the refraction rate will affect the sunrise time.

Then we need to check the explanatory variables of the refraction rate. We give an assumption that the refraction rate is only influenced by the atmospheric pressure and temperature, and the atmospheric pressure is only influenced by the altitude, temperature, and altitude. In order to avoid the complex collinearity between variables, we decided to separate the refraction rate and altitude into temperature, altitude, and humidity. Then we decided to use these factors to build a regression model.

The next step is to build two separated linear models predicting sunrise time. Given that the earth is tilted by 23.5 degrees, the change of latitude from the southern hemisphere and northern hemisphere would affect the sunrise time at a different rate. With our datasets, we were able to observe this fact by plotting the graph below. Due to this non-linear relationship, this feature should be adjusted so that they are linearly related, approximately taking the logarithm function. Finally, our group used this transformed sunrise time as a response variable to build the model. In the next step, our group will discuss specifically how we collected and processed these features used in modeling.

**Step2. Data Collection & Processing**

Our group chose 30 cities from each hemisphere and collected 7 different types of data (Latitude, Longitude, Standard Local Time Zone, Sunrise Time, Temperature, Altitude, and Humidity), based on Oct 10, 2021.

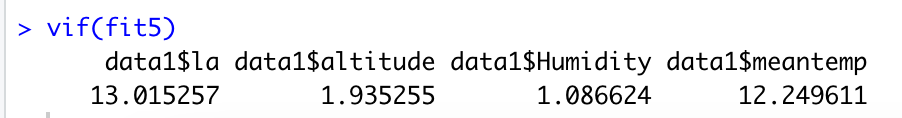
While our group used the latitude without transformation, the longitude is used to make a new derived variable combined with the standard time zone. The below Equation-(a) is a derived variable, and we name this *‘longdiff’*. The reason why we need to derive this variable is that the latitude of a city is not exactly aligned with the latitude line of that time zone. So, the *longdiff* represents the gap between the reference longitude and a city’s longitude. Based on this, our group calculated the real sunrise time using the Equation-(2) to reflect the gap. We name the result feature as ‘adjusted sunrise time(adjsun)’ and use this as our response variable.

**Step3. Modeling**

Let us use the model on the northern hemisphere to give an example. From the *adjsun* and *temperature, latitude, altitude, humidity*, our group built a linear model predicting adjusted sunrise time. The value of *adjsun* field is transformed by taking logarithms, and we will use ‘minutes’ instead of ‘hours’ to avoid confusion. For example, the 5:00 AM is transformed into 300Minutes. The below is our first linear regression:

Linear Model:

In the traditional linear model, all the factors have to be independent of each other. However, on the same date in the northern hemisphere, there is higher latitude when the temperature is low. And we will meet the same problem between altitude and sunrise time. Thus, we have to consider the interaction term, multicollinearity, because the factors are dependent. Multicollinearity refers to the distortion or difficulty in estimating accurate model estimates due to the presence of exact or high correlations between the explanatory variables in a linear regression model. We use vif(Variance Inflation Factor) function to test the multicollinearity. The higher vif test result, the higher colinearity we will have.



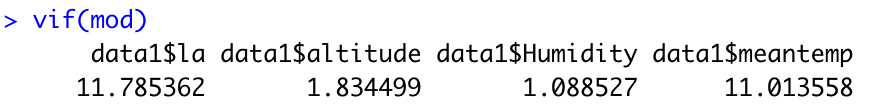
There is clearly collinearity between latitude and meantemp. In order to reduce the effect of collinearity, we decided to use ridge regression to solve this problem.

Ridge regression analysis is a biased regression method dedicated to the analysis of covariance data. It is essentially a modified least-squares estimation method, which obtains more realistic and reliable regression coefficients by abandoning the unbiased nature of the least-squares method, at the expense of losing some information and reducing accuracy.

