Project- Climate Change

Camden Jones

2023-03-24

Data

The data in this report can be found at these links:

Temps

and

TempsApproach3

Study

Approach 1

If there have been no temperature trends over the years, then one would expect for the number of dates in which a record high occurred more recently to be approximately the same as the number of dates in which a record low occurred more recently. On the other hand, if there has been a warming trend, one would expect to see more dates in which a record high has occurred more recently.

1. Use R to find the proportion of dates in the sample where the record high occurred more recently. That is, find p-hat.
2. To determine whether there is evidence that record highs occur more recently than record lows in general, test the hypothesis H0: p=0.5 against H1: p>0.5. Be sure to check assumptions, perform the test, state the p-value, and whether the null hypothesis is rejected or not.
3. Based on the test performed, what is your conclusion about whether there has been a warming trend at the Blue Hill Meteorological Observatory?

Approach 2

We take a different approach. If a record low is equally likely to occur in any year of observation, then the mean year in which a record low is observed would occur at or near the midpoint of the observation period. On the other hand, if there is a warming trend, then the mean year in which a record low is observed would be less than the midpoint of the observation period.

1. What is the midpoint of the observation period? Recall that the data span the years 1893 to 2023.
2. Use the data to test the hypothesis that the mean year in which a record low occurred is equal to the midpoint value of the observation period, against the alternative that the mean year is less than the midpoint. Be sure to check assumptions, perform the test, state the p-value, and whether the null hypothesis is rejected or not.
3. Based on the test performed, what is your conclusion about whether there has been a warming trend at the Blue Hill Meteorological Observatory?

Approach 3

Suppose that a researcher wanted to repeat the test that was performed in Approach #2. However, this researcher only has access to a data set where the historical record high and low temperatures were collected approximately once per month.

1. Should a hypothesis test be performed to determine whether the mean year in which a record low occurred is less than the midpoint value of the observation period? Explain why or why not – use appropriate graphs to back up your answer.

Approach 1

```
library(readr)
Temps<-read.csv("https://www.dropbox.com/s/gir83plkf8bw5gd/climate_change_blue_hill_2023.csv?dl=1")
```

1. Use R to find the proportion of dates in the sample where the record high occurred more recently. That is, find p-hat.

```
RecentHigh<-sum(Temps$H_Year>Temps$L_Year)
TotalTemps<-length(Temps$H_Year)
p_hat<-RecentHigh/TotalTemps
print(p_hat)
```

```
## [1] 0.852459
```

p-hat, or the proportion of dates in the sample where the record high occurred more recently, is equal to 0.852459.

2. To determine whether there is evidence that record highs occur more recently than record lows in general, test the hypothesis H0: p=0.5 against H1: p>0.5.

```
one.samp.prop.test <- function(x, n, p0, direction = "two.sided") {
     # Calculate sample proportion
     p_hat <- x/n
     # Calculate standard error
     se <- sqrt(p0 * (1 - p0) / n)
     # Calculate z-value
     z <- (p_hat - p0) / se
     # Calculate p-value
     if (direction == "two.sided") {
         p_value <- 2 * pnorm(-abs(z))
     } else if (direction == "greater") {
         p_value <- 1 - pnorm(z)
     } else if (direction == "less") {
         p_value <- pnorm(z)
     } else {
         stop("Invalid direction. Please use 'two.sided', 'greater', or 'less'.")
     }
     # Return a list of results
     return(list(p_hat=p_hat,z_value = z, p_value = p_value))
}

RecentHigh<-sum(Temps$H_Year>Temps$L_Year)
```

```r
TotalTemps<-length(Temps$H_Year)

p0<-0.5

p0*TotalTemps>10
```

```
## [1] TRUE
```

```r
(1-p0)*TotalTemps>10
```

```
## [1] TRUE
```

```r
one.samp.prop.test(RecentHigh,TotalTemps,p0,direction='greater')
```

```
## $p_hat
## [1] 0.852459
##
## $z_value
## [1] 5.505586
##
## $p_value
## [1] 1.839713e-08
```

In checking assumptions, we assume the data is collected from a simple random sample of dates, and we can reasonably know that the population of all dates containing record highs or lows in temperature is fairly large. We also know that the dates are divided into two categories: Record Highs and Record Lows. Lastly, when n, or the total number of dates in the sample (61), is multiplied by both p0 (0.5) and (1-p0) (also 0.5), we get 30.5, which is greater than 10. Since all assumptions for a population proportion hypothesis test are satisfied, we are good to proceed with the test. Upon performing the test, we calculate a p-value of 1.839713e-08, or 0.00000001839713.This p-value gives us significant evidence to reject the null hypothesis at alphas, or significance levels, of both 0.05 and 0.01. Therefore, we reject the null hypothesis H0: p=0.5.

