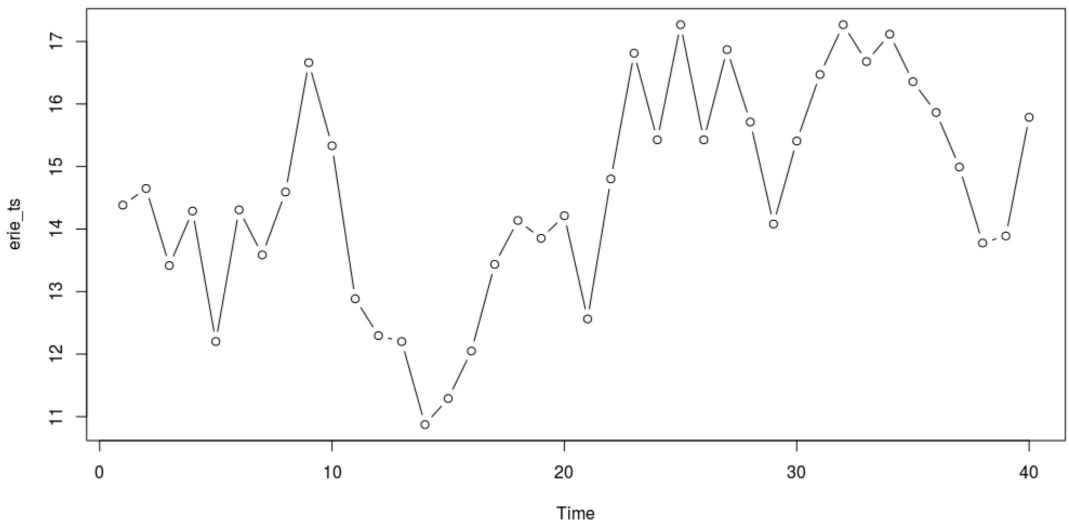
# STAT 463: Homework 2

**Name:**Kyle Salitrik | ***ID#:*** *997543474* | ***PSU ID:*** *kps168*

January 29, 2018

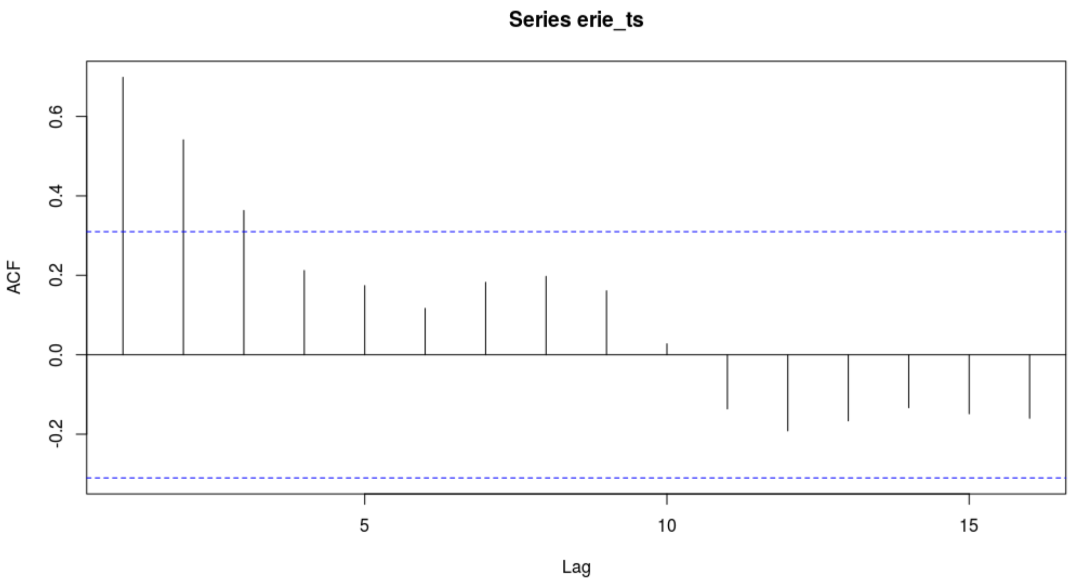
a)




b)

From the plot of the time series, point 14 could potentially be an outlier depending on who is looking at the data, but most likely is not. There is no seasonality to the data because it was collected yearly, however after year 20 the average does appear to increase. The variance is fairly constant from year 20 through 40.