Ridge regression can reduce the mean square error by sacrificing unbiasedness in exchange for a

significant reduction in the variance component. We build a ridge regression model like this:

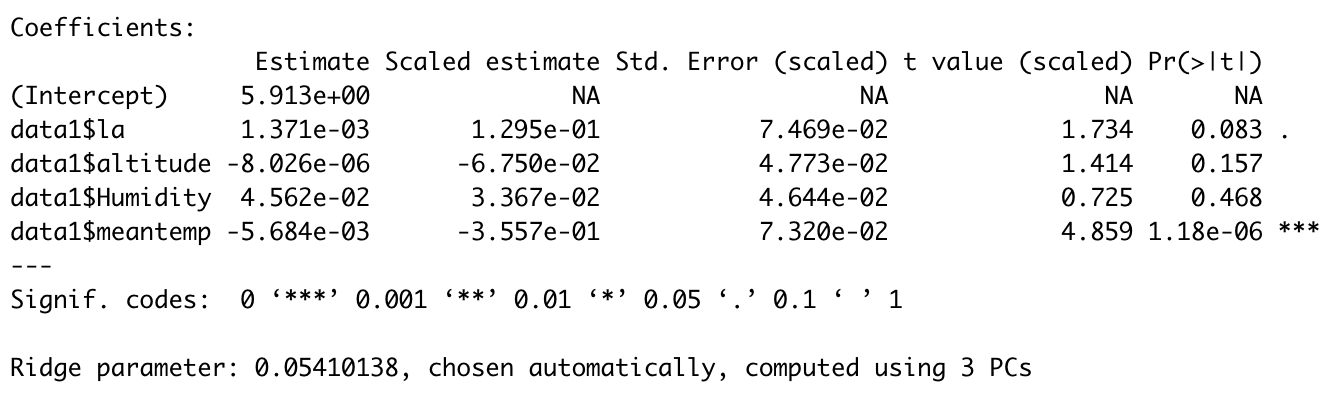
Ridge Regression:

**

Through the ridge regression, the colinearity between the factors has been reduced. Although it is more than 10, it still has multicollinearity, we think the result is better than the traditional model.

*[Part3 Result & Interpretation]*

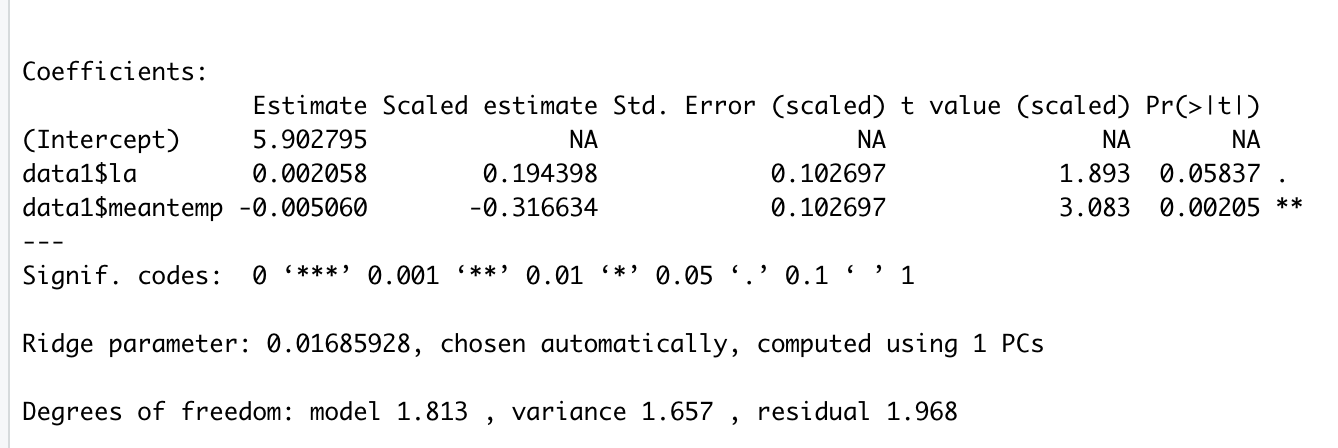
Through the ridge regression, the parameter of different factors are shown below:



Through the result, the P-value of altitude and humidity is significantly bigger than 0.1, so we can think that this is not significant on this model, so we delete these two explanatory factors, to fit another model.

