

# APPLICATIONS OF MATHEMATICS IN FINANCE

CORBIN APPLE, BRIDGET HYLAND, BEN STERLING

**1. Introduction.** It has been well established by The United States Environmental Protection Agency and other independent organizations that average temperature is increasing across the country. This study examines monthly temperature data from two stations: Rye Patch Dam, Nevada, and Salt Lake City International Airport, Utah. These stations were chosen because they are roughly at the same latitude ( $40.498^{\circ}\text{N}$  and  $40.790^{\circ}\text{N}$ ), longitude ( $118.316^{\circ}\text{W}$  and  $111.980^{\circ}\text{W}$ ), and elevation (1260.3m and 1287.8m, for Rye Patch Dam and Salt Lake City respectively). Our data is from the National Centers for Environmental Information (NCEI, a subsidiary of the National Oceanic and Atmospheric Administration) Global Summary of the Month (GSOM) dataset (Lawrimore, Jay H.; Ray, Ron; Applequist, Scott; Korzeniewski, Bryant; Menne, Matthew J. (2016): Global Summary of the Month (GSOM), Version 1. [indicate subset used]. NOAA National Centers for Environmental Information. <https://doi.org/10.7289/V5QV3JJ5>. [access date]), which contains monthly data about major meteorological parameters at many locations across the country. We chose Salt Lake City and Rye Patch Dam because of their similar geographical characteristics. Data for Rye Patch Dam begins in 1935, but data for Salt Lake City only reaches back to 1948, so we used only years from 1948 to 2020 in our analysis. The dataset includes many parameters, but of particular interest to us were average temperature, average precipitation, number of days with thunderstorms, and total minutes of sunshine.

The goal of our first hypothesis was to examine whether the effects of climate change statistically differ between these two locations; if they do, we may be able to infer that the climate change is predominantly man-made. We accomplished this by calculating the monthly temperature anomaly at each location and comparing the mean temperature anomalies of the two locations.

The goal of our second hypothesis was to determine predictors of temperature. We used a multiple linear regression with temperature as the response variable and precipitation, days with thunderstorms, and minutes of sunshine as the dependent variables.

**2. First Hypothesis.** Climate change is a well-documented phenomenon: on average, the global temperature is increasing. We wanted to test the hypothesis that the temperature is increasing at a faster rate in Salt Lake City than in Rye Patch Dam. Practically, we tested the hypothesis that the mean temperature anomaly of Salt Lake City is greater than the mean temperature anomaly in Rye Patch Dam.

Temperature anomaly is used to compare change in temperature over time. A mean temperature is calculated over a long period of time, and this long-term mean is subtracted from more recent observations. For example, it is known that the global average temperature was  $13.9^{\circ}\text{C}$  between 1901 and 2000. If the global average temperature in 2015 was  $14.5^{\circ}\text{C}$ , then the temperature anomaly for that year would be  $14.5 - 13.9 = 0.6^{\circ}\text{C}$ . When analyzing temperature data over many years, it is advisable to use a temperature anomaly rather than absolute temperature because it eliminates seasonal variation within the year, allowing for a more significant result. This is an accepted and widely used method in climate analysis. [CITE SOMETHING HERE]

The GSOM records the monthly average temperature at each location, so we determined temperature anomaly monthly. We calculated a January anomaly by

averaging the January temperatures of each year from 1948 to 1974. Then, we subtracted this long-term mean from each January temperature from 1975 to 2020. We did the same for the other eleven months and for both locations, yielding different anomalies for each. In our data files, this is column CG, labeled TAVG ADJ. We used this anomaly in our comparison in place of the absolute temperature. It served to reduce the variance of each sample to produce a meaningful result.

Because of the large number of observations in each sample, we invoked the central limit theorem to approximate the distribution as normal. The Shapiro-Wilk test for normality gives a p-value of [P] for the Salt Lake City sample and [P] for the Rye Patch Dam sample, which verifies that the normal approximation is appropriate. The normal plots of the temperature anomalies for each station is shown below. The linear nature of the plots further supports our approximation.

Formally, our hypotheses are

$$H_0 : \mu_{SLC} = \mu_{RPD} \text{ vs. } H_a : \mu_{SLC} > \mu_{RPD}.$$

Because the two stations have similar geographical characteristics (i.e. latitude, longitude, and elevation), we determined that this is an observational matched pairs analysis and used the paired t-test at  $\alpha = 0.05$  accordingly. As such, it was not necessary to determine equality of variance between the two samples. We determined the test statistic to be  $t = [T]$ . This is [GREATER THAN / LESS THAN] the critical value  $t_{n-1, \alpha/2} = [T]$ . Also, the p-value is [P], which is [GREATER THAN / LESS THAN] the confidence level  $\alpha = 0.5$ . Based on these results, we reject  $H_0$  and conclude that the mean temperature anomaly at Salt Lake City is significantly greater than the mean temperature anomaly at Rye Patch Dam. In other words, since 1975, the temperature has increased more at Salt Lake City than at Rye Patch Dam. The reason for this is not known, but previous research suggests that Salt Lake City emits a relatively large amount of greenhouse gases per capita, which hastens the effects of climate change there. [CITE SOMETHING HERE]

