Time Series Analysis
Homework Assignment 2
Submitted by Sruthy Annie Santhosh,312213

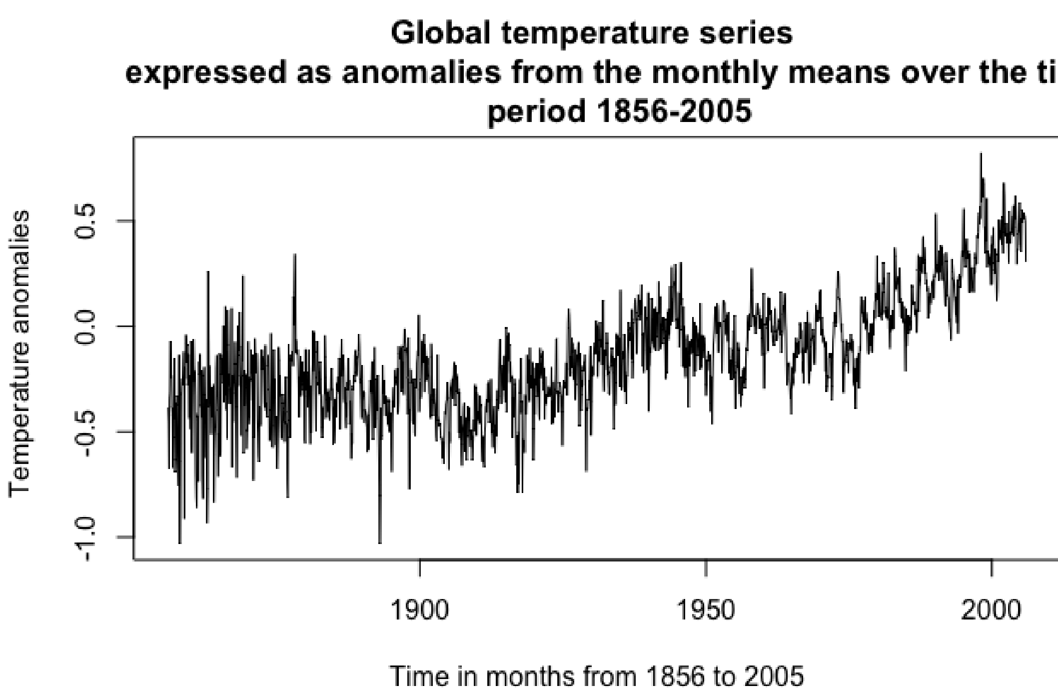
Stationary Processes

In this assignment we are dealing with the data set on the statistic of global temperature series, expressed as anomalies from the monthly means over the period 1856-2005.(source: http://www.maths.adelaide.edu.au/ emac2009/).

(1) To plot the time series

First we need to set the local directory using setwd() and then import the required libraries. Then I imported the given data file using read.table() and stored it in temperature. The scan function is used to read the file.

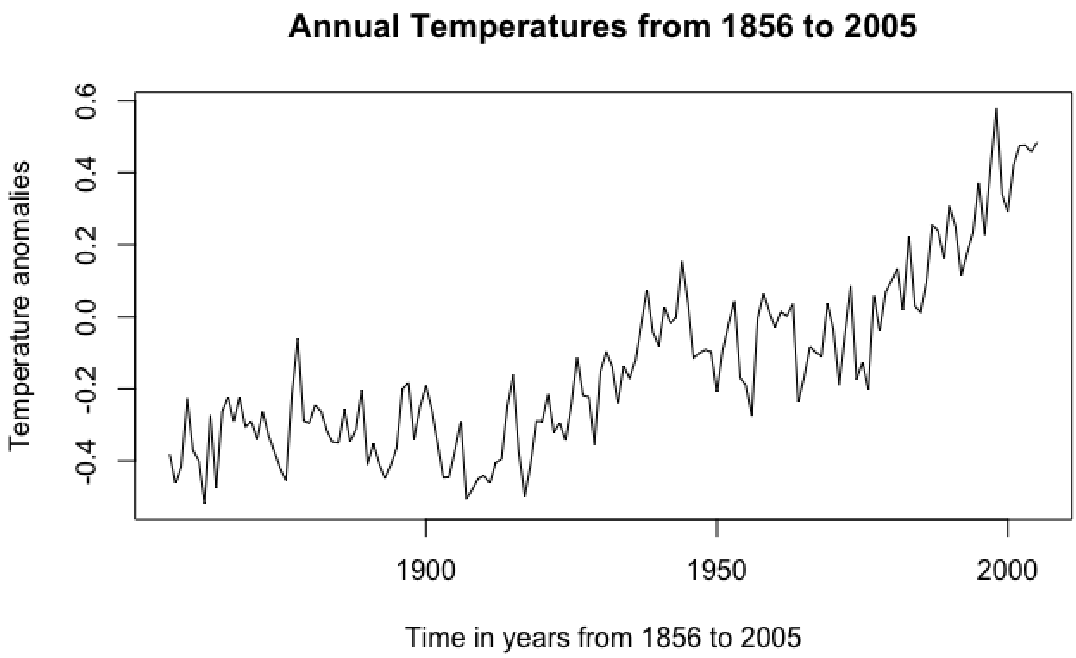
The data is loaded from time 1856 to 2005 to the time series object temp.ts. This ts object is plotted with Time in months along x axis and temperature anomalies on y axis.