Ridge Regression:

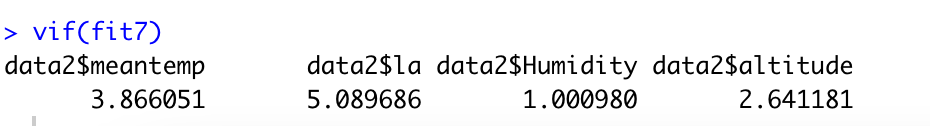
So, the final model for the northern hemisphere is shown below.



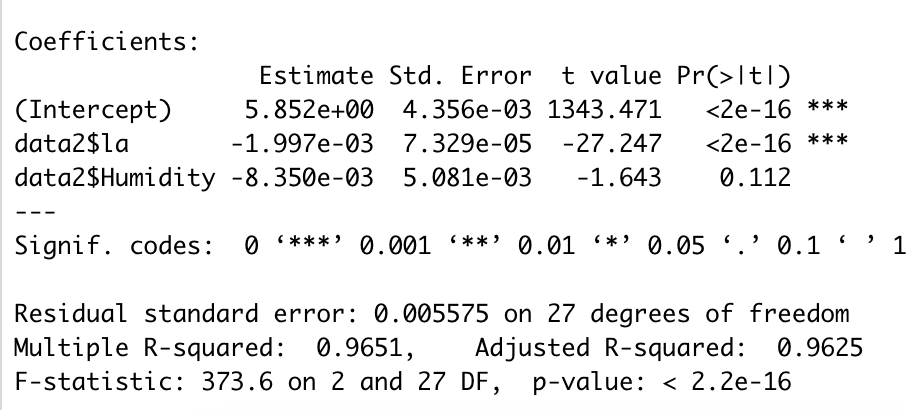
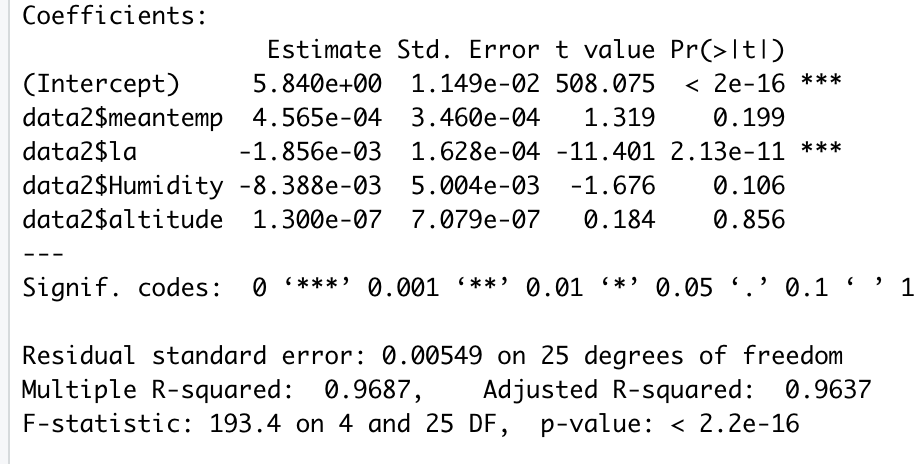
In this model, we can use two parameters to simulate the log(adjsun). We can use meantemp(the mean of temperature in Oct), latitude, to predict the sunrise time.

In the northern hemisphere, if we keep other factors constant, on Oct 10 on the equator, the temperature is 20ºC. When the latitude increases by 1º, the adjsun(local sunrise time) will be delayed by about 1.89 minutes. When the meantemp(adjust temperature) decreases by 1ºC, the *adjsun*(local sunrise time) will be delayed by 0.78 minutes. Thus, through the result on the northern hemisphere, we can find a clear relationship between sunrise time and latitude.

Similarly, the data on the southern hemisphere can be answered in this way, but we find that the collinearity on the southern hemisphere is not significant. The vif test result was less than 10, so we use a linear model to get the answer.



Linear Model:



In the southern hemisphere, with the P-value of 0.199, we thought that *meantemp* do not significant in the result. The same reason for the altitude. So we deleted them and used a reduced model.

