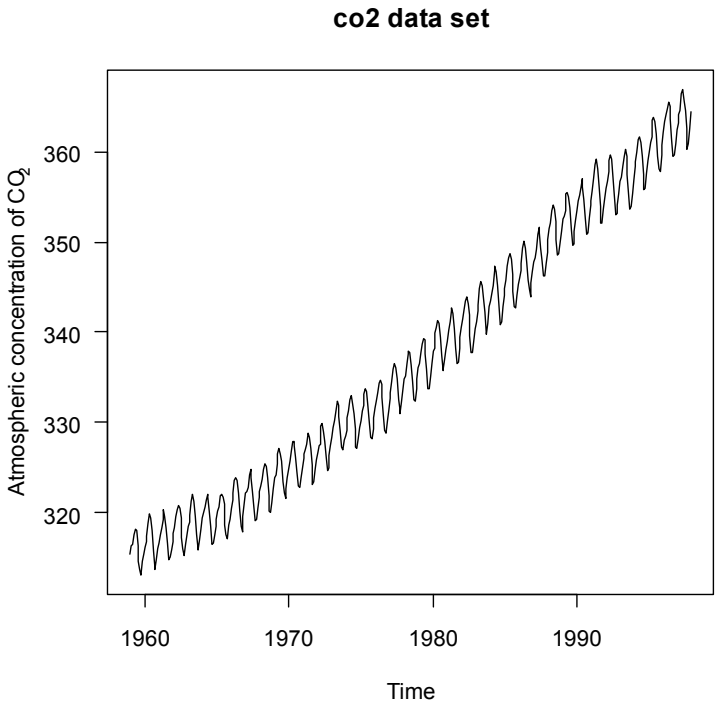
Meaghan Petix
Myat Phyo

Data Set

```
*****************************************************************
*** Atmospheric CO2 concentrations (ppmv) derived from in situ  ***
*** air samples collected at Mauna Loa Observatory, Hawaii      ***
***                                                            ***
*** Source: C.D. Keeling                                       ***
***         T.P. Whorf, and the Carbon Dioxide Research Group  ***
***         Scripps Institution of Oceanography (SIO)          ***
***         University of California                           ***
***         La Jolla, California USA 92093-0444                ***
***                                                            ***
*** June 2004                                                  ***
***                                                            ***
*****************************************************************
```

