The impact of CO2 on the climate of planet Earth
Template STAT651 Project

Suhasini Subba Rao

November 17, 2019

Introduction

Over the past twenty/thirty years there has been growing concern about the increase in global temperatures and possible causes of it. In particular, the impact that the growing amount of man-made CO2 may have on global temperatures. There has been a huge reluctance amongst policy makers to tackle this issue. This is mainly because any real change requires a huge amount of investment in infrastructure and a realignment of economic targets (such as economic growth). In order to detract from the issue, one line of argument that is often used is that climate change does not exist or that it is completely natural and fits with long term periodic trends. The objective of this project is to use basic statistical tools to see what, if any changes, are happening in the climate and possible causes for these changes.

Summary statistics and plots

The data sets we will be using are the yearly northen hemisphere, southern hemisphere and global temperatures from 1850-2004 (155 observations for each data set) and the estimated CO2 emissions over the same period of time. The monthly temperature data can be obtained from `http://www.cru.uea.ac.uk/data`. The CO2 data is monitored in Mauna Loa, Hawaii (`https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html`). In this section we will summarise this data.

In Figure 1 we plot the northern, southern and global temperature anomalies from 1850-2004. We observe that there does appear to be relationship between time and the temperature (the temperatures do not appear to be completely 'random'). Indeed, there appears to be a possible upward trend in the data. The simplest model to explain this relationship is to use a linear model. Therefore all three plots also contain the line of best fit. However, this is a rather crude approximation, as there appears to be a more complex relationship between temperatures and time (possibly periodicities).

In Figure 2 we plot the yearly estimated CO2 emissions over the same time period. As one would expect, since industrialisation there has been a steep rise in CO2 emission from the middle of the 19th century to the present. In order to see whether there is a relationship between CO2 emissions and global temperatures in Figure 2 we have plotted the CO2 emission against temperature and fitted the line of best fit.

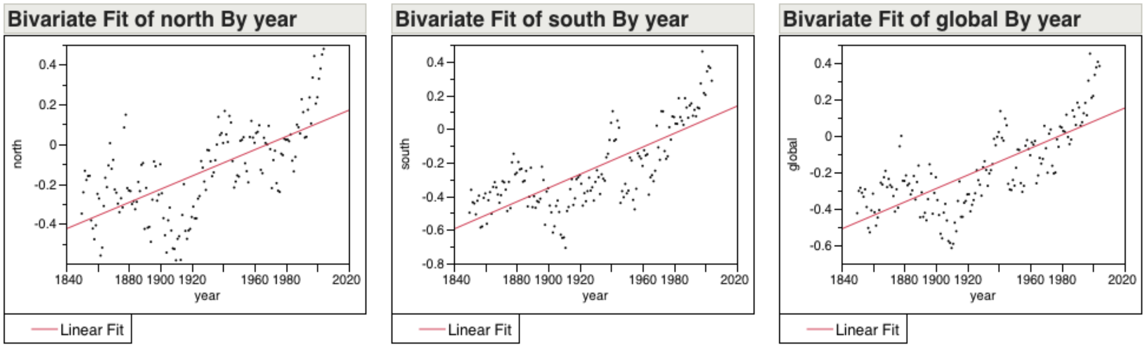


Figure 1: Left: Northern Temp. Middle: Southern Temp. Right: Global temperatures from 1850-2004 and the line of best fit.

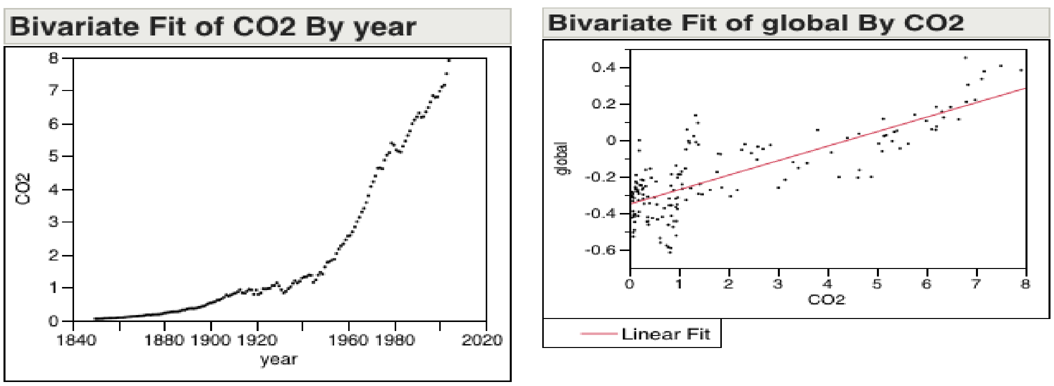
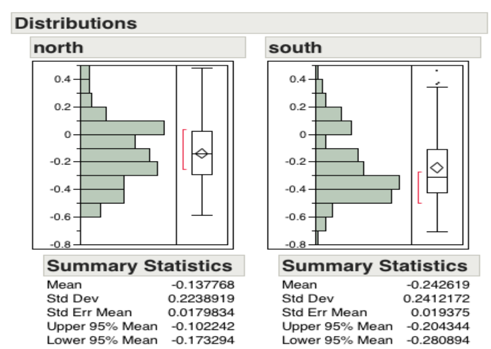