3. Based on the test performed, what is your conclusion about whether there has been a warming trend at the Blue Hill Meteorological Observatory?

Based on our p-value and decision to reject the null hypothesis from question 2, we conclude that there is sufficient evidence to support the hypothesis that there has been a warming trend at the Blue Hill Meteorological Observatory. This is due to the fact that there have been more dates in which a record high in temperature has occurred more recently with statistical significance based on our test.

Approach 2

1. What is the midpoint of the observation period? Recall that the data span the years 1893 to 2023.

```r
midpoint<-(1893+2023)/2
print(midpoint)
```

```
## [1] 1958
```

The midpoint of the observation period is calculated by adding together the years at each end of the range and dividing the resulting sum by 2. Running this calculation gives us a midpoint year of 1958.

2. Use the data to test the hypothesis that the mean year in which a record low occurred is equal to the midpoint value of the observation period, against the alternative that the mean year is less than the midpoint.

```r
# One Sample T Test for the Mean Given Sample Data
one.samp.t.test.data<-function(sample_data, null_mean, direction = "two.sided") {
  # Calculate sample statistics
  sample_mean<-mean(sample_data)
  sample_sd<-sd(sample_data)
  sample_size<-length(sample_data)
  # Calculate the t-value
  t_value <- (sample_mean - null_mean) / (sample_sd / sqrt(sample_size))
  # Calculate the degrees of freedom
  df <- sample_size - 1
  # Calculate the p-value and the critical value for the confidence interval
  if (direction == "two.sided") {
    p_value <- 2 * pt(-abs(t_value), df)
  } else if (direction == "greater") {
    p_value <- 1-pt(t_value, df)
  } else if (direction == "less") {
    p_value <- pt(t_value, df)
  } else {
    stop("Invalid direction. Please use 'two.sided', 'greater', or 'less'.")}
  # Return a list of results
  return(list(sample_mean=sample_mean, sample_sd=sample_sd, sample_size=sample_size, t_value = t_value,
              p_value = p_value))}

one.samp.t.test.data(Temps$L_Year,midpoint,direction='less')
```

```
## $sample_mean
## [1] 1945.213
##
## $sample_sd
## [1] 36.90488
##
## $sample_size
## [1] 61
##
## $t_value
## [1] -2.706112
##
## $p_value
## [1] 0.004425378
```

We carry over the assumption from the proportion test in Approach 1 that the sample of data given is a simple random sample, and the size of this sample, n, is 61, which is greater than 30. Since the assumptions for a hypothesis test for a population mean are satisfied, we are good to proceed with the test regarding the mean of Record Low Years. The null hypothesis is H0: p=1958, while the alternative hypothesis is H1: p<1958, as 1958 is the calculated midpoint of the observation period, or where we would expect the mean year in which a record low is observed to fall if there is no warming trend. Upon performing the test, we calculate a sample mean of 1945.213 and a p-value of 0.004425378. This p-value gives us significant evidence to reject the null hypothesis at alphas, or significance levels, of both 0.05 and 0.01. Therefore, we reject the null hypothesis H0: p=1958.

3. Based on the test performed, what is your conclusion about whether there has been a warming trend at the Blue Hill Meteorological Observatory?

Based on our p-value and decision to reject the null hypothesis from question 2 of Approach #2, we conclude that there is again sufficient evidence to support the hypothesis that there has been a warming trend at Blue Hill Meteorological Observatory. This is due to the fact that the sample mean year in which a record low in temperature was observed is less than the midpoint of the observation range at a statistically significant level.

Approach 3

```
#The below code is used so that the full link is visible.
library(magrittr)
urlRemote<-"https://www.dropbox.com/"
pathDropbox<-"s/bl9cuq9a6f98j3s/"
fileName<-"climate_change_blue_hill_2023_monthly.csv?dl=1"
TempsApproach3<-paste0(urlRemote,pathDropbox,fileName) %>% read.csv(header=TRUE)
```