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Annual	Annual-Fit
1958	-99.99	-99.99	315.71	317.45	317.50	-99.99	315.86	314.93	313.19	-99.99	313.34	314.67	-99.99	-99.99
1959	315.58	316.47	316.65	317.71	318.29	318.16	316.55	314.80	313.84	313.34	314.81	315.59	315.98	316.00
1960	316.43	316.97	317.58	319.03	320.03	319.59	318.18	315.91	314.16	313.83	315.00	316.19	316.91	316.91
1961	316.89	317.70	318.54	319.48	320.58	319.78	318.58	316.79	314.99	315.31	316.10	317.01	317.65	317.63
1962	317.94	318.56	319.69	320.58	321.01	320.61	319.61	317.40	316.26	315.42	316.69	317.69	318.45	318.46
1963	318.74	319.08	319.86	321.39	322.24	321.47	319.74	317.77	316.21	315.99	317.07	318.36	318.99	319.02
1964	319.57	-99.99	-99.99	-99.99	322.23	321.89	320.44	318.70	316.70	316.87	317.68	318.71	-99.99	319.52
1965	319.44	320.44	320.89	322.13	322.16	321.87	321.21	318.87	317.81	317.30	318.87	319.42	320.03	320.09
1966	320.62	321.59	322.39	323.70	324.07	323.75	322.40	320.37	318.64	318.10	319.79	321.03	321.37	321.34
1967	322.33	322.50	323.04	324.42	325.00	324.09	322.55	320.92	319.26	319.39	320.72	321.96	322.18	322.13
1968	322.57	323.15	323.89	325.02	325.57	325.36	324.14	322.11	320.33	320.25	321.32	322.90	323.05	323.11
1969	324.00	324.42	325.64	326.66	327.38	326.70	325.89	323.67	322.38	321.78	322.85	324.12	324.62	324.60
1970	325.06	325.98	326.93	328.13	328.07	327.66	326.35	324.69	323.10	323.07	324.01	325.13	325.68	325.65
1971	326.17	326.68	327.18	327.78	328.92	328.57	327.37	325.43	323.36	323.56	324.80	326.01	326.32	326.32
1972	326.77	327.63	327.75	329.72	330.07	329.09	328.05	326.32	324.84	325.20	326.50	327.55	327.46	327.52
1973	328.54	329.56	330.30	331.50	332.48	332.07	330.87	329.31	327.51	327.18	328.16	328.64	329.68	329.61
1974	329.35	330.71	331.48	332.65	333.09	332.25	331.18	329.40	327.44	327.37	328.46	329.58	330.25	330.29
1975	330.40	331.41	332.04	333.31	333.96	333.59	331.91	330.06	328.56	328.34	329.49	330.76	331.15	331.16
1976	331.74	332.56	333.50	334.58	334.87	334.34	333.05	330.94	329.30	328.94	330.31	331.68	332.15	332.18
1977	332.92	333.42	334.70	336.07	336.74	336.27	334.93	332.75	331.58	331.16	332.40	333.85	333.90	333.88
1978	334.97	335.39	336.64	337.76	338.01	337.89	336.54	334.68	332.76	332.54	333.92	334.95	335.50	335.52
1979	336.23	336.76	337.96	338.89	339.47	339.29	337.73	336.09	333.91	333.86	335.29	336.73	336.85	336.89
1980	338.01	338.36	340.08	340.77	341.46	341.17	339.56	337.60	335.88	336.01	337.10	338.21	338.69	338.67
1981	339.23	340.47	341.38	342.51	342.91	342.25	340.49	338.43	336.69	336.85	338.36	339.61	339.93	339.95
1982	340.75	341.61	342.70	343.56	344.13	343.35	342.06	339.82	337.97	337.86	339.26	340.49	341.13	341.09
1983	341.37	342.52	343.10	344.94	345.75	345.32	343.99	342.39	339.86	339.99	341.16	342.99	342.78	342.75
1984	343.70	344.51	345.28	347.08	347.43	346.79	345.40	343.28	341.07	341.35	342.98	344.22	344.42	344.44
1985	344.97	346.00	347.43	348.35	348.93	348.25	346.56	344.69	343.09	342.80	344.24	345.56	345.90	345.86
1986	346.29	346.96	347.86	349.55	350.21	349.54	347.94	345.91	344.86	344.17	345.66	346.90	347.15	347.14
1987	348.02	348.47	349.42	350.99	351.84	351.25	349.52	348.10	346.44	346.36	347.81	348.96	348.93	348.99
1988	350.43	351.72	352.22	353.59	354.22	353.79	352.39	350.44	348.72	348.88	350.07	351.34	351.48	351.44
1989	352.76	353.07	353.68	355.42	355.67	355.13	353.90	351.67	349.80	349.99	351.30	352.53	352.91	352.94
1990	353.66	354.70	355.39	356.20	357.16	356.22	354.82	352.91	350.96	351.18	352.83	354.21	354.19	354.19
1991	354.72	355.75	357.16	358.60	359.34	358.24	356.17	354.03	352.16	352.21	353.75	354.99	355.59	355.62
1992	355.98	356.72	357.81	359.15	359.66	359.25	357.03	355.00	353.01	353.31	354.16	355.40	356.37	356.36
1993	356.70	357.16	358.38	359.46	360.28	359.60	357.57	355.52	353.70	353.98	355.33	356.80	357.04	357.10
1994	358.36	358.91	359.97	361.26	361.68	360.95	359.55	357.49	355.84	355.99	357.58	359.04	358.88	358.86
1995	359.96	361.00	361.64	363.45	363.79	363.26	361.90	359.46	358.06	357.75	359.56	360.70	360.88	360.90
1996	362.05	363.25	364.03	364.72	365.41	364.97	363.65	361.49	359.46	359.60	360.76	362.33	362.64	362.58
1997	363.18	364.00	364.57	366.35	366.79	365.62	364.47	362.51	360.19	360.77	362.43	364.28	363.76	363.84
1998	365.32	366.15	367.31	368.61	369.29	368.87	367.64	365.77	363.90	364.23	365.46	366.97	366.63	366.58
1999	368.15	368.87	369.59	371.14	371.00	370.35	369.27	366.94	364.63	365.12	366.67	368.01	368.31	368.30
2000	369.14	369.46	370.52	371.66	371.82	371.70	370.12	368.12	366.62	366.73	368.29	369.53	369.48	369.47
2001	370.28	371.50	372.12	372.87	374.02	373.30	371.62	369.55	367.96	368.09	369.68	371.24	371.02	371.03
2002	372.43	373.09	373.52	374.86	375.55	375.41	374.02	371.49	370.71	370.25	372.08	373.78	373.10	373.07
2003	374.68	375.63	376.11	377.65	378.35	378.13	376.62	374.50	372.99	373.00	374.35	375.70	375.64	375.61

Monthly values are expressed in parts per million (ppm) and reported in the 2003A SIO manometric mole fraction scale. The monthly values have been adjusted to the 15th of each month. Missing values are denoted by -99.99. The "annual" average is the arithmetic mean of the twelve monthly values. In years with one or two missing monthly values, annual values were calculated by substituting a fit value (4-harmonics with gain factor and spline) for that month and then averaging the twelve monthly values.