Figure 2: Left: CO2 over time. Right: Plot of CO2 against temperature

Finally, we consider the difference between yearly northern and southern hemisphere temperature anomalies. It is of interest to see whether northern and southern temperatures are the same or not. A histogram and boxplot is given in the figure on the right. There does appear to be some differences in the sample means (-0.137 verses -0.242), however, without a formal analysis it is not possible to say whether these differences are statistically significant.



The Statistical Analysis

In this section, we quantatively analysis the data described in the previous section.

We first analysis how the global temperature anomalies vary over time, in particular how well a linear trend fits the data. The output of the statistical analysis using JMP can be found in Figure 3. The least squares estimate of the linear slope is 0.0036. To understand if the slope is statistical significant and whether the evidence points towards a positive trend we conduct a t-test (testing $H_0 : \beta_1 \leq 0$ against $H_A : \beta_1 > 0$), the t-value is $t = 13.71$ and corresponding p-value is highly significant (less than 0.01%). Based on this, there is strong evidence to suggest that the temperatures have been rising since 1850.

To check that the p-value is accurate, a plot of the *residuals* in the least squares fit is given in Figure 3. Both the histogram and the QQplot of the residuals suggest that the residuals do not deviate much from normality, this together with the sample size (155 observations) suggest that the p-value is accurate (though some concerns with the assumptions are discussed in the conclusions). Finally, to check whether a linear model seems appropriate in the right plot of Figure 3 a scatterplot of residuals against year is given. There does appear to be a 'trend' in this plot (over simple random variation), which suggests the relationship between temperature and time is more complex than just a linear model. Indeed the R^2 is 0.55, therefore a linear trend does not explain a large part of the variability in the temperaures. Finally, we note that as the slope is statistically significant and a 95% confidence interval for the slope is $[0.00036 \pm 1.97 \times 0.00027] = [0.0032, 0.0042]$. This suggests that temperatures are rising by anywhere from 0.0032 to 0.0042 celsius a year.

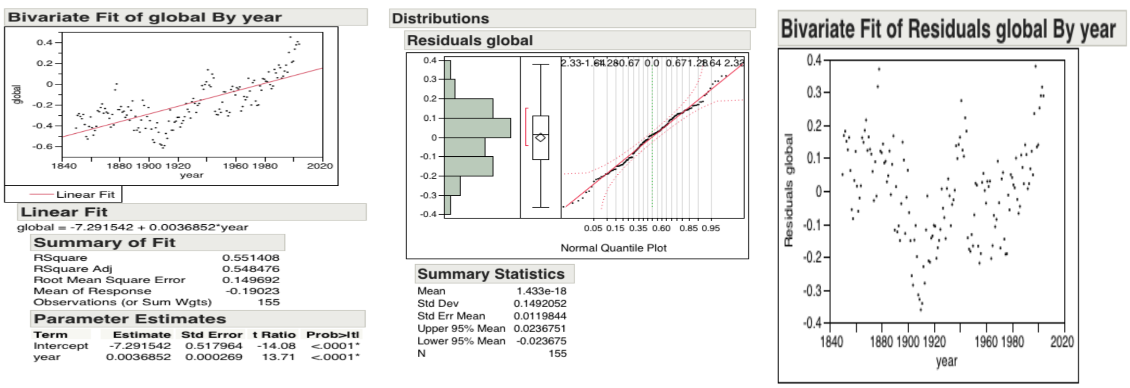


Figure 3: Left: The JMP output. Middle: Distribution of residuals. Right: Scatter plot of residuals against year.