c)




The first 5 or so values almost follow the theoretical curve. The table below summarizes the calculated and theoretical values, the absolute difference and the percentage difference from theoretical:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ACF Values | 0.698 | 0.541 | 0.363 | 0.212 | 0.174 | 0.117 | 0.182 |
| Theoretical Values | 0.698 | 0.488 | 0.341 | 0.238 | 0.1662 | 0.1661 | 0.081 |
| Absolute Difference | 0 | 0.053 | 0.022 | 0.026 | 0.008 | 0.001 | 0.101 |
| Percentage Difference | 0% | 9.829% | 6.117% | 12.220% | 4.534% | 0.634% | 55.540% |

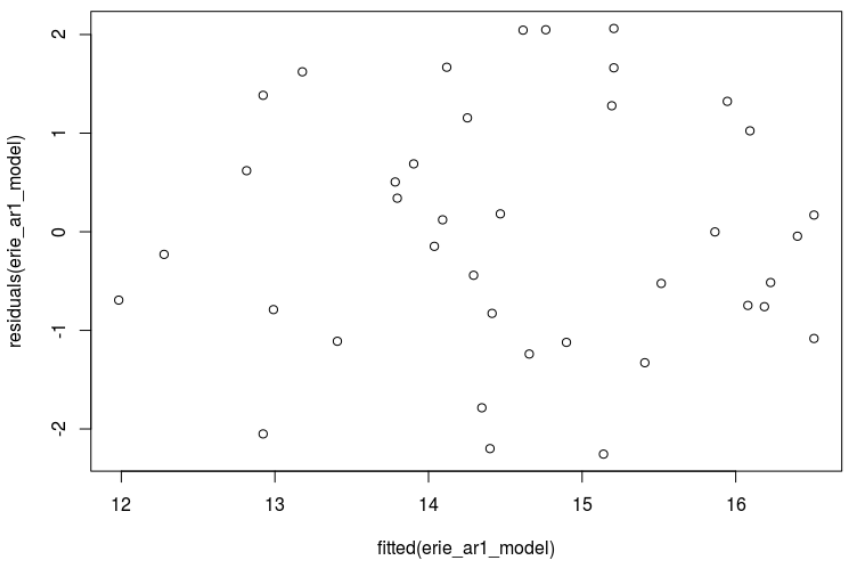
## d)

```
Call:
lm(formula = y[, 1] ~ y[, 2])

Residuals:
     Min       1Q   Median       3Q      Max
-2.25526 -0.80864 -0.04491  1.08912  2.06151

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2878     1.7231   2.488   0.0175 *
y[, 2]        0.7078     0.1176   6.017 5.95e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.25 on 37 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.4946,    Adjusted R-squared:  0.4809
F-statistic:  36.2 on 1 and 37 DF,  p-value: 5.954e-07
```

$$\hat{y} = 4.2878 + 0.7078x$$

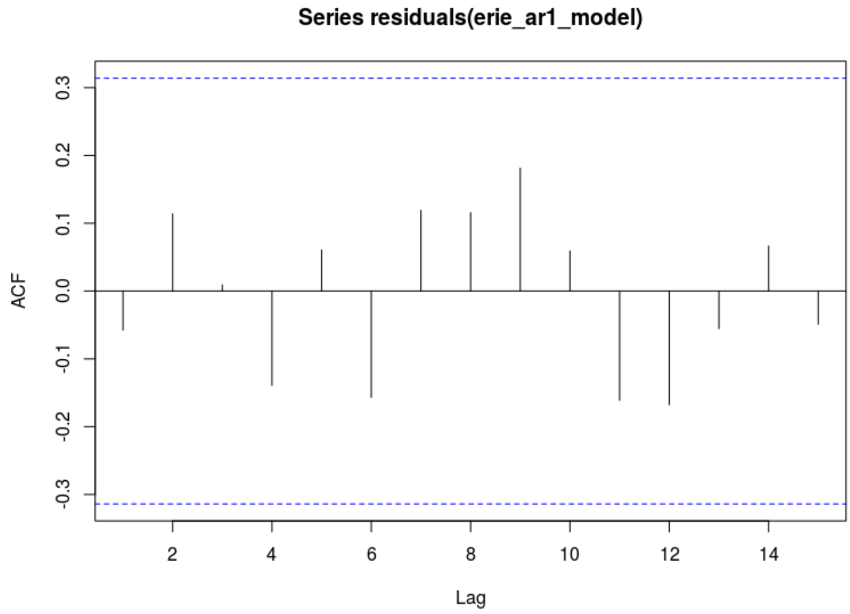
## e)