We observe that temperature values increases over time with irregularities which can be random or seasonal effects.

(2) Use the aggregate function to remove any seasonal effects within each year and plot an annual series (called {xt}) of mean temperatures for the period 1856-2005.

aggregate() on the yearly data divided by 12, removes all the seasonality and provides us with a smooth graph.



We observe that randomness is present throughout. But the overall trend is that of increasing.

(3) Assume that the time series {xt} obtained in step (2) has the following form:
xt = α + βt + rt , where α, β are constants, and {rt} is a random series. Determine the coefficients α and β using the least-square method.

We use time() of the time series object to get the X axis values, Y values - time series obj.
Then using lm() we find the estimate parameters as :

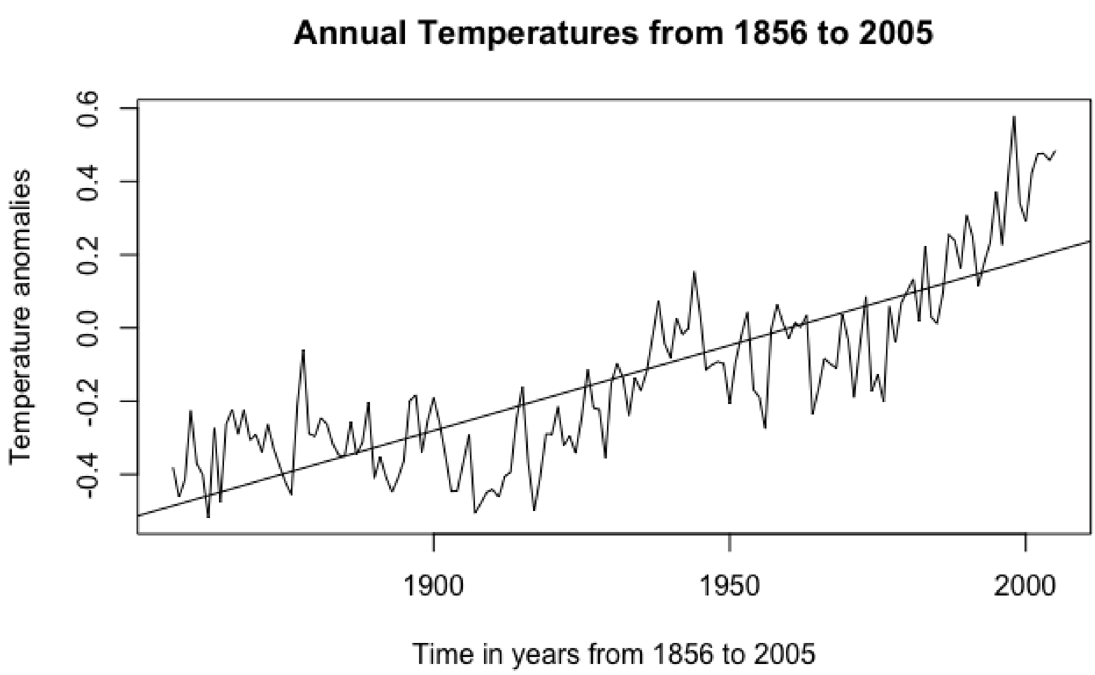
```
> coef(temp.lm)
 (Intercept)         TIME
-9.135592601  0.004660622
>
```

(4) Use the command abline to draw a regression line to the existing plot in step (2).