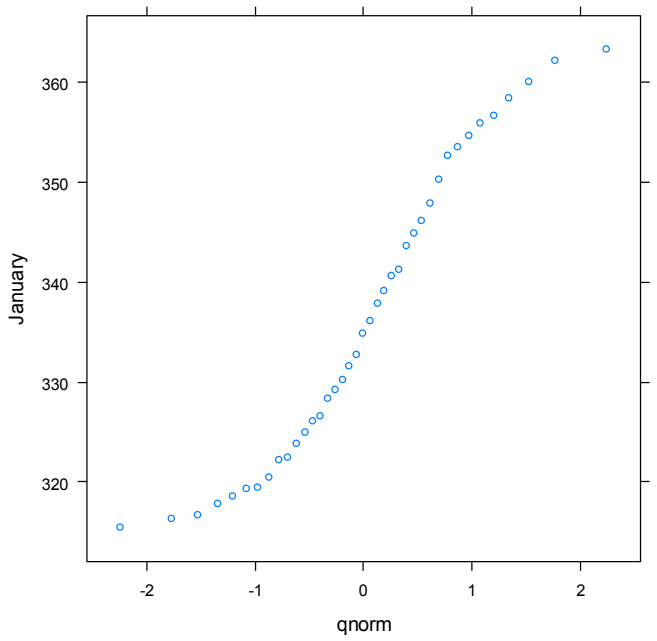
Introduction

For our project, we will be working with Paired t-test, Wilcoxan Signed-Ranked test, and Sign-test. All three of the tests are employed when using paired data, which means that the two sets of measurements are exactly matched for each individual. However, there are some considerable differences between these tests. The Sign-test and Wilcoxon Signed-Rank test are both non-parametric analogs to the Paired t-test; they are typically employed when the paired data seriously violates the underlying assumption of normality for the Paired t-test.

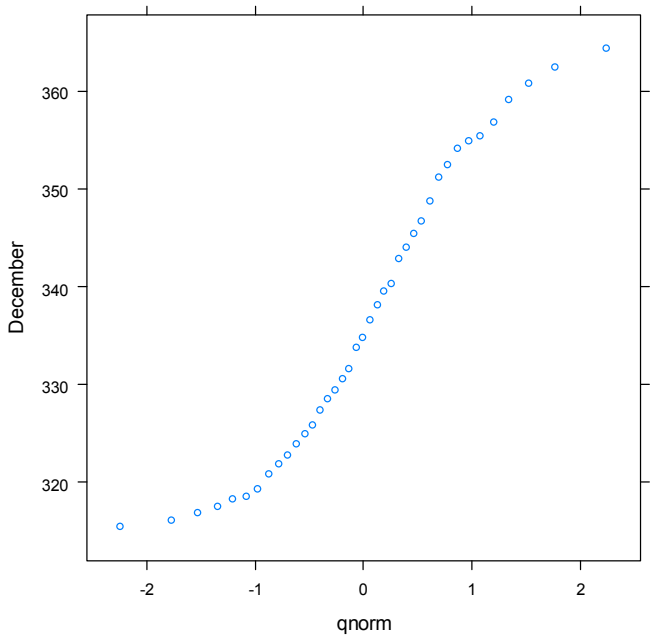
We began our project by looking at the normality of the data with qqmath() function from the R program. Since our data displayed a fairly normal distribution, we predicted that the Paired t-test would give us the most accurate results for our data set. One reason for this is that the Sign test involved loss of information and as a result, lessens the power of the test. Wilcoxon Signed-Rank test utilizes more information than available in the Sign test, but for this particular situation we still believed that the Paired t-test would be preferable since our data is indeed normally distributed.

