Introductory Statistics Project

Brittany Ballwanz & Justin Hageman

12/10/2008

Introduction:

The issue of global warming has become a very important concern in today's society. Global warming carries with it a wide variety of impacts that can affect economic, social, and political landscapes. The manner in which we address global warming depends largely on locating the areas that are being impacted by it and finding out to what extent these areas are being affected. For these reasons, we have decided to look for evidence of global warming in Wisconsin and more specifically the city of Milwaukee. We have decided to compare the average annual temperature over the past ten years to the average annual temperature over the previous thirty years to determine if there is evidence of significant warming in the city of Milwaukee within the past ten years. By comparing the past ten years to the previous thirty years, we can get an indication as to whether the temperature has deviated from its historical values.

It is important to note that any evidence of a temperature increase cannot be definitively linked to global warming as there could be other explanations such as weather cycles or localized pollution. However, evidence of global warming has been found around the globe and we would be interested to see if warming is also occurring within the state of Wisconsin.

Methods:

To collect the data for this analysis, we used the archives of the National Climactic Data Center which operates under the authority of the National Oceanic and Atmospheric Administration (NOAA). The data obtained from NOAA contains average monthly temperatures for the forty years involved in this project. Some years also contained an average annual temperature and for the years that didn't contain this data, we calculated the average annual temperature by using the average monthly temperatures for each specific year.

Randomization is not a part of this project because of the nature of the analysis. We are looking to compare specific years to other specific years and randomizing these years would defeat the purpose of the project.

To analyze the data gathered from NOAA, we used a two sample t-test and a box plot comparing the two different groupings of years.

Results

The following EDA for this study shows higher values of temperature data for the past 10 years as compared to the previous 30 years. Both the mean and the mode for the past 10 years have higher values than the previous 30 years. However, we will give more credibility to the mean because Figure 1 is slightly left skewed. The quartiles for the 10 year sample are also higher than the 30 year sample. This gives us an indication that the t-test will show a higher average annual temperature for the past 10 years as compared to the previous 30 years but we must actually perform this test to find out definitively.

	n	NAs	Valid n	Mean	St. Dev.	Min.	1st Qu.	Median	3rd Qu.	Max.
Temp.10	10	0	10	48.95000	1.177804	46.90	48.15	49.20	49.42	50.9
Temp.30	30	0	30	46.88933	1.653884	43.58	45.55	46.35	47.95	50.2

1.) The test that we preformed was two-sample t-test because a quantitative variable (Temperature) was measured for two populations (past 10 years and previous 30 year) that were independent of each other.

2.) H_0: $\mu_{past10} - \mu_{previous30}=0$. H_A: $\mu_{past10} - \mu_{previous30} > 0$

3.) $\alpha = 0.05$ (testing at 5%)

4.) The data is observational and not random.

5.) The sample distribution on Temp.10 is slightly left skewed (Figure 1) and the Temp.30 sample distribution is approximately normal (Figure 2), the two samples are independent. We must now test to see if the variances are equal.

i. Use Levene's test

ii. H_0: $\sigma_1^2 = \sigma_2^2$ H_A: $\sigma_1^2 \neq \sigma_2^2$

iii. $\alpha = 0.05$

iv. p-value = 0.1819

v. p-value > α so we DNR H_o

vi. Variances are equal

The Levene's Test p-vaule (0.1819) is larger than alpha therefore the variances appear to be equal.

```
      Levene's Test for Homogeneity of Variance
      Df F value Pr(>F)
group  1  1.8492 0.1819
      38
   Two Sample t-test
```

```
data:  Temp by Year
t = 3.6307, df = 38, p-value = 0.0004152
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.103765      Inf
sample estimates:
mean in group Temp.10 mean in group Temp.30
             48.95000              46.88933
```

6.) X_{10}= 48.95 X_{30}= 46.89 X_{10}-X_{30}= 2.06

7.) The t test statistic is 3.63 and with 38 degrees of freedom.

8.) The p-value for this test statistic is 0.00042

9.) p-value < α so we reject H_O.

10.) It appears that the temperature has increased significantly over the past 10 years as compared to the previous 40 years.

11.) We can be 95% confident that the mean temperature has been at least 1.103 degrees warmer for the last 10 years compared to the previous 30 years.

Discussion:

Overall we can conclude from our results that the average annual temperature has increased in the last 10 years compared to the pervious 30 years for the city of Milwaukee. The results allow us to say with 95% confidence that the average annual temperature over these ten years has increased by at least 1.103 degrees. Such an increase may not be directly noticeable to humans but it could potentially have an impact on things like growing seasons and annual precipitations, both of which impact humans directly. It is important to continue to monitor this change in average annual temperature to determine if it is still increasing so we can begin to assign a cause to it. This analysis could have been improved if more historical temperature data had been available. Having more historical data would have allowed us to see get a better picture of past climate changes to compare current changes to. Unfortunately, going back much further than forty years required payment to acquire the data and records were not kept prior to about 100 years ago.