Summary() function outputs the results of the linear regression.

```
> summary(temp.lm)

Call:
lm(formula = temp.annual ~ TIME)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.29749 -0.09620 -0.00188  0.08956  0.40334 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -9.1355926  0.5043511  -18.11   <2e-16 ***
TIME         0.0046606  0.0002612   17.84   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1385 on 148 degrees of freedom
Multiple R-squared:  0.6827,	Adjusted R-squared:  0.6805 
F-statistic: 318.4 on 1 and 148 DF,  p-value: < 2.2e-16
```

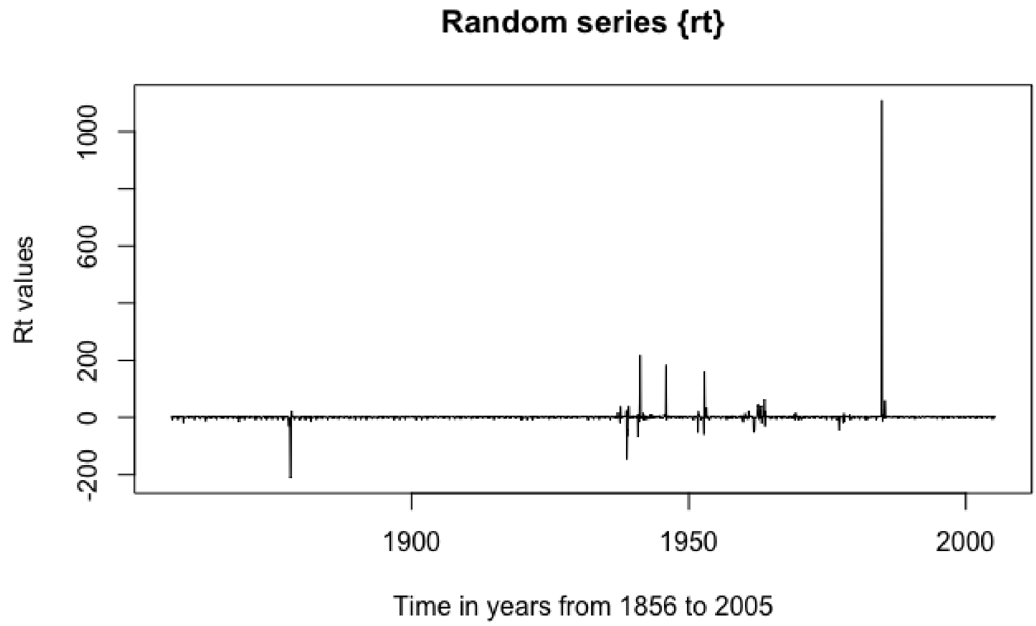
abline() function - adds lines to graph. Hence we can pass the result of lm to obtain the regression line.



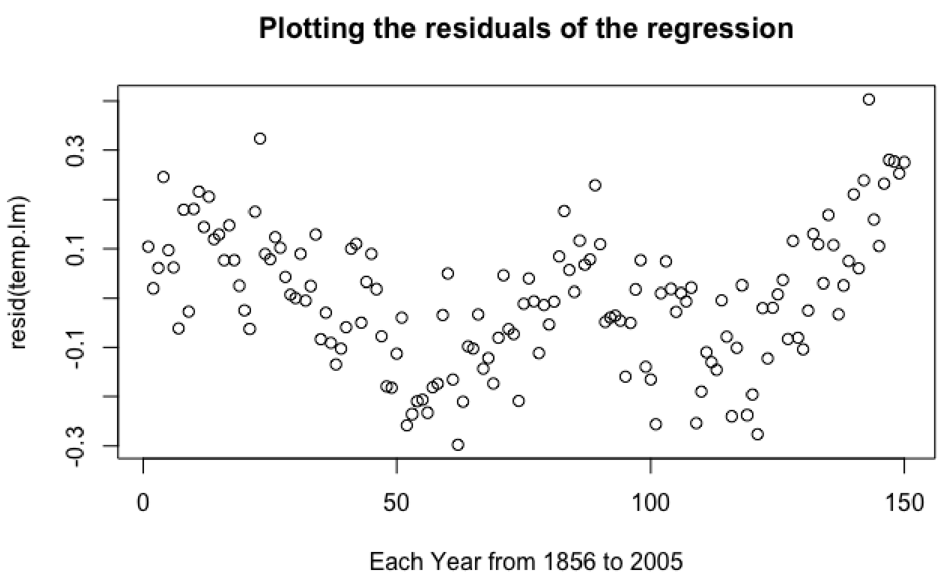
We can see that the line is also increasing.

(5) Plot the random series {rt}.