Linear Model:

The result is shown on the right. So the final model is:

Linear Model: .

In the southern hemisphere, if we keep other factors constant, on the equator, humidity is 80%. When latitude increases by 1ºC, the sunrise time will be delayed for 0.67 minutes. When humidity increases 10%, the sunrise time will be delayed by 0.28 minutes.

We chose Harare, to check this model. The location of Harare is 17.8216S,31.0492E, sunrise time is 5:30, time zone is GMT+2, altitude is 4888 feet, humidity is 40%, and mean temperature is 24ºC. Through the regression model, the log(adjsun)=5.78, then the *adjsun* is 323.76, about 6:23AM. We used the data that we collected to calculate the *adjsun*. The real *adjsun* is 334.19, about 6:34AM. The error is around 11 minutes, 3.4%, so we think this result is acceptable. We can conclude that the model predicts the sunrise time.

Based on our model, if we constrain the influential factor and explanatory factors and only consider the collinearity between temperature and latitude, combine the temperature and altitude to simplify this model, do not consider the collinearity between vectors after we use the ridge regression and the vif test result approximately less than 10. Then we can give a conclusion.

After getting the model to predict the sunrise time, we find that the latitude has a significant influence on the sunrise time. In the northern hemisphere, when the latitude increases by 1º, the sunrise time will be delayed about 1.89 minutes. In the southern hemisphere, when latitude increases by 1º, the sunrise time will be delayed about 0.67 minutes. Through the result, we can provide the conclusion that the latitude will influence the sunrise time. Then, based on our model, we give the answer to the original question that the earth is round.

**Code**

----------------------R Code -----------------------

library(readxl)

library(car)

library(carData)

library(ridge)

#Northern Hemisphere

data1<-read\_excel("sunrisedata.xlsx", sheet=1)

plot(data1$la, log((data1$adjsun)-340))

fit5<-lm(log(data1$adjsun)~data1$la+data1$altitude+data1$Humidity+data1$meantemp,data=data1)

fit5

summary(fit5)

vif(fit5)

#Ridge Regreesion#

mod<-linearRidge(log(data1$adjsun)~data1$la+data1$altitude+data1$Humidity

+data1$meantemp,data=data1)

mod

summary(mod)

vif(mod)

#Regression with the reduced model#

mod1<-linearRidge(log(data1$adjsun)~data1$la

+data1$meantemp,data=data1)

mod1

summary(mod1)

#Southern Hemisphere

data2<-read\_excel("sunrisedata.xlsx", sheet=2)

plot(data2$la, log(data2$adjsun))

fit7<-lm(log(data2$adjsun)~data2$meantemp+data2$la+data2$Humidity+data2$altitude,data=data2)

summary(fit7)

vif(fit7)

#The reduced model

fit8<-lm(log(data2$adjsun)~data2$la+data2$Humidity,data=data2)

summary(fit8)

vif(fit8)

**Reference:**

1.HK Observatory 2017, accessed 02 Oct 2021,

<<https://www.hko.gov.hk/tc/education/astronomy-and-time/astronomy/00493-effect-of-atmosph-eric-refraction-on-the-times-of-sunrise-and-sunset.html>>

2.En.wikipedia.org. 2019, accessed 02 Oct 2021,

[https://en.wikipedia.org/wiki/Sunrise\_equation](https://link.zhihu.com/?target=https%3A//en.wikipedia.org/wiki/Sunrise_equation) .

3.TimeAndDate.com. 2021, accessed 06 Oct 2021, <https://www.timeanddate.com/sun/china/beijing>

4.WorldWeatherOnline 2021, accessed 07 Oct 2021,

<https://www.worldweatheronline.com/parque-nacional-los-alerces-weather/chubut/ar.aspx>

5. Google Earth 2021, accessed 12 Oct 2021,

<https://earth.google.com/web>

6.WeatherUnderground 2021, accessed 12 Oct 2021,

<https://www.wunderground.com/history/daily/sj/ny/ENAS/date/2021-10-10>