A Study on the Differences in Climate between Cities of Similar Latitudes

Dylan Way, Jeff Rand, Seren Dances

May 31, 2022

1 Abstract

The purpose of this study is to investigate the monthly rainfall totals from January, 2005 to December, 2018 for several coastal cities that lay near the 48^{th} parallel, and to determine if the rainfall totals between such cities are statistically significant. We chose the cities of (1) St. Johns, Newfoundland, (2) Bellingham, Washington, (3) Sovetskaya Gavan, Russia, and (4) Brest, France. For this purpose, we analyzed the data with a blocking design using the months as the blocks, and the cities as the treatments. Methods used for determining significance include the one way ANOVA and the Friedman test. We find that the rainfall in all four cities is not normally distributed, and thus we must rely only on the nonparametric tests, from which we conclude that the monthly rainfall totals between cities is significantly different, despite these cities roughly equal proximity to an ocean and nearly equal latitude.

2 Introduction

2.1 Objective

The purpose of this study is to compare the rainfall totals of several cities of similar latitudes for significant difference. The four cities included in this study are all coastal cities of between the 48^{th} and 49^{th} parallel. We expect to see differences in their local rainfall totals due to differences in the relative direction of weather systems over each city. In short, our hypothsis test is as follows, where μ_n represents the mean of the monthly rainfall for city n:

H_0 : The monthly rainfall totals in each of the four cities are equal.

H_a : At least one pair of the cities have unequal rainfall totals.

2.2 Data Collection

We collected data for the following four cities: (1) St. Johns, Newfoundland, (2) Bellingham, Washington, (3) Sovetskaya Gavan, Russia, and (4)

Brest, France. Monthly rainfall data was only available for every city from January, 2005 to December, 2018. So, our analysis was restricted to data from that 14-year period. We standardized all our monthly data to inches, and sorted it into a blocking design where the months of the year were treated as blocks, and the cities as treatments.

3 Methodology

We tried several approached to investigate the possible differences in monthly rainfall. We thought it best to first use the Friedman test, since this would require no parametric assumptions. Unfortunately, because this test is done by ranking, we could only use one observation for each cell (meaning one observation for each unique block and treatment). Our data provided 14 such observations for each cell. But first, the boxplots of the means and medians of each cell are as follows:

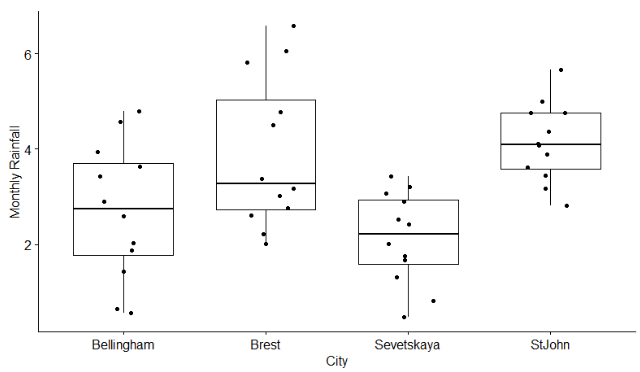

Figure 1: A boxplot of the medians of the monthly rainfall in each city.

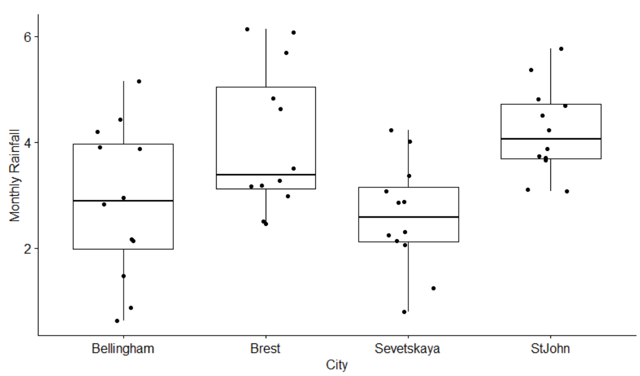

Figure 2: A boxplot of the means of the monthly rainfall in each city.

A Friedman test of the mean monthly rainfall of each city yielded the following results:

```
Friedman chi-squared = 13.2, df = 3, p-value = 0.004223
```

Figure 3: A Friedman test on the mean of each cell

A Friedman test of the median monthly rainfall for each city yielded similar results:

```
Friedman chi-squared = 15.3, df = 3, p-value = 0.001577
```

Figure 4: A Friedman test on the median of each cell

As you can see, there is a statistically significant difference in both the mean and median data. For the sake of comparison, we also conducted an ANOVA test of the data, which does accept multiple observations in each cell, so no simplification of the data was required. The following is a boxplot of the original data:

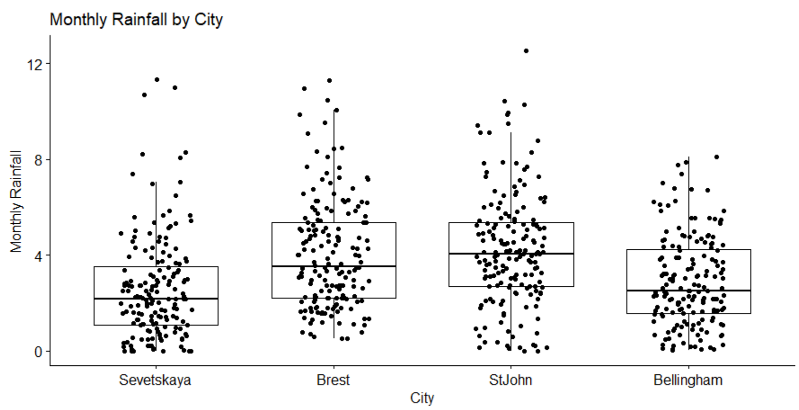

Figure 5: A boxplot of the original monthly rainfall in each city.