The plot of residuals vs fitted values looks alright for a linear fitted model. The variance is mostly constant, there does not appear to be any trends in the data and the residuals themselves are fairly small ($\pm 2$) compared to the values of the water level (10.87 minimum, so less than a 20% error at most for the residuals).



## f)

From the plot of the residual ACF, because all of the residuals are within the confidence band, they appear to be affects of noise in the data.



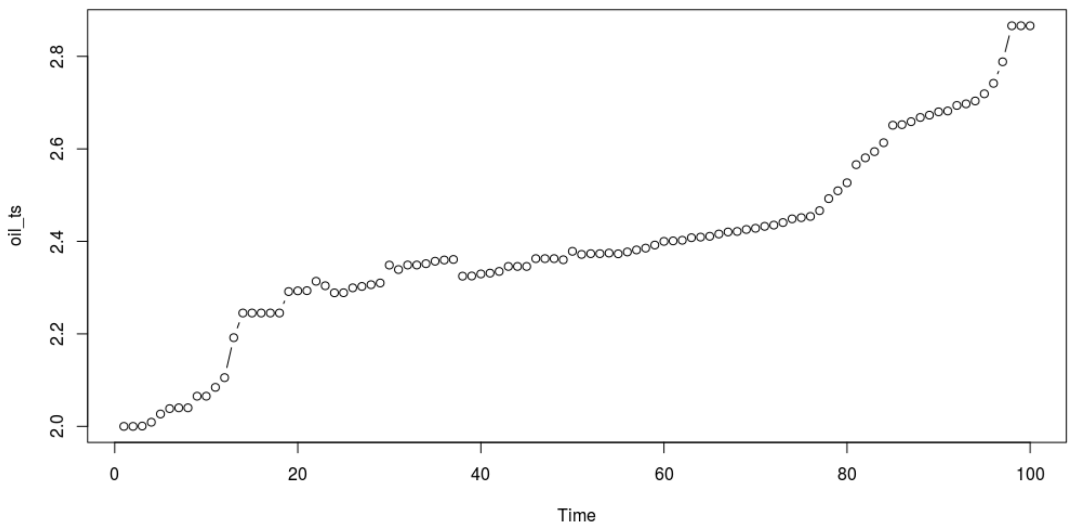

## g)

$\hat{y}_{t=41} = 4.2878 + 0.7078 * 15.787 = 15.4618386$

## PROBLEM 2

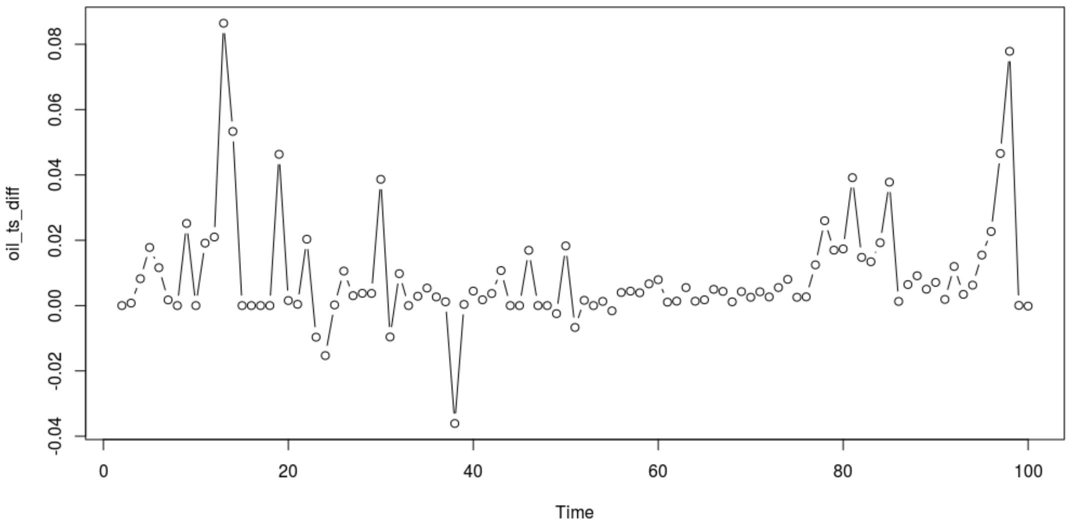
## a)

Based on the plot of the time series below, one piece of evidence that makes it safe to assume the series is not stationary is a non-constant mean.