We can use the decompose function to get the randomness in the time series. This randomness can be plotted as follows:



We see that the randomness usually near 0 but has a spike in between 1950 - 2000.
We can also extract the residuals after performing the linear regression on the time series. This can be plotted as :



(6) Carry out the turning point*, Box-Pierce, Ljung-Box and difference sign* tests for the random series {rt}. For the tests with *, the Package randtests are required. Based on the above tests, give comments on the results. In the box tests, calculate all p-values for lag = 1, 2, . . . , 24.

We need to install the library randtests to do the turning point and difference sign tests for random series rt.

Here we are taking the threshold value as 0.05. A p-value of 0.05 or lower is generally considered statistically significant.

```
> turning.point.test(resid(temp.lm))

        Turning Point Test

data:  resid(temp.lm)
statistic = -2.6627, n = 150, p-value = 0.007752
alternative hypothesis: non randomness
```

We can see that the value for turning point test statistic is -2.6627 and the corresponding p-value is 0.007752. This p value is smaller than a threshold α = 0.05. Hence we can reject the i.i.d hypothesis at this level 0.05. Thus the result is statistically significant if threshold=0.05

Box-pierce Test for p-values for lag = 1…24

```
        Box-Pierce test

data:  resid(temp.lm)
X-squared = 64.371, df = 1, p-value = 9.992e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 99.342, df = 2, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 130.39, df = 3, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 164.74, df = 4, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 189.31, df = 5, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 212.99, df = 6, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 236.66, df = 7, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 254.47, df = 8, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 267.3, df = 9, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 282.51, df = 10, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 292.38, df = 11, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 296.8, df = 12, p-value < 2.2e-16
```

```
        Box-Pierce test

data:  resid(temp.lm)
X-squared = 299.25, df = 13, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 304.1, df = 14, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 307.56, df = 15, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 308.53, df = 16, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 309.65, df = 17, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 311.43, df = 18, p-value < 2.2e-16
```

```
        Box-Pierce test

data:  resid(temp.lm)
X-squared = 311.63, df = 19, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 311.64, df = 20, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 311.79, df = 21, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 312.32, df = 22, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 315.39, df = 23, p-value < 2.2e-16


        Box-Pierce test

data:  resid(temp.lm)
X-squared = 318.76, df = 24, p-value < 2.2e-16
```

Here we see that for Box Pierce test, all the p-values are less than $2.2e^{-16}$. This indicates a significant result. For a threshold $\alpha = 0.05$, we can reject the i.i.d hypothesis.

Ljung Box test

```
	Box-Ljung test

data:  resid(temp.lm)
X-squared = 65.667, df = 1, p-value = 5.551e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 101.58, df = 2, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 133.69, df = 3, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 169.45, df = 4, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 195.21, df = 5, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 220.2, df = 6, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 245.36, df = 7, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 264.42, df = 8, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 278.26, df = 9, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 294.77, df = 10, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 305.57, df = 11, p-value < 2.2e-16


	Box-Ljung test

data:  resid(temp.lm)
X-squared = 310.43, df = 12, p-value < 2.2e-16
```

```
        Box-Ljung test

data:  resid(temp.lm)
X-squared = 313.14, df = 13, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 318.56, df = 14, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 322.46, df = 15, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 323.57, df = 16, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 324.85, df = 17, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 326.89, df = 18, p-value < 2.2e-16
```

```
        Box-Ljung test

data:  resid(temp.lm)
X-squared = 327.13, df = 19, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 327.14, df = 20, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 327.32, df = 21, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 327.94, df = 22, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 331.62, df = 23, p-value < 2.2e-16

        Box-Ljung test

data:  resid(temp.lm)
X-squared = 335.69, df = 24, p-value < 2.2e-16
```

Here also we see that for Box Ljung test, all the p-values are less than $2.2e^{-16}$. This indicates a significant result. For a threshold $\alpha = 0.05$, we can reject the i.i.d hypothesis.

Difference Sign Test

```
> difference.sign.test(resid(temp.lm))

        Difference Sign Test

data:  resid(temp.lm)
statistic = 0.42286, n = 150, p-value = 0.6724
alternative hypothesis: nonrandomness

> difference.sign.test(temp.annual)

        Difference Sign Test

data:  temp.annual
statistic = 1.2686, n = 150, p-value = 0.2046
alternative hypothesis: nonrandomness
```

For the Difference Sign test, the statistic value is 1.286 and the p value is 0.6724. This value is not less than the threshold value of 0.05. The Difference sign test does not detect a significant contradiction with the IID hypothesis. But this does not prove the IID hypothesis.