We then randomly selected 10% of the data from each station to remove from the dataset and treat as missing. Although our dataset contains ample missing values on its own, we removed more values in order to evaluate the effects of missing data. Our analysis ignored months in which temperature data for Salt Lake City *or* Rye Patch Dam was missing. This resulted in a loss of [N] data points, leaving us with a new sample size of [N] for each station. Performing the Shapiro-Wilk test again, the distribution of the observations [WAS STILL NORMAL / WAS NO LONGER NORMAL / WAS STILL NORMAL AT RYE PATCH DAM BUT NOT AT SALT LAKE CITY / WAS STILL NORMAL AT SALT LAKE CITY BUT NOT AT RYE PATCH DAM]. The new test statistic was [T], which is [GREATER THAN / LESS THAN] the critical value  $t_{n-1, \alpha/2} = [T]$ . The p-value was [P], which is [GREATER THAN / LESS THAN] the confidence level  $\alpha = 0.5$ . Therefore, we [REJECT H0 / CANNOT REJECT H0] and [CONCLUDE / DO NOT HAVE SUFFICIENT EVIDENCE TO CONCLUDE] that the mean temperature anomaly at Salt Lake City is greater than the mean temperature anomaly at Rye Patch Dam. The addition of more missing values reduced the sample size and resulted in a higher p-value, [BUT DID NOT CHANGE / AND CHANGED] the outcome of the hypothesis test.

### 3. Hypothesis 1. def

```
library(ggpubr);
library(ggplot2);
library(latex2exp);
library(car);
```

```

# Disable pdf generation:
pdf(NULL);

data_Salt_Lake = read.csv("GSOY/GSOM_Salt_Lake_updated_monthly_adjustment_v3.csv"); #1287.8m elevation
data_rye_patch = read.csv("GSOY/GSOM_rye_patch_updated_dates_monthly_adjustment_v3.csv"); #1260.3m elevation

Salt_Lake_Diff = (data_Salt_Lake$TAVG.ADJ)[1:length(data_Salt_Lake$TAVG.ADJ)];
rye_patch_Diff = (data_rye_patch$TAVG.ADJ)[1:length(data_rye_patch$TAVG.ADJ)];

#print("Means:")
#print(Salt_Lake_mean)
#print(rye_patch_mean)

#print("Latest temperature:")
#print(tail(Salt_Lake_Temp_Data))
#print(tail(rye_patch_Temp_Data))

for(i in 1:length(Salt_Lake_Diff)){
  if(is.na(Salt_Lake_Diff[i]) | is.na(rye_patch_Diff[i])){
    Salt_Lake_Diff[i] = NA;
    rye_patch_Diff[i] = NA;
  }
}

#png(file = "Missing value Salt.png");
#plot(is.na(Salt_Lake_Diff));

#png(file = "Missing value Rye.png");
#plot(is.na(rye_patch_Diff));

# Remove missing data (values of NA) here:
Salt_Lake_Diff = Salt_Lake_Diff[!is.na(Salt_Lake_Diff)];
rye_patch_Diff = rye_patch_Diff[!is.na(rye_patch_Diff)];

# Print mean, variance, and sample size for each dataset:
print("Salt_Lake");
print(paste("\U03BC", "=", mean(Salt_Lake_Diff, na.rm=TRUE)));
print(paste("\U03C3", "=", sqrt(var(Salt_Lake_Diff, na.rm=TRUE))));
print(paste("n=", length(Salt_Lake_Diff)));
# New line:
cat("\n");

print("rye_patch");
print(paste("\U03BC", "=", mean(rye_patch_Diff, na.rm=TRUE)));
print(paste("\U03C3", "=", sqrt(var(rye_patch_Diff, na.rm=TRUE))));
print(paste("n=", length(rye_patch_Diff)));
# New line:
cat("\n");

# Perform the Shapiro Tests:
shapiro.test(Salt_Lake_Diff);
shapiro.test(rye_patch_Diff);

# Visualize with QQ plots:
png(file = "Salt_Lake_Diff_QQ_Plot.png");

#qqPlot(x=Salt_Lake_Diff, distribution="norm")
qqnorm(Salt_Lake_Diff)
qqline(Salt_Lake_Diff);
dev.off();

png(file = "Rye_Patch_Diff_QQ_Plot.png");

#qqPlot(x=Salt_Lake_Diff, distribution="norm")
qqnorm(rye_patch_Diff)
qqline(rye_patch_Diff);
dev.off();

#Perform two sided t-test with unequal variances:
t.test(rye_patch_Diff, Salt_Lake_Diff, paired=TRUE);

```

## REFERENCES