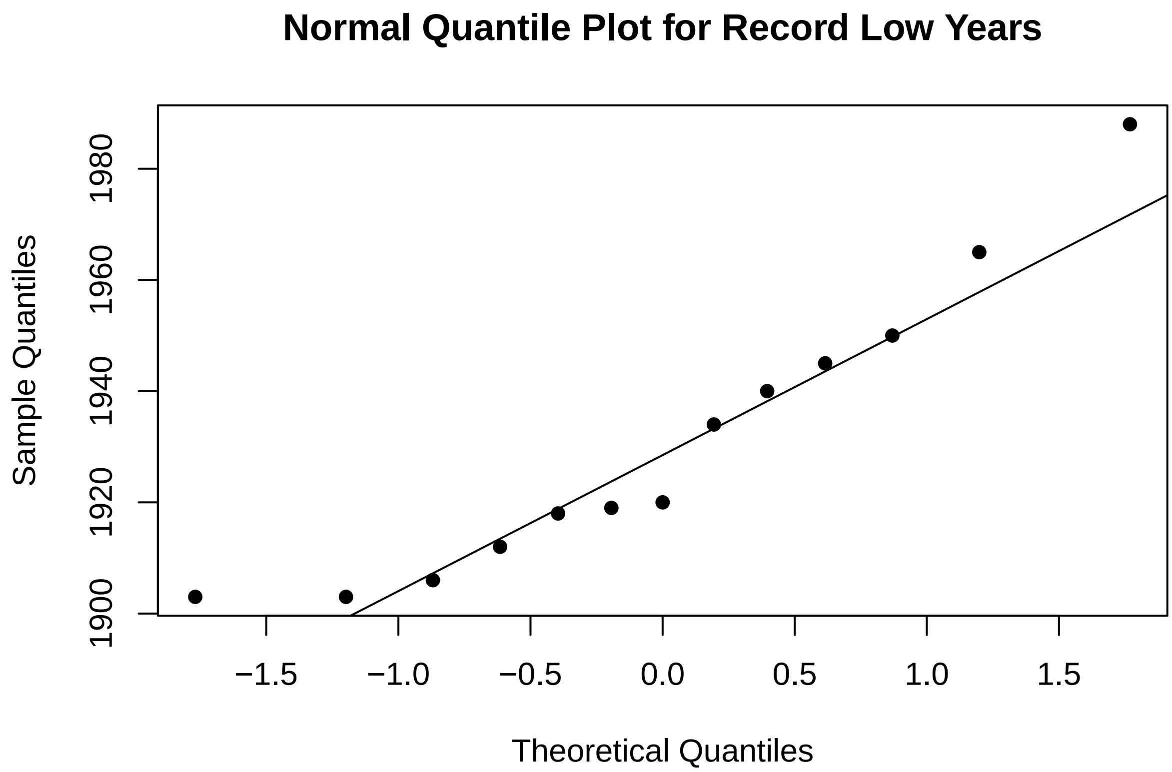
1. Should a hypothesis test be performed to determine whether the mean year in which a record low occurred is less than the midpoint value of the observation period? Explain why or why not – use appropriate graphs to back up your answer.

```
length(TempsApproach3$L_Year)
```