Assessing Data Normality

```
x<-read.csv(file="C:/Rdata/co2.csv")
x
attach(x)
#load package “lattice”
qqmath(January)
```




```
qqmath(December)
```



Paired t-test

```
x<-read.csv(file="C:/Rdata/co2.csv")
x
```

```
   Year January December
1  1959  315.42   315.43
2  1960  316.27   316.03
3  1961  316.73   316.85
4  1962  317.78   317.53
5  1963  318.58   318.20
6  1964  319.41   318.55
7  1965  319.27   319.25
8  1966  320.46   320.87
9  1967  322.17   321.80
10 1968  322.40   322.74
11 1969  323.83   323.95
12 1970  324.89   324.96
13 1971  326.01   325.85
14 1972  326.60   327.39
15 1973  328.37   328.48
16 1974  329.18   329.41
17 1975  330.23   330.59
18 1976  331.58   331.52
19 1977  332.75   333.68
20 1978  334.80   334.78
21 1979  336.05   336.56
22 1980  337.84   338.04
23 1981  339.06   339.44
24 1982  340.57   340.32
25 1983  341.20   342.82
26 1984  343.52   344.04
27 1985  344.79   345.38
28 1986  346.11   346.72
29 1987  347.84   348.78
30 1988  350.25   351.18
31 1989  352.60   352.37
32 1990  353.50   354.07
33 1991  354.59   354.89
34 1992  355.88   355.33
35 1993  356.63   356.78
36 1994  358.34   359.05
37 1995  359.98   360.74
38 1996  362.09   362.38
39 1997  363.23   364.34
```

```
attach(x)
X1=January
X2=December
d=X1-X2
n=length(d)
n
#n=39
dbar=mean(d)
dbar
#dbar=-0.2638462
sd=sqrt(var(d))
sd
#sd=0.500482
```

#Hypotheses:

Null Hypothesis (H0): No difference in mean between populations X1 and X2
Alternative Hypothesis (H1): Difference in mean between populations X1 and X2

#Performing the Paired t-test:

```
t.test(X1,X2,alternative="two.sided",paired=T,conf.level=0.95)
```

#Output:

```
        Paired t-test

data:  X1 and X2
t = -3.2923, df = 38, p-value = 0.002154
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4260836 -0.1016087
sample estimates:
mean of the differences
           -0.2638462
```

#Critical Value:

```
n=39
a=0.05
C1=qt(a/2,n-1)
C1    <- Two-sided lower critical value
C2=qt(1-a/2,n-1)
C2    <- Two-sided upper critical value
OR
C1=qt(a/2,n-1)
C2=abs(C1)    <- C2 is the absolute value of C1
```

#Output:

```
C1 =  -2.024394
C2 =   2.024394
```

#Decision Rule

IF |t| > C THEN REJECT H0, OTHERWISE ACCEPT H0
Therefore, in this case:
|t| > C so we reject H0 (null hypothesis) and conclude that there is a difference in means between populations X1 and X2

#Probability Value

```
#t < 0
t=-3.2923
P=2*pt(t,n-1)
P
```

#Output:

```
P=0.002153345
a=0.05
```

P < a, therefore reject H0 and conclude that there is a difference in means between populations X1 and X2

Wilcoxon Signed-Rank Test:

```
x<-read.csv(file="C:/Rdata/co2(d).csv")
x
```

```
   Year January December X.VALUE.
1  1959  315.42   315.43    -0.01
2  1960  316.27   316.03     0.24
3  1961  316.73   316.85    -0.12
4  1962  317.78   317.53     0.25
5  1963  318.58   318.20     0.38
6  1964  319.41   318.55     0.86
7  1965  319.27   319.25     0.02
8  1966  320.46   320.87    -0.41
9  1967  322.17   321.80     0.37
10 1968  322.40   322.74    -0.34
11 1969  323.83   323.95    -0.12
12 1970  324.89   324.96    -0.07
13 1971  326.01   325.85     0.16
14 1972  326.60   327.39    -0.79
15 1973  328.37   328.48    -0.11
16 1974  329.18   329.41    -0.23
17 1975  330.23   330.59    -0.36
18 1976  331.58   331.52     0.06
19 1977  332.75   333.68    -0.93
20 1978  334.80   334.78     0.02
21 1979  336.05   336.56    -0.51
22 1980  337.84   338.04    -0.20
23 1981  339.06   339.44    -0.38
24 1982  340.57   340.32     0.25
25 1983  341.20   342.82    -1.62
26 1984  343.52   344.04    -0.52
27 1985  344.79   345.38    -0.59
28 1986  346.11   346.72    -0.61
29 1987  347.84   348.78    -0.94
30 1988  350.25   351.18    -0.93
31 1989  352.60   352.37     0.23
32 1990  353.50   354.07    -0.57
33 1991  354.59   354.89    -0.30
34 1992  355.88   355.33     0.55
35 1993  356.63   356.78    -0.15
36 1994  358.34   359.05    -0.71
37 1995  359.98   360.74    -0.76
38 1996  362.09   362.38    -0.29
39 1997  363.23   364.34    -1.11
```

```
attach(x)
```