We now analysis the relationship between the yearly CO2 and temperatures. The JMP output and relevant plots are give in Figure 4. The least squares estimate of the slope is 0.079. Again, we conduct a t-test, to access whether this slope is statistically significant ($H_0 : \beta_1 \leq 0$ against $H_A : \beta_1 > 0$) and obtain a p-value of less than 0.01%. This suggests that there is a possible linear relationship between CO2 and temperature anomalies.

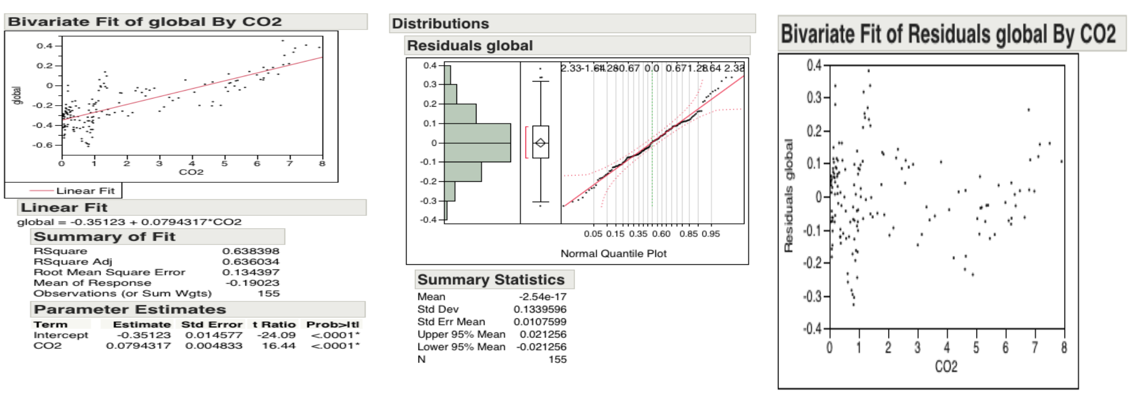


Figure 4: Left: The JMP output. Middle: Distribution of residuals. Right: Scatter plot of residuals against year.

Comparing the R^2 for both time and CO2 (0.55 verses 0.80) we see that C02 is better at explaining the variation in the temperatures than time. Furthermore, if we were to do a multiple regression (see the figure on the right), where we fit a linear model with <u>both</u> time and CO2 as covariates to the temperature data, we see that the time coefficient is not significant (p-value of 10%) where as the CO2 coefficient is still highly significant (though for multiple regression models one needs to the F-test to really test which variables are significant). Before determining a causal relationship, we need to cautious. Any two time series with a trend, can give rise to spurious correlation. However, the physics do suggest that CO2 may play some role in the temperature increase.

Response global

Summary of Fit

RSquare	0.642172
RSquare Adj	0.637464
Root Mean Square Error	0.134132
Mean of Response	-0.19023
Observations (or Sum Wgts)	155

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	4.9078204	2.45391	136.3928
Error	152	2.7347067	0.01799	**Prob > F**
C. Total	154	7.6425271		<.0001*

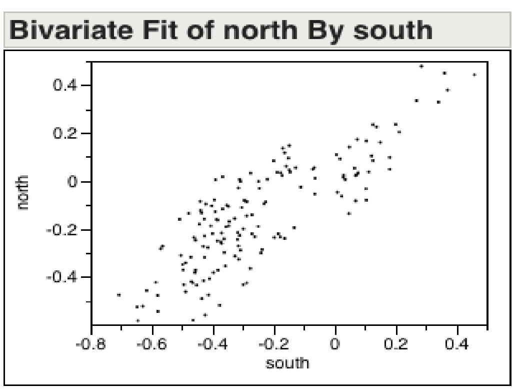
Parameter Estimates

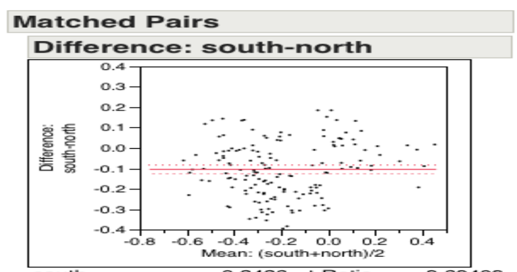
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.644113	1.021088	-1.61	0.1094
year	0.0006838	0.00054	1.27	0.2073
CO2	0.0671703	0.010818	6.21	<.0001*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
year	1	1	0.02885011	1.6035	0.2073
CO2	1	1	0.69367044	38.5555	<.0001*