Appendix

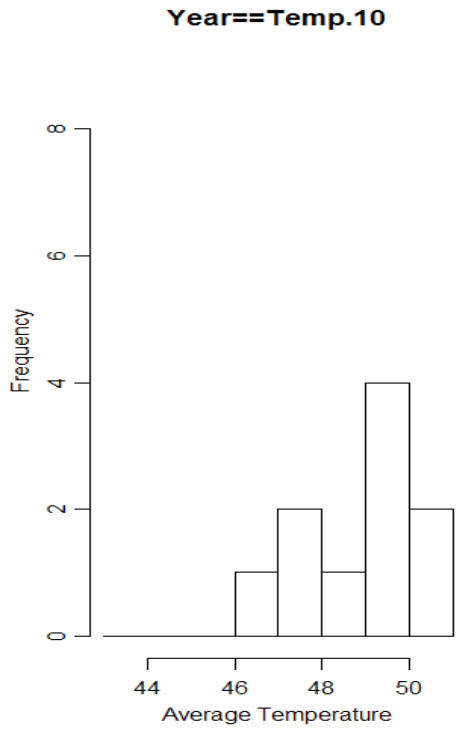

Figure 1

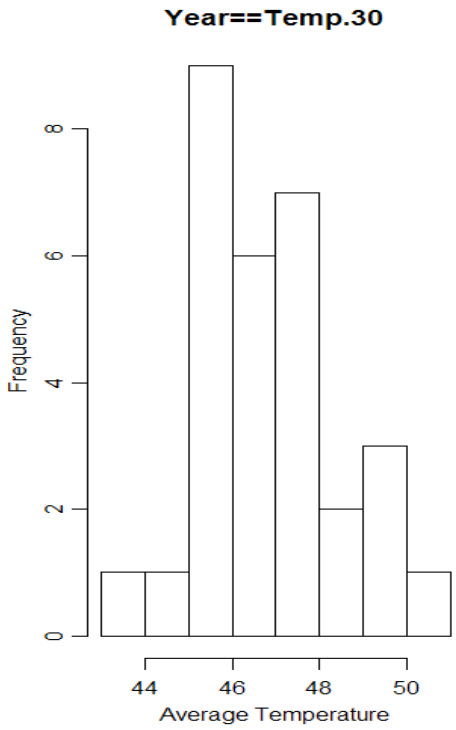

Figure 2

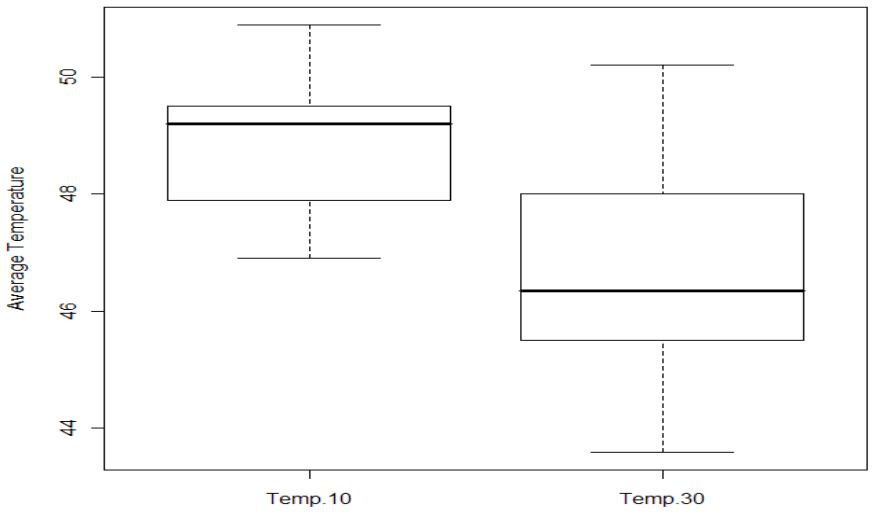

Figure 3

R-appendix

```
library(NCStats)
temp<-read.table("Past.txt",header=TRUE)
temp
str(temp)
attach(temp)
Summarize(Temp~Year)
levene.test(Temp~Year)
t2<-t.test(Temp~Year,alt="greater",var.equal=TRUE,conf.level=.95)
t2
plot(t2)
hist(Temp~Year,ylab="Frequency",xlab="Average Temperature")
boxplot(Temp~Year,ylab="Average Temperature")
```

Works Cited

DS3220 - Surface Data, Monthly - US , Data Access." NCDC Climate Data Online. 9 Dec. 2008 <http://www1.ncdc.noaa.gov/pub/orders/CDO4151471644873.html>.