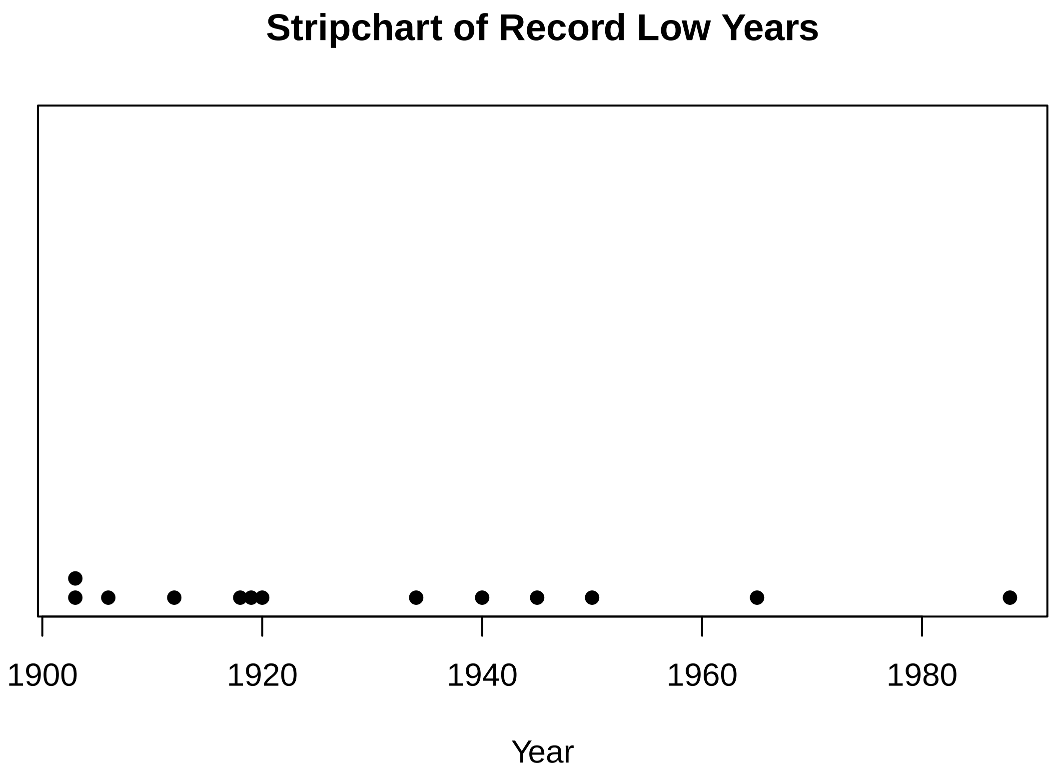
```
## [1] 13
```

While it may be fair to assume that the data set given is taken from a simple random sample, the sample size, n, is only 13, which is less than 30. For this reason, in order to satisfy the assumptions needed to conduct a hypothesis test for a population mean, it would need to be reasonable for us to assume that the distribution of population data for Record Low Years is approximately normal. This is addressed below.

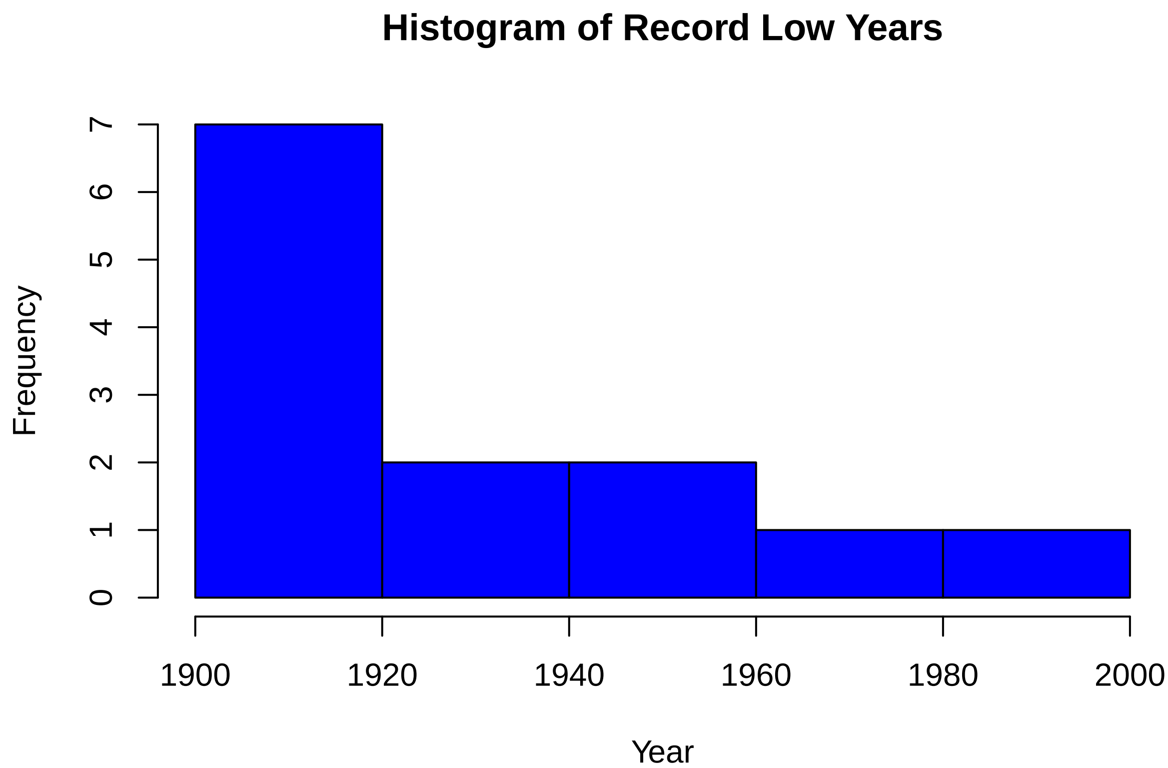
```
qqnorm(TempsApproach3$L_Year,pch=16,main="Normal Quantile Plot for Record Low Years");
qqline(TempsApproach3$L_Year)
```




```
stripchart(TempsApproach3$L_Year,main="Stripchart of Record Low Years",xlab="Year",pch=16,
method="stack",offset=0.5,at=0)
```




```
hist(TempsApproach3$L_Year, col="blue",main="Histogram of Record Low Years",xlab="Year")
```



As shown above by the Normal Quantile Plot, Stripchart, and Histogram of the distribution of the sample data for Record Low Years, this data set is skewed to the right. This means that we cannot reasonably assume that the distribution of population data for Record Low Years is approximately normal. For this reason, the assumptions needed to conduct a hypothesis test for a population mean are not satisfied. A hypothesis test should not be performed on this data set to determine whether the mean year in which a record low occurred is less than the midpoint value of the observation period, as the results would be inconclusive.