Finally, we consider the differences between northern and southern temperatures. As noted in the previous section a difference of 0.105 degrees is seen between the averages, to see whether this is statistically significant we do a matched paired t-test (since there is a clear matching between the yearly temperatures between the northern and southern hemispheres, see Figure 5 where a clear linear trend is seen). The p-value (testing $H_0 : \mu_N - \mu_S = 0$ against $H_A : \mu_N - \mu_S \neq 0$) is less than 0.01% which strongly suggests that there is on average a difference between northern and southern temperatures. The 95% confidence interval for this mean difference is $[-0.12, -0.08]$ degrees. On explanation for the difference in temperatures between the two hemipheres is the amount of land mass verses the amount of ocean, and possibly the axis of spin.





south	-0.2426	t-Ratio	-9.69403
north	-0.1378	DF	154
Mean Difference	-0.1049	Prob > |t|	<.0001*
Std Error	0.01082	Prob > t	1.0000
Upper 95%	-0.0835	Prob < t	<.0001*
Lower 95%	-0.1262		
N	155		
Correlation	0.8349		

Figure 5: Left: Plot of yearly North against Southern Temperatures. Right: Matched test output

Conclusions

Our preliminary statistical suggests that global temperature anomalies have been increasing over the past 150 years (with a yearly rise of between 0.0032-0.0042 degrees). It has been well documented that man-made CO2 has been on the the rise and this can be seen in the CO2 data. By including the yearly CO2 emission as a covariate in the regression, there is evidence to suggest that the rise can be at least partially explained by the rise in CO2 emissions. To understand how the rise in CO2 is the driving force behind the increase in temperatures a knowledge of the thermodynamics behind the system, which are used in climate computer models. Another factor of interest, is that there is evidence to suggest that the mean temperatures in the northern and southern hemispheres are different.

We mention some possible drawbacks with the basic analysis. Both these methods assume that the observations are independent. However, the data is observed over a period of several years. There is no reason to suppose that the temperatures in any one year are completely independent of the temperatures in the past or the future. Without accounting for these dependencies the estimated standard errors (which are instrumental in analysis) are probably incorrect. Furthermore, it is not clear that the linear model is a correct model to use. A more complex model which allows for nonlinearity may be required. Again the standard errors that are used depend on the correct model being linear. There are methods for correcting for these issues, but they are beyond the scope of this class.

The evidence strongly suggests that the climate is changing and the global temperatures are rising. One possible explanation for the increase in temperatures, is the rise CO2 in our atmosphere (as seen in our analysis). Climate models (mathematical models which simulate how the temperatures could change in the future, computer models are not made up they are based on our understanding of thermodynamics, Newton's law of motion etc) predict a rise of over 2 degrees in the next 50 years. So far, the oceans have absorbed a large amount of the CO2 (`https://www.pmel.noaa.gov/co2/story/What+is+Ocean+Acidification%3F`). Besides the negative impact this has on the oceanic ecosystem there is evidence that the oceans are reaching its saturation point. It is inclear how the CO2 in the atmosphere will change once this point has been reached.

The rise in temperatures will have a substantial impacts on where we live, how we eat and on the quality of life. Therefore, it is prudent to take action to curb this rise. This can be done through policies places by the governments. However, even ordinary people can do their bit, by taking simple measures. Such as increasing the temperature on their AC by one degree, using less throwaway items, walking/cycling instead of taking the car etc. The benefits are great, at the very least they will reduce the amount of nanoparticles in the air (these are ultrafine particles less than 100nm) and are known to cause very serious health issues.

There are people who say this is all a hoax and for whatever reason claim that the rises are completely natural (I am unsure where the man-made CO2 goes in their theory). But the evidence strongly suggests otherwise. So you have to ask yourself, do you take action to curb your CO2 emissions, at the "slight" risk that everything will be fine in the end. Or completely ignore all the warning signs in the hope that all will be fine in the end, but pay a very steep price if the temperatures do exceed the forecasted 2 degrees.