## b)

Observing the plot of the first differences of the data, with the exception of a few values that have a significantly large absolute value, the first difference is fairly constant at 0. The potential outliers referenced are at approximately points 13,14, 20, 30, 40, 97 and 98. However it is worth noting that the largest difference is 0.08 which is fairly close to 0.

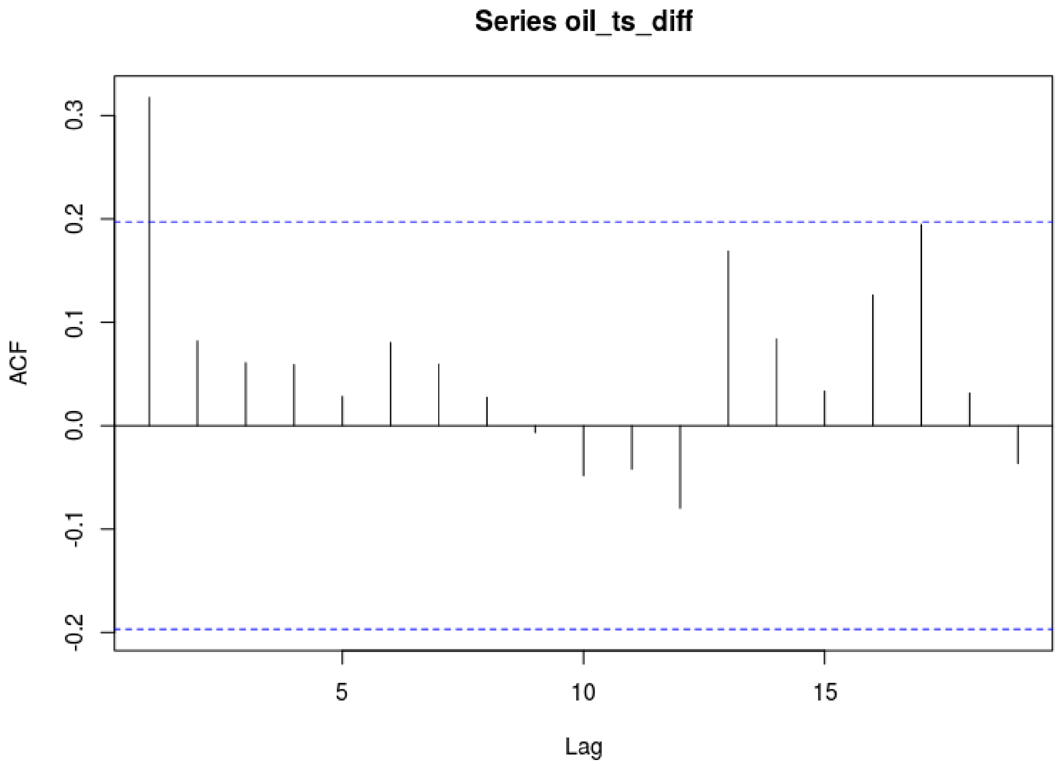



## c)

The R output is as follows:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 0.317 | 0.082 | 0.061 | 0.059 | 0.028 | 0.080 | 0.059 | 0.027 | -0.007 | -0.048 | -0.042 | -0.080 |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 | | | | | |
| 0.169 | 0.084 | 0.033 | 0.126 | 0.194 | 0.031 | -0.037 | | | | | |

The plot below shows the ACF for the first differences. The plot doesn't appear to follow an AR(1) model because the drop off in magnitude seems too steep and the values are not constantly decreasing.