#Hypotheses:

Null Hypothesis (H0) - no population ordinal difference in median
Alternative Hypothesis (H1) - population ordinal difference in median

#Performing the Wilcoxon Signed-Rank Test:

```
wilcox.test(January,December,alternative="two.sided",paired=T)
```

#Output:

```
    Wilcoxon signed rank test with continuity correction

data:  January and December
V = 181, p-value = 0.003617
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(January, December, alternative = "two.sided",  :
  cannot compute exact p-value with ties
```

#Critical Value of the Test

```
a=0.05
C=qnorm(1-a/2,0,1)
C
```

#Output:

```
C =  1.959964
```

#Probability Value

P = 0.003617

P < a, therefore reject H0

Since we reject the null hypothesis, we accept the alternative hypothesis that there is a population ordinal difference in median.

Sign Test:

```
x<-read.csv(file="C:/Rdata/co2(s).csv")
x
```

```
   Year January December     d
1  1959  315.42   315.43 minus
2  1960  316.27   316.03  plus
3  1961  316.73   316.85 minus
4  1962  317.78   317.53  plus
5  1963  318.58   318.20  plus
6  1964  319.41   318.55  plus
7  1965  319.27   319.25  plus
8  1966  320.46   320.87 minus
9  1967  322.17   321.80  plus
10 1968  322.40   322.74 minus
11 1969  323.83   323.95 minus
12 1970  324.89   324.96 minus
13 1971  326.01   325.85  plus
14 1972  326.60   327.39 minus
15 1973  328.37   328.48 minus
16 1974  329.18   329.41 minus
17 1975  330.23   330.59 minus
18 1976  331.58   331.52  plus
19 1977  332.75   333.68 minus
20 1978  334.80   334.78  plus
21 1979  336.05   336.56 minus
22 1980  337.84   338.04 minus
23 1981  339.06   339.44 minus
24 1982  340.57   340.32  plus
25 1983  341.20   342.82 minus
26 1984  343.52   344.04 minus
27 1985  344.79   345.38 minus
28 1986  346.11   346.72 minus
29 1987  347.84   348.78 minus
30 1988  350.25   351.18 minus
31 1989  352.60   352.37  plus
32 1990  353.50   354.07 minus
33 1991  354.59   354.89 minus
34 1992  355.88   355.33  plus
35 1993  356.63   356.78 minus
36 1994  358.34   359.05 minus
37 1995  359.98   360.74 minus
38 1996  362.09   362.38 minus
39 1997  363.23   364.34 minus
```

```
attach(x)
```

```
n=length(January)
n
#n=39
p=0.5
PLUS=length(January[d=="plus"])
PLUS
#PLUS=12
MINUS=length(January[d=="minus"])
MINUS
#MINUS=27
```

#Hypotheses:

Null Hypothesis (H0) - no population ordinal difference in median
Alternative Hypothesis (H1) - population ordinal difference in median

#Performing the Sign Test:

```
binom.test(PLUS,n,p, alternative="two.sided",conf.level=0.95)
```

#Output:

```
  Exact binomial test

data:  PLUS and n
number of successes = 12, number of trials = 39, p-value = 0.02370
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.1701959 0.4756914
sample estimates:
probability of success
          0.3076923
```

#Exact Method Test Statistic:

C = 12
where C = # of di's where di is +

#Probability Value

a=0.05
P=0.02370

P < a, therefore reject H0 (null hypothesis) and conclude that there is a population ordinal difference in median

#Alternative calculation using the Binomial function pbinom():

Since C=12 and n/2=19.5,
C < n/2

```
C=12
p=0.5
P=2*pbinom(C,n,p)
#P = 0.02370270
```

Conclusion

All three tests brought us to the same conclusion that there was a significant difference between populations X1 and X2, which were the months of January and December, respectively. The Paired t-test showed us there was a difference in mean between the two populations, whereas the other two tests showed us there was a population ordinal difference in median. Although all three tests led us to the same conclusion, we have the greatest confidence in the Paired t-test because we were able to reject our null hypothesis with both the decision rule for critical value and probability value. In conclusion, we believe that there is a difference in means between population X1 and X2, which means there is a significant difference in the collected data values of January and those collected nearly a year later in December.