Next, the ANOVA summary of the data is as follows:

```
             Df Sum Sq Mean Sq F value   Pr(>F)    
block         1   90.9   90.87   19.40 1.23e-05 ***
group         3  328.4  109.48   23.38 2.13e-14 ***
Residuals   667 3123.7    4.68                     
```

Figure 6: The ANOVA summary of the data

We tested this assumption of normality on the data from each city individually, and the results were as follows:

```
For Brest, France:
W = 0.94374, p-value = 3.292e-06

For Sovetskaya Gavan, Russia:
W = 0.8818, p-value = 2.824e-10

For St. John's, Newfoundland:
W = 0.97048, p-value = 0.001197

For Bellingham, Washington:
W = 0.9611, p-value = 0.0001234
```

Figure 7: Shapiro-Wilk tests for each of the four cities.

Despite the significant results observed in the one-way ANOVA table, the results of the previous tests for normality show that the assumption that the rainfall data is normally distributed clearly violated in all four cities.

4 Conclusion

The results herein are unambiguous: Even when, via blocking, we adjust for the seasonality which may exist between different months, we find there exist statistically significant differences in the rainfall totals between these cities. Additionally, when comparing the results of the Friedman tests to that of the one-way ANOVA, we find similarly significant results despite the clear evidence implying the data violates the normality assumption within the ANOVA. These results were unsurprising, based on what we know about weather patterns in each city. Brest and Bellingham are located on the western edge of their respective continents. As a result, the prevailing winds blow moist, oceanic air into the city for most of or all of the year. St. John's and Sovetskaya Gavan are located on the opposite end of their respective continents. In these regions, the prevailing winds are much more inconsistent and blow from the land into the city more often than not, reducing precipitation.

5 Future Work

While we did expect to see differences in overall precipitation, it is possible that the several cities may have unique peak rainfall months. Additionally, we suspect that the results of our data may be affected by autocovariance, as this is time-series data. While taking the mean or median of each month may have helped reduce the impact of autocovariance on our data set, it likely failed to remove it altogether. Future studies could take this into account by fitting an ARMA model to the data, then performing the same tests on the resulting residuals.

6 References

[1] Weather Spark, Weather Data Download for Sovetskaya Gavan, https://weatherspark.com/download/144086/Download-Sovetskaya-Gavan%E2%80%99-Russia-Weather-Data.

[2] National Centers for Environmental Information, National Oceanic and Atmospheric Administration, https://www.ncdc.noaa.gov/.

[3] Government of Canada, https://climate.weather.gc.ca/historical_data/search_historic_data_e.html

7 R Code

```
#rainfull is our complete data set
rainfull<-read.csv(file="rainraw2.csv",header=TRUE)

#Anova test of the data
aov(data ~ group + block, data=data)

#Here we're separating the full data set into separate chunks based on city
#Sevetskaya has been abbreviated as Russ for the sake of simplicity
rainRuss<-rainfull[1:168,"data"]
rainBrest<-rainfull[169:336,"data"]
rainJohn<-rainfull[337:504,"data"]
rainBell<-rainfull[505:672,"data"]

#An Index Vector to apply mean and median to the data in the next step
indexl<-rep(1:12,14)

#tapply is reducing each data set into a vector of 12 values,
#one for the mean of each month.
```

```
Russmean<-tapply(rainRuss,indexl,mean)
Brestmean<-tapply(rainBrest,indexl,mean)
Johnmean<-tapply(rainJohn,indexl,mean)
Bellmean<-tapply(rainBell,indexl,mean)

#Identical code to above using median instead of mean
Russmed<-tapply(rainRuss,indexl,median)
Brestmed<-tapply(rainBrest,indexl,median)
Johnmed<-tapply(rainJohn,indexl,median)
Bellmed<-tapply(rainBell,indexl,median)

#Combining the monthly means into a matrix with block and gorup information
datamean<-matrix(c(rep(city,each=12),rep(1:12,4),Bellmean,Brestmean,
    Russmean,Johnmean),nrow=48,ncol=3)
colnames(datamean)<-c("City","Month","Data")

#Performing the Friedman Test on the Data in datamean
friedman.test(datamean[,"Data"],blocks=as.factor(datamean[,"Month"]),
    groups=as.factor(datamean[,"City"]))

#Combining the monthly means into a matrix with block and group information
datamed<-matrix(c(rep(city,each=12),rep(1:12,4),Bellmed,Brestmed,
    Russmed,Johnmed),nrow=48,ncol=3)
colnames(datamed)<-c("City","Month","Data")

#Performing the Friedman Test on the Data in datamed
friedman.test(datamed[,"Data"],blocks=as.factor(datamed[,"Month"])
    ,groups=as.factor(datamed[,"City"]))

#Creating the boxplots for the means, medians, and original data:
library(ggpubr)
ggboxplot(data, x = "group", y = "data", ylab="Monthly Rainfall",
    add = 'jitter',main="Monthly Rainfall by City")
ggboxplot(datamedian, x = "City", y = "Data", ylab="Monthly Rainfall",
    add = "jitter")
ggboxplot(datamean, x = "City", y = "Data", ylab="Monthly Rainfall",
    add = "jitter")
```