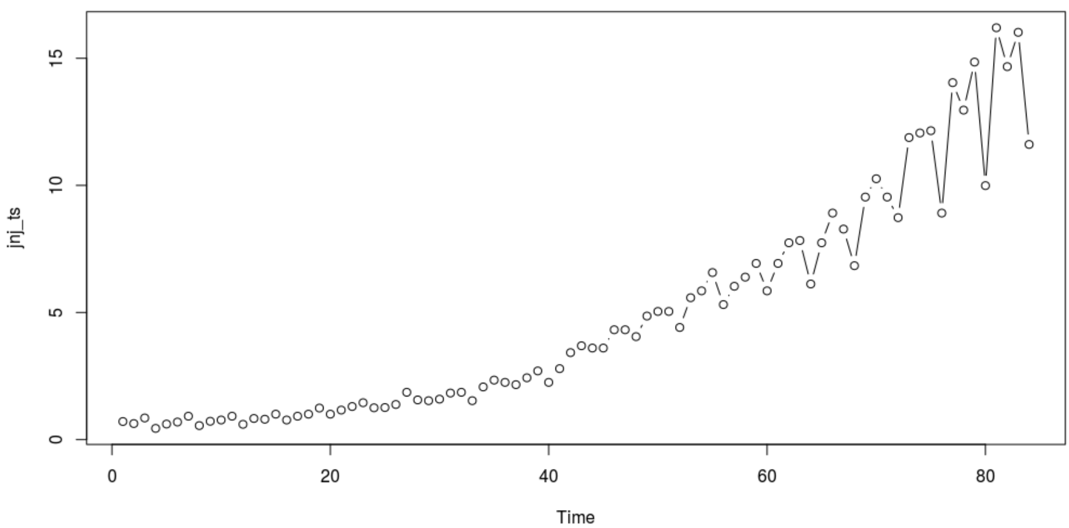



## d)

The MA(1) model could potentially work because, as explained in class, the MA(1) model ACF plot should have a single spike in theory. However in practice the noisy data can cause small spikes further in the series. Because all of the values are within the confidence interval, they are considered to be close enough to 0 to be insignificant as true data and written off as noise.

## PROBLEM 3

## a)

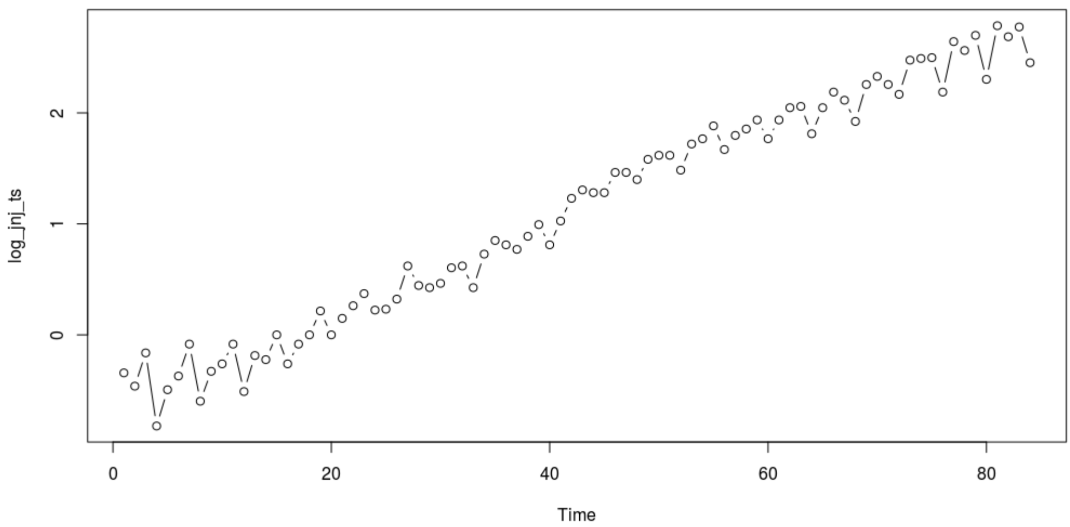


## b)

For the profit data, there appears to be a constant quadratic or exponential trend to the plot over time. Up until around t = 50, the variance is relatively constant, however it also begins to increase as time goes on.

## c)

Taking the logarithm of the profit values, the exponential trend is replaced with a linear one and the variance seems to be constant throughout the entire time period observed.




## d)

If using a linear regression, the following model would most likely be a sufficient fit:

$$log(\hat{y}) = \beta_0 + \beta_1 * \text{profit}$$